Title: Spatial dynamics of the tumor microenvironment linked to emerging resistance in *EGFR*-mutated lung cancer

Authors: 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]

Affiliations:

[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.

*e-mail: shibahara.daisuke.805@m.kyushu-u.ac.jp

Abstract: Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) provide therapeutic benefit in *EGFR* mutation–positive non–small cell lung cancer, but some individuals develop early resistance. We performed spatial transcriptomics analysis of pre- and posttreatment tumor samples from the same patients to explore the underlying mechanisms of such early resistance. The proportion and activation of fibroblasts increased in association with the development of early resistance, whereas a distinct tumor cell cluster showed activation of tumor necrosis factor–α signaling via the NF-κB pathway even before treatment. Also in the early resistance sample, specific tumor cell clusters interacted with immune and stromal cells. Immature tertiary lymphoid structures (TLSs) were enriched in the early resistance sample, whereas mature TLSs were observed in the long-term response sample. These findings implicate tumor heterogeneity and an inflammatory tumor microenvironment in early EGFR-TKI resistance, providing insight into potential therapeutic strategies to improve treatment outcomes.

Introduction

Epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) show durable therapeutic efficacy in most individuals with non–small cell lung cancer (NSCLC) harboring activating mutations of the *EGFR* gene[1]. Nevertheless, resistance to EGFR-TKIs inevitably emerges in such patients after treatment initiation[2,3]. Whereas several mechanisms of resistance to EGFR-TKIs have been identified, the underlying cause of resistance remains unknown in 40% to 50% of cases[4–6]. Furthermore, some patients develop early resistance during EGFR-TKI treatment, whereas others achieve a long-term response, with the reasons for these divergent outcomes also remaining unclear. There is therefore a pressing need to elucidate the precise mechanisms of EGFR-TKI resistance not only for the development of effective therapies to overcome such resistance but also to extend the duration of the response to EGFR-TKI treatment. Increasing evidence has suggested that the tumor microenvironment (TME) plays a pivotal role in shaping the response to EGFR-TKIs[7–12]. However, spatial analyses of the TME based on paired pre- and posttreatment tumor specimens from the same patients with *EGFR*-mutated NSCLC have been limited[13].

Spatial transcriptomics analysis has emerged as a powerful approach to characterization of the TME[13–18]. As an array-based method, Visium HD (10× Genomics) allows both comprehensive mRNA capture, without limiting the analysis to specific target genes, as well as analysis of an entire specimen section. We have now performed a detailed analysis with Visium HD of the TME in paired pre- and posttreatment tumor specimens from patients with *EGFR*-mutated NSCLC. Furthermore, we converted Visium HD spot data into single-cell data with the use of StarDist, a deep learning–based method for detection and segmentation of nuclei and cells. By

integrating the spatial coordinates derived from StarDist into the single-cell data, we were able to conduct spatial transcriptomics analysis at the single-cell level while preserving the actual spatial relations between cells. Our findings provide insight into the spatial dynamics of the TME, highlighting a contribution of an inflammatory TME and tumor heterogeneity to the early onset of EGFR-TKI resistance in *EGFR*-mutated NSCLC.

Results

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 posttreatment tumor specimens from four individuals with *EGFR*-mutated NSCLC (Fig. 1A and fig S1). These individuals were categorized as one with early resistance (ER), one with a standard response (SR), and two with a long-term response (LR) on the basis of progression-free survival (PFS)[2], and their clinical characteristics are provided in table S1. All samples were processed with the Visium HD platform, which provides a spatial resolution of 2 µm per spot. To reconstruct single cells from the data, we applied StarDist[19], a deep learning–based tool that allows the detection and segmentation of nuclei from hematoxylin-eosin (H&E)–stained images generated alongside the spatial transcriptomics data (Fig. 1B and fig. S2). With the use of 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). In addition, 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 (fig. S3, A to C).

A total of 814,636,020 reads were 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 on the basis of coordinates for their location on the image, and the sample names were linked to the data (fig. S3D). The integrated data for the eight samples were then processed for quality control, filtering, normalization, dimensionality reduction, batch correction, nearest-neighbor graph building, and Leiden clustering, and they were visualized by uniform manifold approximation and projection (UMAP). Cells were broadly categorized into three main lineages—epithelial cells, stromal cells, and immune cells—together with a small number of blood cells (Fig. 1D and Fig. S3E). The epithelial cell population was further divided into tumor cells, basal cells, club cells, alveolar cells type 2, and goblet/ciliated cells (Fig. 1E and Fig. S4A). Immune and stromal cells were clustered into lymphocytes, plasmablasts, macrophages, fibroblasts, endothelial cells, and smooth muscle cells (Fig. 1, F and G, and fig. S4, B and C). In addition, lymphocytes were subdivided by type, plasmablasts were classified by the immunoglobin (Ig) they produced, and macrophages were distinguished from mixed cells (Fig. S4, D to I). The results of these clustering analyses were integrated with the polygonal coordinate data obtained from StarDist, thereby allowing visualization of distinct cell populations on spatial coordinates (Fig. 1, H to K, and fig. S5 and S6).

Changes in the TME between pre- and posttreatment samples

To investigate changes in tumor cell characteristics between before and after treatment, we performed differential gene expression (DGE) analysis following pseudo–bulk processing, comparing tumor cells from pre- and posttreatment samples (Fig. 2A). Genes with significantly higher expression in the pretreatment group showed a tendency to be associated with a favorable prognosis for NSCLC patients in the Kaplan-Meier Plotter database (fig. S7, A to D), whereas those with higher expression in the posttreatment group tended to be associated with a poorer prognosis (fig. S7. E to H). Gene set enrichment analysis (GSEA) revealed significant enrichment of gene sets associated with the cell cycle and cell proliferation in posttreatment tumor cells (Fig. 2, B and C, and fig. S7. I and J). These results suggested that tumor cells had an enhanced proliferative activity after the development of EGFR-TKI resistance.

We next compared the proportions of stromal and immune cell clusters between pre- and posttreatment samples. The proportion of stromal cells, including fibroblasts and endothelial cells, was increased in the ER_post sample compared with the ER_pre sample (Fig. 2D). DGE analysis of fibroblasts from the ER samples revealed increased expression of collagen-related genes after treatment. (Fig. 2E). GSEA further revealed that a gene set related to extracellular matrix productionwas significantly enriched in fibroblasts of the ER_post sample (Fig. 2F). In addition, collagen gene expression in fibroblasts was significantly increased for the ER_post sample compared with all other samples (Fig. 2G and fig. S7K). These results suggested that the activity of fibroblasts was increased after EGFR-TKI treatment, contributing to the changes in the TME associated with early resistance.

Analysis of the T cell cluster revealed that CD4-expressing T cells were present in slightly higher numbers than CD8-expressing or regulatory T cells, although low

expression levels of *CD4*, *CD8A*, and *FOXP3* impeded the classification of these cells (Fig. 2H and fig. S7L), possibly as a result of the filtering effect of the analytic methods applied. For 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], suggestive of an immune-suppressive phenotype that was consistently observed across all samples (Fig. 2I and fig. S7M).

Spatial distribution of tumor cell heterogeneity associated with early resistance or a longer-term response

To explore factors that might contribute to early resistance, we categorized tumor cells from all pretreatment samples into two groups: ER_pre and SLR_pre (the latter comprising SR_pre, LR1_pre, and LR2_pre) for DGE analysis (Fig. 3A). GSEA revealed significant enrichment of gene sets related to the inflammatory response, tumor necrosis factor–α (TNF-α) signaling via the nuclear factor–κB (NF-κB) pathway, the interleukin-6 (IL-6) pathway, and chemokine signaling pathways in ER_pre tumor cells (Fig. 3, B and C, and fig. S8, A and B), indicating that the TME of ER_pre was characterized by active inflammation.

To better understand the factors that contribute to the differences between samples, we performed a comparative analysis of tumor cell clusters. We classified and labeled each tumor cell cluster on the basis of DGE for cancer-related genes[23–28] as discriminatory markers (Fig. 3, D and E, and fig. S8. C and D). Application of inferCNV for analysis of copy number variation (CNV) revealed a tendency for consistent CNV alterations in each tumor cell cluster in comparison with other normal epithelial cell clusters as a reference group (fig. S8E). To quantify their distribution, we

visualized the tumor cell clusters across samples and mapped them spatially for each sample (Fig. 3, F and G). The tc_S100A6 and tc_AQP1 clusters were present in all pre- and posttreatment samples, accounting for a high proportion of tumor cells in most samples. The tc_S100A6 cluster manifested significantly higher expression of *CCND1* and *ERBB2* compared with the other clusters (fig. S8, F anf G), whereas the tc_AQP1 and tc_TMSB4X clusters showed increased expression of *MET* and *CD74*, respectively (fig. S8, H and 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 the G2-M checkpoint, compared with the other clusters (Fig. 4, A and B). In contrast, gene sets related to the inflammatory response, epithelial-mesenchymal transition (EMT), and KRAS signaling were not enriched (Fig. 4C). Similar DGE analysis and GSEA for the tc_AQP1 cluster revealed upregulation of gene sets associated with mechanistic target of rapamycin complex 1 (mTORC1) and IL-6–JAK–STAT3 signaling, both of which have been implicated in EGFR-TKI resistance[4,29–31], in addition to those related to EMT and the inflammatory response (Fig. 4, D to F).

We next analyzed distinct clusters that showed differences between samples. The tc_TMSB4X cluster was most pronounced in the ER_pre sample (Fig. 3F), suggestive of a potential role in early resistance. DGE analysis and GSEA for this cluster revealed significant enrichment of gene sets associated with metabolic pathways, whereas those related to cell proliferation showed relatively low expression (Fig. 4, G to I). The tc_MMP7 cluster was prominent in both pre- and posttreatment ER samples but was essentially absent from the other samples (Fig. 3F), also suggestive of a potential role in early resistance. DGE analysis and GSEA for this cluster showed significant enrichment of pathways related to EMT, the inflammatory response, IL-6–JAK–STAT3

signaling, KRAS signaling, and TNF-α signaling via NF-κB (Fig. 4, J to L). The tc_SCGB3A1 cluster was most abundant in LR1_pre (Fig. 3F), possibly indicating a contribution to long-term response. DGE analysis and GSEA for this cluster revealed downregulation of gene sets related to E2F targets and the G2-M checkpoint, suggestive of a relatively slow cell cycle (fig. S9, A to C). The tc_AZGP1 cluster was characteristic of the SR_pre sample (Fig. 3F), but no significant enrichment of gene sets was observed for this cluster (fig. S9, D to F), possibly as a result of its extremely low cell number.

To investigate differences in the TME across tumor cell clusters, we quantified the spatial distribution of surrounding cell types within the ER_pre sample. For each tumor cell cluster, we calculated the proportion of cell types within concentric annular regions extending up to 1000 μm from each tumor cell (fig. S9G). 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). Of note, the tc_MMP7 cluster showed the highest proportion of immune cells within its surrounding microenvironment, consistent with the enrichment of genes associated with the inflammatory response in this cluster revealed by GSEA (Fig. 4L). Together, these findings suggested that this cluster is characterized by an inflammatory state.

Maturity of tertiary lymphoid structures in the TME

To examine the spatial relations of each cluster within the TME, we performed a neighborhood enrichment analysis based on the spatial map coordinates annotated with cell types. This analysis revealed a close spatial relation between T cells, B cells, and follicular dendritic cells (FDCs), particularly in the ER_pre and LR1_pre samples (Fig.

5A). Visualization of these immune cell populations on the spatial map revealed that they were distributed across several regions of the ER_pre sample, whereas only one region was observed in the LR1_pre sample (fig. S10, A and B). In the ER_pre sample, these regions were characterized by a structure in which B cells and FDCs were intermixed and sparsely surrounded by T cells (Fig. 5B and fig. S10C). In contrast, in the LR1_pre sample, FDCs were positioned around B cells, with T cells surrounding the outer perimeter (Fig. 5C and fig. S10D). Moreover, *CXCL13* expression was increased within these structures (Fig. 5, B and C), indicating that they corresponded to tertiary lymphoid structures (TLSs). On the basis of previous observations[32], we identified the poorly organized regions in the ER_pre sample as immature TLSs and the well-structured region in the LR1_pre sample as a mature TLS.

