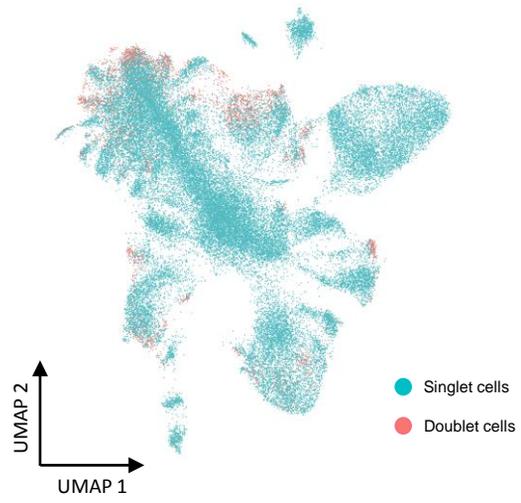
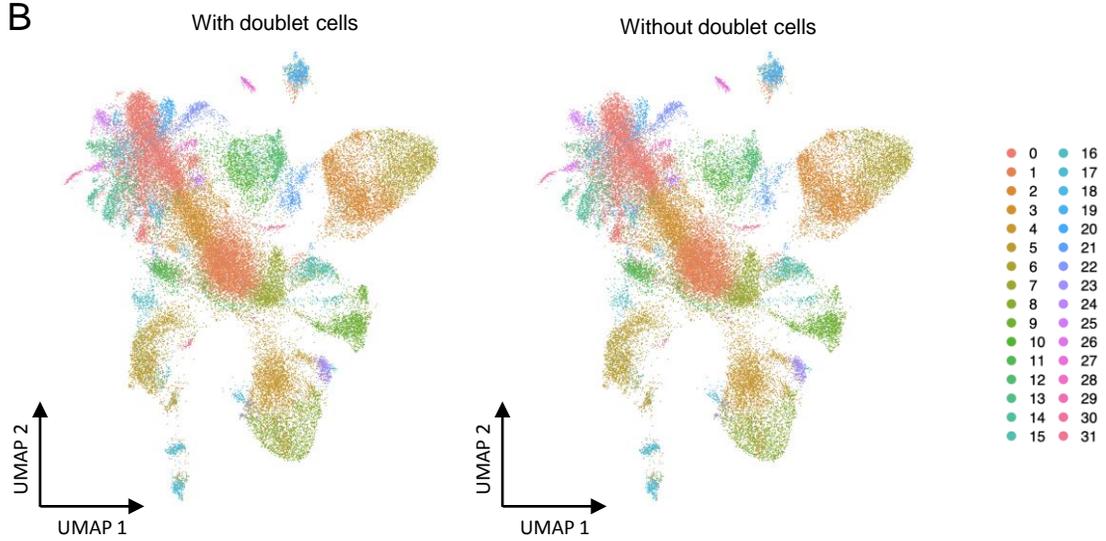


