iScience, Volume *25*

Supplemental information

Establishment of a reference single-cell RNA sequencing dataset for human pancreatic adenocarcinoma

Ryota Chijimatsu, Shogo Kobayashi, Yu Takeda, Masatoshi Kitakaze, Shotaro Tatekawa, Yasuko Arao, Mika Nakayama, Naohiro Tachibana, Taku Saito, Daisuke Ennishi, Shuta Tomida, Kazuki Sasaki, Daisaku Yamada, Yoshito Tomimaru, Hidenori Takahashi, Daisuke Okuzaki, Daisuke Motooka, Takahito Ohshiro, Masateru Taniguchi, Yutaka Suzuki, Kazuhiko Ogawa, Masaki Mori, Yuichiro Doki, Hidetoshi Eguchi, and Hideshi Ishii

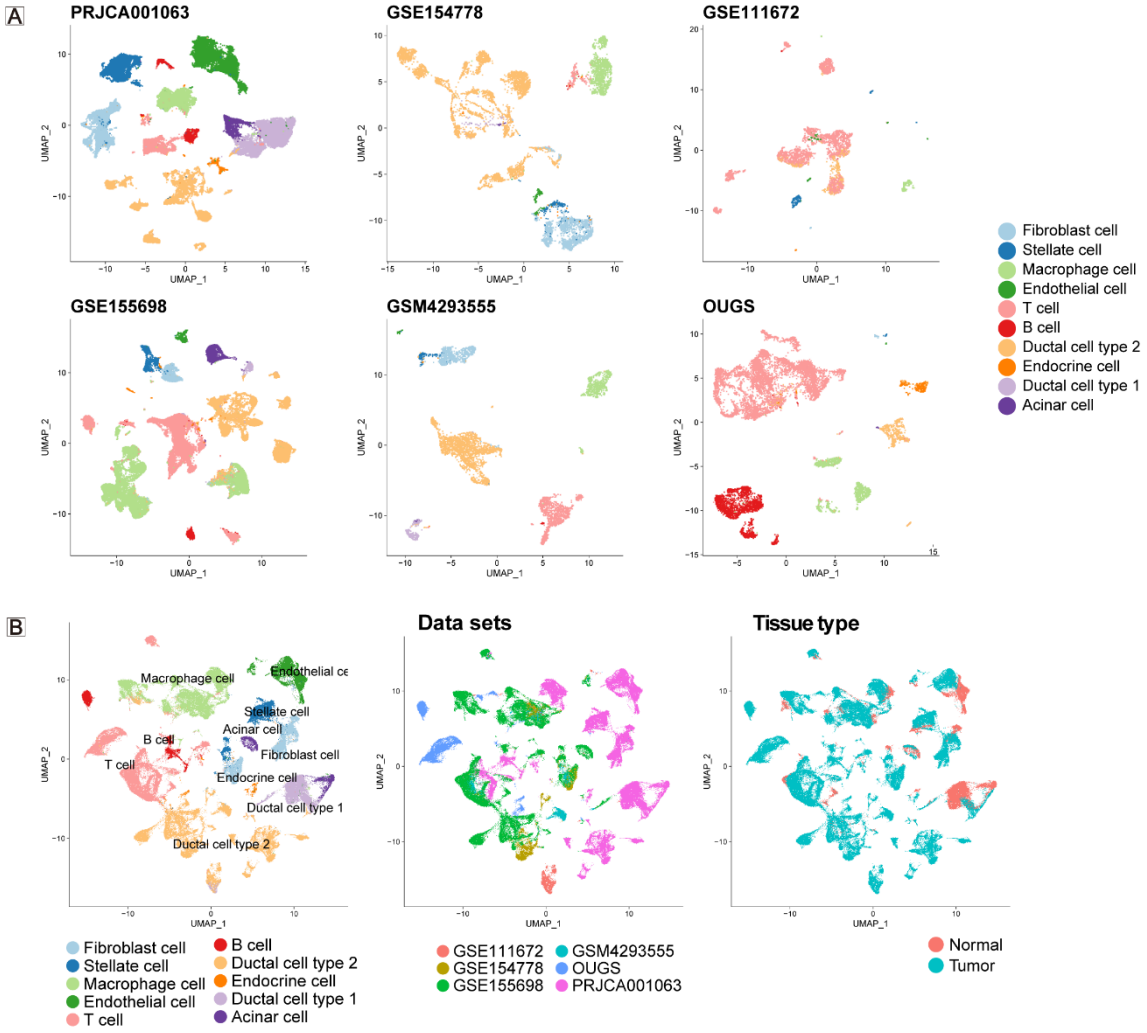

Figure S1. Variety of each scRNAseq dataset and their batch effect. Related to Figure 1.
(A) UMAP plot color-coded according to cell type from each scRNAseq dataset.
(B) Simple data mixture indicates the batch effect among datasets. Merged UMAP plot color-coded according to cell type, dataset, and tissue type.

Figure S2, Related to Figure 1.

(A, B) Stacked bar chart (A) and heatmap (B) showing the proportion of 10 cell types in each individual. The normal pancreas sample is highlighted in red.

(C) Raincloud plots showing the detected number of genes and their UMI count at each cellular level and in each dataset.