Enrichment of MIF- and PDGF-mediated cell-cell interactions in the TME associated with early resistance

To further elucidate the characteristics of the TME, we analyzed intercellular interactions with the use of COMMOT. This Python library calculates cell-cell interactions on the basis of 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 (table S2). Pathway comparisons across samples revealed increased expression of genes associated with the macrophage migration inhibitory factor (MIF) pathway in ER and SR samples (Fig. 6, A and B). MIF is a ligand of CD74-CD44 and CD74-CXCR4 complexes and regulates the inflammatory response[33]. To quantify ligand-receptor interactions for the MIF pathway, we calculated signal strength on the basis of the spatial proximity of ligand- and

receptor-expressing cells and their expression levels. Examination of the ER_pre and LR1_pre spatial signal vectors revealed that the signals on the spatial map were directed toward the regions containing the TLSs and inflammatory cell infiltration (Fig. 6, C and D). Signal intensity and direction were further analyzed by cell type (Fig. 6, E to G). Both ER_pre and LR1_pre samples manifested bidirectional signaling, predominantly among T cells, FDCs, and B cells. Tumor cells showed limited signals in the LR1_pre sample, whereas those in the ER_pre sample manifested signals directed toward macrophages and TLS-related cell types. Some tumor cell clusters showed activation of this pathway via an autocrine mechanism. These findings suggested that tumor cells actively stimulated macrophages and TLSs via the MIF pathway in the ER_pre sample.

Given that the proportion of fibroblasts and endothelial cells was increased after treatment for the ER samples (Fig. 2D), we next focused on the platelet-derived growth factor (PDGF) pathway, which is associated with fibroblast proliferation[34]. Signals were not localized to specific regions within the ER_pre and LR1_pre samples (Fig. 6, H and I). Bidirectional signaling was observed between fibroblasts and endothelial cells in both samples (Fig. 6, J to L). In the ER_pre sample, tumor cells, including tc_TMSB4X and tc_MMP7 clusters, showed signals directed toward stromal cells. In contrast, tumor cell clusters in the LR1_pre sample showed fewer signals directed at stromal cells. These results suggested that tumor cells interacted 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 identified, few studies have compared cases of early resistance with long-

term responders[3,35,36]. We have now shown that cell proliferation in residual tumors was increased after 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 such as tc_S100A6 and tc_AQP1 were detected both before and after osimertinib treatment, suggesting that they developed resistance through activation of downstream or alternative pathways. Indeed, genes related to resistance mechanisms such as *CCND1*, *ERBB2*, and *MET* were upregulated in these clusters, with the amplification of these genes having being associated with EGFR-TKI resistance[4]. In contrast to tc_S100A6 and tc_AQP1, the tc_TMSB4X cluster was detected predominantly in the ER_pre sample and manifested low cell cycle activity, suggestive of the association of a quiescent state with EGFR-TKI resistance[40]. We found that the MIF pathway was active in the ER samples, with the tc_TMSB4X cluster manifesting activation of this pathway via an autocrine mechanism. Of note, this cluster showed significantly higher *CD74* expression compared with the other clusters. *CD74* has previously been associated with resistance to EGFR-TKIs[41], suggesting that the tc_TMSB4X cluster may serve as a reservoir for resistant cells. Furthermore, this cluster showed increased expression of genes related to bile acid metabolism, potentially reflecting suppression of immune cell responses[42]. In contrast, the tc_SCGB3A1 cluster, which also manifested a slow cell cycle, was largely restricted to the LR1_pre sample, having disappeared after osimertinib treatment, suggestive of characteristics of low proliferation and low malignancy. The tc_MMP7 cluster was prominent in both ER_pre and ER_post samples and showed high expression of genes related to TNF-α signaling via NF-κB. Given that the activation of TNF-α signaling by osimertinib has been shown to contribute to EGFR-TKI resistance[43–45], such signaling

may constitute an early resistance mechanism. Furthermore, spatial distribution analysis revealed that the tc_MMP7 cluster was in an immunologically hot state due to a large number of plasmablasts infiltrating the surrounding area. A recent study found that $CXCR1^{+}$ neutrophils contribute to early EGFR-TKI resistance through TNF-α signaling via NF-κB[13]. However, in our analysis, the number of neutrophils was extremely low and thus not included in the assessment. Together, our findings reveal that tumor heterogeneity contributes to EGFR-TKI resistance.

Increasing evidence implicates interactions between tumor cells and surrounding stromal or immune cells in mechanisms of EGFR-TKI resistance[9,10,46]. Cancer-associated fibroblasts have been identified as key regulators of the TME[47]. Our comparative analysis revealed an increase in the proportion of stromal cells in the ER samples after osimertinib treatment, with no marked change apparent in the SR samples and a decrease apparent in the LR samples. Of note, the proportion of fibroblasts also showed a marked increase that was accompanied by increased extracellular matrix production and upregulation of collagen genes including *COL1A1* and *COL3A1* in the ER_post sample relative to the ER_pre sample, suggestive of a potential association with early resistance[31]. *PDGFRA* and *PDGFRB*, which encode receptors for PDGF signaling, are known markers of tumor-associated fibroblasts and regulate angiogenesis[47]. PDGF signaling was active between tumor cells and stromal cells in the ER_pre sample, suggesting that tumor cells interact with stromal cells via the PDGF pathway and thereby contribute to the formation of an early-resistant TME.

We found that genes related to the inflammatory response were significantly upregulated in tumor cells of the ER-pre sample compared with SLR-pre samples. We further showed that the TME of the ER-pre sample was characterized by the presence of

immature TLSs and infiltration of B cells and plasmablasts, whereas the TME of the LR1_pre sample manifested a mature TLS with reduced immune cell infiltration. Immature TLSs have been implicated in tumor promotion, whereas mature TLSs appear to be tumor suppressive[32,48,49]. Our data suggest that the extent of TLS maturation might be a determinant of early EGFR-TKI resistance. *EGFR*-mutated NSCLC is generally characterized by low immunogenicity, which is often described as an "immune cold" state and 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]. Of note, individuals with a short response to EGFR-TKIs have shown better responses to immunotherapy[55], suggesting that the TME of such tumors may be characterized by an inflammatory state. Our finding that genes related to the inflammatory response were enriched in the ER_pre sample is consistent with this notion. Together, these observations indicate that an inflammatory TME can be 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 issue by detecting and segmenting cell nuclei, reconstructing read count data and cell coordinate data to improve data reliability. Whereas 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 have become possible and the accuracy of single-cell data and spatial coordinates has improved. For further improvements in precision, multi-omics approaches are required, however.

Our study has several limitations. First, the size of the bronchoscopic lung biopsy specimens was small compared with 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 between before and after treatment. Second, enhancement of 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 are warranted to collect additional clinical data for validation of our findings.

In conclusion, our study suggests that an inflammatory TME and tumor heterogeneity contribute to early EGFR-TKI resistance in *EGFR*-mutated NSCLC. Of note, upregulation of resistance-related pathways was apparent in certain tumor cell clusters even before treatment onset.

Methods

Sample preparation

Formalin-fixed paraffin-embedded (FFPE) block specimens were collected from four *EGFR*-mutated NSCLC patients both before and after treatment with osimertinib at multiple institutions. The biopsy methods for the specimens were surgical biopsy, bronchoscopic biopsy, surgical pleural biopsy, and single-hole thoracoscopic pleural biopsy. Clinical data for the patients are provided in table S1. The patients were classified as showing early resistance (ER, $n = 1$), a standard response (SR, $n = 1$), or a long-term response (LR, $n = 2$) on the basis of the duration of PFS. The initial specimens were obtained to determine treatment. Written informed consent was obtained from all patients for specimen collection and research use. This study was

performed in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of Kyushu University Hospital (approval no. 23297-01).

Visium HD data generation

Two capture areas of a Visium HD slide (10× Genomics) were used, with four samples being assigned to each capture area. A total of eight samples consisting of paired pre- and posttreatment specimens from four patients (ER_pre, ER_post, SR_pre, SR_post, LR1_pre, LR1_post, LR2_pre, LR2_post) was analyzed, with ER and SR samples being assigned to one capture area and LR1 and LR2 samples to the other. RNA was extracted from the FFPE tissue sections for RNA quality assessment, and the DV200 (percentage of RNA fragments of >200 nucleotides) value was measured with the Agilent 4200 TapeStation instrument. The DV200 value for all samples was ≥30%. FFPE tissue sections were then depleted of paraffin and subjected to H&E staining according to the Visium HD FFPE Tissue Preparation protocol (Visium HD FFPE Tissue Preparation Handbook, CG000684 Rev A). The tissue sections were imaged with an Evident VS200 microscope in accordance with the 10× Genomics protocol (Visium HD Spatial Applications Image Guidelines CG000688 Rev A). Gene expression libraries were prepared from the H&E-stained sections according to the 10× Genomics protocol (Visium HD Spatial Gene Expression Reagent Kits User Guide CG000685 Rev A). The quality of the prepared libraries was evaluated with the Agilent 4200 TapeStation instrument, and all libraries were confirmed to have a sufficient concentration. The Visium libraries were sequenced with the Illumina NovaSeq X Plus platform with a paired-end configuration (read 1, 150 bp; read 2, 150 bp). Sequencing data (FASTQ

files) were processed and analyzed with Space Ranger (version 3.0.0) software for assessment of data quality and generation of spatial gene expression profiles.

Analysis environment

Analyses were conducted with 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). For some analyses, R Script was executed in Python. The primary R packages and their versions used in this study were tradeSeq (1.8.0), edgeR (3.36.0), and clusterExperiment (2.14.0). Whereas the libraries were generally applied as described in their documentation, modifications were made to Commot library to improve visual clarity in plots.

Detection and segmentation of nuclei

Although Visium HD allows for the use of 8-μm square spots as single cells, these spots do not accurately reflect true cellular coordinates. For preservation of cellular coordinates as much as possible, StarDist was applied for detection and segmentation of nuclei[19]. Specifically, Python code was developed on the basis of the Analysis Guide provided by 10× Genomics (https://www.10xgenomics.com/jp/analysis-guides/segmentation-visium-hd). Running of this code generated coordinate data for individual nuclei represented as polygons. Count data for each 2-μm spot within the

polygon area were relabeled as corresponding to a single cell. In addition, the centroid coordinates of each polygon were added to the count data, together with metadata content, thereby generating a data structure compatible with libraries commonly used in spatial transcriptomics analysis.

Sample labeling on spatial coordinates

In typical spatial transcriptomics analysis, all samples within a single capture area are assumed to be the same. However, in the present study, four different samples were present within each capture area. Given that this situation is not typically supported by existing tools, it was necessary to label samples manually on the basis of spatial coordinates. With the use of the spatial coordinate data within the data structure, samples were divided into eight groups by referencing images and assigning coordinates manually. Given that there was an overlapping area between LR1_pre and LR1_post specimens, a blank area was generated to avoid sample contamination.

Major clustering

The newly generated data from nuclei segmentation was analyzed with a pipeline similar to that for single-cell RNA-seq analysis with the Scanpy package. Cells with <20 counts or containing >10% mitochondrial genes as well as genes with <20 total counts were excluded from the analysis. Cells with a cross-sectional area of ≥900 μm^2 on the image were excluded given the possibility that they might actually have been composed of multiple cells. After filtering, the data were calculated for highly variable genes (HVGs) and normalized with Pearson residuals, which adjusts for library size and overdispersion. Principal component analysis (PCA) for the top 4000 HVGs allowed

calculation of principal components up to the top 35. Harmony integration was applied with the use of the HarmonyPy library through Scanpy with the sample labels in order to account for batch effects across samples. The nearest-neighbors distance matrix and the neighborhood graph of the observations were then calculated with a local neighborhood size of 100 and 35 principal components. An 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, given the high noise level in cells, it was possible that the differences between cells and noise would not be distinguishable. Instead, we analyzed the expression of HVGs in each cluster and compared the expression levels of known cell markers to classify the clusters into three major cell types: epithelial, stromal, and immune (table S3).

