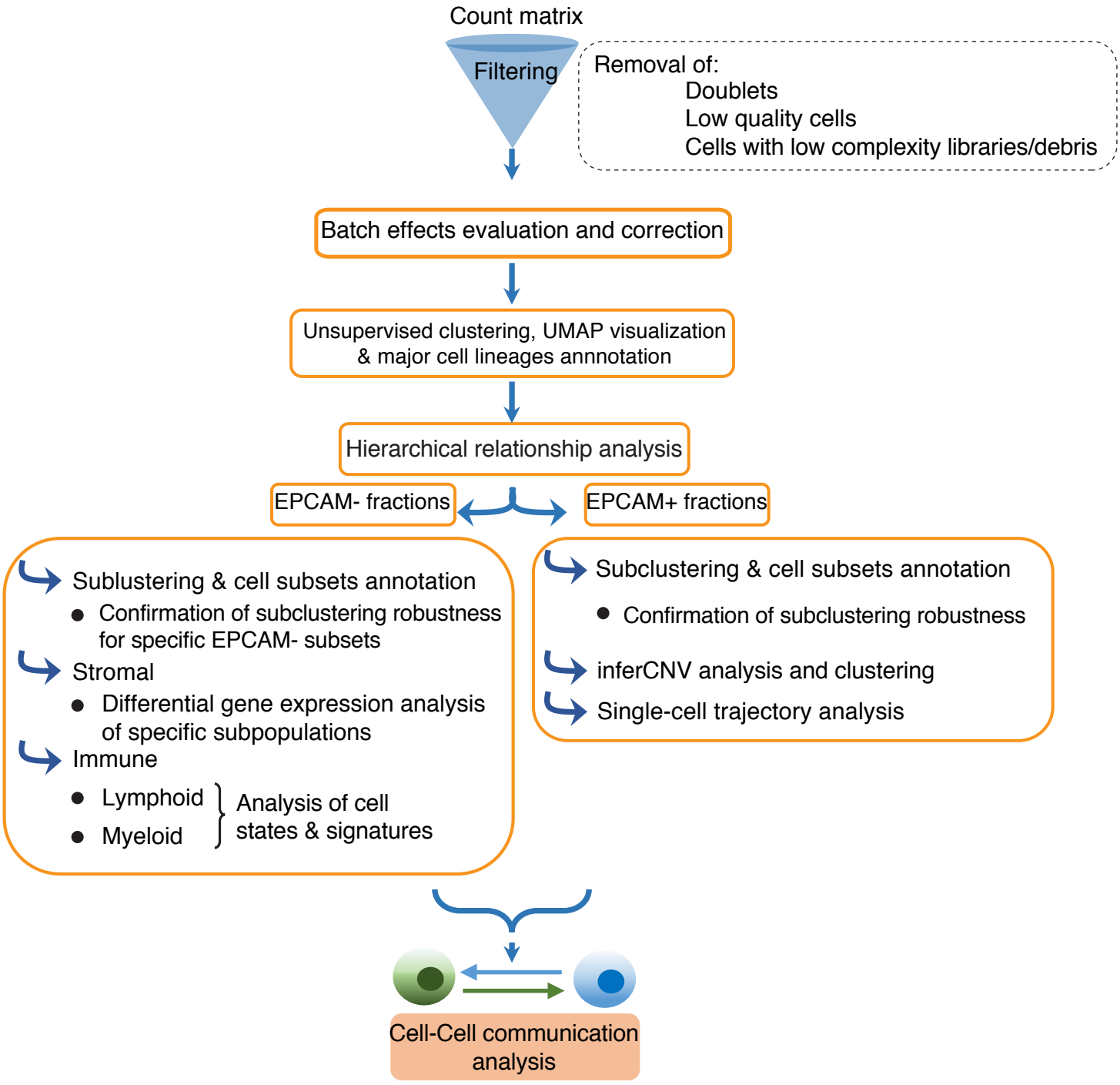
Resolving the spatial and cellular architecture of lung adenocarcinoma by multi-region single-cell sequencing