# Supplementary Figure 1

A

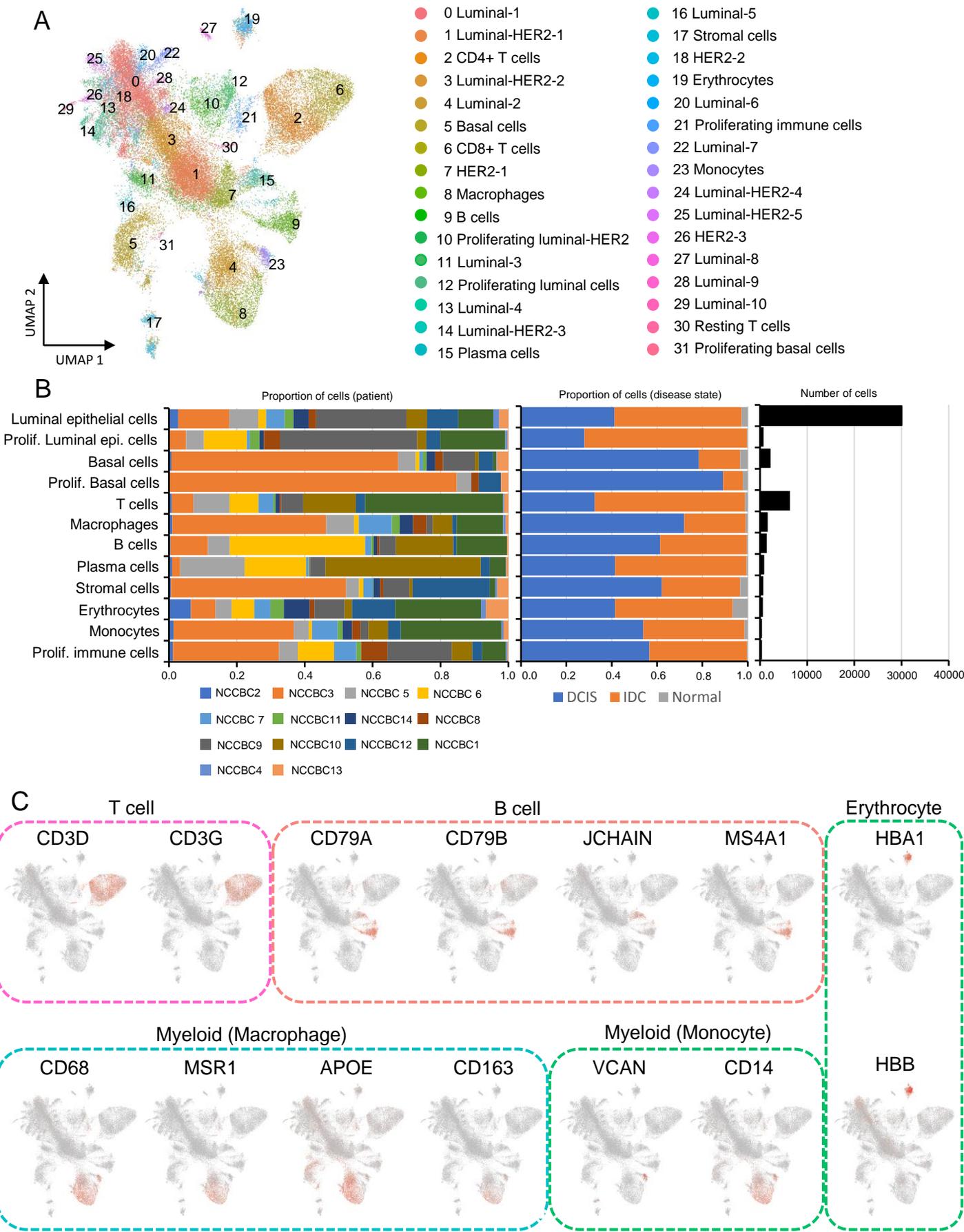


B



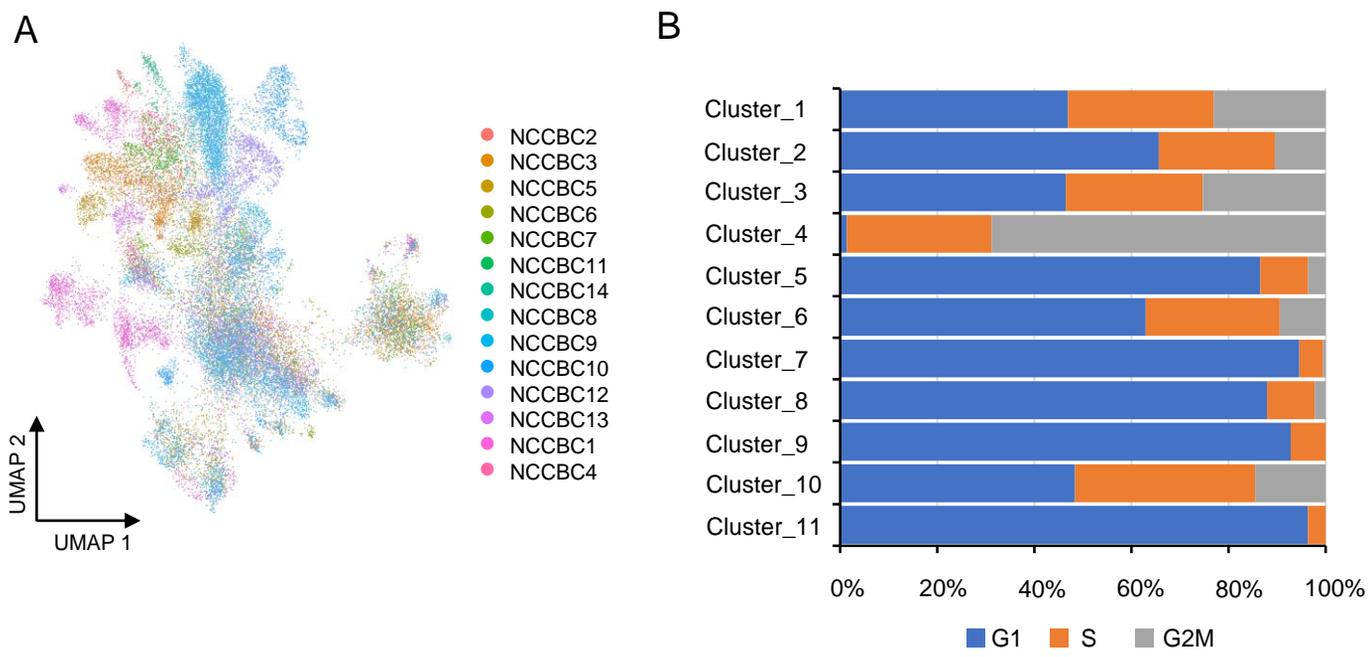
**Supplementary Figure 1 | DoubletFinder analysis of all cells. A. DoubletFinder analysis. B. Umap analysis with/without doublet cells in all samples.**