Subcluster analysis

Subclustering was performed for further exploration of the epithelial, stromal, and immune cell populations identified in the initial clustering. For each cell type cluster, the analysis was started again from the raw count data. The data were filtered to the same threshold as for the major clustering. HVGs were calculated within some subclusters and PCA was performed, followed by the construction of neighborhood graphs, UMAP embedding, and clustering. The distinction between tumor cells and normal cells was evaluated with inferCNVpy, which is based on inferCNV, in addition to the pathological images. Normal airway epithelial cells obtained from the clustering results were set as nonmalignant cells and analyzed. Furthermore, additional clustering was performed for tumor cells. 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 tables S3 and S4.

DGE Analysis

DGE analysis was performed with generated pseudo–bulk samples and the edgeR package[56,57]. Single-cell RNA-seq data were aggregated at the sample level for each cell type of interest. Samples with <30 cells of the target cell type were excluded. For comparisons between ER and SLR groups as well pre- versus posttreatment comparisons of the tc_AZGP1 cluster, two random replicates were generated for each group because the edgeR library requires biological replicates for statistical modeling. For other comparisons, pseudo–bulk aggregation was performed without generation of additional replicates. After pseudo–bulk aggregation, DGE analysis was performed with the edgeR library[57]. Low-expression genes were first filtered out, retaining only those with sufficient counts across experimental groups to improve statistical power. Library size normalization was performed with 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, allowing the estimation of both biological variation and group-specific effects. The quasi–likelihood variance estimation method was adopted to generate a gene list for DGE. Adjusted *P* values were calculated with the Benjamini-Hochberg method to control the false discovery rate (FDR). 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 construction of volcano plots and GSEA.

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[58]. Patients in the database were divided into high- and low-expression groups on the basis of median expression values, and overall survival curves were plotted with no follow-up restrictions.

Integration of cell type and spatial coordinate data

The data set for this analysis was generated by integration of single-cell analysis for cell type identification with spatial coordinate data derived from StarDist. The StarDist-based coordinate data, which represent new spot coordinates, were transformed on the basis of 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 eight sample data sets, the spatial neighbors function within Squidpy was used to calculate the spatial neighbors of each spot. The neighborhood enrichment function was then applied to calculate the z-scores with predefined cell cluster labels. For standardization of the enrichment scores across all samples, the minimum and maximum values were first calculated across all data sets. 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 relations of tumor cells were analyzed by integration of single-cell cluster information with spatial coordinate data. For each of the eight sample data sets, the spatial positions of tumor cells were identified. The composition of surrounding cells was quantified by calculation of the pairwise distances between each tumor cell and all other cell types within the tissue. For analysis of the spatial distribution, concentric annular regions extending up to 1000 μm from each tumor cell were defined, with each region increasing in 20-μm increments. These regions served as distance bins, allowing categorization of surrounding cells on the basis of their proximity to tumor cells. The proportion of each cell type within these bins was then calculated to assess the spatial distribution relative to tumor clusters. This method allowed a quantitative evaluation of neighborhood enrichment patterns, providing insight into the TME and its cellular composition.

Cell-cell communication analysis

Cell-cell interactions were analyzed with COMMOT, a tool that allows comprehensive modeling of both direct and indirect interactions between different cell types while incorporating the spatial organization of cells within tissues[59]. Smooth muscle cells, normal airway epithelial cells, and blood cells were excluded from this analysis to allow

focus on the TME. In addition, genes expressed in ≥0.5% of all cells were first filtered. The gene list was further refined by extraction of 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 on the basis of the distance between sending and receiving cells as well as gene expression intensity.

Gene expression analysis and statistics

Gene expression scores were calculated by a 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 with control gene sets of 5000 randomly selected genes to ensure robust scoring. Pairwise comparisons of the MIF score and of gene expression in tumor cells between samples were performed with the unilateral Mann-Whitney U test. Visualization was conducted with Seaborn for violin and strip plots annotated with the *P* value. The significance level for GSEA after DGE analysis was set as a *P* value of <0.05 and FDR of 0.25.

Data availability

All data associated with this study are in the paper or the Supplementary Materials. The Visium HD data set for this study is available in the GEO repository under accession number GSE288758.

Acknowledgments

We thank CyberomiX (Kyoto, Japan) for assistance with spatial transcriptomics analysis via Visium HD platform.

Author contributions

D.S., K.T., and I.O. made substantial contributions to the conception of the work. Y.K., K.N., N.N., M.K., and K.A. provided samples and clinical information for this work. S.N., D.S., K.T, M.H., R.I., K.O., Y.Y., E.I., and I.O. made significant contributions to data interpretation. S.N. and D.S. analyzed the data and drafted the original manuscript. S.N. and D.S. reviewed and edited the manuscript. Y.O. and I.O. supervised the study.

Competing interests

D.S. has received honoraria from Eli Lilly Japan K.K. K.A. has received honoraria from AstraZeneca, MSD, Chugai Pharmaceutical, Ono Pharmaceutical, Bristol-Myers Squibb, Takeda Pharmaceutical, Taiho Pharmaceutical, and Amgen. Y.Y. has received honoraria from AstraZeneca. K.T. 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. E.I. has received honoraria from AstraZeneca and Chugai Pharmaceutical. I.O. 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. All other authors declare no competing interests.

References

1. Ramalingam, SS *et al.* Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *New England Journal of Medicine* **382,** 41–50 (2020).
2. Soria, J-C *et al.* Osimertinib in Untreated EGFR-Mutated Advanced Non–Small-Cell Lung Cancer. *New England Journal of Medicine* **378,** 113–125 (2018).
3. Oxnard, GR *et al.* Assessment of Resistance Mechanisms and Clinical Implications in Patients With EGFR T790M–Positive Lung Cancer and Acquired Resistance to Osimertinib. *JAMA Oncol* **4,** (2018).
4. Leonetti, A *et al.* Resistance mechanisms to osimertinib in EGFR-mutated non-small cell lung cancer. *Br J Cancer* **121,** 725–737 (2019).
5. Maynard, A *et al.* Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. *Cell* **182,** 1232-1251.e22 (2020).
6. Oh, S *et al.* Transcriptomic Heterogeneity of EGFR-Mutant Non–Small Cell Lung Cancer Evolution Toward Small-Cell Lung Cancer. *Clinical Cancer Research* **30,** 4729–4742 (2024).
7. Madeddu, C *et al.* EGFR-Mutated Non-Small Cell Lung Cancer and Resistance to Immunotherapy: Role of the Tumor Microenvironment. *Int J Mol Sci* **23,** 6489 (2022).
8. Chen, S *et al.* Changes of tumor microenvironment in non-small cell lung cancer after TKI treatments. *Front. Immunol.* **14,** (2023).

9. Yang, L *et al.* Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer* **10,** e003534 (2022).
10. Matsumoto, Y *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* **110,** 3244–3254 (2019).
11. Yang, C-Y *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* **124,** 110–122 (2020).
12. Hsu, K-H *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* **12,** 9753 (2022).
13. Wang, H *et al.* CXCR1+ neutrophil infiltration orchestrates response to third-generation EGFR-TKI in EGFR mutant non-small-cell lung cancer. *Sig Transduct Target Ther* **9,** 1–15 (2024).
14. Zhang, H *et al.* Spatial Positioning of Immune Hotspots Reflects the Interplay between B and T Cells in Lung Squamous Cell Carcinoma. *Cancer Res* **83,** 1410–1425 (2023).
15. De Zuani, M *et al.* Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat Commun* **15,** 4388 (2024).
16. Peng, H *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* **13,** e1155 (2023).

17. Wang, S *et al.* Features of tumor-microenvironment images predict targeted therapy survival benefit in patients with EGFR-mutant lung cancer. *J Clin Invest* **133,** e160330
18. Dagogo-Jack, I *et al.* B-Cell Infiltrate in the Tumor Microenvironment Is Associated With Improved Survival in Resected Lung Adenocarcinoma. *JTO Clinical and Research Reports* **4,** (2023).
19. Schmidt, U, Weigert, M, Broaddus, C & Myers, G. Cell Detection with Star-Convex Polygons. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds Frangi, AF et al.) 265–273 (Springer International Publishing, 2018). 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* **22,** 231–237 (2010).
21. Matsubara, E *et al.* The Significance of SPP1 in Lung Cancers and Its Impact as a Marker for Protumor Tumor-Associated Macrophages. *Cancers (Basel)* **15,** 2250 (2023).
22. Song, Y *et al.* Complement C1q induces the M2-polarization of tumor-associated macrophages in lung adenocarcinoma. *Genes Dis* **11,** 101093 (2023).
23. Wei, X & Dong, J. Aquaporin 1 promotes the proliferation and migration of lung cancer cell in vitro. *Oncology Reports* **34,** 1440–1448 (2015).
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* **11,** 4634–4649 (2019).
25. Yang, Z *et al.* TMSB4X: A novel prognostic marker for non-small cell lung cancer. *Heliyon* **9,** e21505 (2023).

26. Yamada, K *et al.* MMP-7 expression is associated with a higher rate of tumor spread through air spaces in resected lung adenocarcinomas. *Lung Cancer* **175,** 125–130 (2023).
27. Mazumdar, J *et al.* HIF-2α deletion promotes Kras-driven lung tumor development. *Proc Natl Acad Sci U S A* **107,** 14182–14187 (2010).
28. Albertus, DL *et al.* AZGP1 Autoantibody Predicts Survival and Histone Deacetylase Inhibitors Increase Expression in Lung Adenocarcinoma. *Journal of Thoracic Oncology* **3,** 1236–1244 (2008).
29. Li, L *et al.* Ibrutinib reverses IL-6-induced osimertinib resistance through inhibition of Laminin α5/FAK signaling. *Commun Biol* **5,** 1–14 (2022).
30. Fang, 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* **9,** 1258–1267 (2020).
31. Yamazaki, S *et al.* Collagen type I induces EGFR□TKI resistance in EGFR□mutated cancer cells by mTOR activation through Akt□independent pathway. *Cancer Sci* **109,** 2063–2073 (2018).
32. Bruno, TC. B cells to the forefront of immunotherapy. *Nature* **577,** 474–476 (2020).
33. Calandra, T & Roger, T. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat Rev Immunol* **3,** 791–800 (2003).
34. Yao, L *et al.* Temporal control of PDGFRα regulates the fibroblast-to-myofibroblast transition in wound healing. *Cell Rep* **40,** 111192 (2022).
35. Thress, KS *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* **12,** 1588–1594 (2017).

36. Isomoto, K *et al.* Impact of EGFR-TKI Treatment on the Tumor Immune Microenvironment in *EGFR* Mutation–Positive Non–Small Cell Lung Cancer. *Clinical Cancer Research* **26,** 2037–2046 (2020).
37. Jiang, C *et al.* High CEP55 expression is associated with poor prognosis in non-small-cell lung cancer. *Onco Targets Ther* **11,** 4979–4990 (2018).
38. Luo, A *et al.* The clinical significance of FSCN1 in non-small cell lung cancer. *Biomedicine & Pharmacotherapy* **73,** 75–79 (2015).
39. Li, Y *et al.* Cyclin B1 expression as an independent prognostic factor for lung adenocarcinoma and its potential pathways. *Oncol Lett* **24,** 441 (2022).
40. Lindell, E, Zhong, L & Zhang, X. Quiescent Cancer Cells—A Potential Therapeutic Target to Overcome Tumor Resistance and Relapse. *Int J Mol Sci* **24,** 3762 (2023).
41. Kashima, Y *et al.* Single-cell analyses reveal diverse mechanisms of resistance to EGFR tyrosine kinase inhibitors in lung cancer. *Cancer Res* **81,** 4835–4848 (2021).
42. Varanasi, SK *et al.* Bile acid synthesis impedes tumor-specific T cell responses during liver cancer. *Science* **387,** 192–201 (2025).
43. Gong, K *et al.* TNF-driven adaptive response mediates resistance to EGFR inhibition in lung cancer. *J Clin Invest* **128,** 2500–2518
44. Nam, YW *et al.* EGFR inhibits TNF-α-mediated pathway by phosphorylating TNFR1 at tyrosine 360 and 401. *Cell Death Differ* **31,** 1318–1332 (2024).
45. Ibusuki, R *et al.* TP53 gain-of-function mutations promote osimertinib resistance via TNF-α–NF-κB signaling in EGFR-mutated lung cancer. *NPJ Precis Oncol* **8,** 60 (2024).