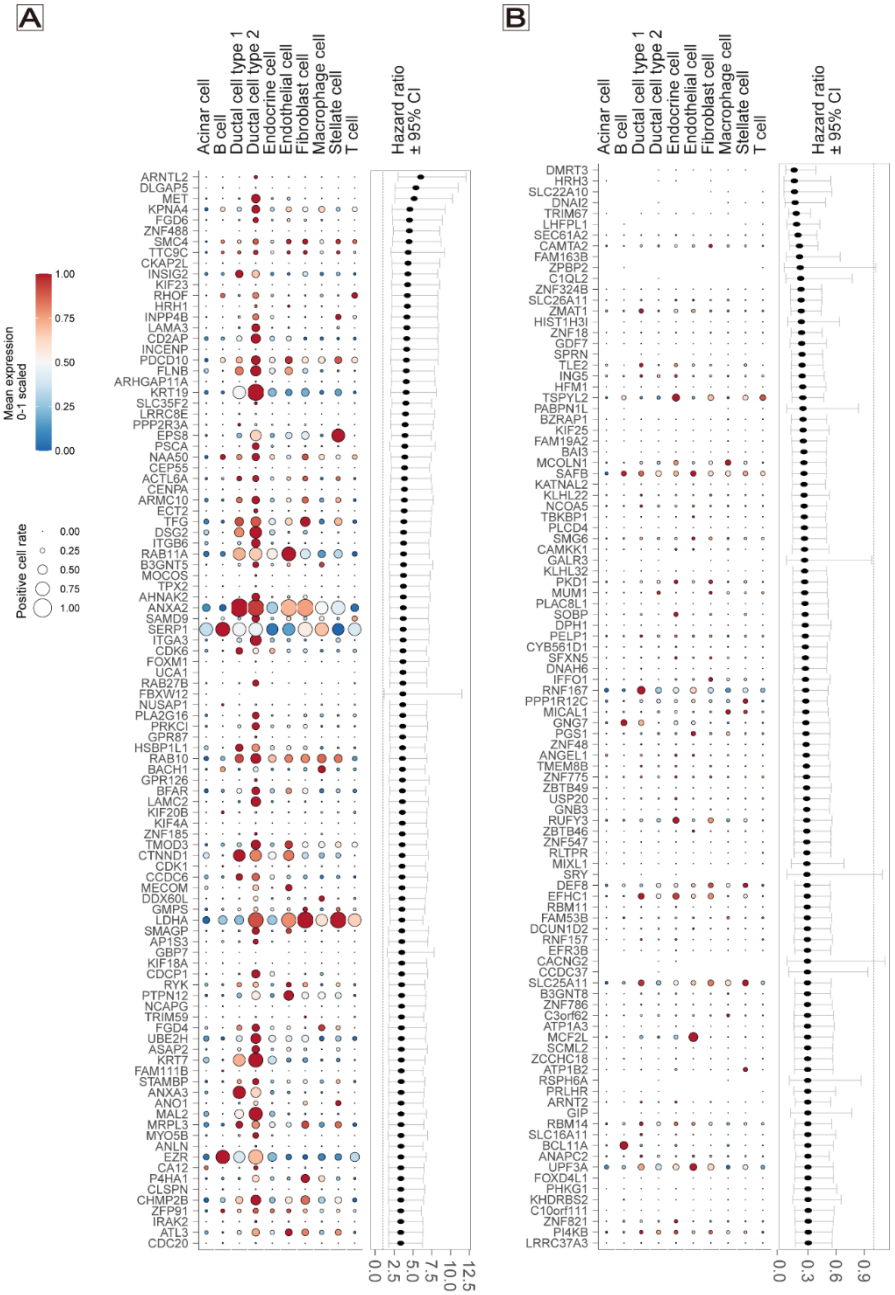

Figure S3, Related to Figure 1

(A, B) Bubble plot showing the cell origin of poor-prognosis (A) or good-prognosis (B) markers in TCGA-PAAD, with genes indicated in rows and cell types in columns. The size of each bubble represents the rate of cells expressing a gene, and the color represents the scaled average expression in their cell-type cluster. Forest plot showing the hazard ratio with 95% confidence interval, which were calculated in the Cox proportional hazards regression model, and the value stems from the comparison between 25% and 75% tiles from four categorized group based on gene expression.