SUPPLEMENTARY DATA FILE 2

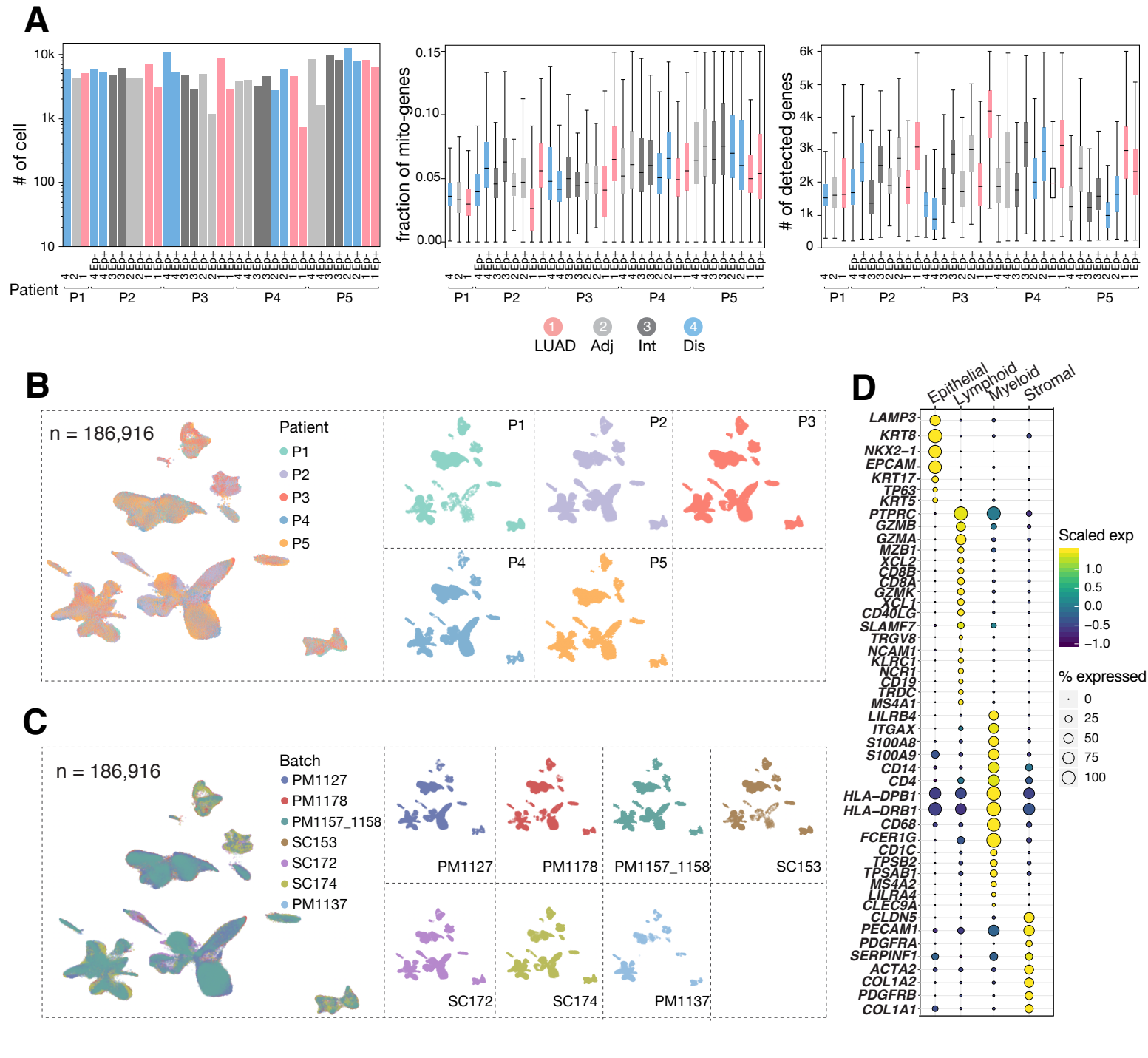
Supplementary Figures S1-S7

Supplementary Fig. S1



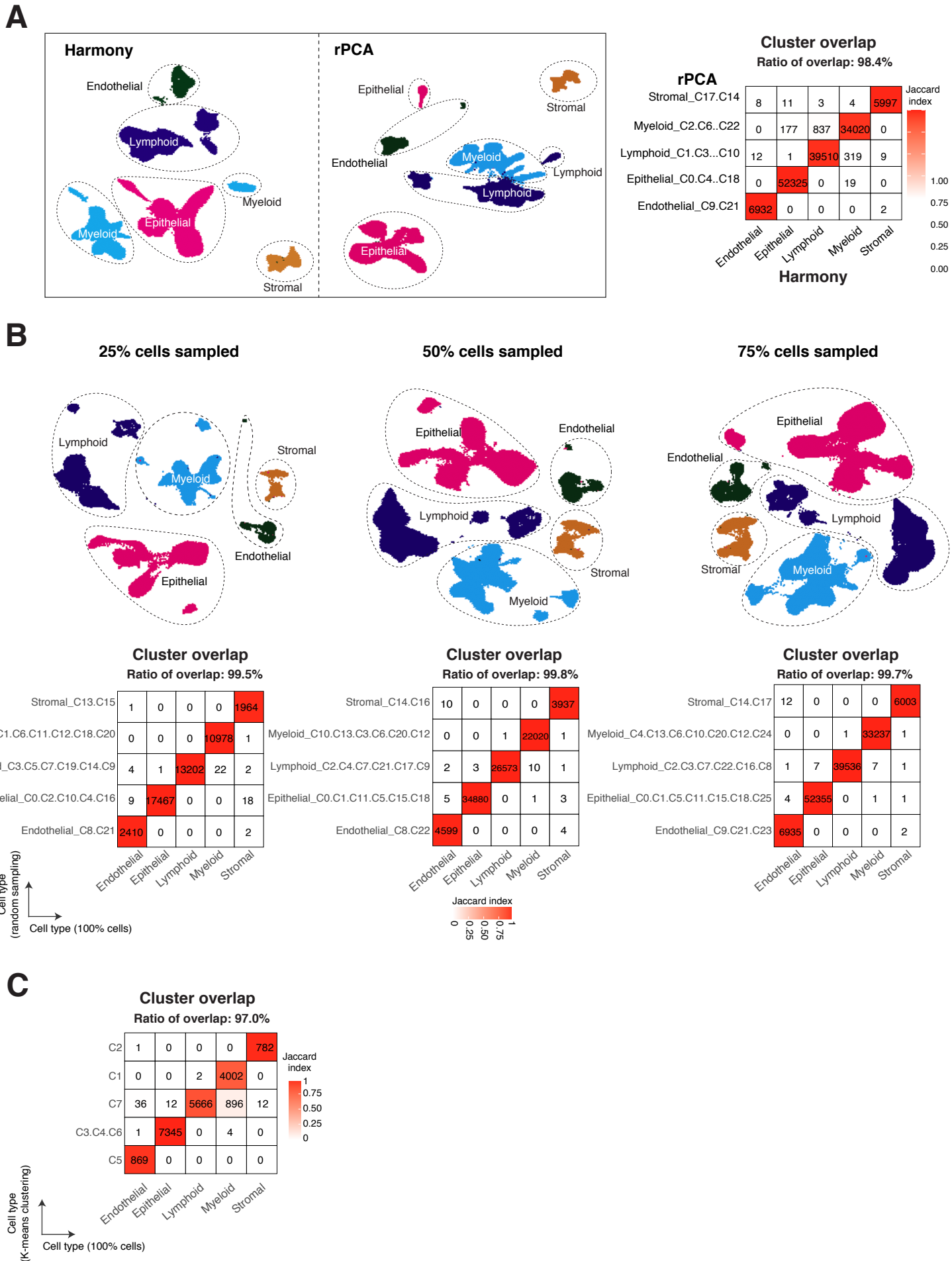
Supplementary Fig. S1. Schematic view of the bioinformatics analysis workflow.
An overview of the quality control steps and analyses done.