# Supplementary Figure 2



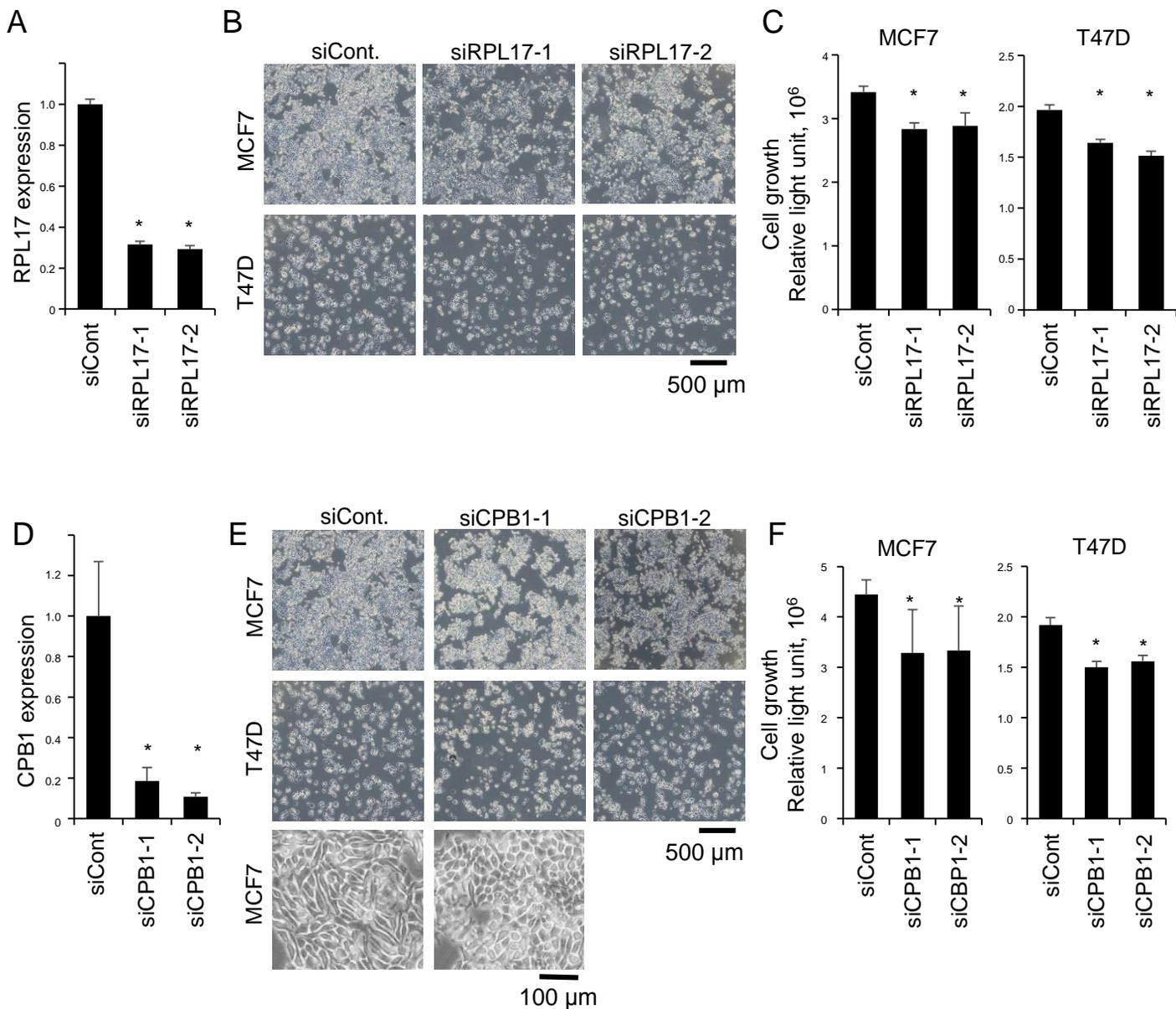
**Supplementary Figure 2 | scRNA-seq analysis of DCIS, IDC, and normal breast tissue.** **A.** UMAP of 7 DCIS samples, 6 IDC samples and 1 normal breast tissue sample. According to the gene expression profiles in each cell, the cells were divided into 32 distinct clusters. By marker expression patterns, each cluster was characterized as shown. **B.** Proportions of cells from each patient (x-axis) showing the proportions of cell types (y-axis). The cell numbers in each cluster are shown (right graph). **C.