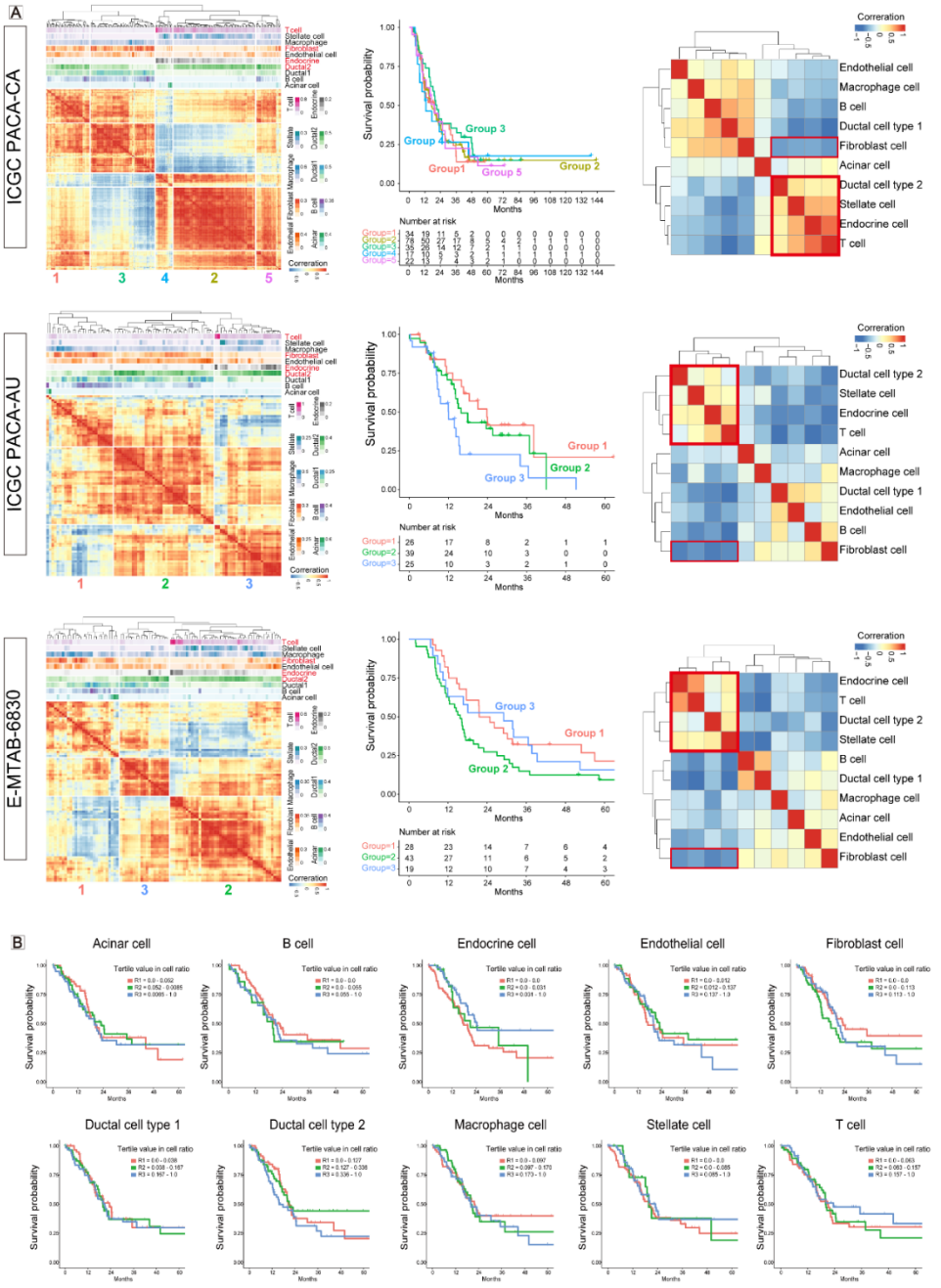

Figure S4, Related to Figure 2. (A) Cellular decomposition for an additional three cohorts that hold RNAseq data. Heatmap of cell ratio and hierarchical clustering based on the correlation among patients showing the heterogeneity of PDAC. Unlike TCGA-PAAD, Kaplan–Meier curves for patient group indicate that the good-prognosis group does not present as malignant cells coexisting with good-prognosis factors, such as T cells. Cellular correlation is comparable in all cohorts: positive/negative correlation of stellate cells, endocrine cells, T cells, and fibroblasts. (B) Kaplan–Meier curve for patients classified into tertile groups based on single-cell type ratio, showing no significant differences.

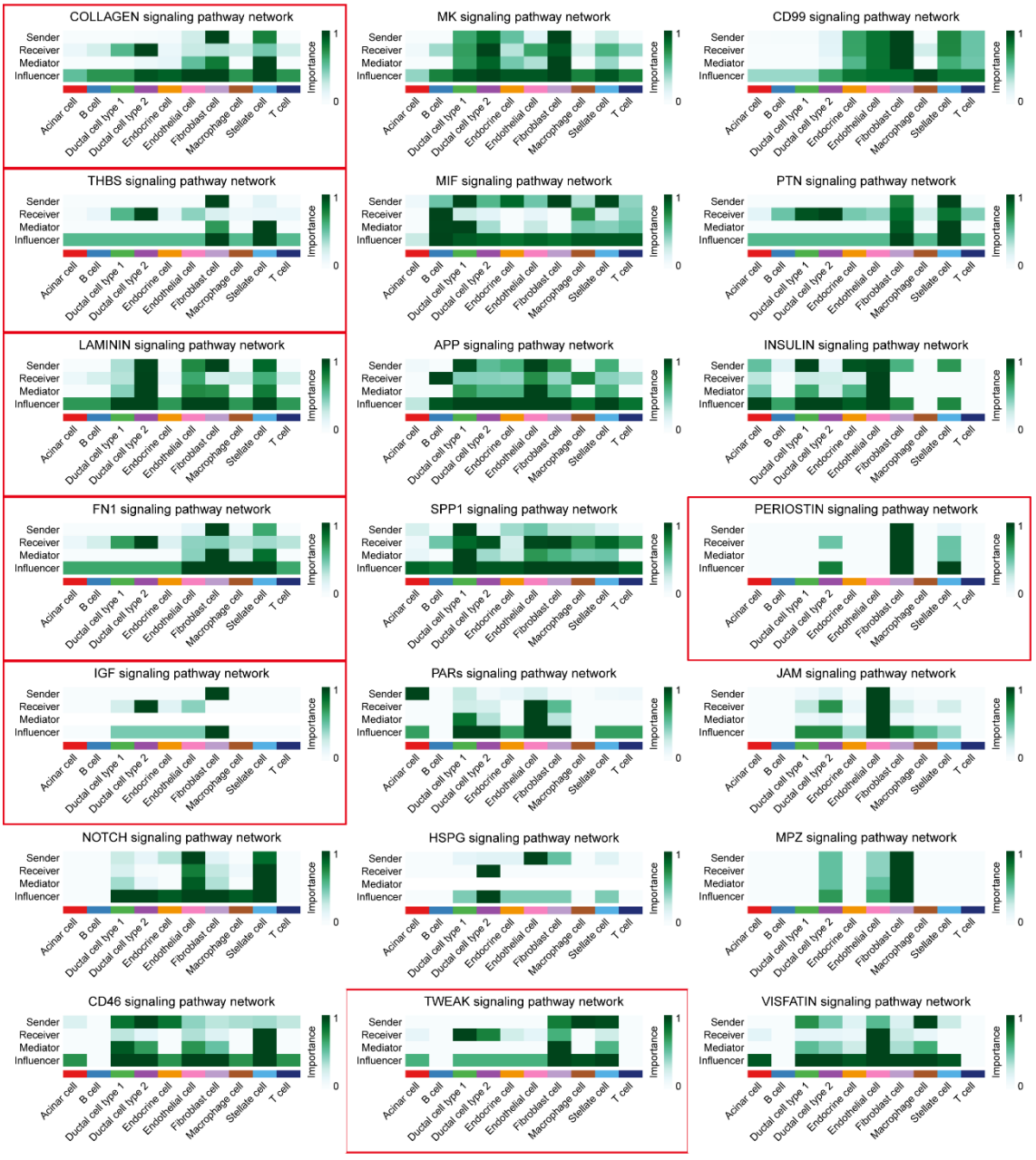

Figure S5, Related to Figure 3

Heatmap summarizing the role of each cell type in each signaling pathway. The plot with the red box is a remarkable pathway in the interaction between CAFs and malignant cells.

