

Supplementary figures

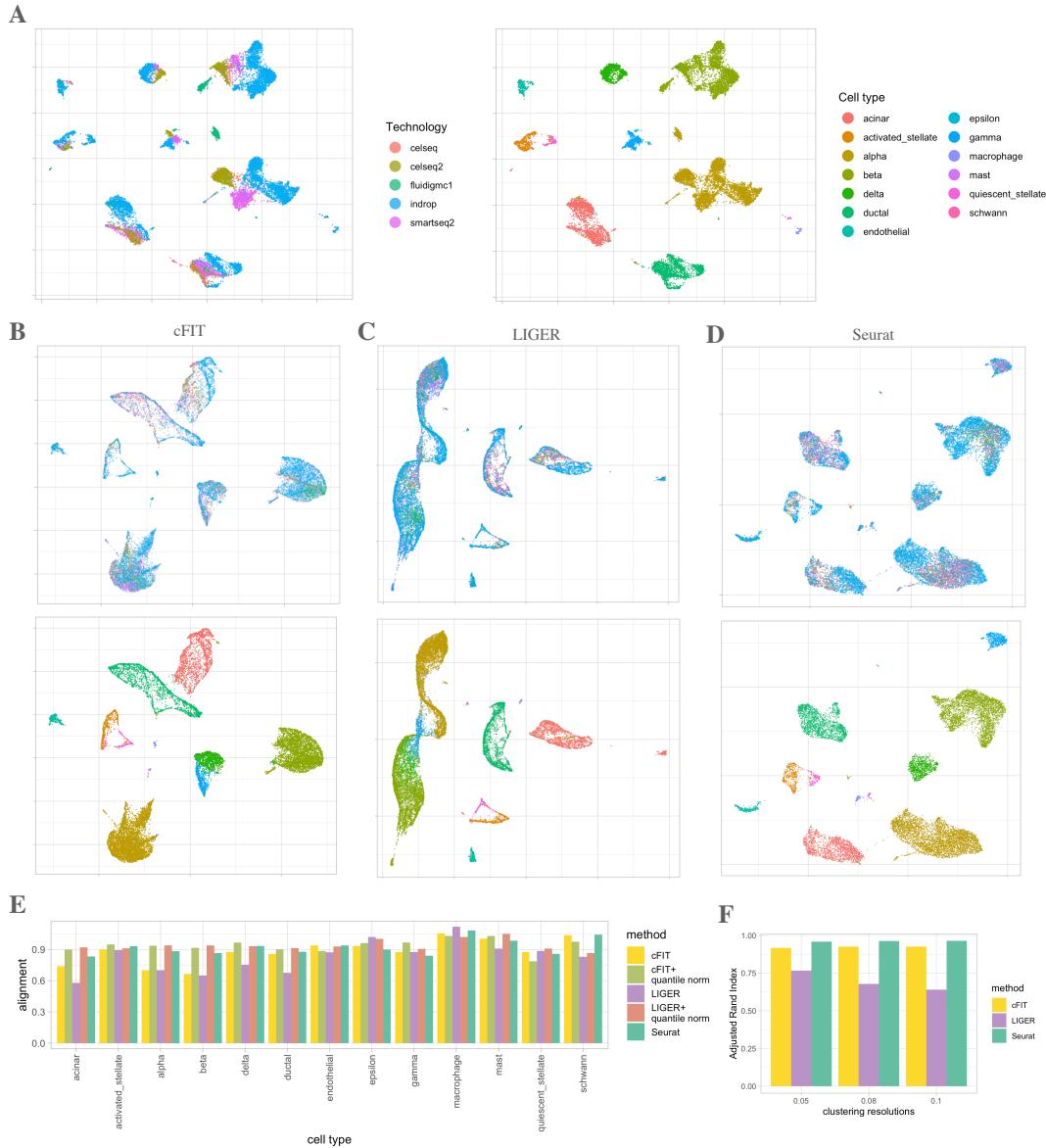


Figure S1: Integration of pancreatic islet cell data sets. (A) UMAP plot of five pancreatic islet cell data sets colored by data sets (left) and by cell type (right). (B-D) UMAP plot after integration by cFIT (B), LIGER (C), and Seurat (D), colored by either data set (top panels) and cell type (bottom panels). (E) Alignment scores are calculated for integration results from five methods, including cFIT, cFIT with quantile normalization for post-processing, LIGER (no quantile normalization), LIGER with quantile normalization and Seurat, to evaluate how well the pancreatic islet cell data sets are mixed together. (F) Evaluate the clustering accuracy of identifying cell types on integrated data from three different methods.