46. Alsaed, B *et al.* Intratumor heterogeneity of EGFR expression mediates targeted therapy resistance and formation of drug tolerant microenvironment. *Nat Commun* **16,** 28 (2025).
47. Luo, H *et al.* Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. *Nat Commun* **13,** 6619 (2022).
48. Helmink, BA *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* **577,** 549–555 (2020).
49. Petitprez, F *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577,** 556–560 (2020).
50. Ukita, M *et al.* CXCL13-producing CD4+ T cells accumulate in the early phase of tertiary lymphoid structures in ovarian cancer. *JCI Insight* **7,** e157215
51. Germain, C *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* **189,** 832–844 (2014).
52. Cabrita, R *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577,** 561–565 (2020).
53. Laumont, CM *et al.* Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat Rev Cancer* **22,** 414–430 (2022).
54. Leong, TL & Bryant, VL. B cells in lung cancer—not just a bystander cell: a literature review. *Transl Lung Cancer Res* **10,** 2830–2841 (2021).
55. Yoshida, H *et al.* Nivolumab in non-small-cell lung cancer with EGFR mutation. *Annals of Oncology* **29,** 777–778 (2018).

56. Murphy, AE & Skene, NG. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nat Commun* **13,** 7851 (2022).

57. Robinson, MD, McCarthy, DJ & Smyth, GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

58. Győrffy, B. Transcriptome-level discovery of survival-associated biomarkers and therapy targets in non-small-cell lung cancer. *British Journal of Pharmacology* **181,** 362–374 (2024).

59. Cang, Z *et al.* Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* **20,** 218–228 (2023).

Figures

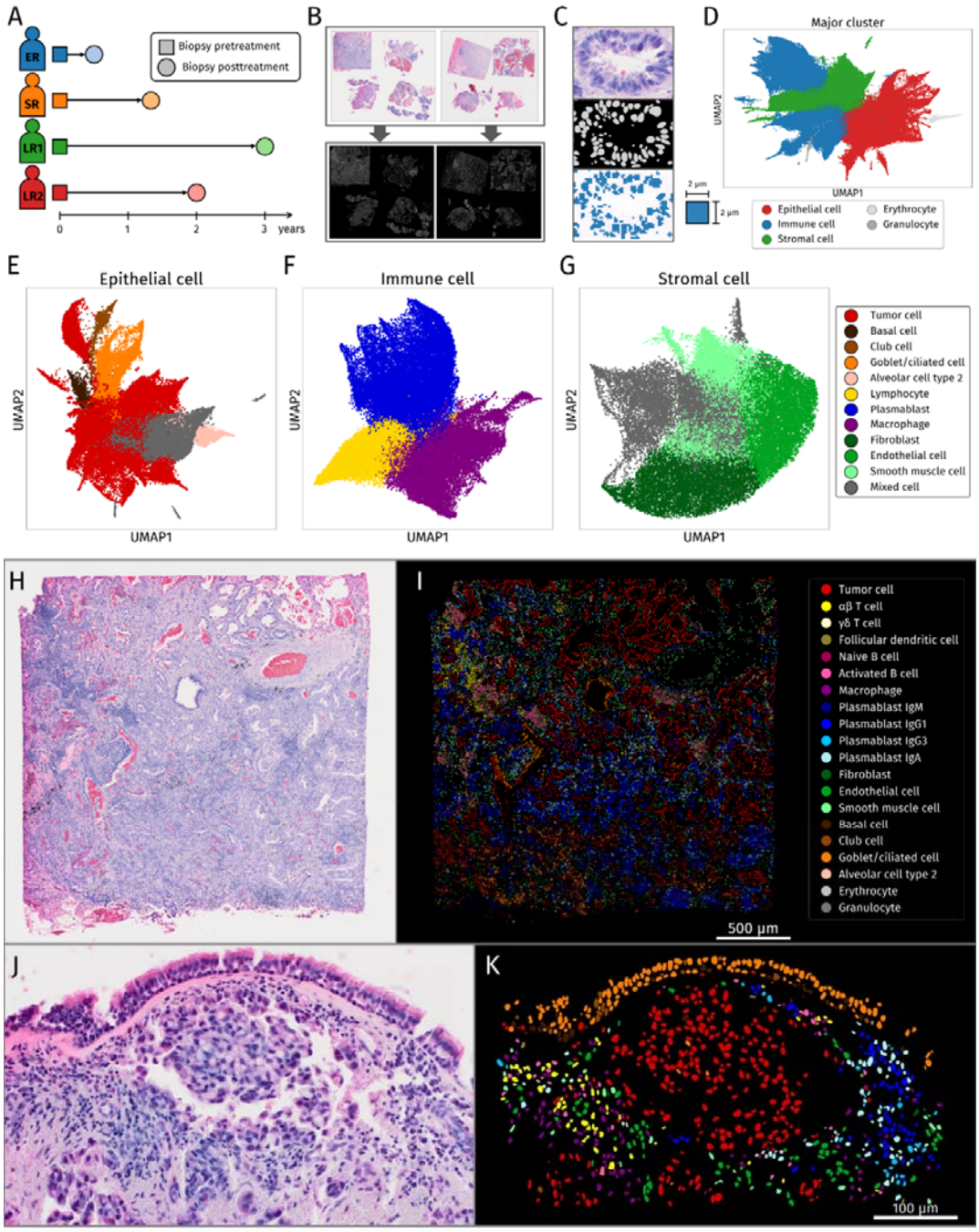


Fig. 1. Analysis overview, clustering, and visualization of cell type and spatial location. (**A**) Biopsy interval for pre− and post−osimertinib treatment and patient classification. (**B**) Detection and segmentation of cell nuclei with StarDist on high-

resolution tissue images loaded with the Visium HD protocol. (**C**) Visualization of detection and segmentation by StarDist and 2-μm spot extraction for identified cell nuclei positions. (**D**) UMAP of annotated cell types from single-cell RNA-seq analysis based on cell nuclei identified by StarDist. (**E** to **G**) UMAP with epithelial cells (E), immune cells (F), and stromal cells (G) reclustered and labeled for cell type. (**H** and **I**) H&E-stained pathological image of an area of the ER_pre sample (H) and image of the same area with cell type identification reflected in polygonal coordinate data obtained by StarDist (I). (**J** and **K**) H&E-stained pathological image for a magnified region of the LR2_pre sample (J) and image of the same region with cell type identification 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.

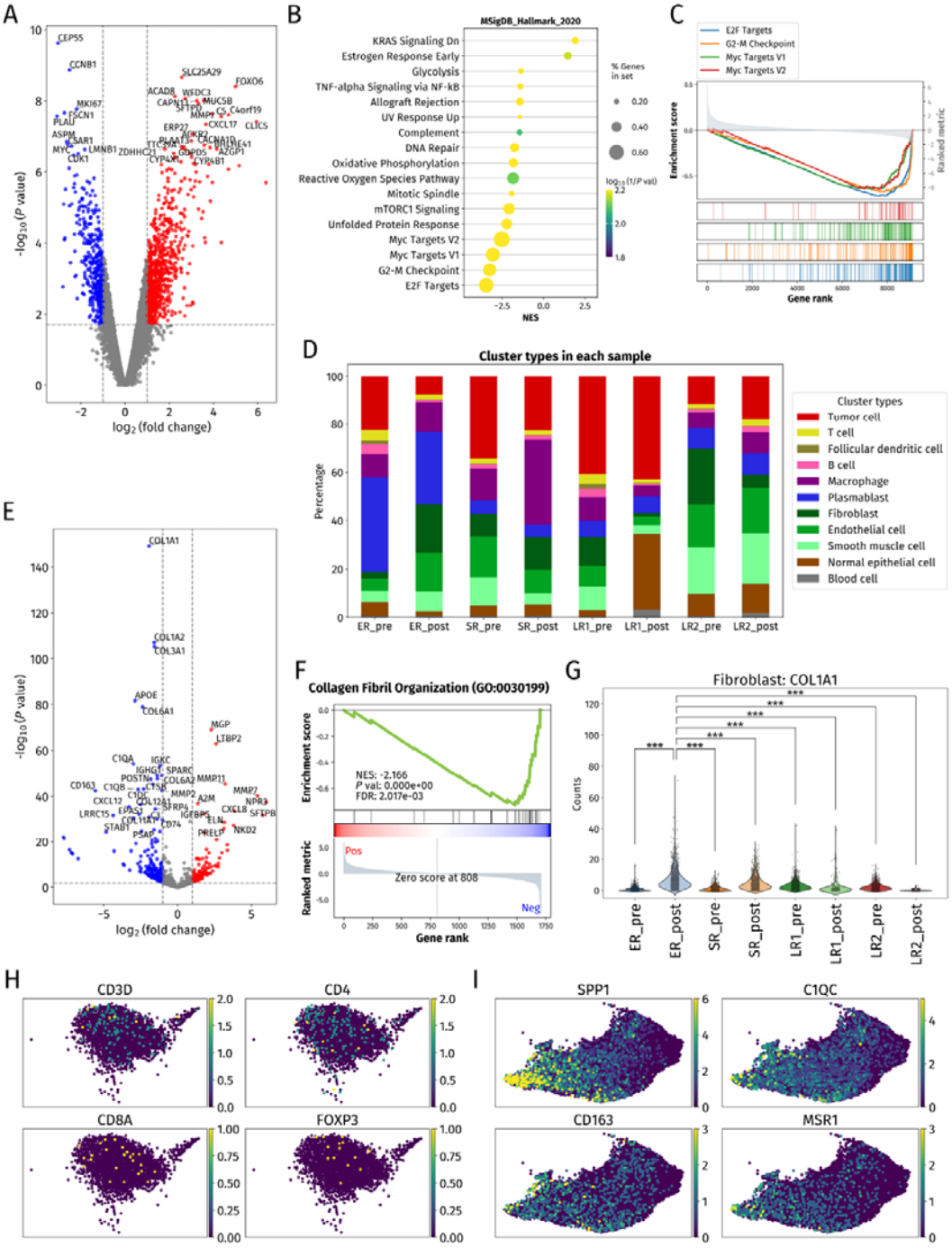


Fig. 2. Comparative analysis of samples obtained before and after treatment. (**A**) Volcano plot comparing tumor cell clusters from all samples dividing into two groups corresponding to before and after treatment (All_pre vs All_post). The red and blue dots

indicate significantly upregulated genes of All_pre and All_post, respectively. The vertical dashed lines represent the fold change of -2 and 2, and the horizontal dashed line represents the *P* value of 0.05. (**B**) Dot plot for GSEA with Hallmark gene sets based on DGE analysis of tumor cells for All_pre vs All_post. (**C**) GSEA plots for gene sets with high negative values of the normalized enrichment score (NES) for tumor cells in All_pre vs All_post. (**D**) Stacked bar chart showing the proportion of cell types within each sample. (**E**) Volcano plot for comparison of fibroblasts extracted from ER_pre and ER_post samples. (**F**) GSEA plot of the only significant gene set among Hallmark gene sets for fibroblasts in ER_pre vs ER_post. FDR, false discovery rate. (**G**) Violin plot and strip plot showing raw counts for the *COL1A1* gene in fibroblasts of each sample. ***$P < 0.0005$ (one-sided Mann-Whitney U test). (**H**) Expression levels of *CD3D*, *CD4*, *CD8A*, and *FOXP3* related to T cell subtypes shown on the UMAP plot of T cell parts extracted from lymphocyte clusters. (**I**) Expression levels of genes related to tumor-associated macrophages (*SPP1*, *C1QC*) and M2 macrophages (*CD163*, *MSR1*) on the UMAP plot of macrophages.