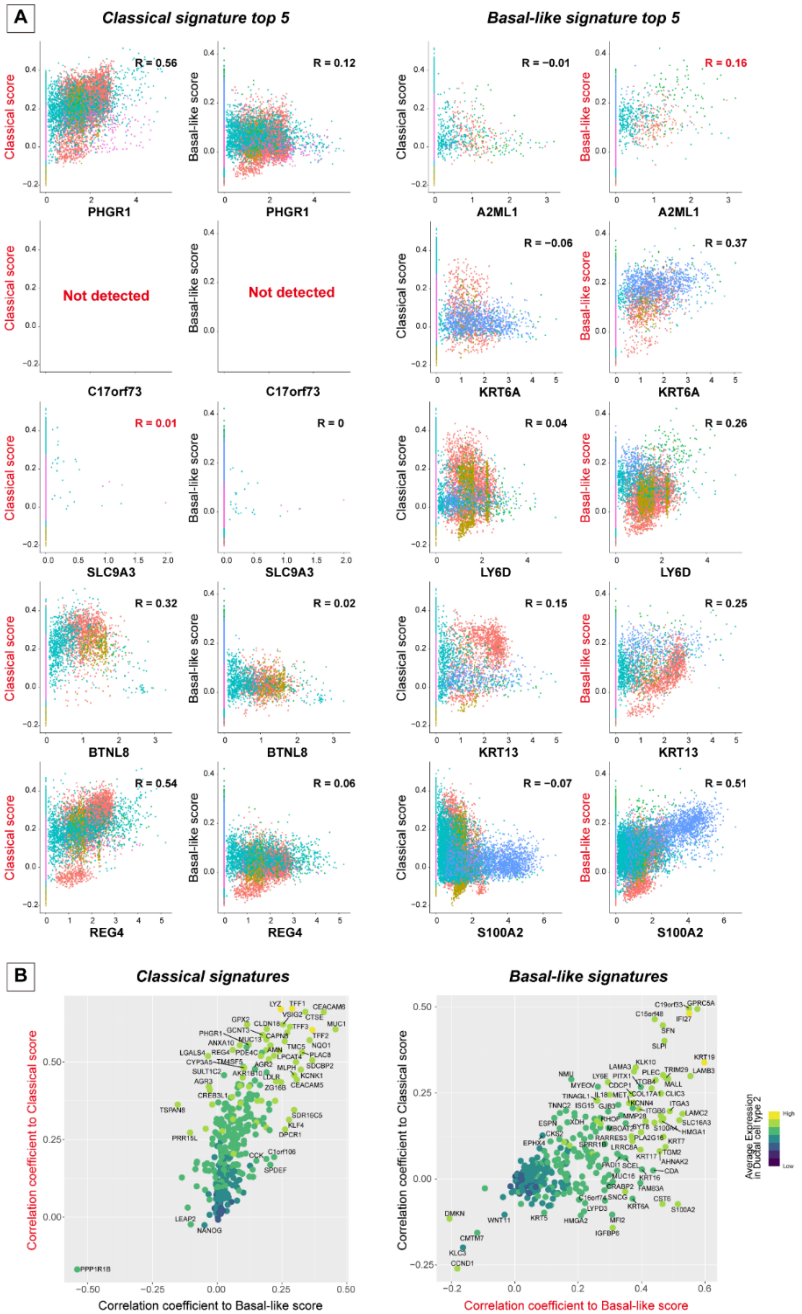

Figure S6. Refinement of the tumor-subtype signature. Related to Figure 4. (A) Pearson's correlation analysis between subtype scores and the expression of top weighted signatures in scRNAseq data, indicating that the signatures defined in the NMF method for TCGA-PAAD were inadequate for application to scRNAseq data. (B) Scatter diagrams of the correlation coefficient of each signature gene defined as (A) showing that some genes contributed to the two scores, whereas others exhibited no correlation. Each dot indicates the subtype signature gene, and the color indicates the expression levels in the malignant cell clusters of scRNAseq data.

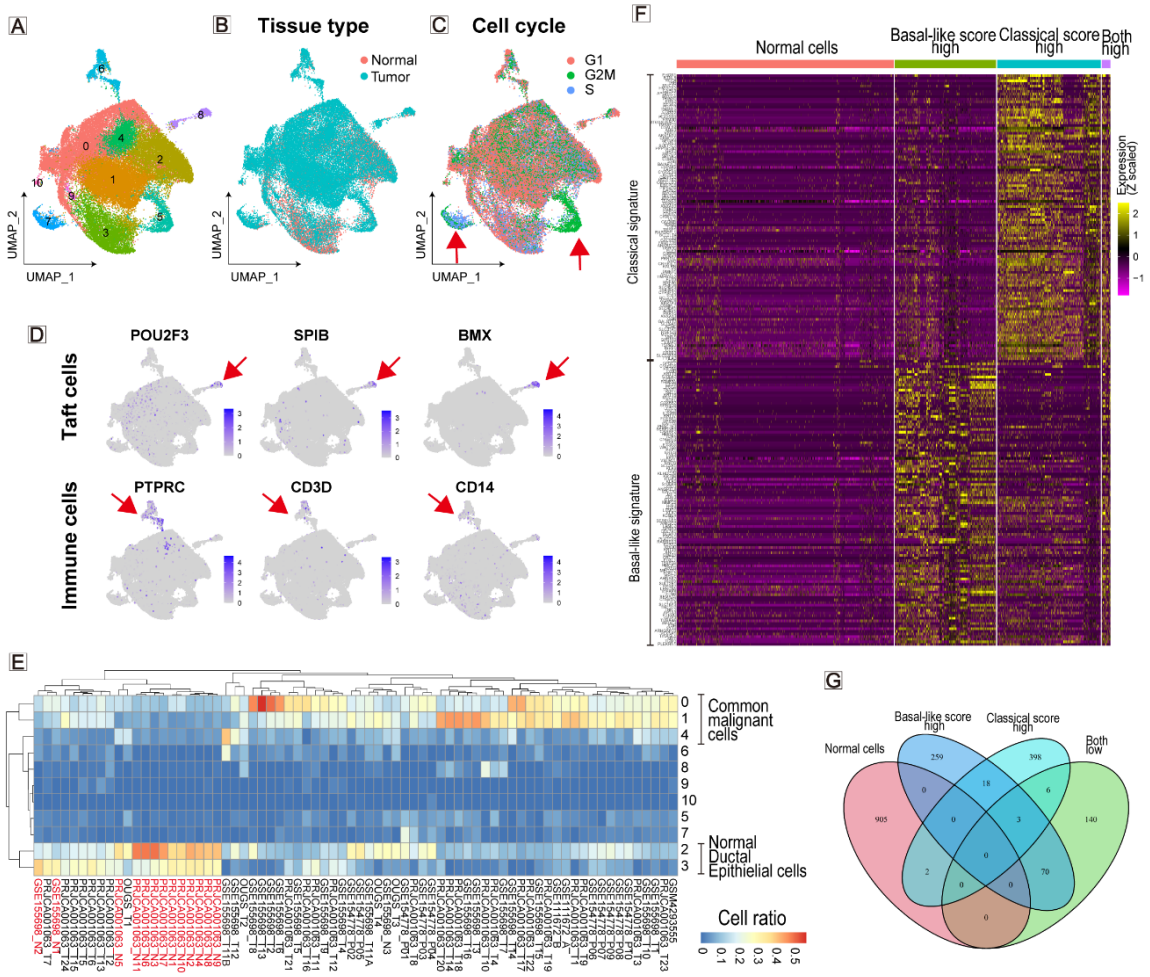

Figure S7, Related to Figure 4.

