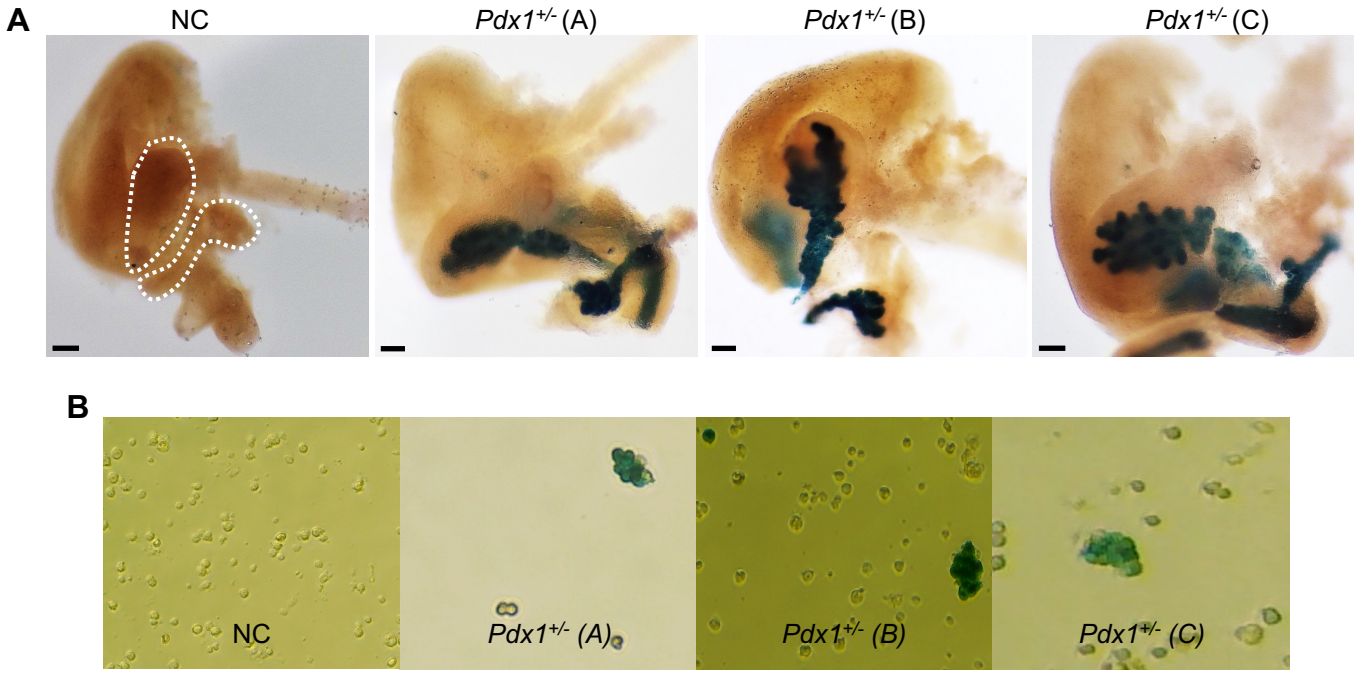
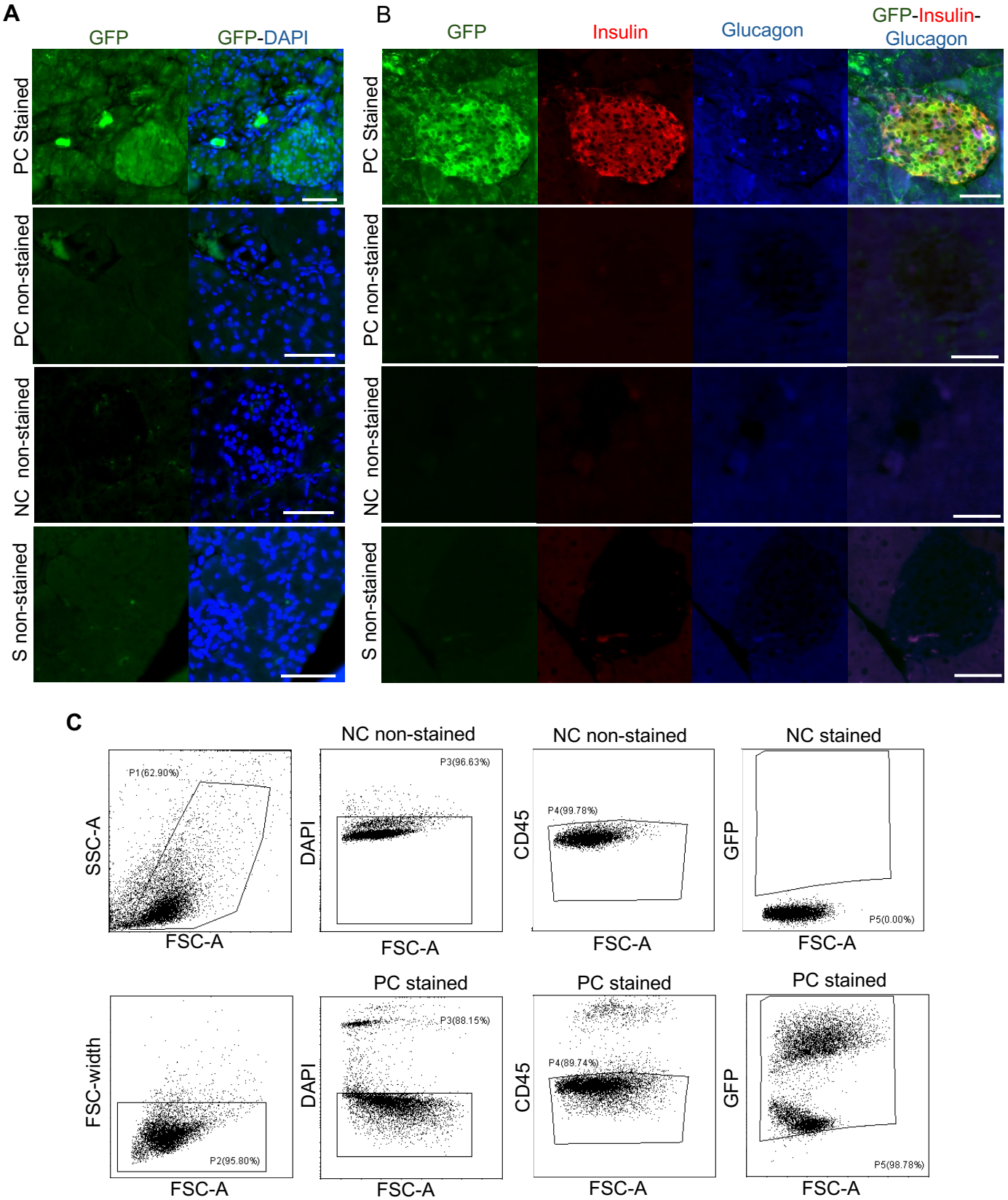
Supplementary Figure 1