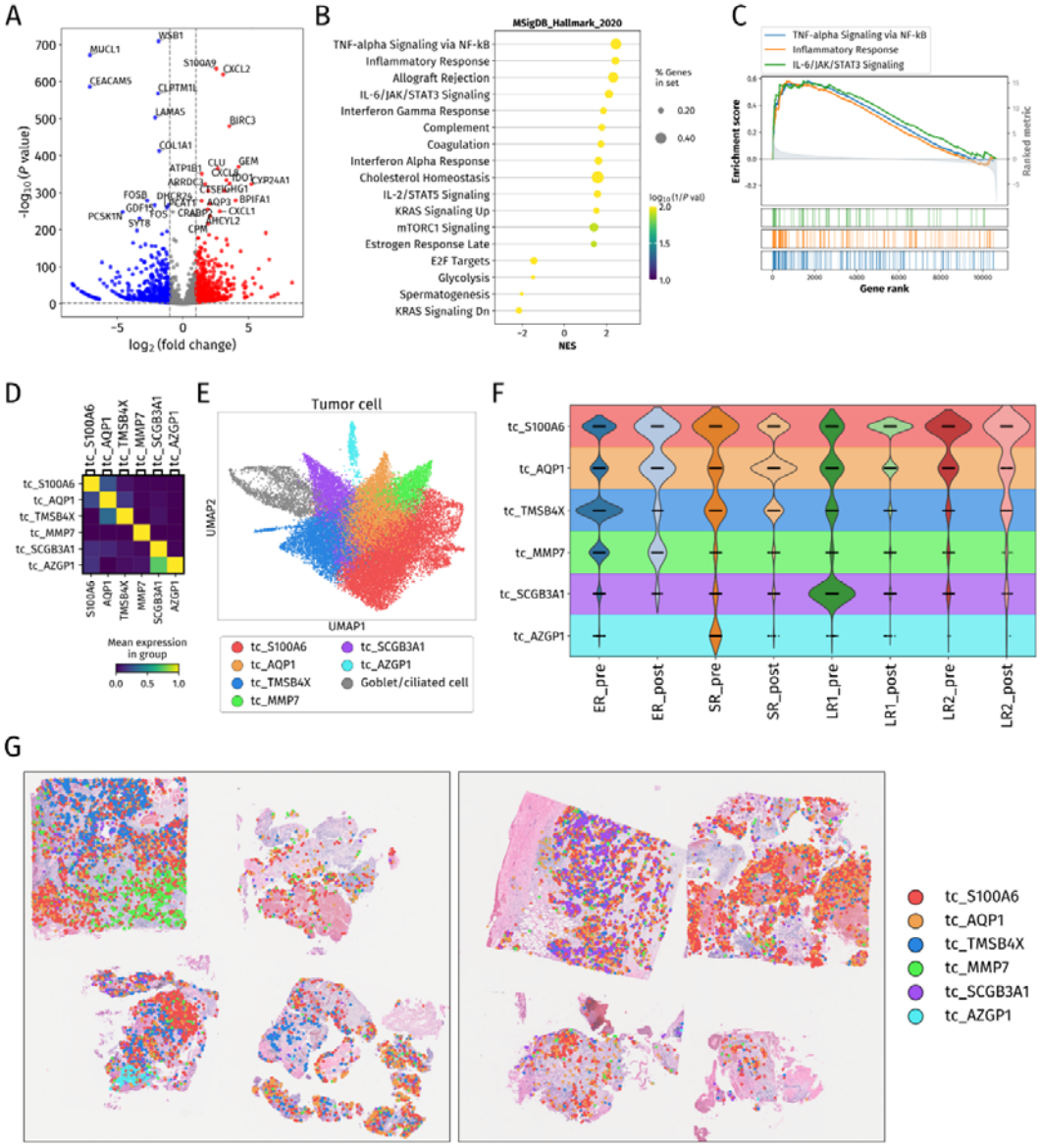


Fig. 3. Analysis exploring mechanisms underlying tumor heterogeneity and early resistance. (**A**) Volcano plot for DGE analysis results obtained by extracting tumor cell clusters from pretreatment samples and dividing them into two groups: ER_pre and SLR_pre. (**B**) Dot plot for GSEA with Hallmark gene sets based on DGE analysis of ER_pre vs SLR_pre. (**C**) GSEA plots for gene sets with high positive values of NES for ER_pre vs SLR_pre. (**D**) Matrix plot showing clustering of tumor cells based on

differentially expressed cancer-related genes. (**E**) UMAP plot with tumor cells reclustered and labeled for differentially expressed genes. (**F**) Violin plot for the distribution of tumor cell clusters in each sample (**G**) Image of tumor cell cluster positions plotted on H&E-stained tissue images for Slide 1 and Slide 2.

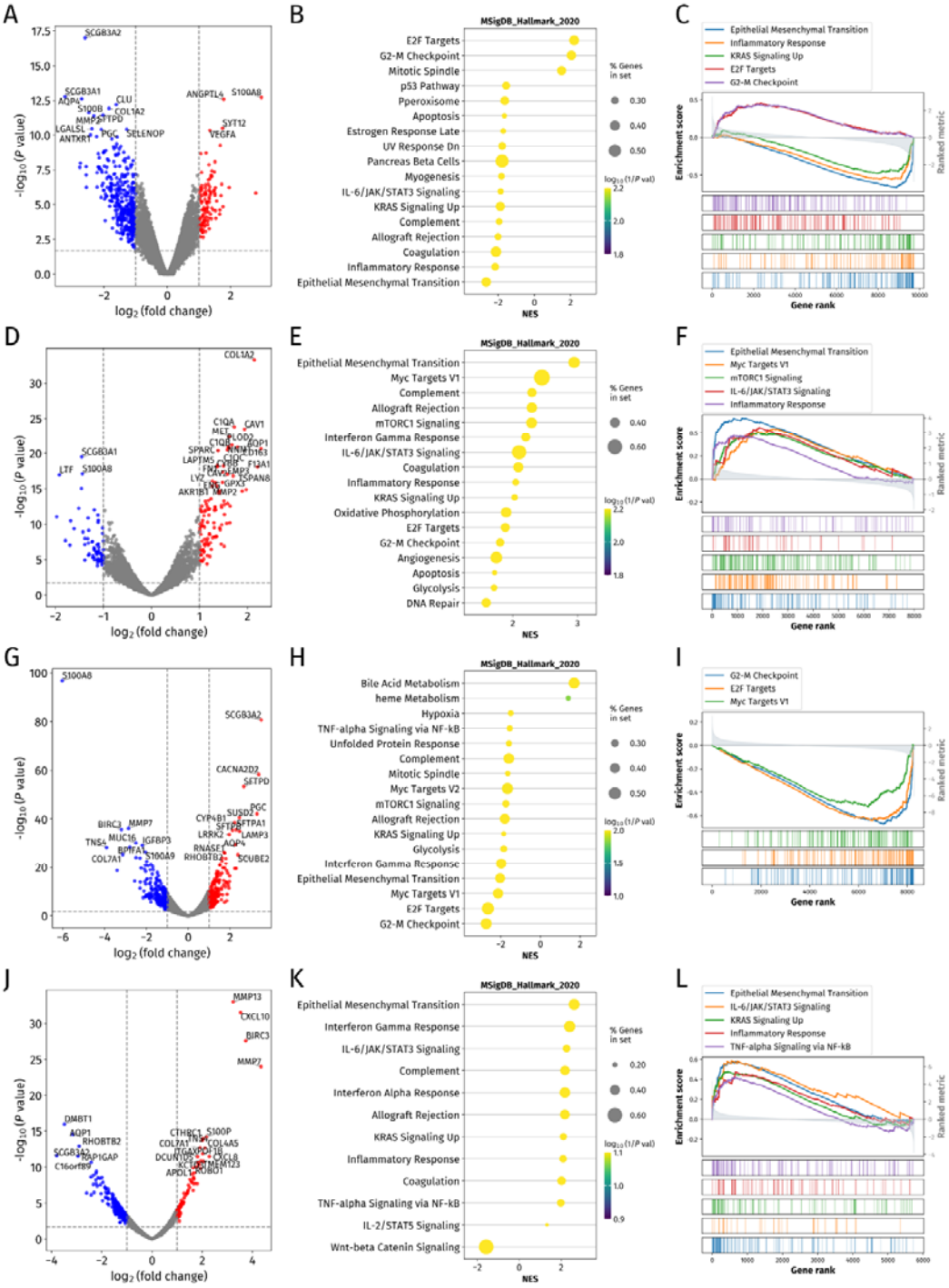


Fig. 4. Analysis of individual tumor cell clusters. (**A**) Volcano plot of DGE analysis results for the tc_S100A6 tumor cell cluster versus the corresponding other tumor cell clusters (tc_rest). (**B**) Dot plot for GSEA with Hallmark gene sets based on DGE

analysis of tc_S100A6 vs tc_rest. (**C**) GSEA plots for gene sets with high absolute values of NES for tc_S100A6 vs tc_rest. (**D**) Volcano plot of DGE analysis results for tc_AQP1 vs tc_rest. (**E**) Dot plot for GSEA with Hallmark gene sets based on DGE analysis of tc_AQP1 vs tc_rest. (**F**) GSEA plots for gene sets with high positive values of NES for tc_AQP1 vs tc_rest. (**G**) Volcano plot of DGE analysis results for tc_TMSB4X vs tc_rest. (**H**) Dot plot for GSEA with Hallmark gene sets based on DGE analysis of tc_TMSB4X vs tc_rest. (**I**) GSEA plots for gene sets with high negative values of NES for tc_TMSB4X vs tc_rest. (**J**) Volcano plot of DGE analysis results for tc_MMP7 vs tc_rest. (**K**) Dot plot for GSEA with Hallmark gene sets based on DGE analysis of tc_MMP7 vs tc_rest. (**L**) GSEA plots for gene sets with high positive values of NES for tc_MMP7 vs tc_rest.

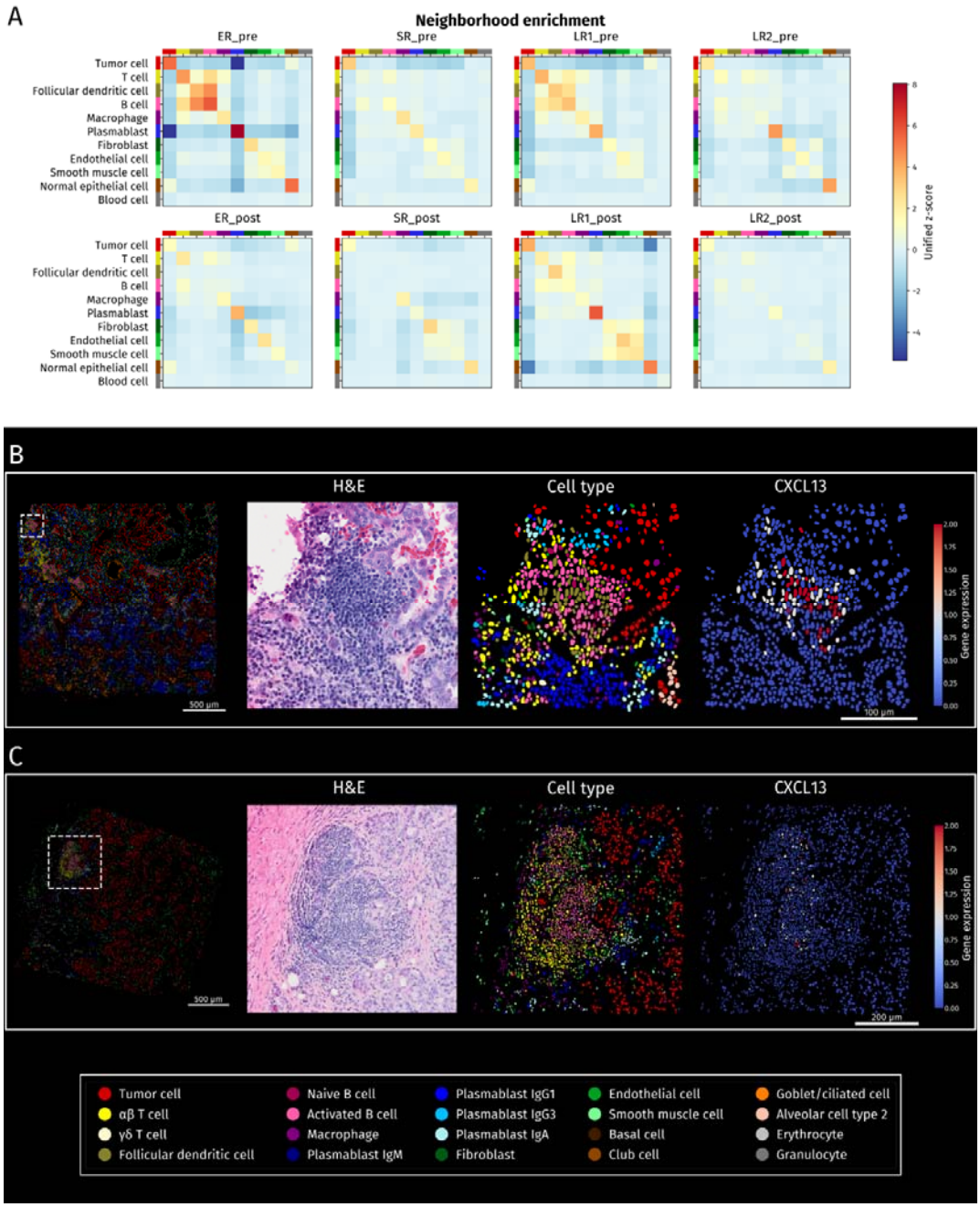


Fig. 5. Spatial neighborhood enrichment analysis and TLS identification. (**A**) Heat maps for the results of spatial neighborhood enrichment analysis for each sample. A total of 5000 surrounding cells was counted for each cell, and the proportion of each cell type among these cells was expressed as a z-score normalized relative to all samples. (**B**

and **C**) Enlarged images of a TLS in the ER_pre (B) and LR1_pre (C) samples. The area enclosed by the white dashed line is shown in the images of H&E staining, cell types, and expression of *CXCL13*.