(A–C) The UMAP of ductal cells after batch correction among individuals, color-coded according to cluster (A), tissue type (B), and cell-cycle stage (C). Cell clusters were defined using an unsupervised clustering method, such as the Seurat pipeline. The stage of the cell cycle was calculated by the Seurat function “CellCycleScoring”. The data showed the normal ductal cell fraction in tumor samples as cluster 2, and clusters 5 and 7 had a biased feature during the cell cycle (red arrows). (D) Marker expression involving taft cells and immune cells is plotted on UMAP. (E) Heatmap of the proportion of ductal cell clusters in each sample indicating the minor population (clusters 9 and 10) not representative in PDAC. The normal pancreas sample is highlighted in red. Clusters 2 and 3 reflect normal ductal cell features. (F) Heatmap showing the expression levels of modified subtype signatures at the cellular level. (G) Venn plot of differentially expressed genes among the data classified according to modified subtype scores.

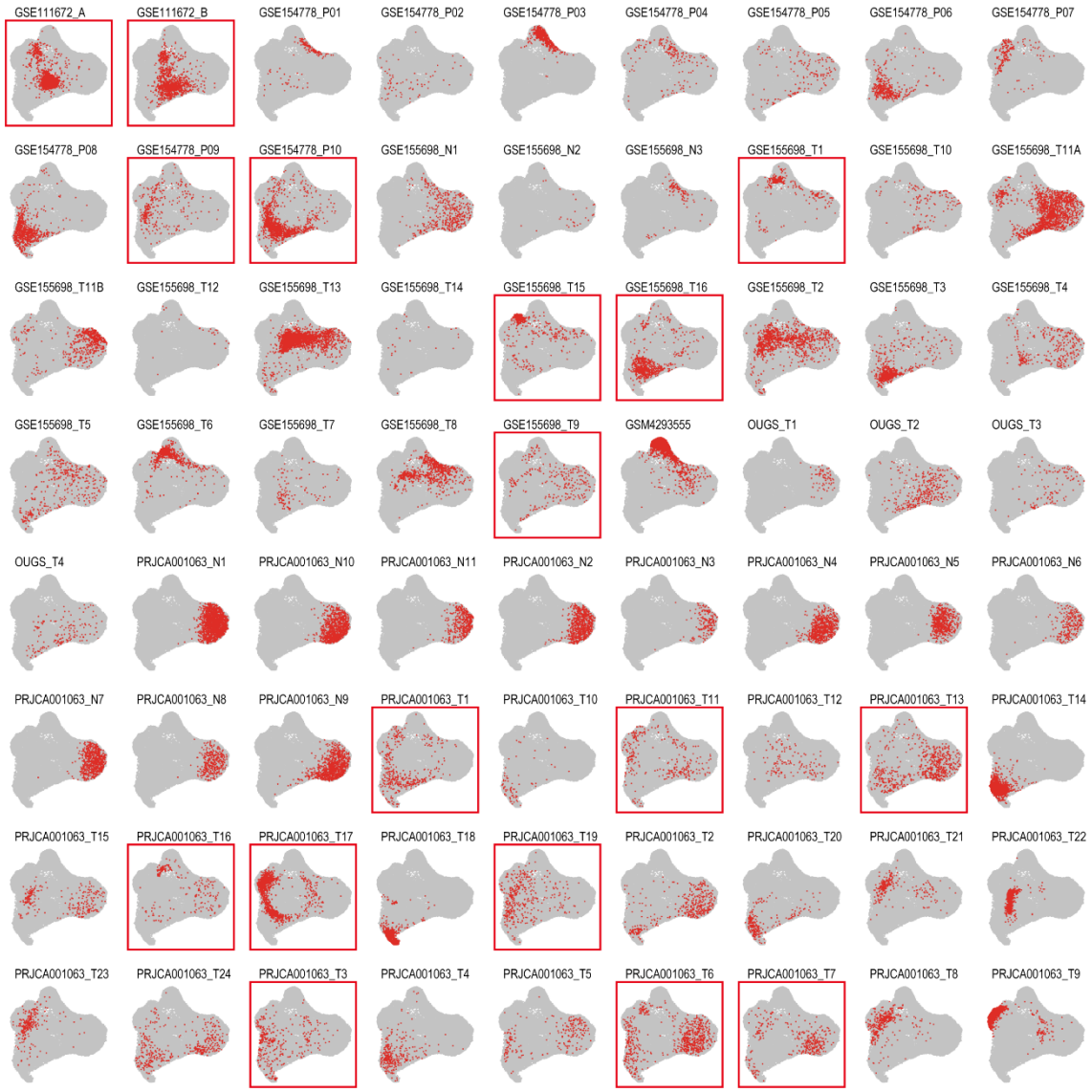

Figure S8. Related to Figure 4.
The distribution on UMAP of each highlighted individual indicates the co-presence of two subtypes of cells in same patient (plots with a red frame).