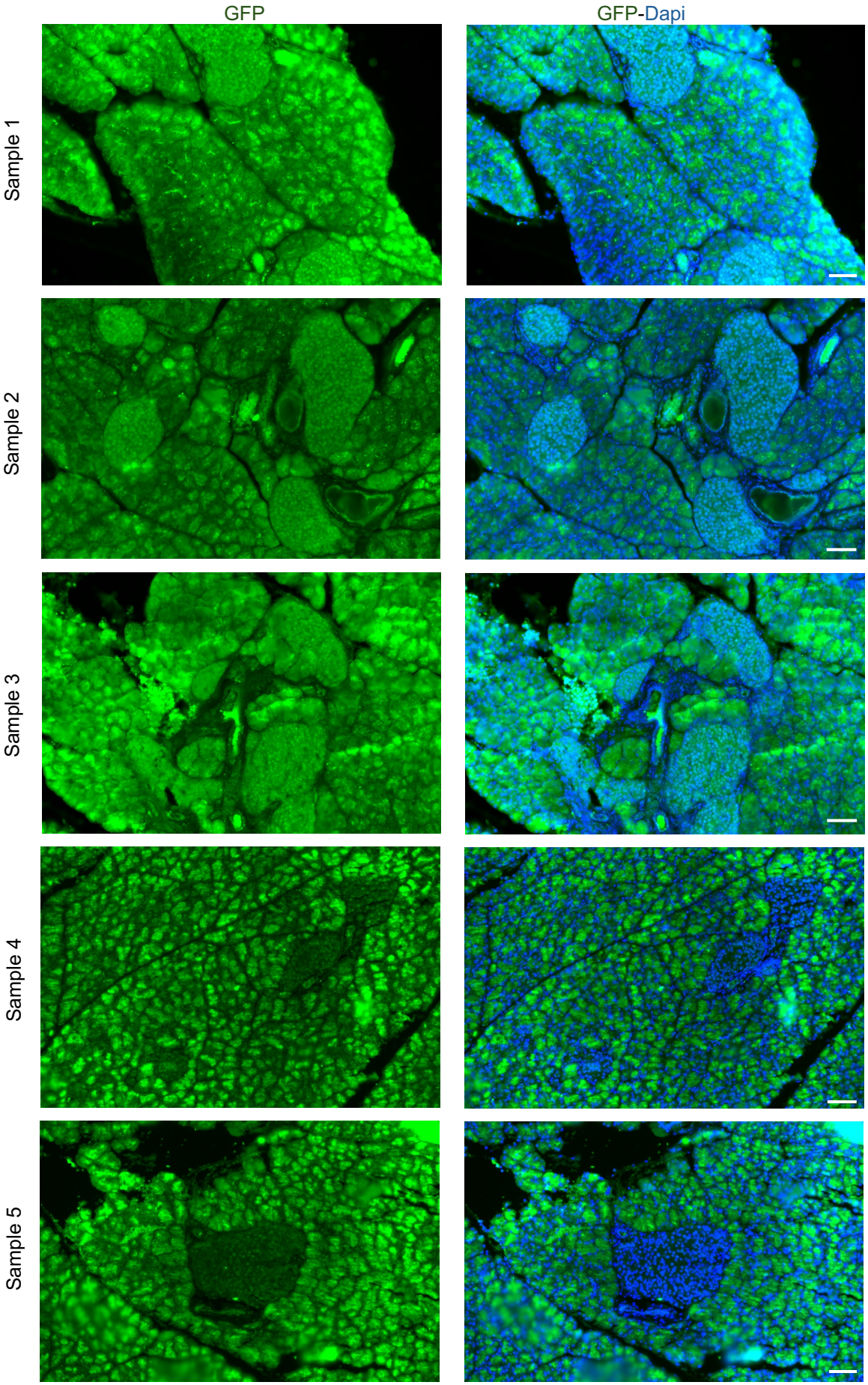
Supplementary Figure 1 | Validation of embryonic pancreatic bud isolation using *Pdx1* reporter activity.
(A) Whole-mount β-galactosidase staining of E12.5 embryos derived from negative control (wild-type) shown as NC and *Pdx1*$^{+/-}$ mice. The dashed outline indicates the anatomical region corresponding to the dorsal and ventral pancreatic bud. Blue staining marks endogenous *Pdx1* transcriptional activity within the developing pancreatic buds. No staining was observed in wild-type embryos, whereas strong reporter activity was detected in *Pdx1*$^{+/-}$ embryos, confirming localization of the pancreatic progenitor domain. Scale bar, 0.1 mm. (B) β-galactosidase staining of dissociated single-cell suspensions derived from microdissected tissues. Robust blue staining was detected in samples obtained from *Pdx1*$^{LacZ/+}$ embryos, whereas anatomically matched cells from wild-type controls tissue showed no detectable signal, confirming accurate isolation of Pdx1^{+} pancreatic progenitor cells. Wild