Supplementary Fig. S2



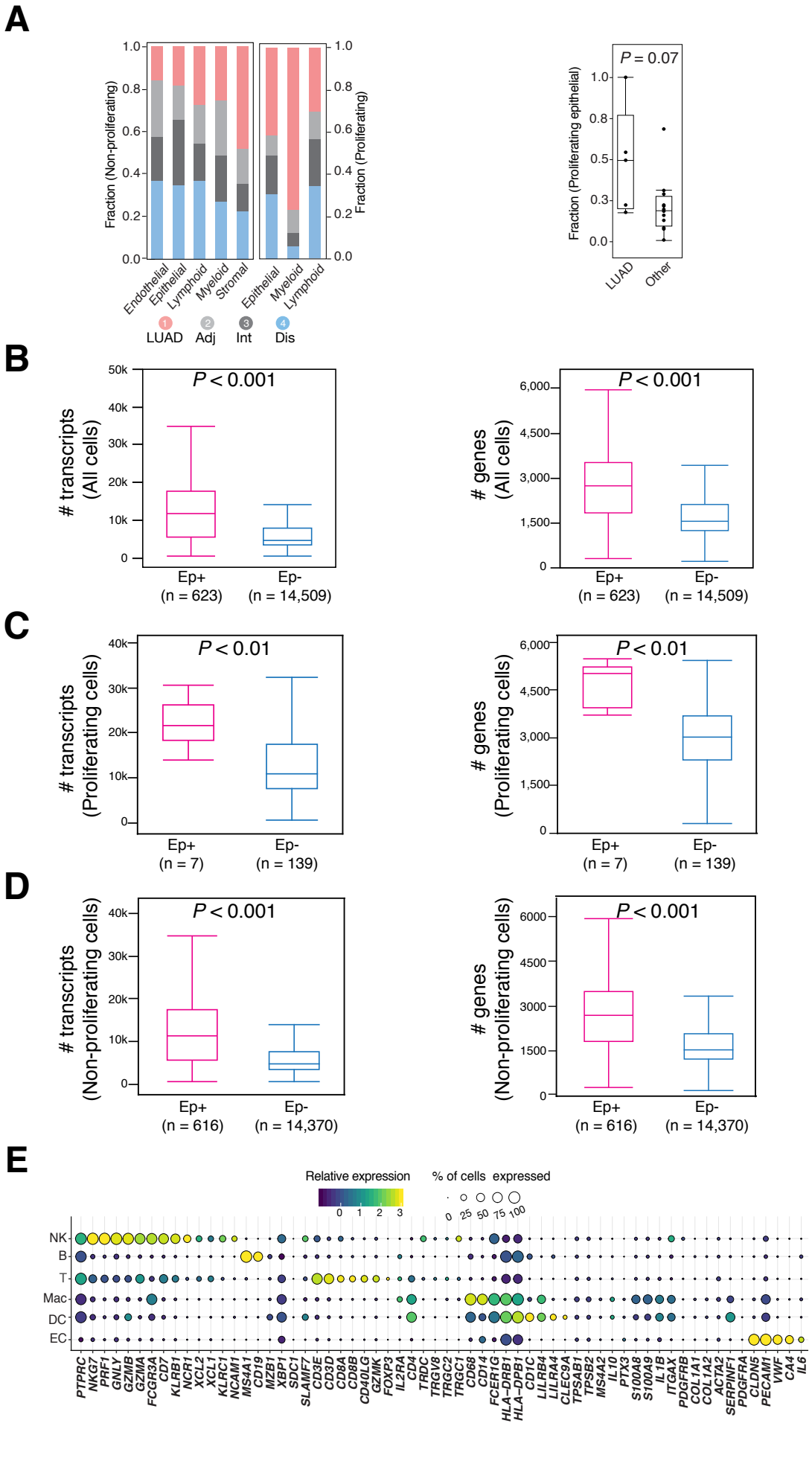
Supplementary Fig. S2. Quality control metrics and expression of major cell lineage markers across the spatial LUAD scRNA-seq dataset. A, Statistical summary of cells passing quality control (QC) and showing cell number (left), fraction of mitochondrial genes (middle), and the number of detected genes (right) per sample. Dis, distant normal; Int, intermediate normal; Adj, adjacent normal; LUAD, tumor tissue. **B-C,** UMAP plots showing cells colored by patient ID (B) and library/sequencing batch (C). **D,** Bubble plot showing the percentage of cells expressing lineage markers (indicated by the size of the circle) as well as their scaled expression levels (indicated by the color of the circle) across all cells (related to main Fig. 1D and Fig. 1E).