** UMAP plots with the expression of typical markers. CD3 and CD3G for T cells; CD79A, CD79B, JCHAIN, and MS4A1 for B cells; CD68, MSR1, APOE, and CD163 for myeloid cells (macrophages); VCAN and CD14 for myeloid cells (monocytes); and HBA1 and HBB for erythrocytes.

# Supplementary Figure 3



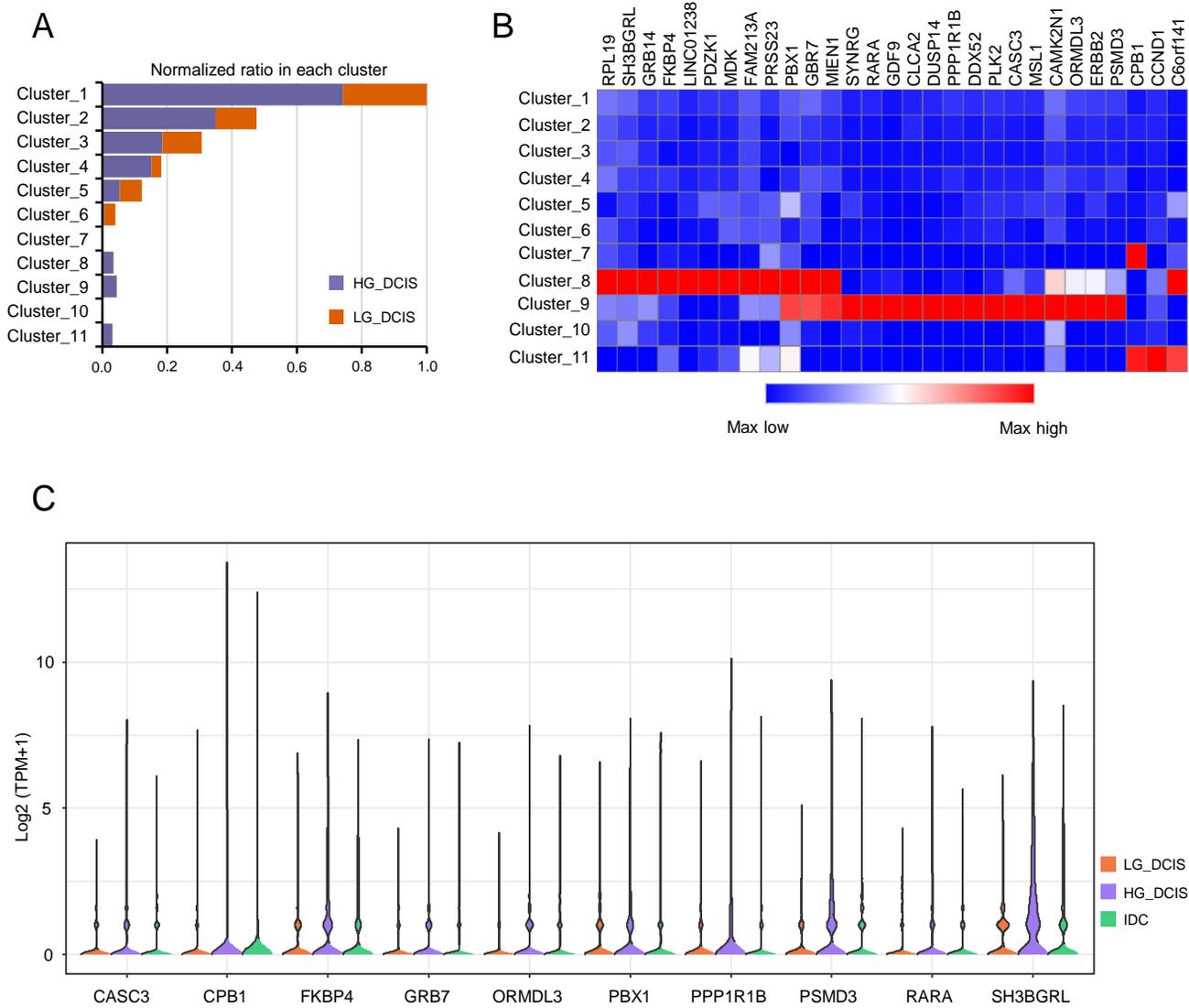
**Supplementary Figure 3 | Analysis of luminal cell populations from scRNA-seq data.**  
**A.** UMAP plot of luminal epithelial components in each patient. **B.** Module analysis of the cell cycle in each cluster with the Seurat R package.