Supplementary Figure 2



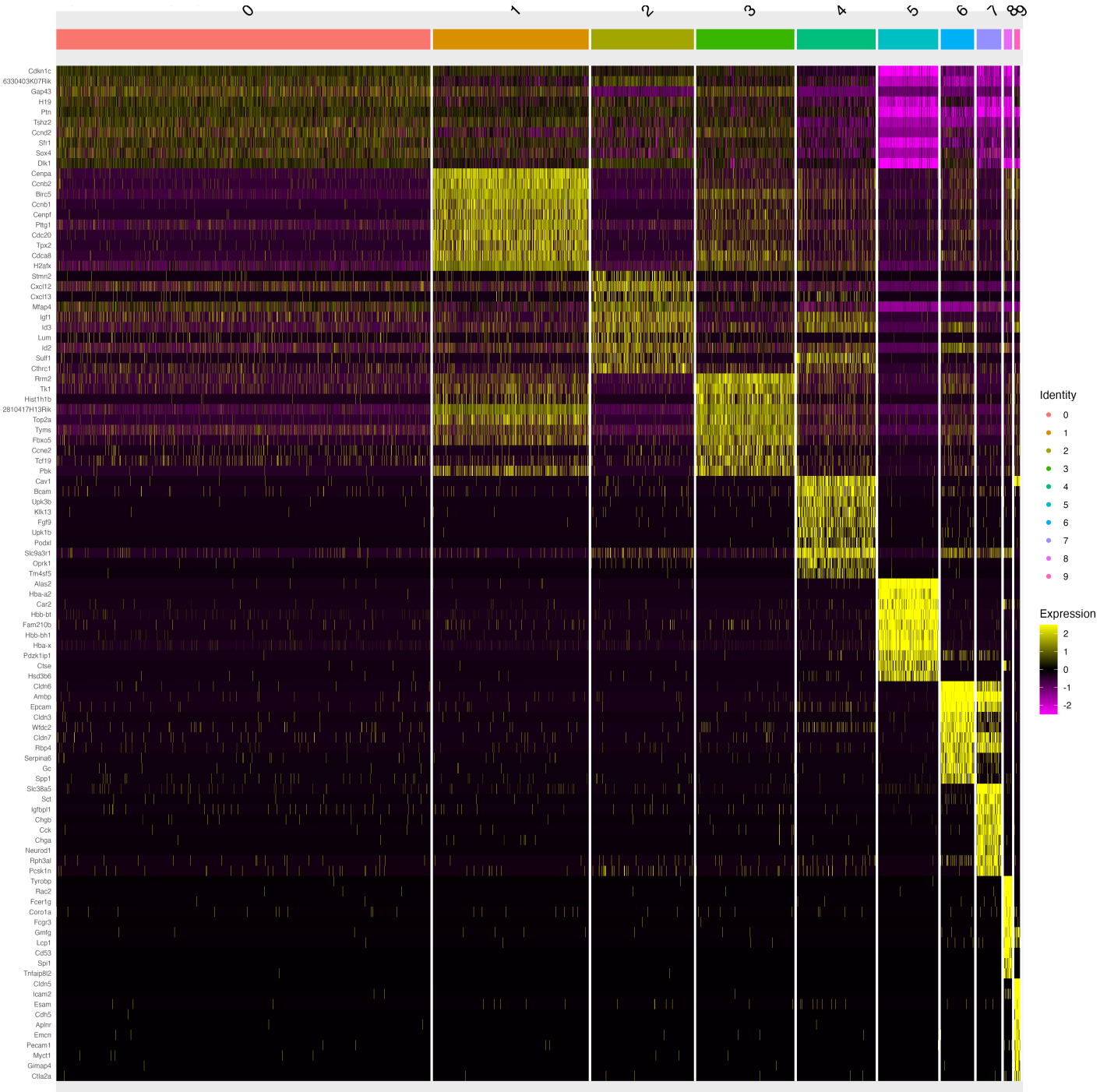
Supplementary Figure 2 | Validation of immunostaining specificity and flow cytometric gating strategy
(A) Immunofluorescence controls for GFP detection. Positive control (PC) pancreatic sections derived from GFP$^+$ tissue show robust GFP signal, whereas no-primary antibody controls in sample (S non-stained) and non-transplanted negative control (NC) tissues exhibit no detectable fluorescence, confirming staining specificity. Nuclei were counterstained with DAPI, Scale bars=50 μm. (B) Validation of endocrine marker staining. GFP, insulin, and glucagon immunostaining demonstrates expected signal detection in positive control sections, while omission of primary antibodies abolishes signal in control samples, confirming antibody specificity, Scale bars=50 μm. (C) Flow cytometry gating strategy used for quantification of donor-derived cells. Sequential gates were applied to exclude debris (FSC/SSC), doublets (FSC-width vs FSC-area), dead cells (DAPI$^+$), and CD45$^+$ hematopoietic cells prior to GFP analysis. Representative plots from negative controls and positive control samples are shown.