Supplementary Fig. S3



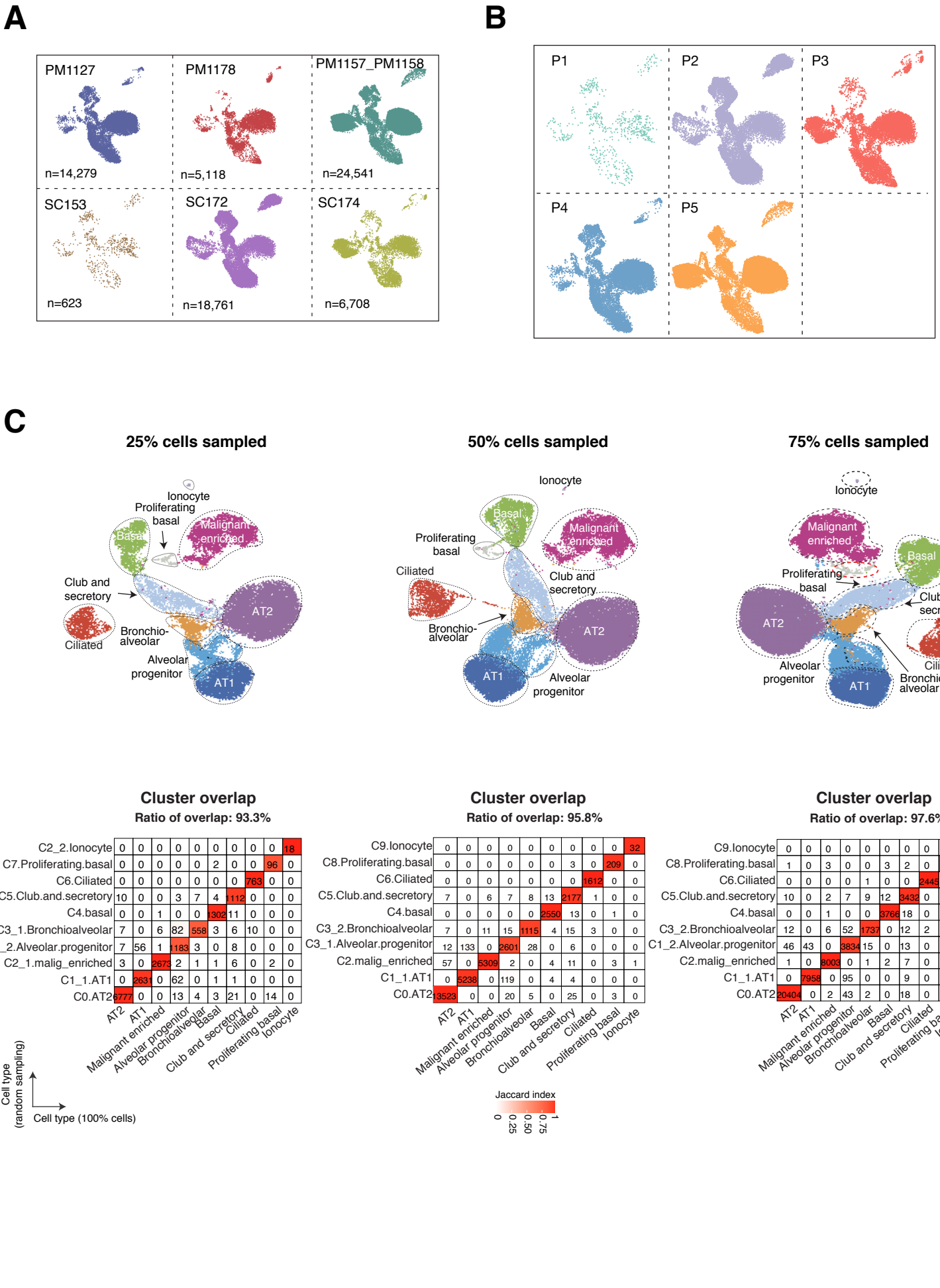
Supplementary Fig. S3. Analysis of clustering robustness of major cellular lineages. **A,** UMAP plots showing cells colored by major lineages and identified with Harmony (left) or rPCA (middle). The right part of panel A shows a heatmap depicting the extent of cluster assignment overlap between rPCA (rows) and Harmony results (columns) as quantified by Jaccard index. **B**, UMAP plots (top) and the corresponding cluster overlap indices (bottom) when using 25% (left), 50% (middle), and 75% (right) of randomly sampled cells. The heatmaps show cluster overlap indices in randomly sampled results (rows) versus when using all 186,916 cells (columns). **C,** Heatmap depicting the extent of cluster assignment overlap between k-means clustering (columns) and Harmony (rows) as quantified by Jaccard index.

