Cell Reports Medicine, Volume *5*

Supplemental information

Modeling type 1 diabetes progression using machine learning and single-cell transcriptomic measurements in human islets

Abhijeet R. Patil, Jonathan Schug, Chengyang Liu, Deeksha Lahori, Hélène C. Descamps, the Human Pancreas Analysis Consortium, Ali Naji, Klaus H. Kaestner, Robert B. Faryabi, and Golnaz Vahedi

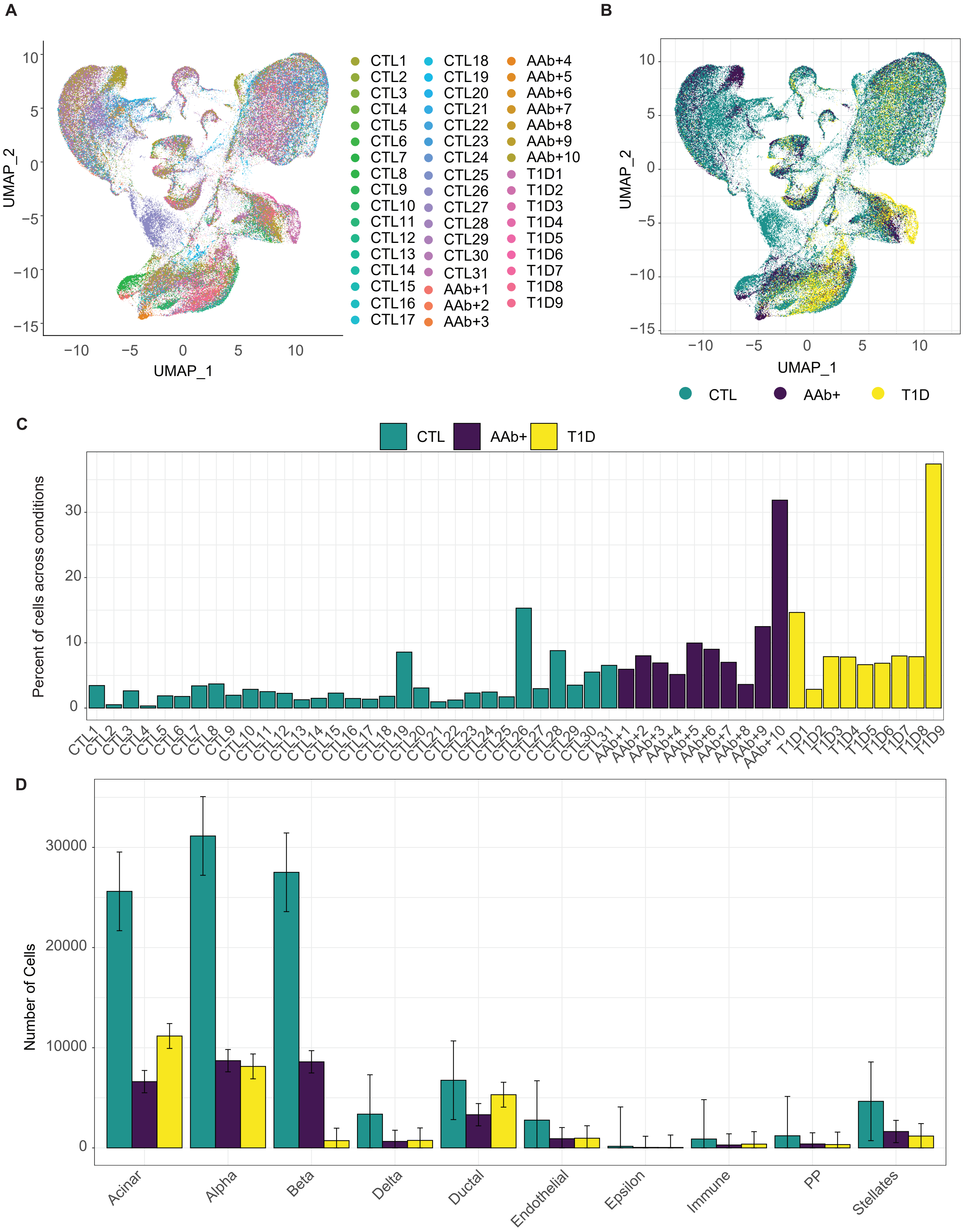


Figure S1. UMAPs and Bar charts showing donor statistics in all conditions, Related to Figure 1

(A) UMAP showing the cells by donor distribution from all conditions.