Suppementary Figure 3

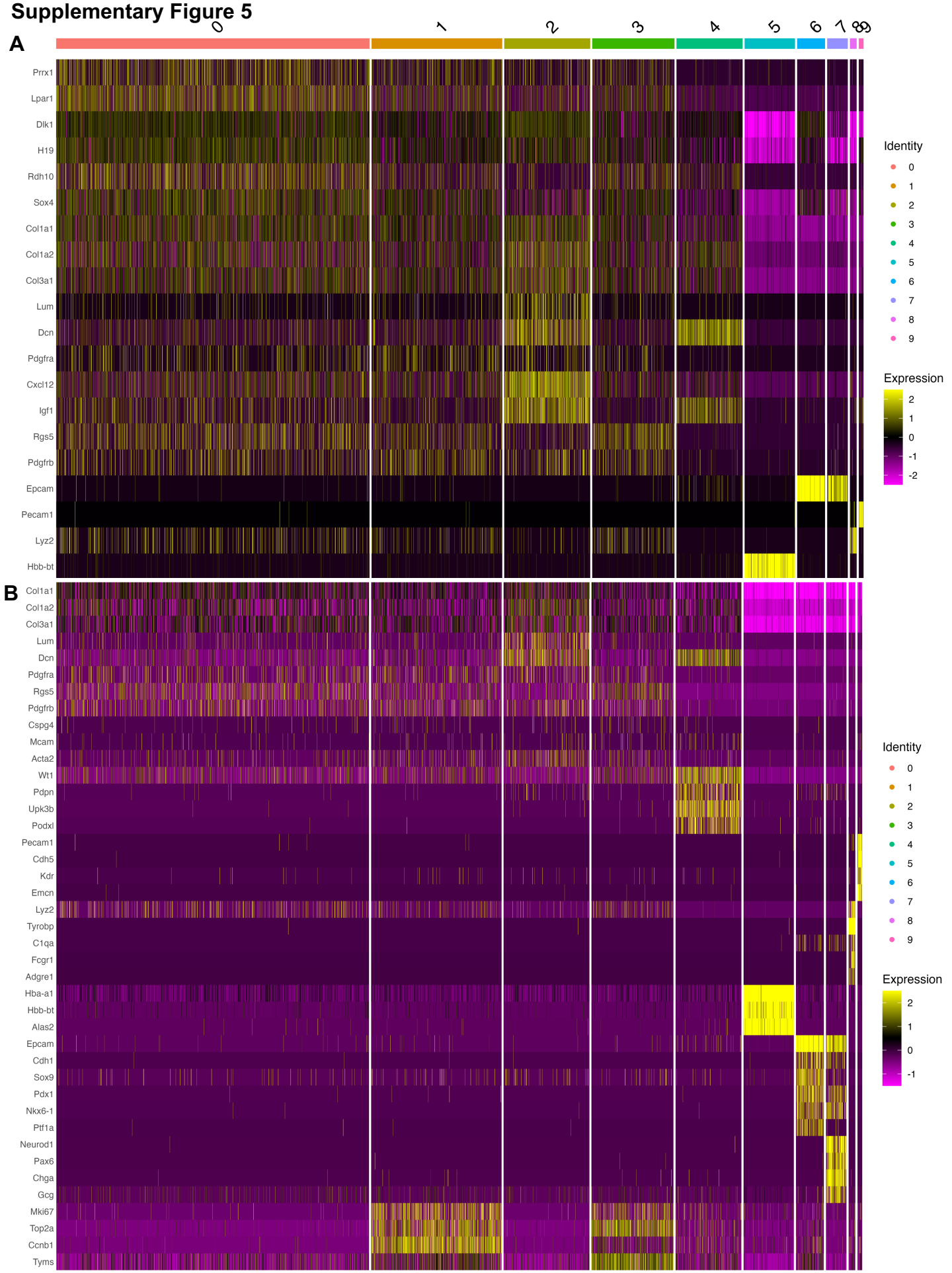


Supplementary Figure 3 | Representative GFP⁺ donor contribution in transplanted non-KO pancreata.
Fluorescence images of pancreatic tissue from five representative transplanted non-KO recipients (Samples 1–5) showing GFP⁺ donor-derived cells (green). Right panels show merged GFP and DAPI nuclear staining (blue). Donor contribution varies among individuals, with some samples showing broader multilineage integration and others displaying predominantly acinar labeling. Scale bars, 50 μm.