Supplementary Fig. S4



Supplementary Fig. S4. Analysis of expression markers among epithelial and immune cell fractions. A, Stacked bar plots showing the relative cell fraction of spatial samples for major lineages considering non-proliferating cells (left) and proliferating ones (middle). Box plot showing fraction of proliferating epithelial cells in LUAD tissues versus normal spatial samples (right). *P* – value was calculated using Wilcoxon rank sum test. **B,** Boxplots showing the number of detected transcripts (left) and the number of detected genes (right) in epithelial cells versus in all non-epithelial lineages including lymphoid, myeloid, and stromal cells. **C,** Boxplots showing the number of detected transcripts (left) and the number of detected genes (right) in proliferating epithelial cells versus in proliferating cells of non-epithelial lineages including lymphoid, myeloid, and stromal cells. **D,** Boxplots showing the number of detected transcripts (left) and genes (right) in non-proliferating epithelial cells versus in non-proliferating and non-epithelial lineages including lymphoid, myeloid, and stromal cells. **E,** Bubble plot showing the percentage of cells expressing lineage markers (indicated by the size of the circle) as well as their scaled expression levels (indicated by the color of the circle) across selected cell types (related to main Fig. 1G and Fig. 1H). NK; natural killer cell, DC; dendritic cell, EC; endothelial cell.

