Supplementary Materials and Methods

1. Single-cell RNA sequencing analysis

The data obtained were analyzed using the Seurat R package version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). After dataset normalization, we sequentially applied the FindVariableFeatures, FindIntegrationAnchor, and Integrate Data functions for data integration. Scaling, principal component analysis (PCA), and RunUMAP analysis were performed using the integration slot. The FindNeighbours and FindClusters functions were used to identify clusters. To confirm the cell types of each cluster, we employed the FindConservedMarkers function on the RNA slot of the Seurat object. Subsequently, each cluster was isolated, and dimension reduction and sub-clustering were performed using various Seurat functions, including FindVariableFeatures, Scale Data, PCA, and RunUMAP. The number of clusters of interest was determined by adjusting the options in the FindNeighbors and FindClusters functions. The barcodes of cells with perturbations in chromosomal gene expression were identified using the InferCNV package version 1.6.0 (https://github.com/broadinstitute/inferCNV). We used epithelial cell barcodes from tumor tissues as the reference, comparing them to cell clusters obtained from normal lung tissues and nonepithelial cell clusters from tumor tissues. The InferCNV package employed a non-supervised clustering method, which designated cells with clear chromosomal gene expression perturbation as lung cancer cells. Differentially expressed gene analysis identified upregulated genes to be mainly related to stem cell markers.