Supplementary Figure 4



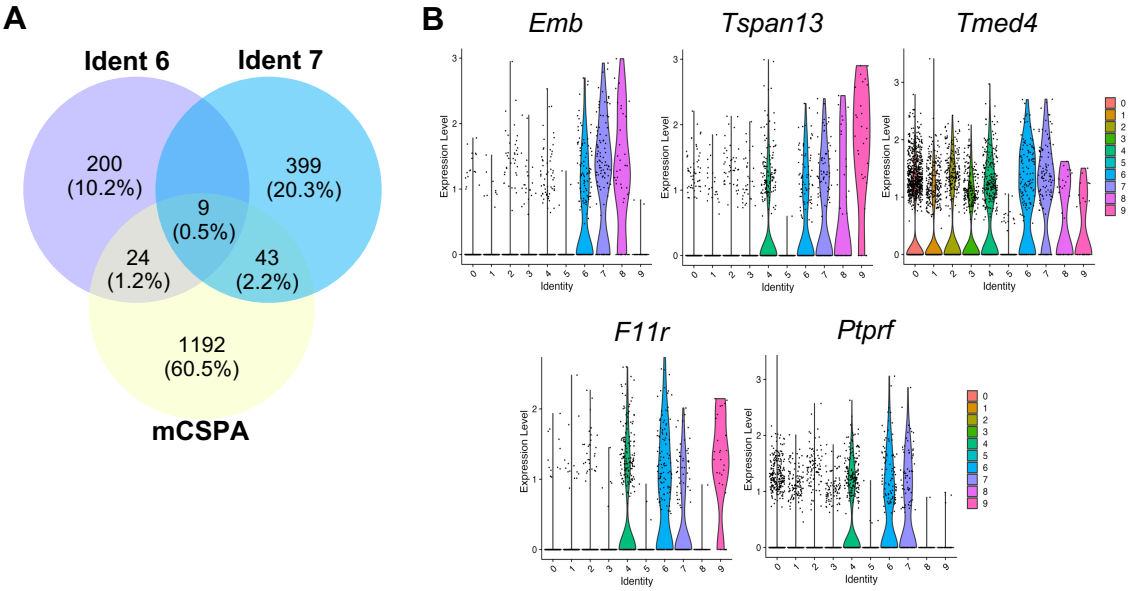
Supplementary Figure 4 | Top differentially expressed genes per cluster in the GSM2699156 dataset.
Heatmap showing the top differentially expressed genes identified for each cluster following unsupervised clustering of embryonic mouse pancreas scRNA-seq data (GSM2699156). Expression values are scaled (z-score) across cells. Distinct expression patterns confirm clear transcriptional separation among the 10 clusters.



Supplementary Figure 5 | Lineage marker validation and epithelial/progenitor marker expression in GSM2699156 clusters.

(A) Heatmap showing expression of core lineage markers used for cluster annotation. (B) Heatmap showing epithelial and pancreatic progenitor markers used for cluster resolution, supporting identification of pancreatic epithelial and endocrine progenitor populations.

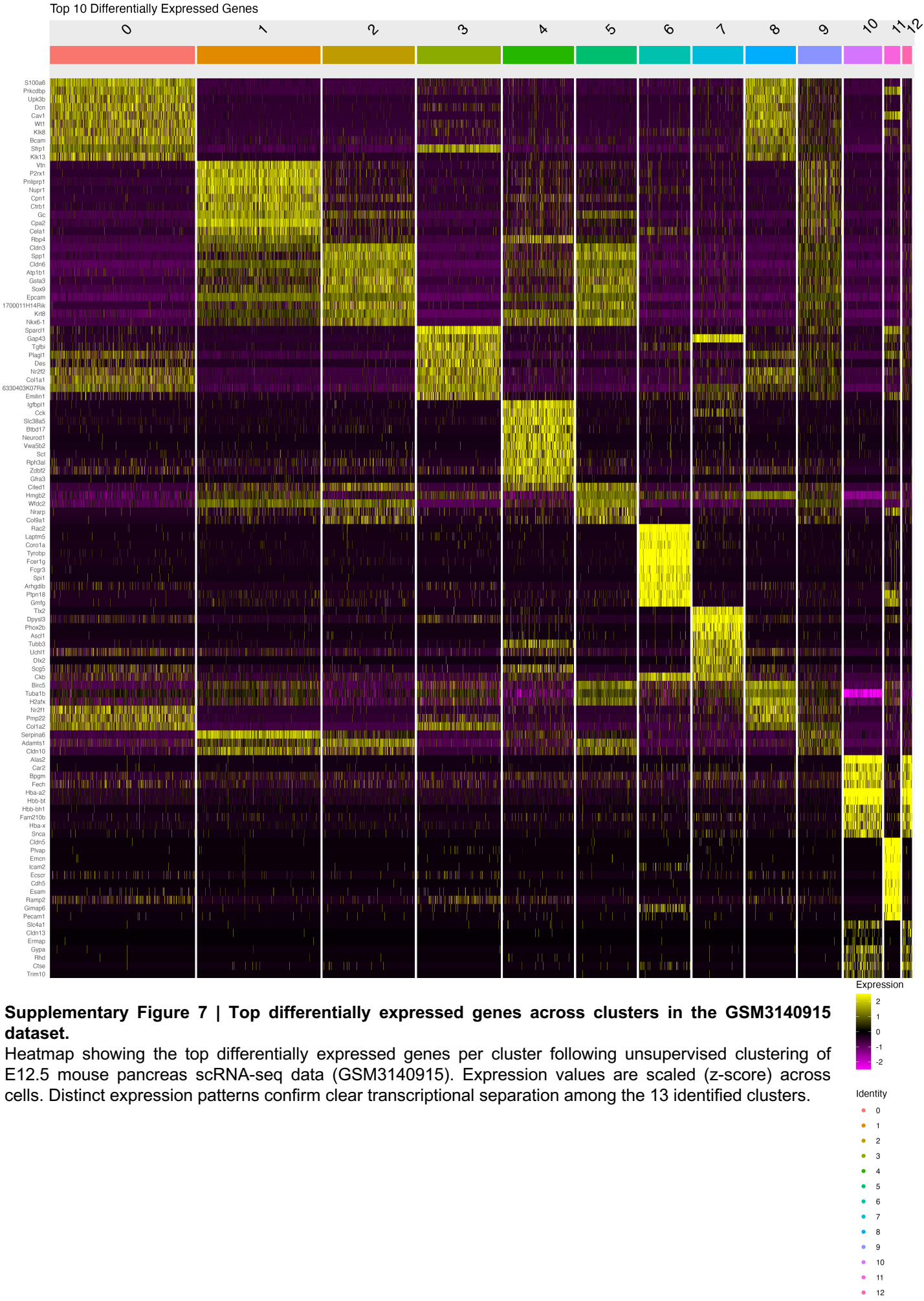
Supplementary Figure 6



Supplementary Figure 6 | Identification of candidate surface markers enriched in Pdx1-expressing epithelial clusters.