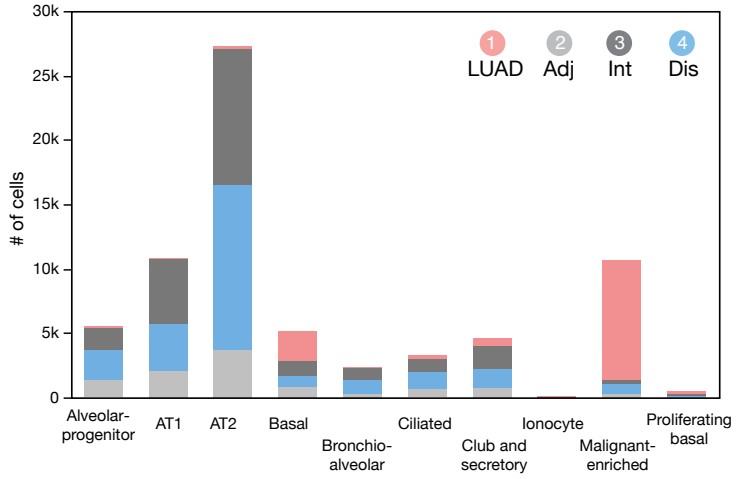
Supplementary Fig. S5



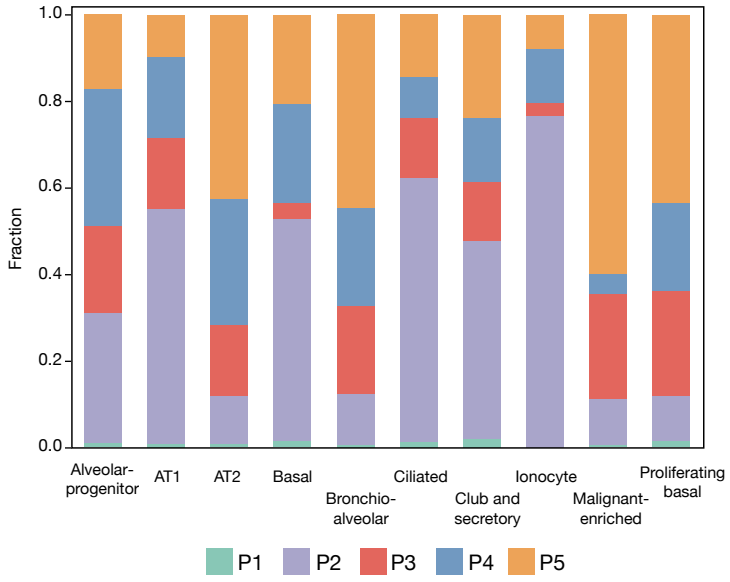
Supplementary Fig. S5. Analysis of robustness of epithelial cell clustering. A, UMAP view showing EPCAM+ cells colored by library/sequencing batch. **B,** UMAP plot showing EPCAM+ cells colored by patient ID. **C,** UMAP plots (top) showing clustering results when using 25% (top left), 50% (top middle), and 75% (top right) of randomly sampled epithelial cells. The corresponding heatmaps (bottom) show cluster overlap indices in randomly sampled results (rows) versus when using all 70,030 epithelial cells (columns).