# Supplementary Figure 4



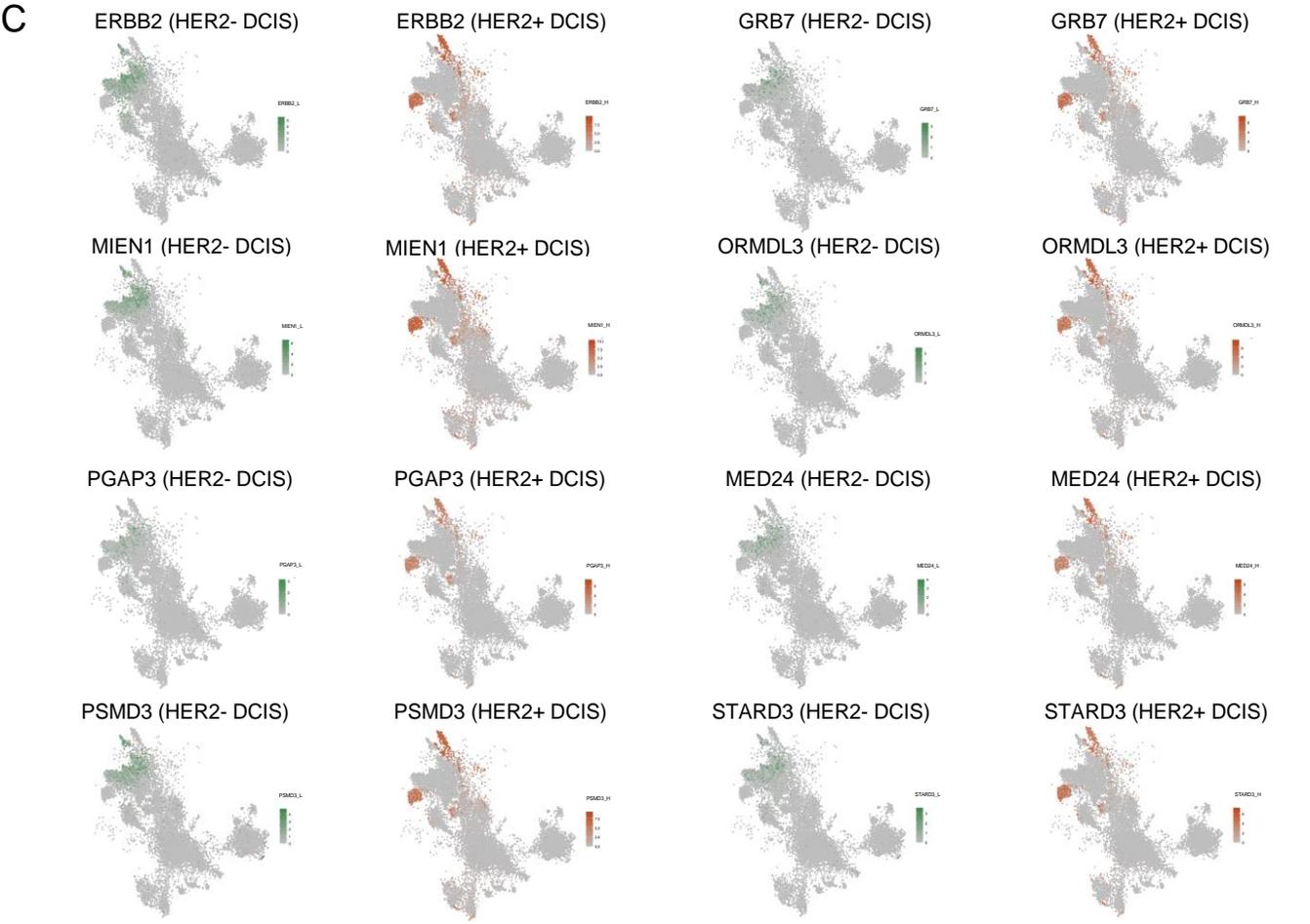
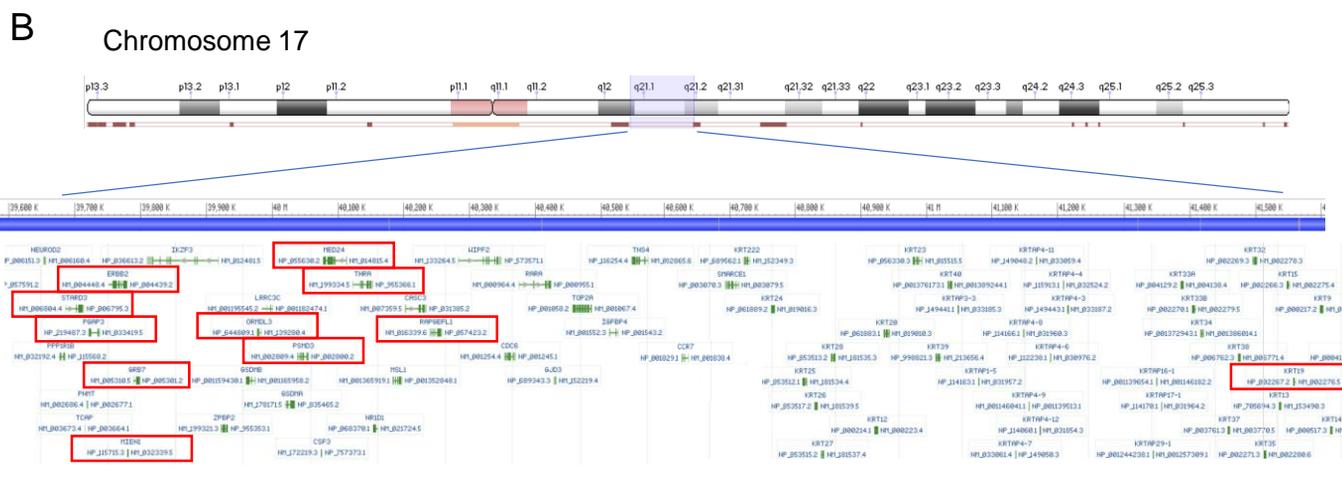
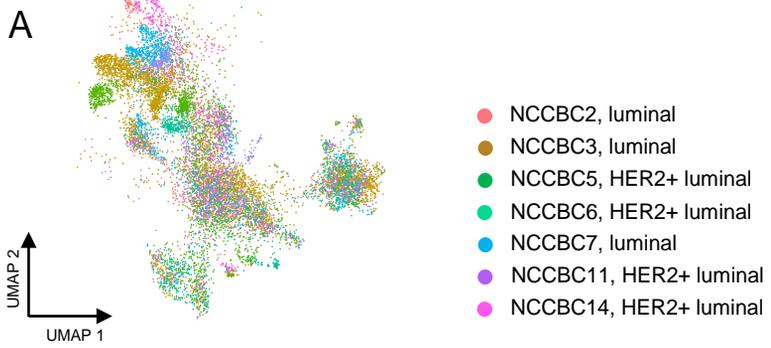
**Supplementary Figure 4 | Functional analysis of RPL17 and CPB1 in MCF7 and T47D cells.** **A.** siRNA knockdown efficiency. RPL17 expression levels in the siCont, siRPL17-1, and siRPL17-2 groups of MCF7 cells are shown. **B.** Representative images of siCont, siRPL17-1, and siRPL17-2 cells (top: MCF7, bottom: T47D). Scale bar: 500  $\mu$ m. **C.** Viability of siCont, siRPL17-1, and siRPL17-2 cells, as measured by a CellTiter Glo assay (left: MCF7, right: T47D). \*:  $p < 0.05$ . **D.** siRNA knockdown efficiency. CPB1 expression levels in the siCont, siCPB1-1, and siCPB1-2 groups of MCF7 cells are shown. **E.** Representative images of siCont, siCPB1-1, and siCPB1-2 cells (top: MCF7, middle: T47D, bottom: high magnification image of MCF7). Scale bar: 500  $\mu$ m, 100 $\mu$ m. **F.** Viability of siCont, siCPB1-1, and siCPB1-2 cells, as measured by a CellTiter Glo assay (left: MCF7, right: T47D). \*:  $p < 0.05$ .