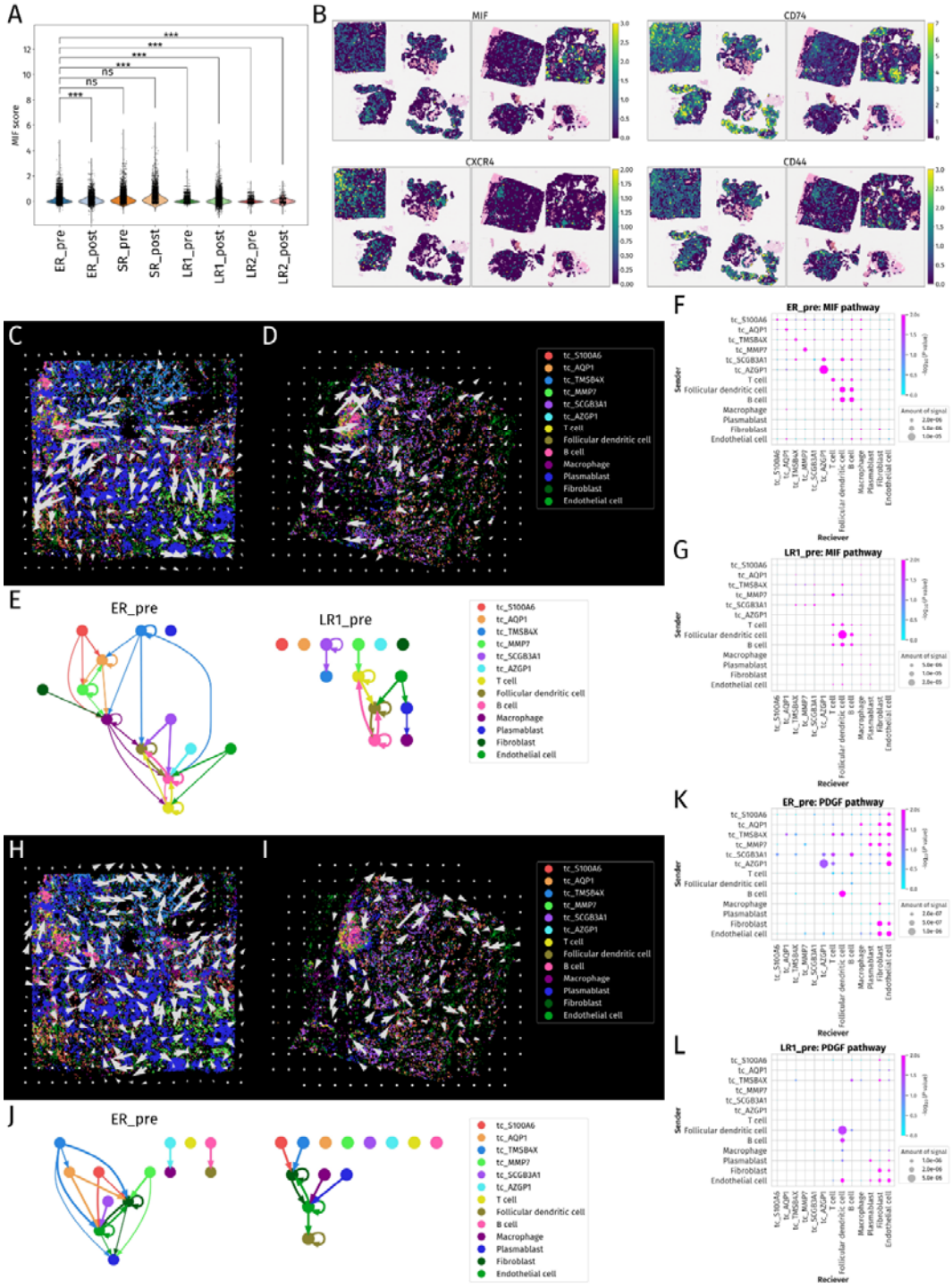


Fig. 6. Cell-cell interaction analysis with COMMOT. (**A**) Violin plot and strip plot of MIF scores for each sample. ***$P < 0.0005$; ns, not significant (one-sided Mann-Whitney U test). (**B**) Images of the expression levels of *MIF*, *CD74*, *CD44*, and *CXCR4*,

genes that constitute the MIF pathway and were expressed in ≥0.5% of all cells in each sample. (**C** and **D**) Images of spatial directions of cell-cell communication by the MIF pathway in ER_pre (C) and LR1_pre (D) samples. (**E**) Visualization of significant MIF pathway signal networks summarized by cell type in ER_pre and LR1_pre samples. (**F** and **G**) Dot plots of significant MIF pathway signal networks summarized by cell type in the ER_pre (F) and LR1_pre (G) samples. (**H** and **I**) Images of spatial directions of cell-cell communication by the PDGF pathway in ER_pre (H) and LR1_pre (I) samples. (**J**) Visualization of significant PDGF pathway signal networks summarized by cell type in ER_pre and LR1_pre samples. (**K** and **L**) Dot plots of significant PDGF pathway signal networks summarized by cell type in the ER_pre (K) and LR1_pre (L) samples.