(B) UMAP showing the cells by conditions such as CTL, AAb+ and T1D.

(C) Bar chart showing the distribution of percent of cells across each donor in CTL, AAb+ and T1D conditions.

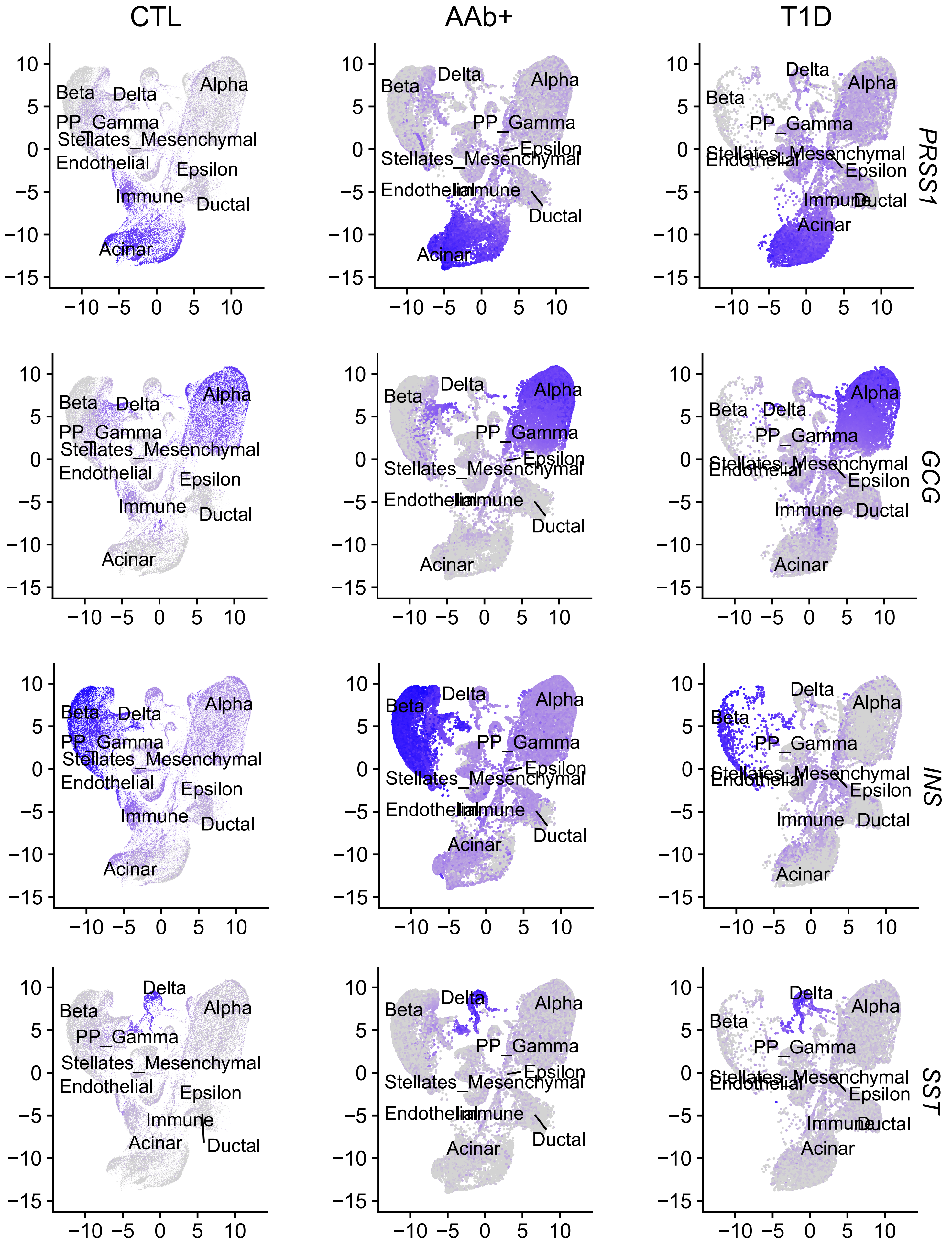


Figure S2. Multiple feature plots (UMAPs) split by CTL, AAb+, and T1D conditions showing the cell type-specific expression of marker genes. Acinar cells (*PRSS1* high), Alpha cells (*GCG* high), Beta cells (*INS* high), Delta cells (*SST* high), Related to Figure 1E.