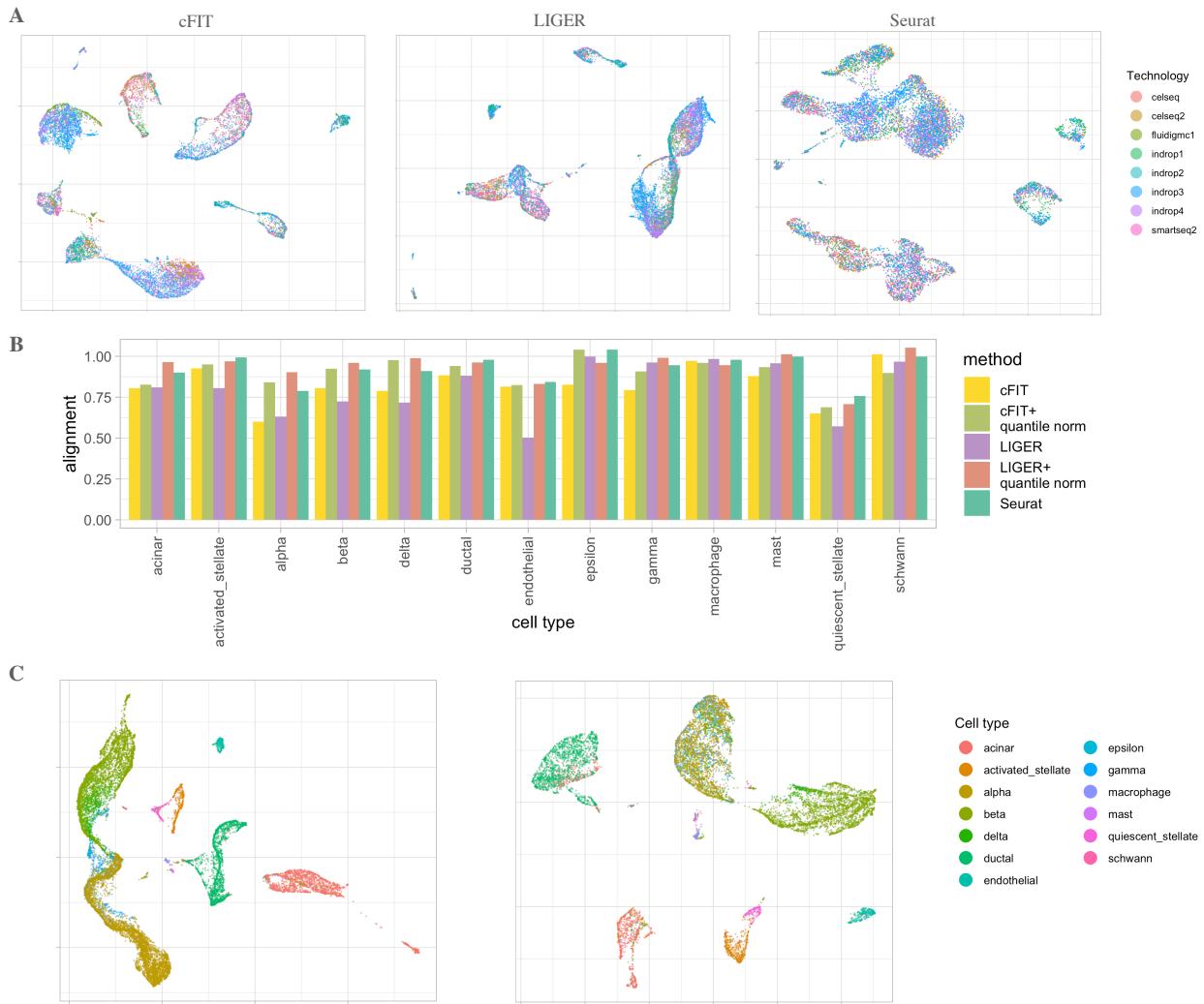


Figure S2: Integration of Pancreatic islet cell data sets with removed cell types from each data sets. (A) UMAP of integrated data colored by technologies, from different integration procedures. (B) Alignment scores calculated from integrated data. (C) Integration results of pancreatic islet data sets using LIGER with quantile normalization as post processing step, on original data sets (left panel), and on modified data sets by removing cell types (right panel).