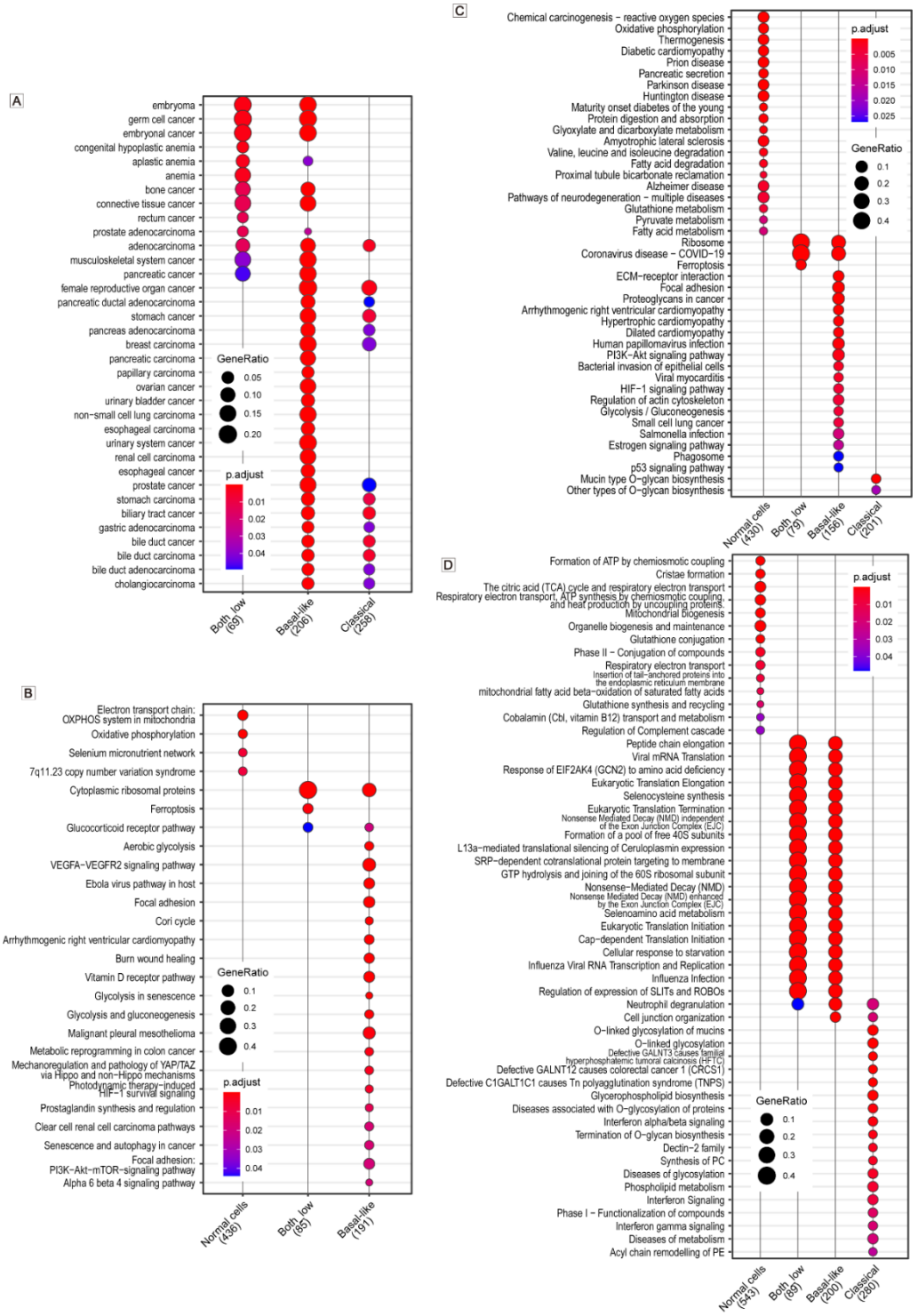

Figure S9. Related to Figure 4.

(A–D) Enrichment analysis of differentially expressed genes in ductal cell subtypes using a disease database (A), WikiPathway (B), KEGG (C), and Reactome (D). The dot size indicates the gene ratio of input DEGs to the given gene set of the pathway, and the color indicates the adjusted *P* values.

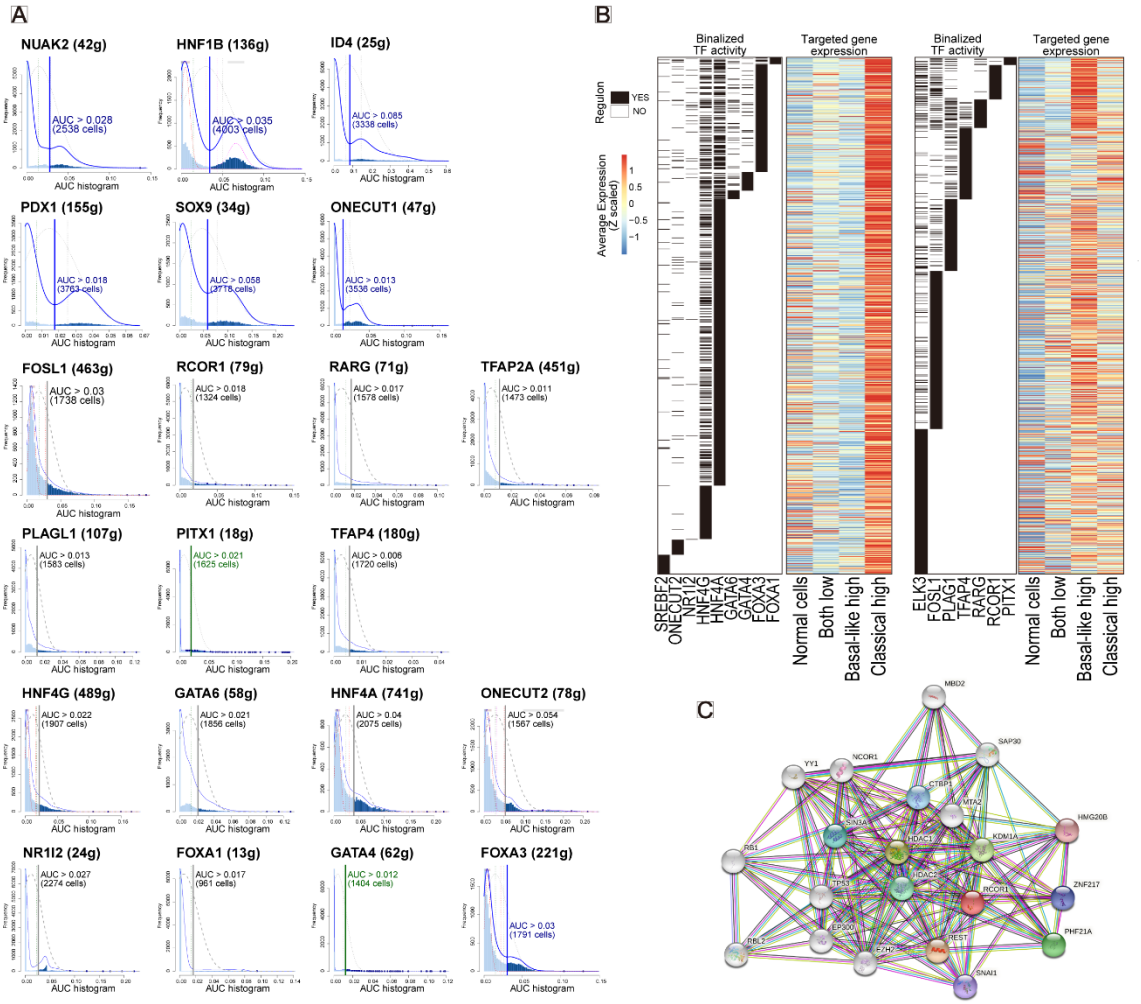

Figure S10. Related to Figure 4.