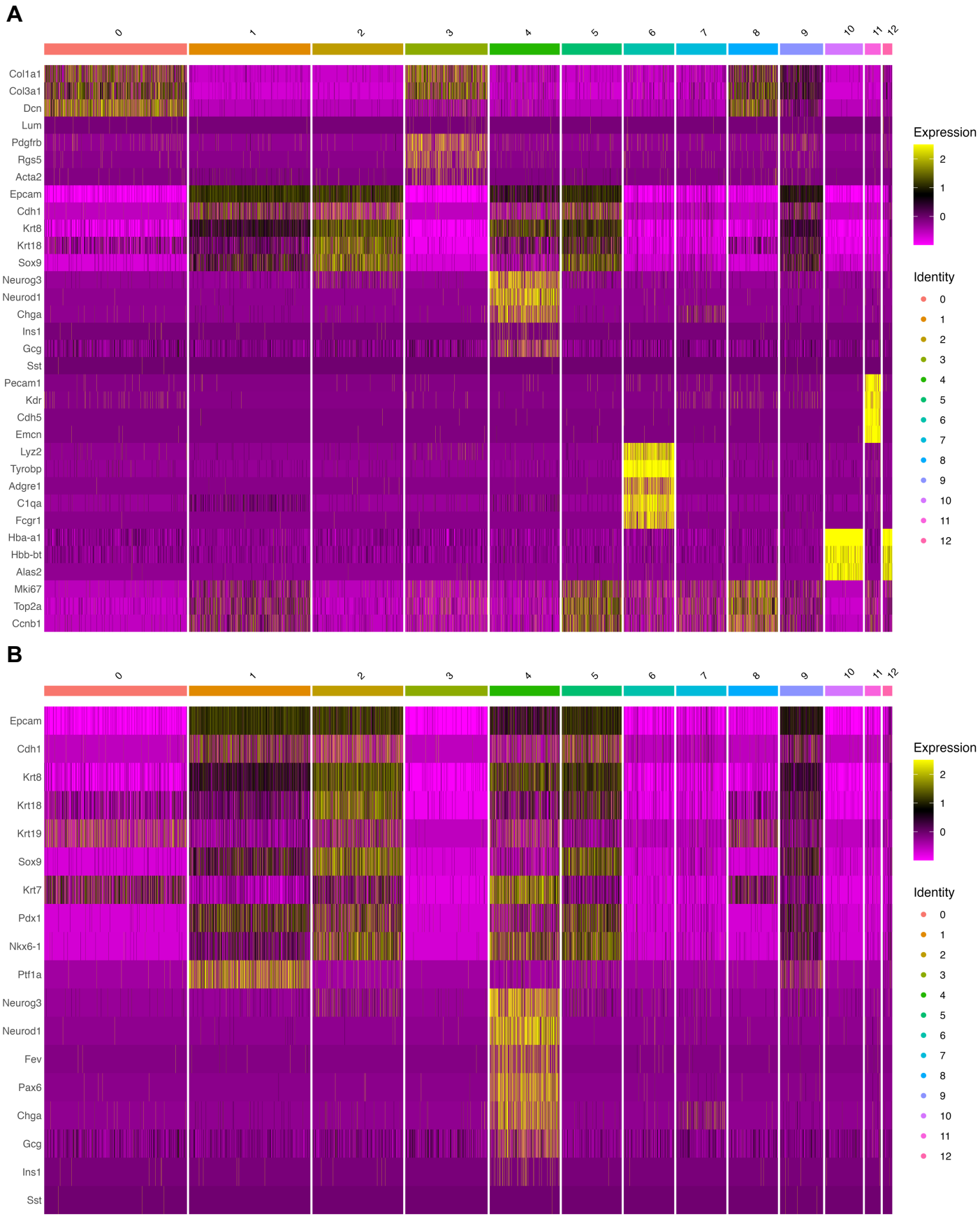
(A) Venn diagram showing overlap between genes enriched in epithelial clusters 6 and 7 (avg-$\log_2$FC > 1, adjusted P < 0.001) and surface proteins listed in the Cell Surface Protein Atlas (mCSPA), identifying candidate surface markers. (B) Violin plots showing expression of selected candidates (*Emb, Tspan13, Tmed4, F11r, and Ptprf*) across clusters. These genes show broader expression across multiple cell populations compared with the epithelial-restricted markers presented in Fig. 3.