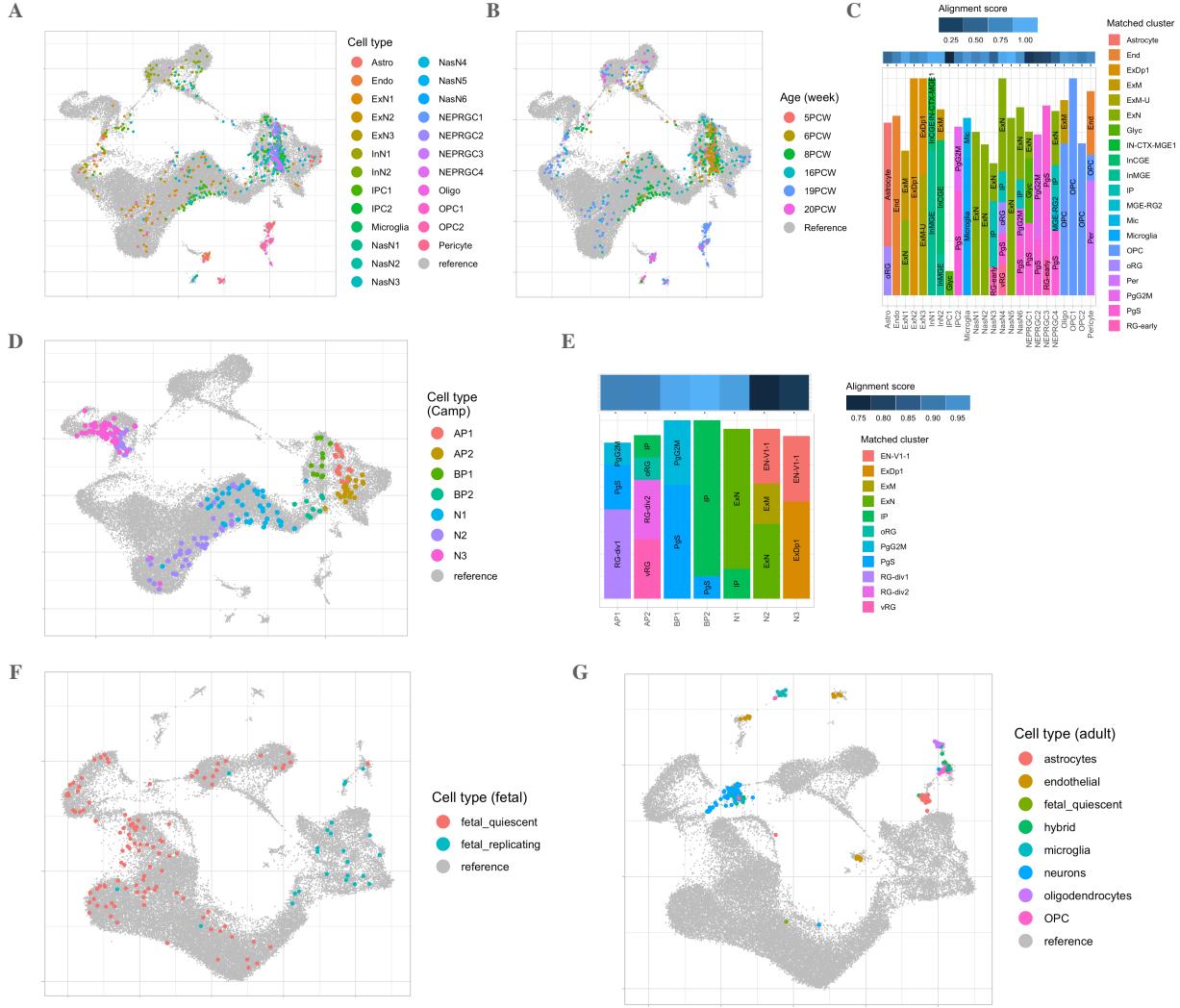


Figure S3: Transfer the learned gene expression signature learned from the integration of Nowakowski and Polioudakis data to smaller scRNA-seq data sets from the human fetal brain. (A-C) Transfer results on Li data (Li et al., 2018), (A) UMAP plot of the factor loading of Li cells obtained from transfer learning, where the Li cells were colored by the cell labels from the while the reference cells (cells from Nowakowski and Polioudakis data sets) were marked in grey in the background. (C) Bar plot showing the composition of assigned cell types labels through transferring for each original cell group. (D,E) Transfer results on Camp data (Camp et al., 2015), (D) UMAP plot of the factor loading of Camp data on top of reference cells, and the bar plot shows the labeling of cells in each original cell group(E). (F,G) Transfer results on Darmanis data (Darmanis et al., 2015) for fetal cells (F) and adult cell (G).