(A) Histograms of the area under the curve in the recovery curve showing the cutoff value to binarize the regulon activity, which was defined in the R package “SCENIC”.

(B) The relationship between subtype-specific transcription factors and their downstream gene expression is shown in a heatmap, with genes indicated in rows. The black or white colors indicate whether the gene is located downstream of the regulon or not, and continuous color values indicate the average expression.

(C) Protein association network of ROCR1 from the STRING database.

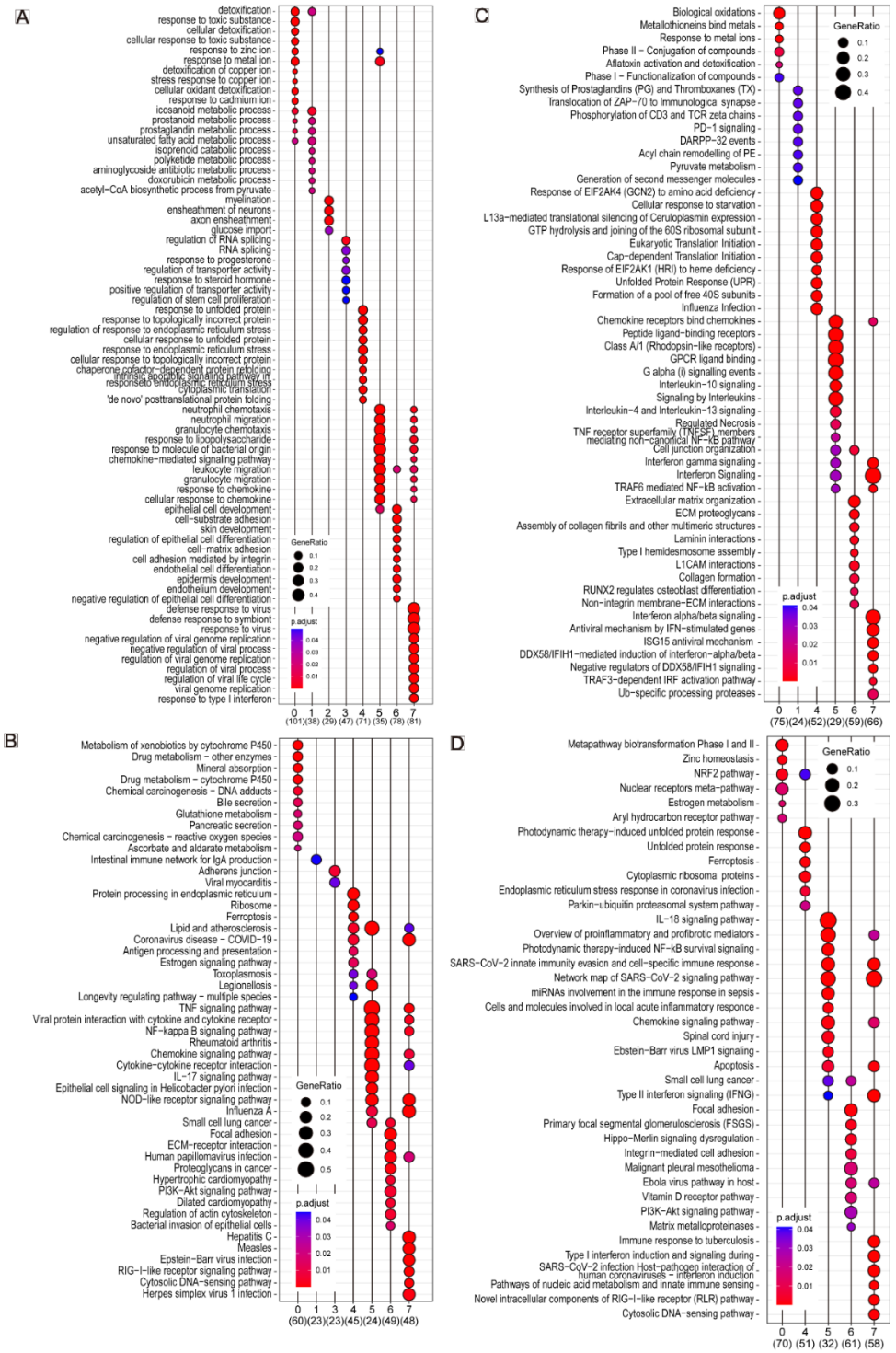

Figure S11. Related to Figure 5. (A–D) Enrichment analysis of differentially expressed genes in malignant cell clusters using Gene Ontology Biological Process (A), KEGG (B), KEGG (C), Reactome (C), and WikiPathway. The dot size indicates the gene ratio of input DEGs to the given gene set of the pathway, and the color indicates the adjusted *P* values.