# Supplementary Figure 5



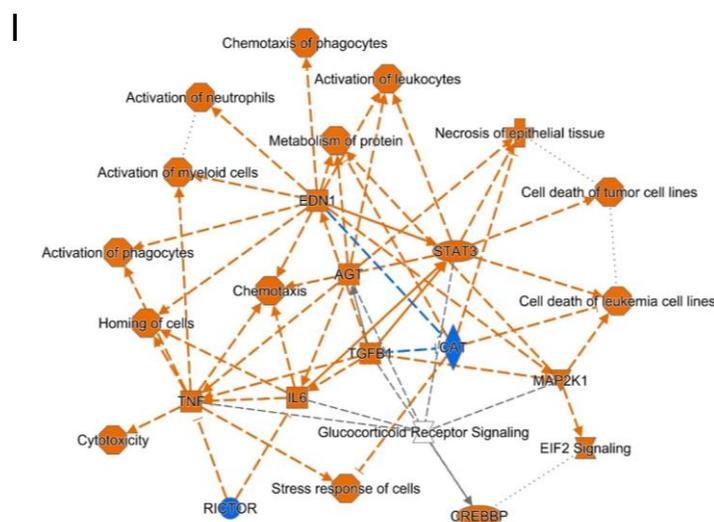
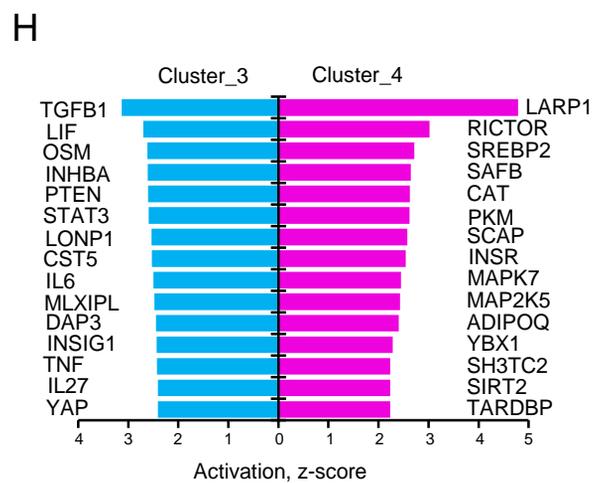
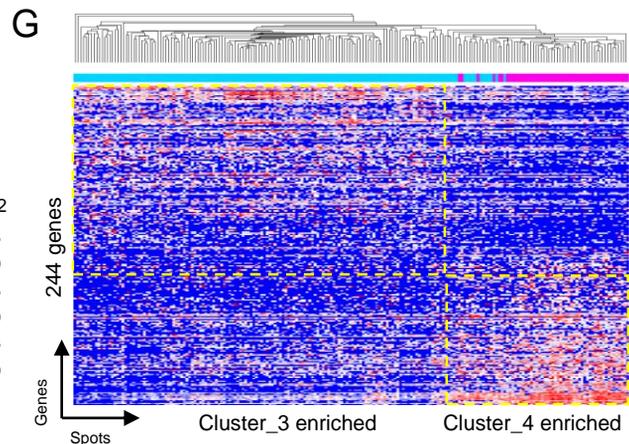
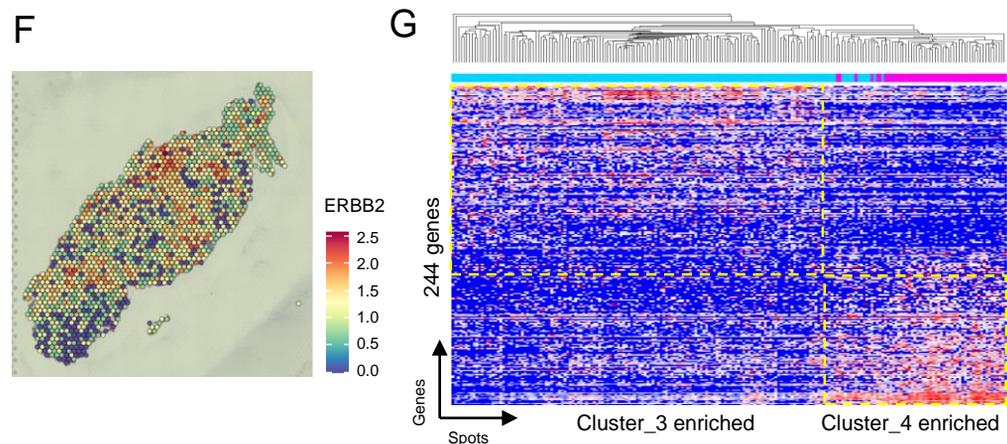
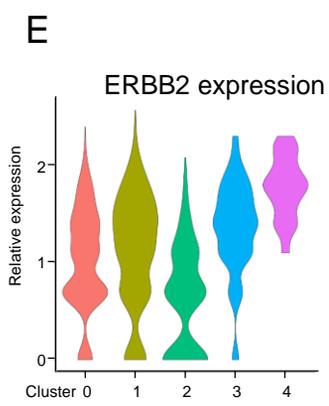
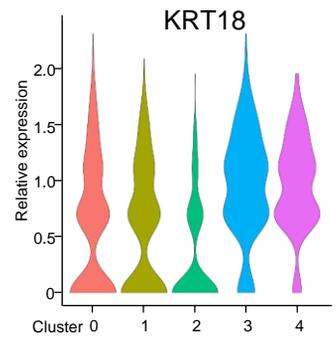
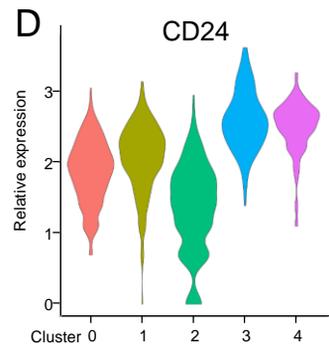
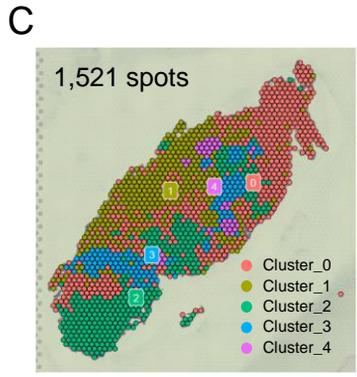
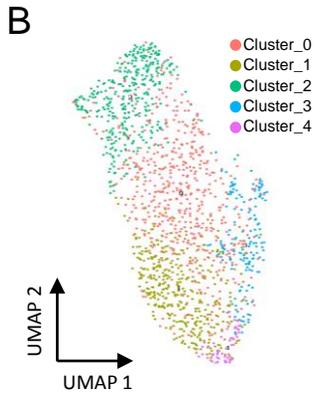
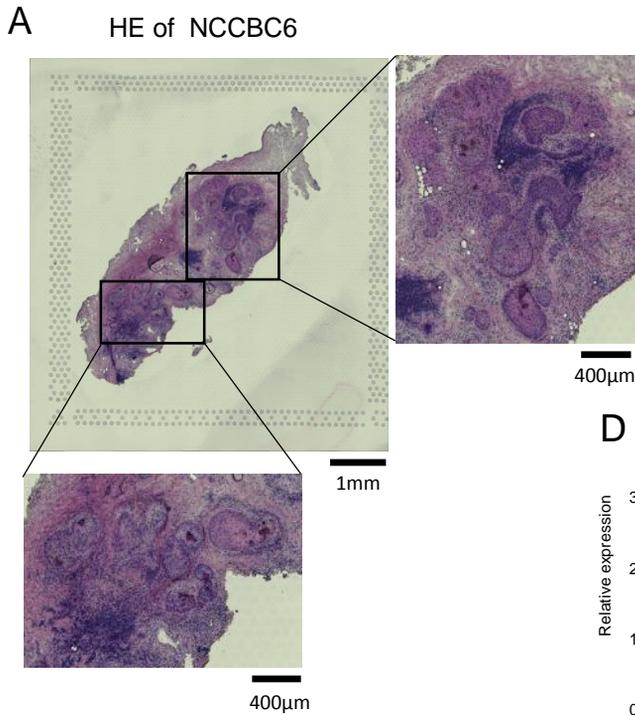
**Supplementary Figure 5 | Comparison between high-grade DCIS and low-grade DCIS. A.** Cell proportion of high-grade and low-grade DCIS luminal epithelial cells in this study. **B.** Specific gene expression in Clusters 8, 9, and 11. **C.** Selected gene expression gradually elevated during the tumor progression.