Supplementary Figure 7



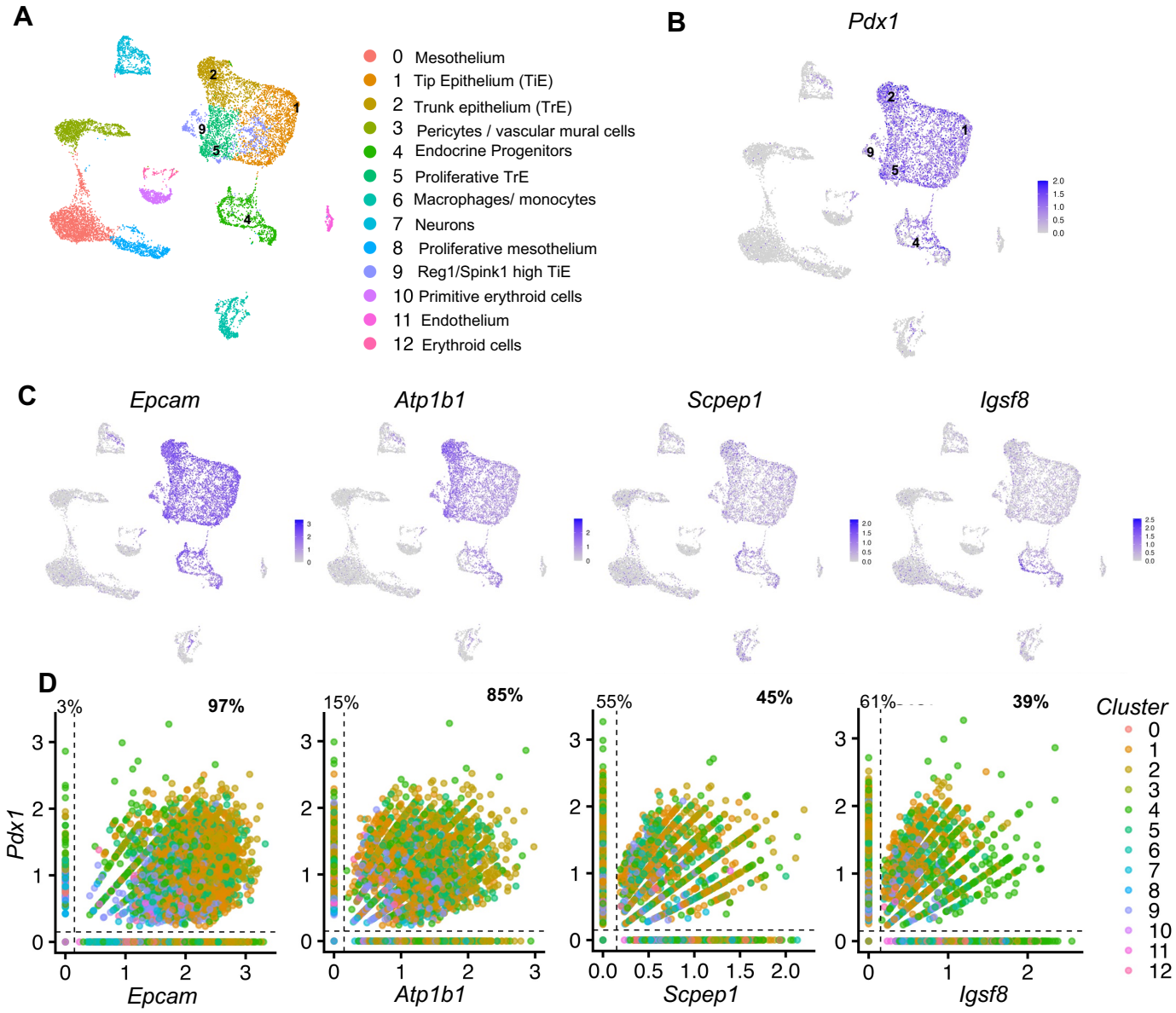
Supplementary Figure 7 | Top differentially expressed genes across clusters in the GSM3140915 dataset.
Heatmap showing the top differentially expressed genes per cluster following unsupervised clustering of E12.5 mouse pancreas scRNA-seq data (GSM3140915). Expression values are scaled (z-score) across cells. Distinct expression patterns confirm clear transcriptional separation among the 13 identified clusters.

Supplementary Figure 8



Supplementary Figure 8 | Lineage marker–based annotation of GSM3140915 clusters.
(A) Heatmap showing expression of core lineage markers used for cluster annotation. (B) Heatmap showing epithelial and pancreatic progenitor markers used to refine epithelial cluster identity.