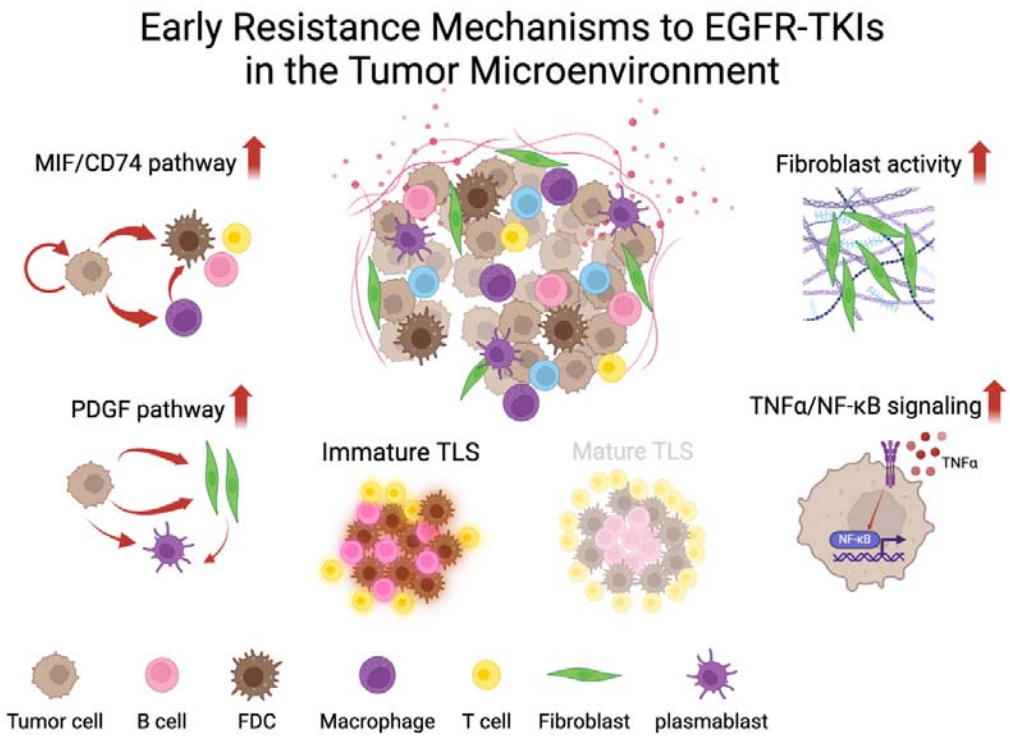
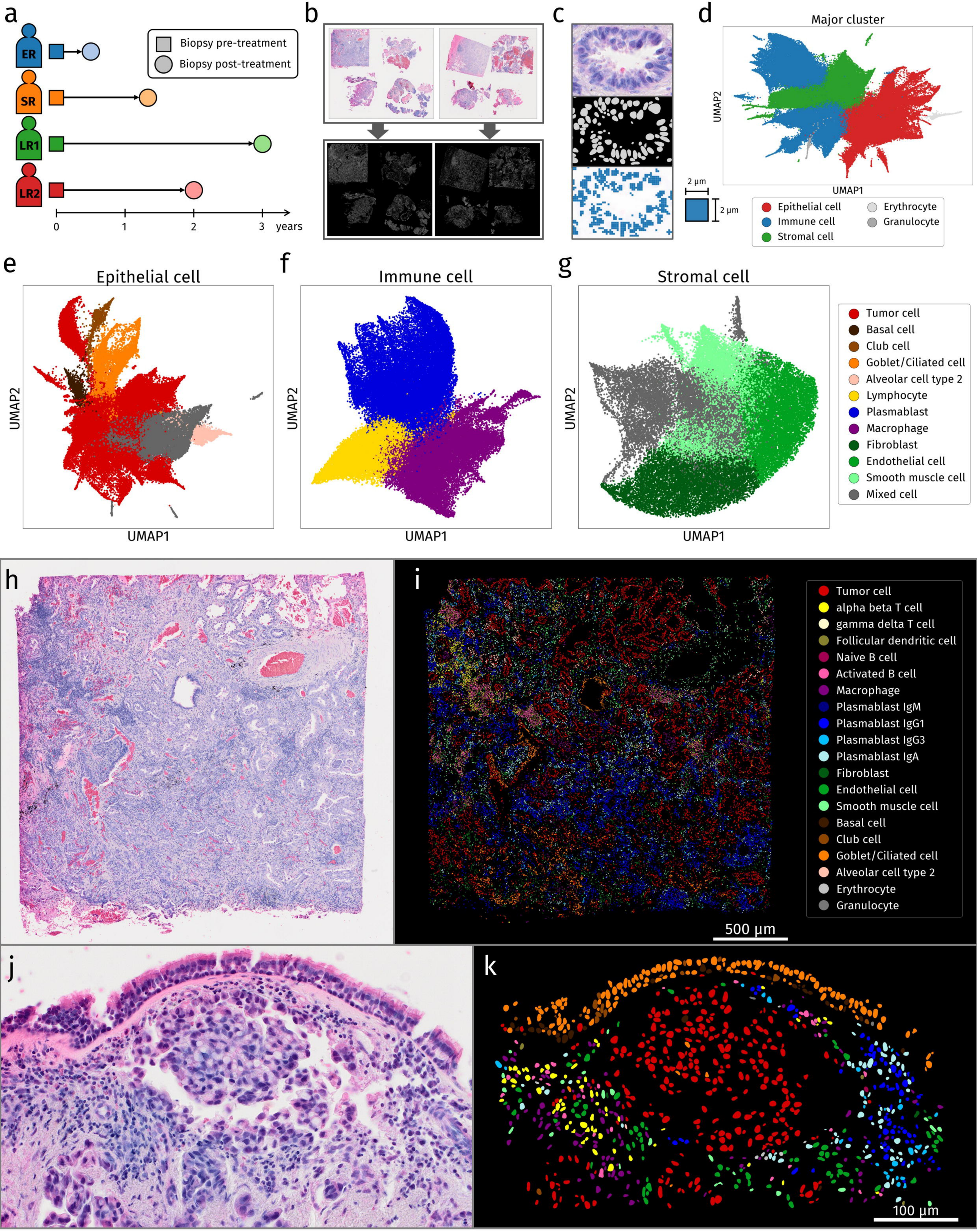
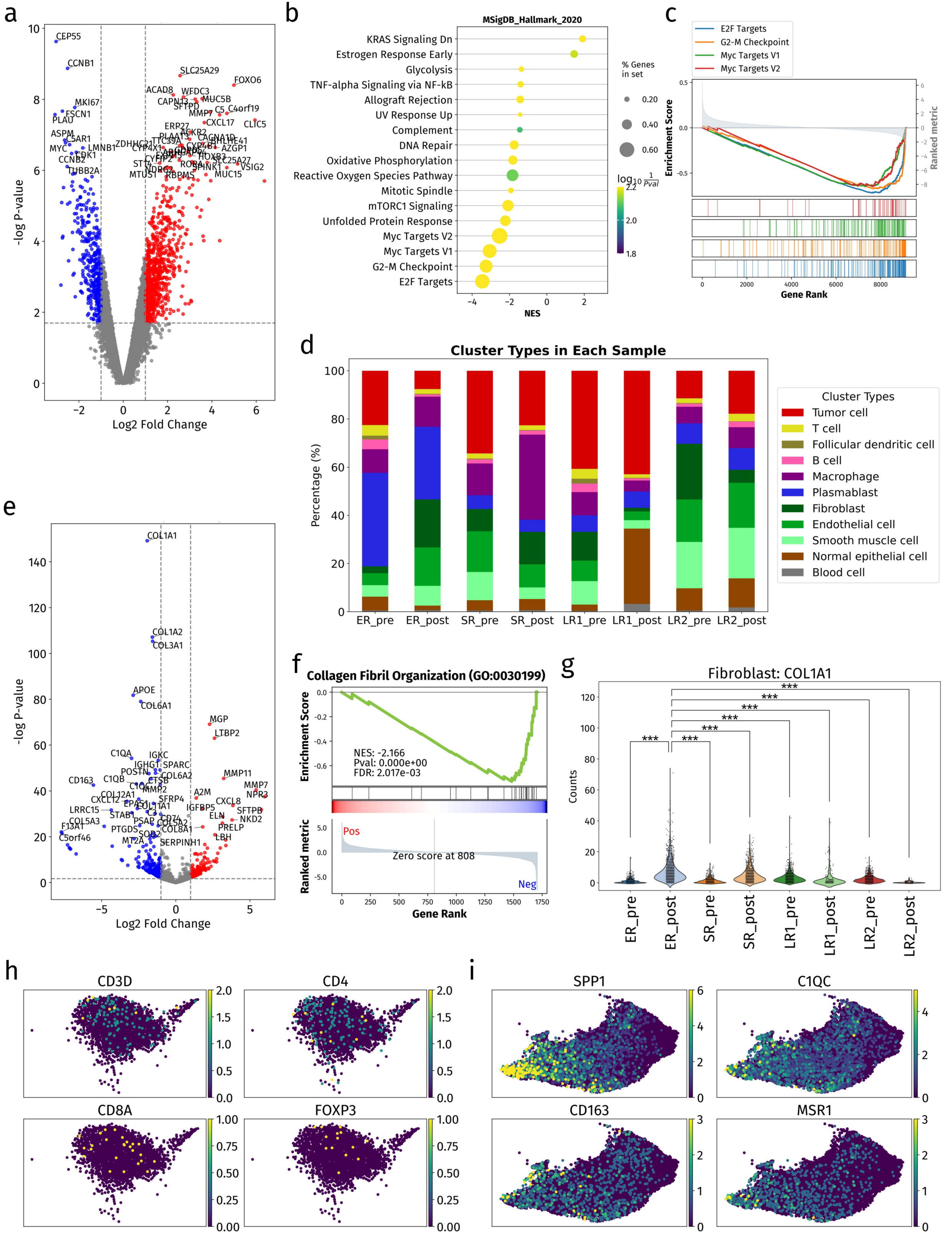
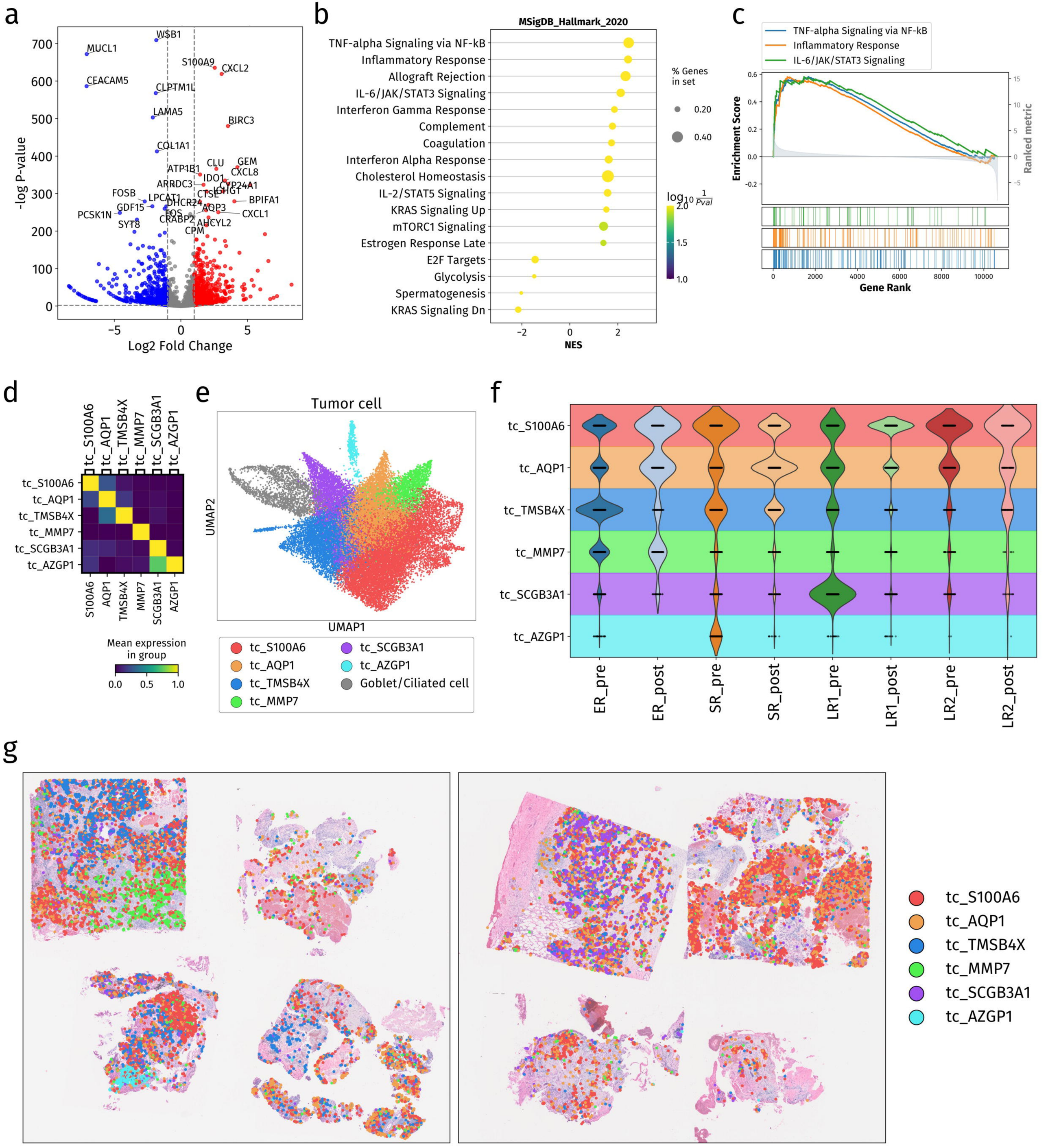
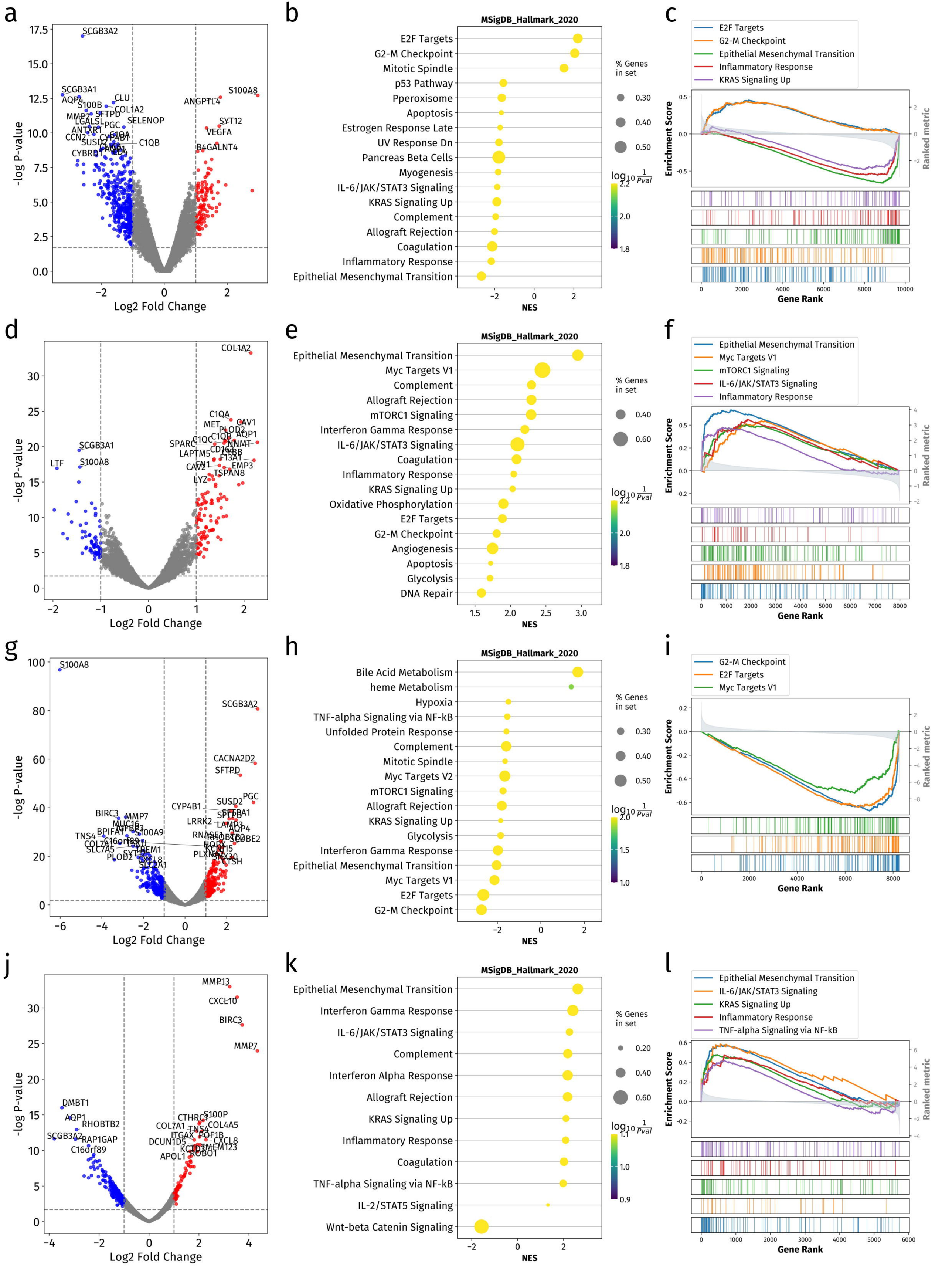