# Supplementary Figure 6



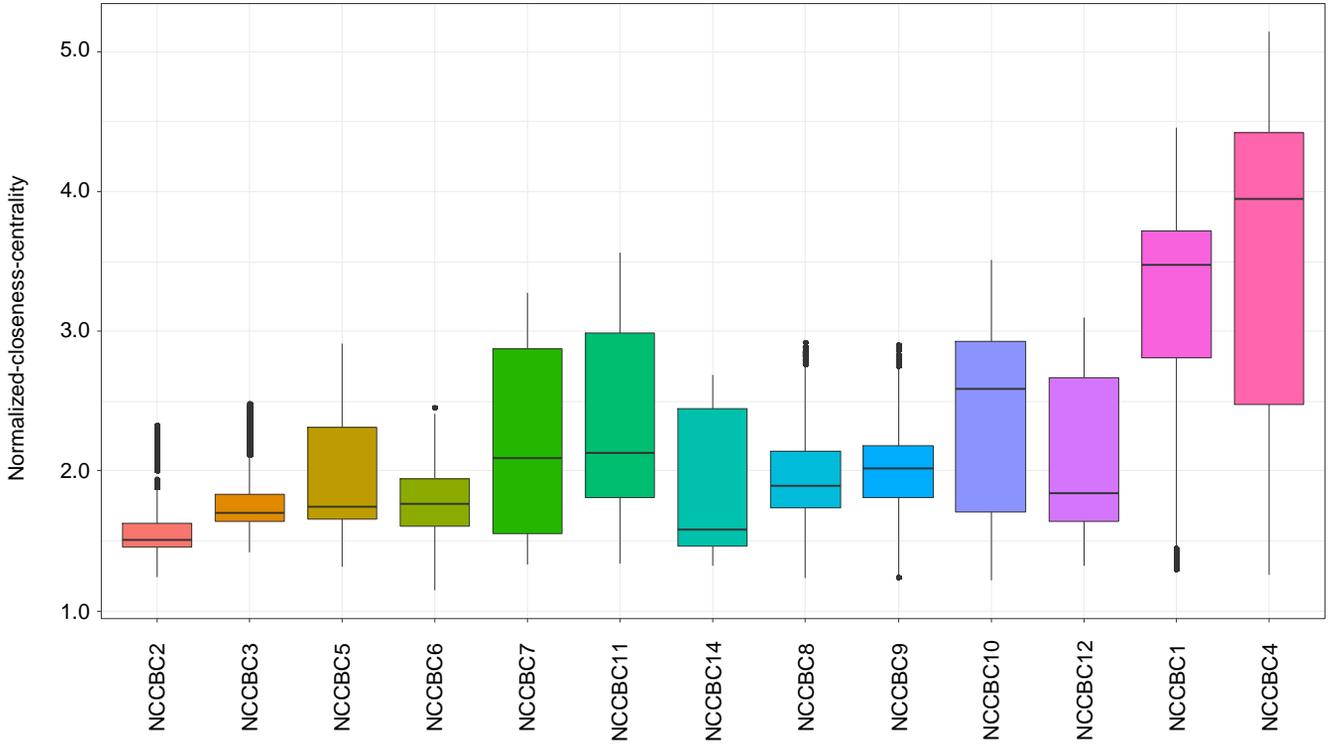
**Supplementary Figure 6 | Analysis of HER2+ luminal DCIS cells from scRNA-seq data.** **A.** UMAP plot of the luminal epithelial components in all DCIS samples. Patient ID and HER2 status are shown. **B.** Chromosome 17q12-q21 information. The red rectangles indicate genes that were detected in this analysis. **C.** UMAP plots of 8 genes that were highly expressed in Clusters 8 and 9 and located at the 17q12-q21 locus. Green: HER2- luminal DCIS, brown: HER2+ luminal DCIS.