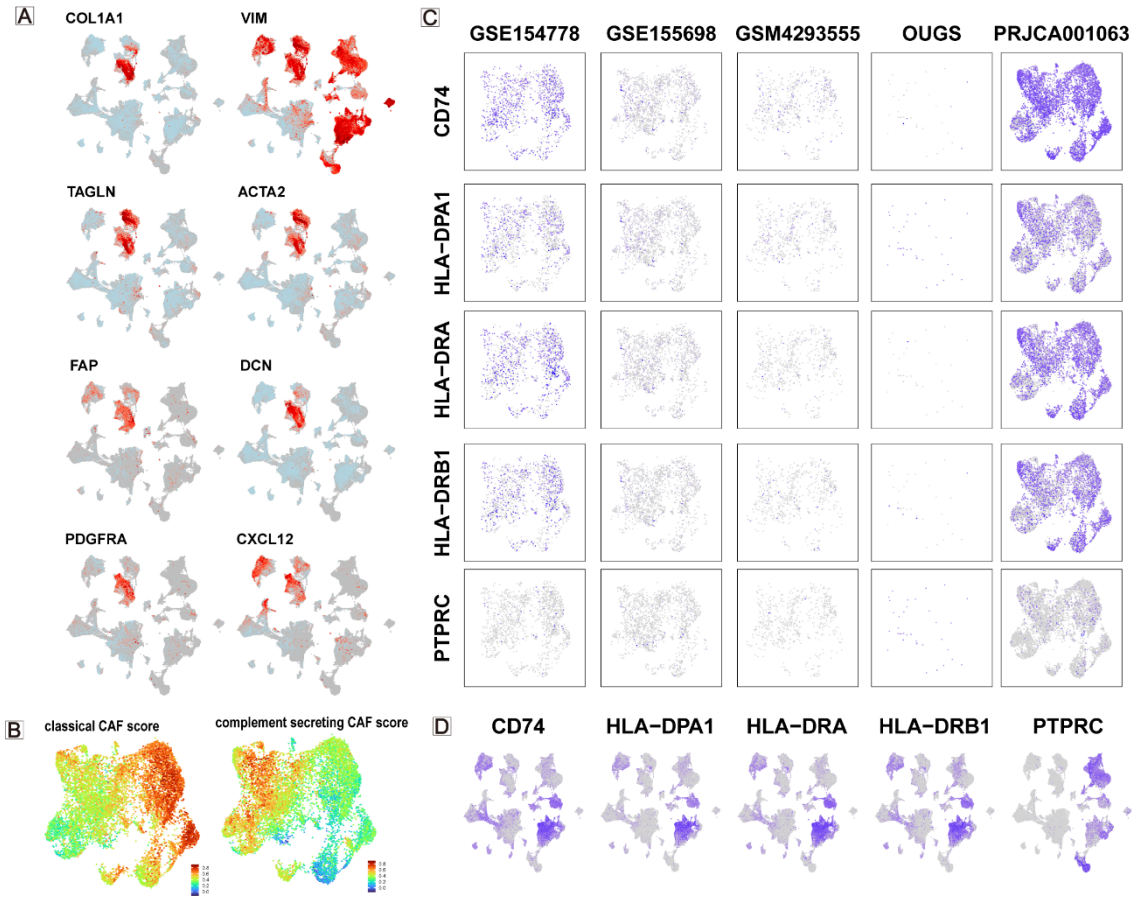

Figure S12. Related to Figure 6.
(A) UMAP from all cell types showing the expression levels of marker genes for pan-CAFs, inflammatory CAFs, and myofibroblast CAFs.
(B) UMAP plots showing the supposed classical CAF score and complement-secreting CAF score.
(C, D) Marker expression of antigen-presenting CAFs in the UMAP from CAF subsets (C) and all cell types (D).

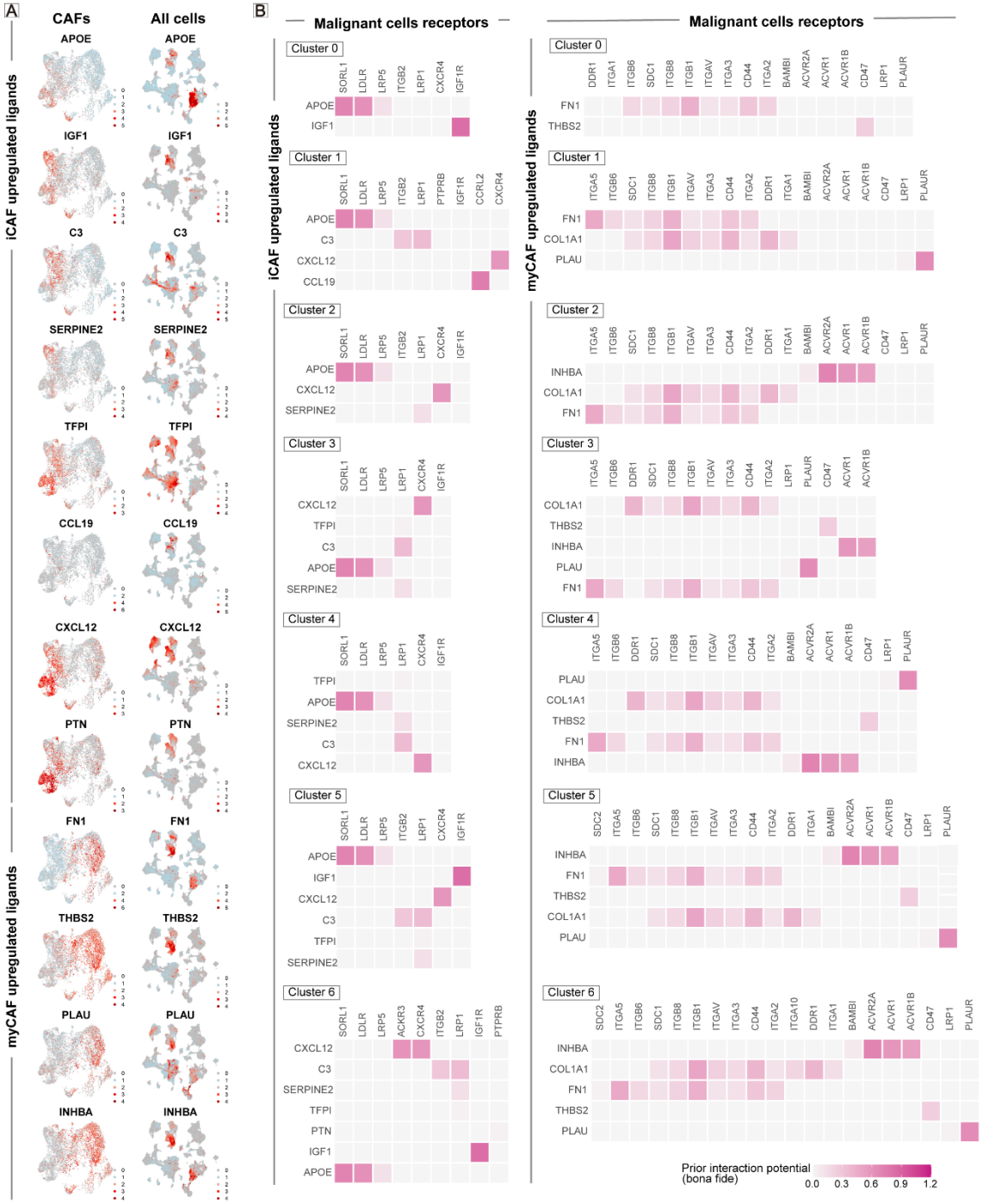

Figure S13. Related to Figure 6

(A) UMAP of the CAF subset or all cell types showing the expression levels of upregulated genes in iCAFs and myCAFs.

(B) Heatmap showing ligand–receptor interactions with putative potential in all combinations of two CAF types and six malignant clusters.