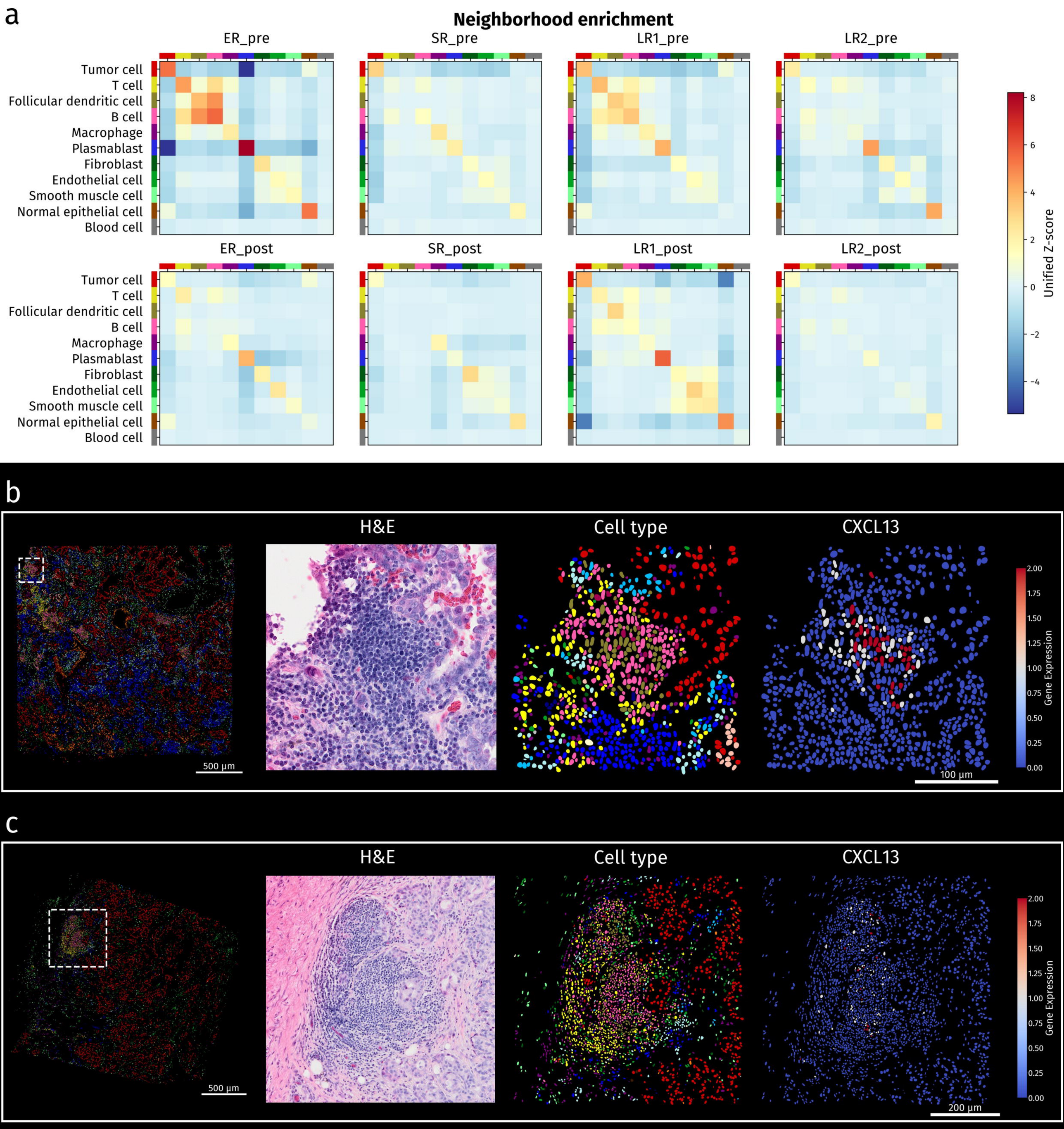
Fig. 7. Proposed model of early resistance to EGFR-TKIs in the TME. Schematic diagram summarizing findings from the analysis of this study

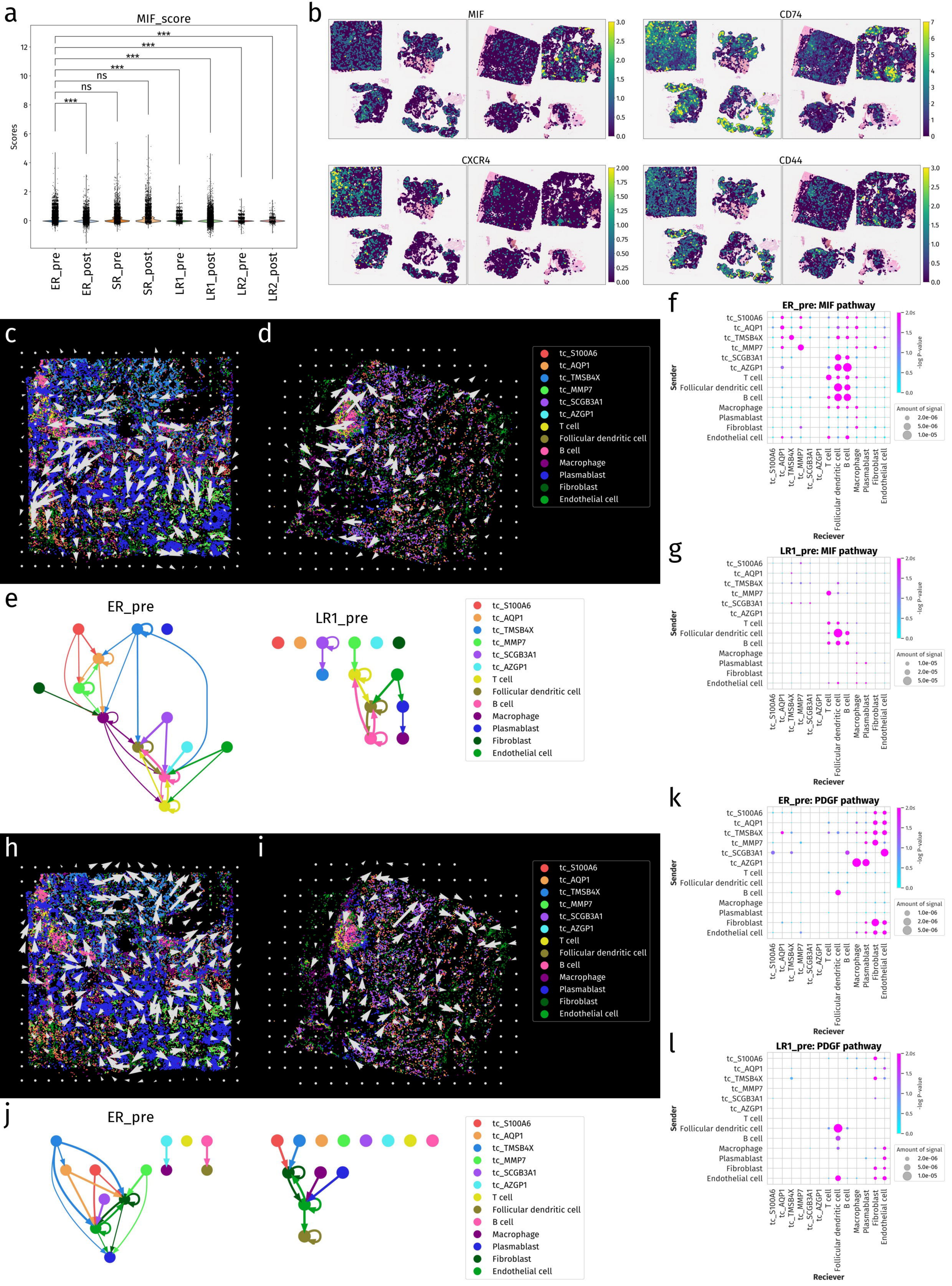
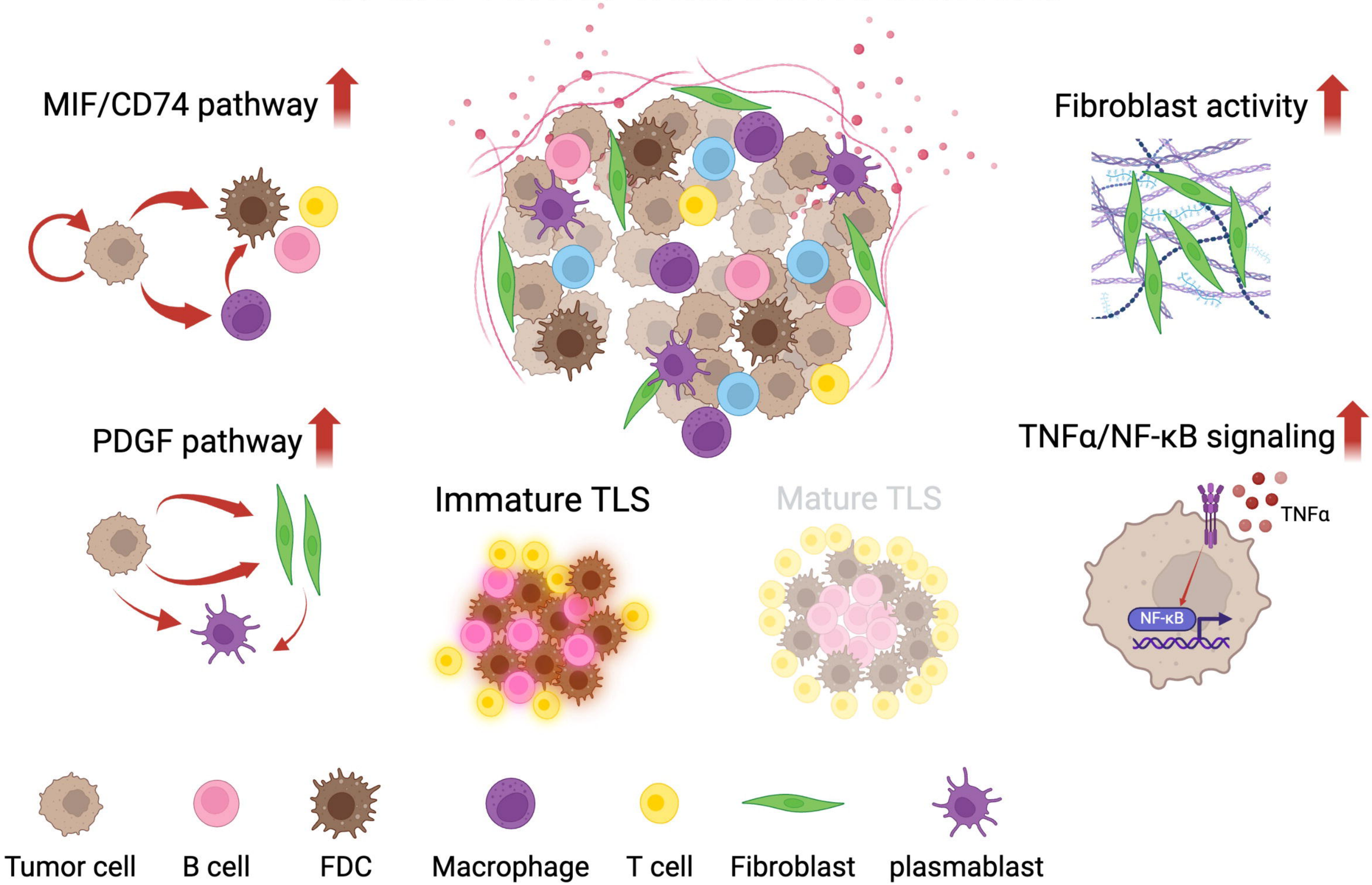

a
ER
SR
LR1
LR2
Biopsy pre-treatment
Biopsy post-treatment
0
1
2
3
years
b
c
2 µm
2 µm
d
Major cluster
UMAP2
UMAP1
Epithelial cell
Immune cell
Stromal cell
Erythrocyte
Granulocyte
e
Epithelial cell
UMAP2
UMAP1
f
Immune cell
UMAP2
UMAP1
g
Stromal cell
UMAP2
UMAP1
Tumor cell
Basal cell
Club cell
Goblet/Ciliated 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