# Supplemental Figure 7



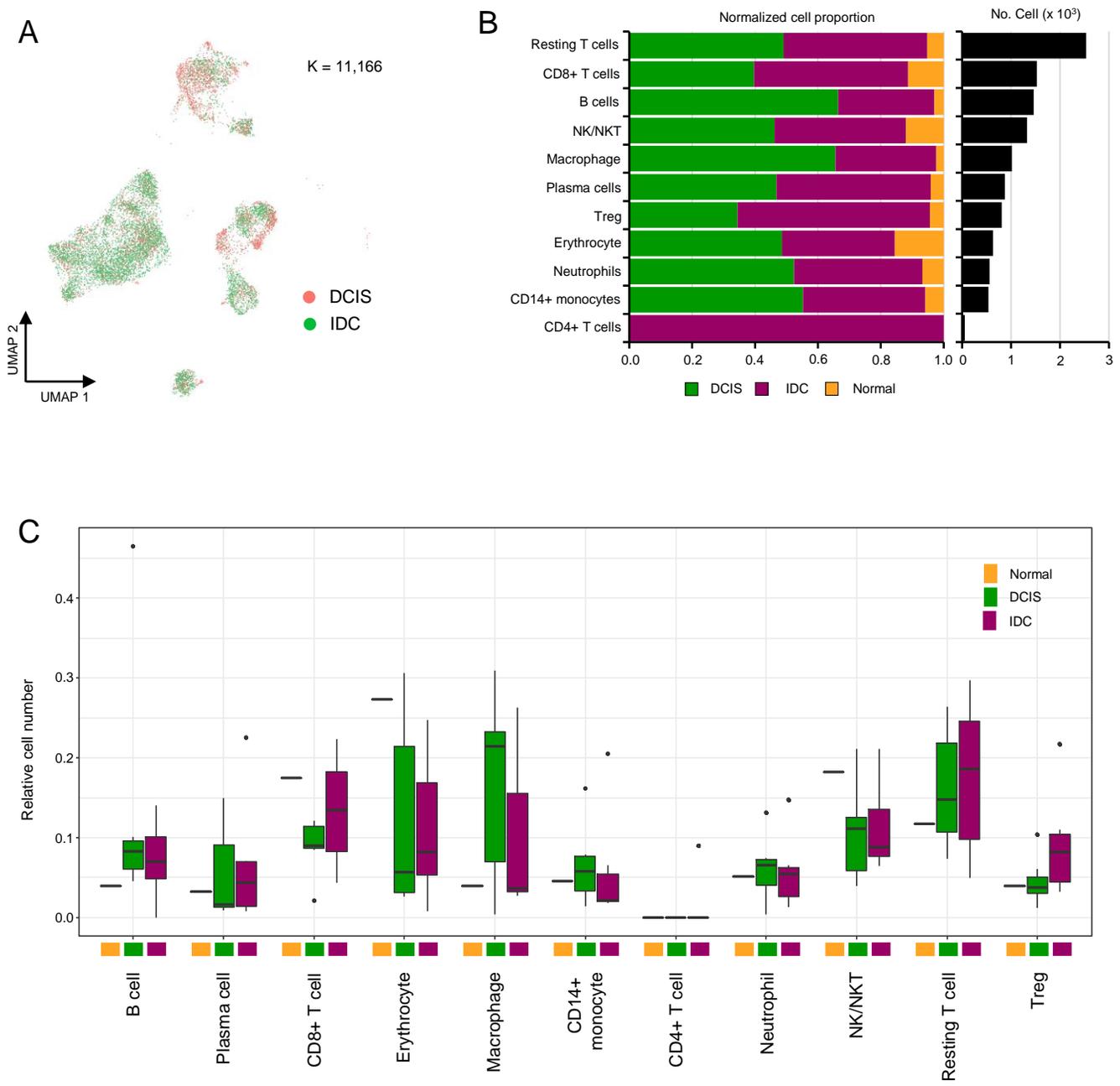
**Supplementary Figure 7 | Spatial transcriptome analysis of the HER2+ DCIS sample (NCCBC6).** **A.** H&E staining images of NCCBC6 on a barcoded slide. **B.** UMAP plot based on spatial transcriptome data of NCCBC6. Based on the gene expression profiles, the cells were divided into 5 clusters. **C.** Spatial information (NCCBC6) of the clusters based on UMAP (Supplementary Figure 7B). **D.** Violin plot of CD24 and KRT18 expression (luminal cell markers) based on cluster. **E.** Violin plot of ERBB2 expression based on cluster. **F.** Spatial information of ERBB2 expression in NCCBC6. **G.** Heatmap showing differentially expressed genes (244 genes, fold change > 1.1 and  $p < 0.05$ ) between Clusters 3 and 4. **H.** Analysis of upstream regulators by IPA. Using 244 genes, upstream regulator genes were predicted. The top 15 significantly activated upstream regulators are shown for Clusters 3 (blue) and 4 (pink). **I.** Signal interaction network with significant enrichment in Cluster 3. Orange: activated, blue: inhibited.

# Supplementary Figure 8



**Supplementary Figure 8** | Boxplots of closeness centrality in the luminal cell population in each DCIS and IDC sample.

# Supplementary Figure 9



**Supplementary Figure 9 | scRNA-seq analysis of the immune cell population in DCIS, IDC and normal breast tissue. A.** UMAP plot of the immune cell population with patient disease states. **B.** Proportions of cells (x-axis) showing the proportions of DCIS, IDC, and normal cells in each immune cell cluster based on Figure 6A (y-axis). The cell numbers in each cluster are shown (right). **C.** Boxplots of immune cell numbers for DCIS, IDC, and normal breast tissue.