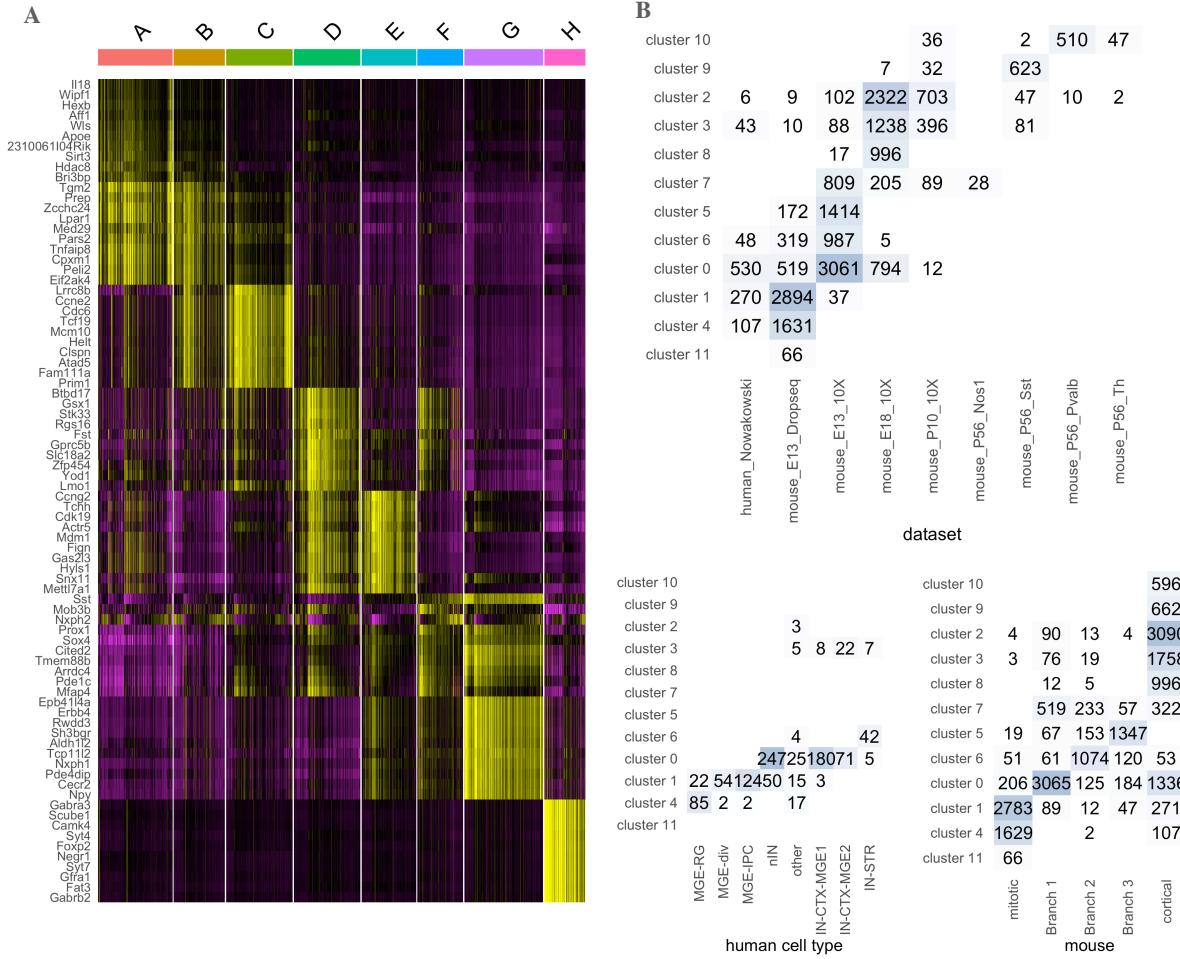


Figure S4: Integration of MGE progenitor cells and interneuron precursors from human and mouse. (A) Heatmap of top 10 marker genes per cluster identified from the integrated data of Nowakowski MGE cells and Drop-Seq mouse MGE cells. (B) Examine the source of cells from each cluster in the full integration of mouse MGE cells, interneurons precursors from mouse embryos (E13, E18), postnatal (P10) and adult stage (P56), human MGE cells and cortical interneurons. Top: clusters versus data sets to which the cells belong. We split the cells from adult mouse data set by known interneuron types for illustration purpose. Bottom left panel: cluster versus human cell types form (Nowakowski et al., 2017). Bottom right panel: cluster versus the mouse labels (branch labeling from (Mayer et al., 2018)).