Supplementary Fig. S6

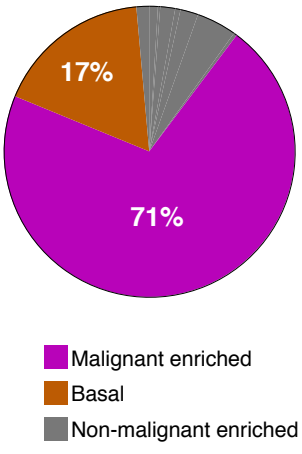
A



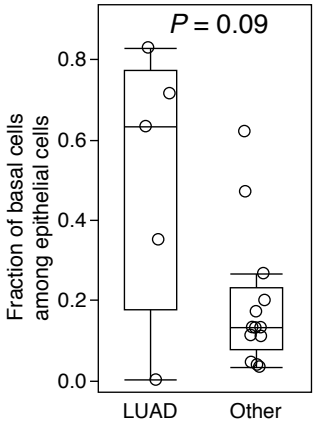
B



C

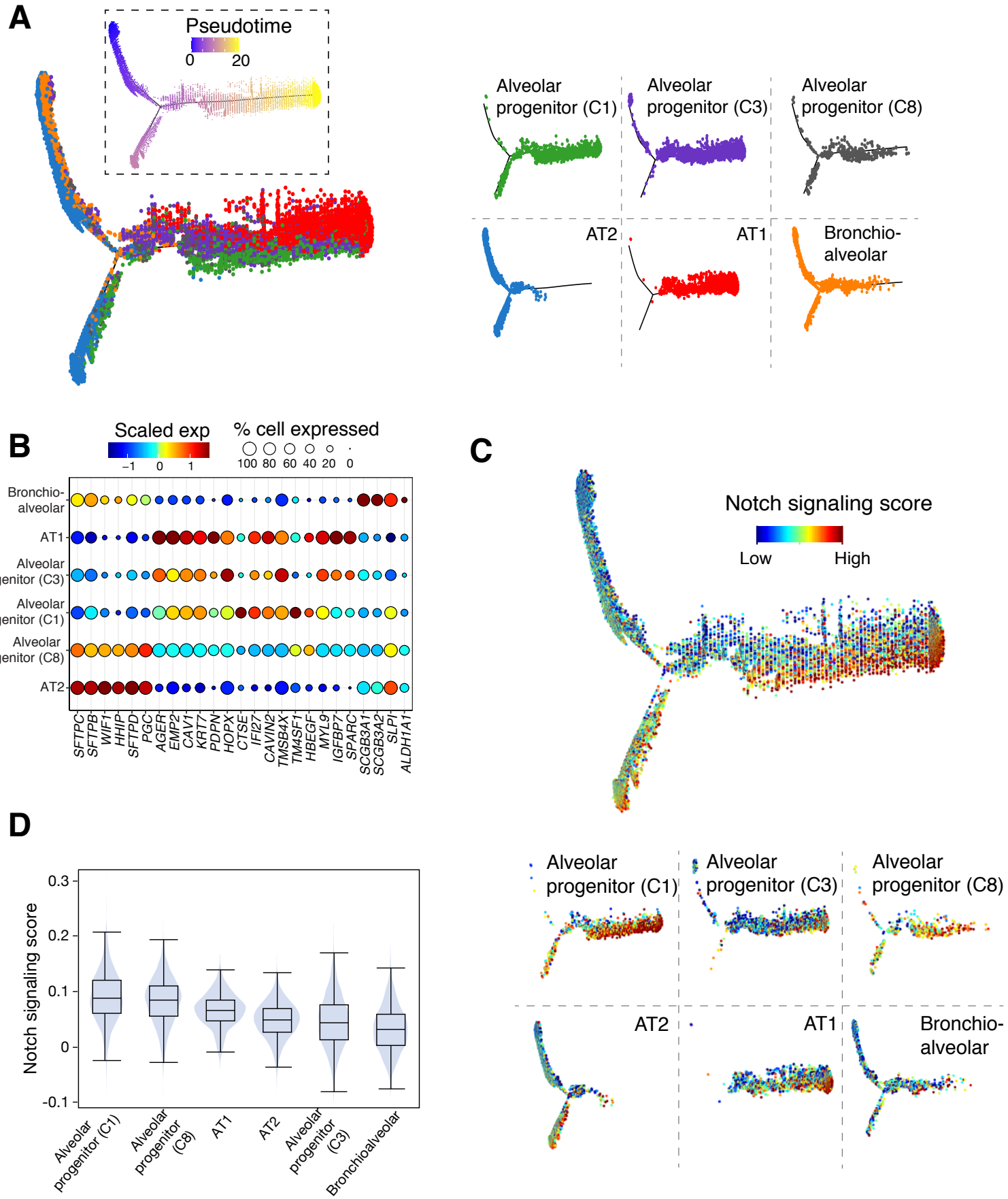


D



Supplementary Fig. S6. Cellular distribution of epithelial lineage clusters. A, Bar plot showing absolute numbers of cells for each lung epithelial cell lineage. Dis, distant normal; Int, intermediate normal; Adj, adjacent normal; LUAD, tumor tissue; AT1, alveolar type 1; AT2, alveolar type 2. **B**, Stacked bar plot showing relative fractions of individual epithelial subclusters derived from each patient. **C.** Pie chart showing the fractional distribution of epithelial cells from the LUADs by epithelial lineage cluster. **D.** Box plot showing fraction of basal cells among epithelial cells and from LUADs versus other normal spatial samples. *P* – value was calculated using Wilcoxon rank sum test.

Supplementary Fig. S7



Supplementary Fig. S7. Trajectory analysis of alveolar cells. A, Potential developmental trajectory for alveolar cells inferred by pseudotime analysis. Cells were ordered by pseudotime (dotted box) and colored by alveolar cell state. **B,** Bubble plots showing the percentage (indicated by the size of the circle) of cells expressing markers of alveolar states shown in trajectory analysis from panel C as well as their scaled expression levels (indicated by the color of the circle) **C,** Pseudotime trajectory showing cells colored by Notch signaling signature score. **D,** Violin plots showing Notch signaling signature score among alveolar cell states in trajectory analysis from panel B.