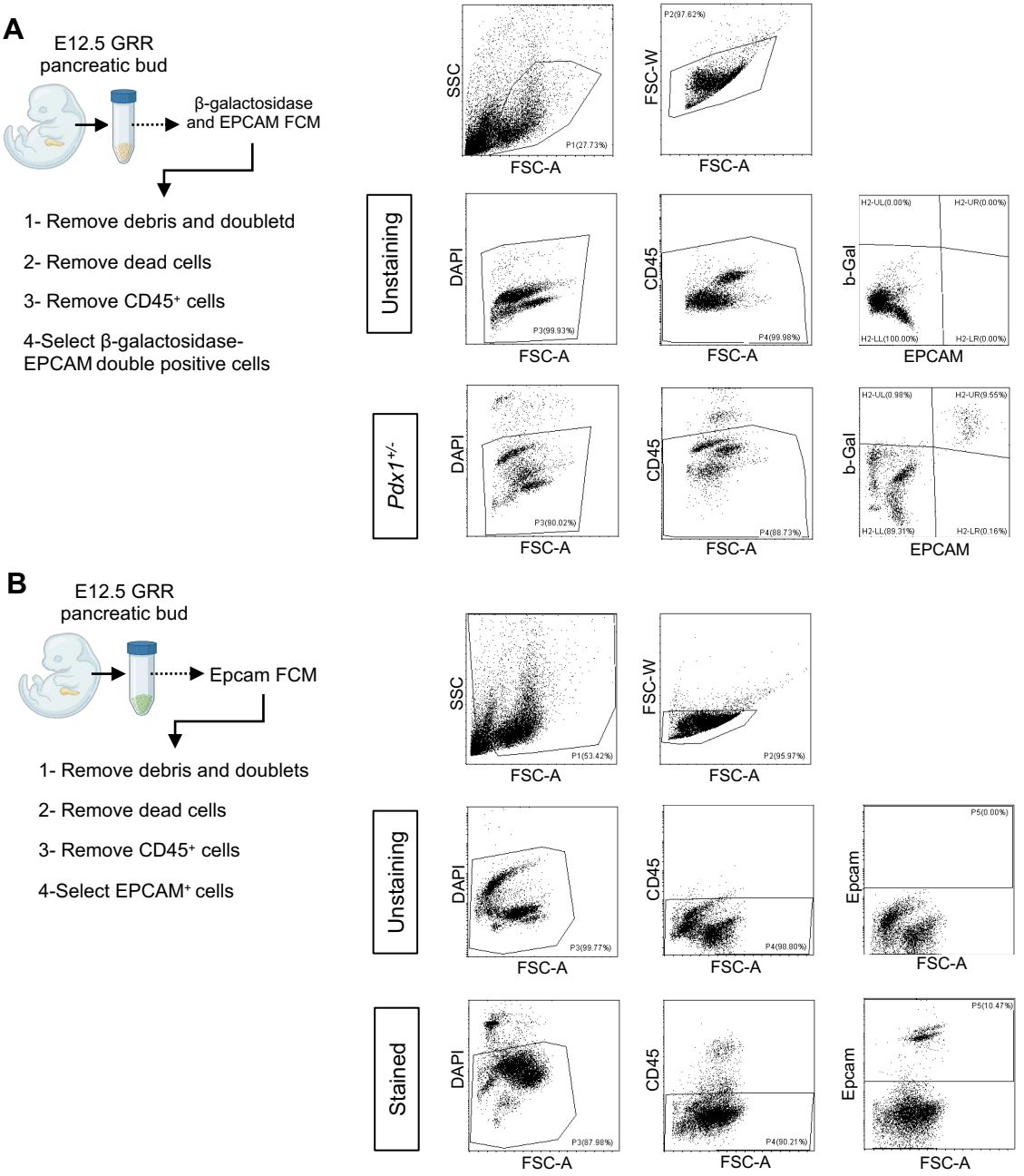
Supplementary Figure 9

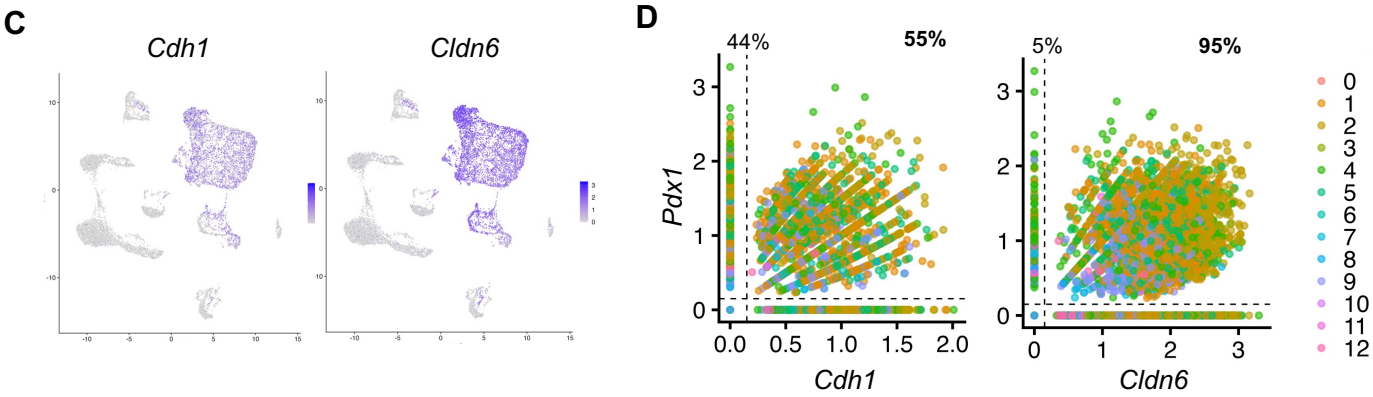


Supplementary Figure 9 | Validation of candidate surface markers in an epithelial-enriched E12.5 pancreas scRNA-seq dataset (GSM3140915).

(A) UMAP visualization showing cluster annotation of the epithelial-enriched dataset. (B) UMAP feature plot showing distribution of *Pdx1* expression across clusters. (C) UMAP feature plots showing expression of candidate surface markers (*Epcam, Atp1b1, Scpep1, and Igsf8*). (D) Pseudo-flow-cytometry plots comparing normalized expression of each candidate marker with *Pdx1*. Percentages indicate the fraction of $Pdx1^{+}$ cells expressing each marker, demonstrating the highest coverage by *Epcam*.

Supplementary Figure 10





Supplementary Figure 10 | Experimental validation of Epcam and comparison with Cdh1 and Cldn6.
(A) Flow-cytometry gating strategy for detection of β-galactosidase and EPCAM in dissociated E12.5 *Pdx1*$^{LacZ/+}$ pancreatic buds using Spider-Gal and anti-EPCAM staining. (B) Flow-cytometry gating strategy for detection of EPCAM^{+} cells in dissociated E12.5 GRR pancreatic buds. (C) UMAP feature plots showing expression of *Cdh1* and *Cldn6* in the epithelial-enriched GSM3140915 dataset. (D) Pseudo-flow-cytometry plots showing normalized expression of Pdx1 versus *Cdh1* or *Cldn6*. Percentages indicate the fraction of *Pdx1*$^{+}$ cells expressing each marker, showing lower coverage by *Cdh1* (~55%) and higher overlap for *Cldn6* (~95%).