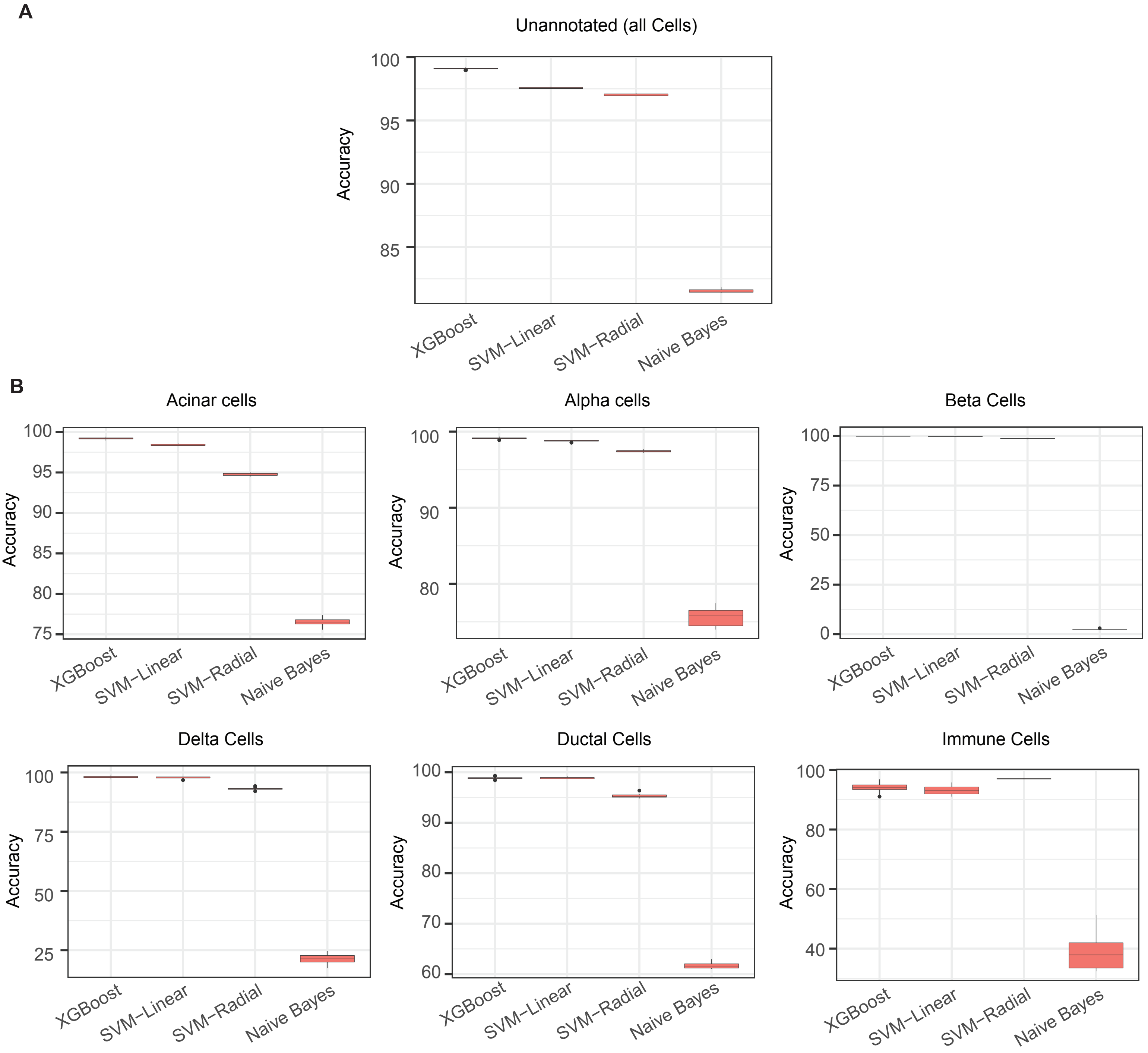


Figure S3. Comparison of classification accuracy of XGBoost, SVM linear, SVM radial, and naïve Bayes in
(A) unannotated (all cells)
(B) different annotated cell types. Related to Figure 2.

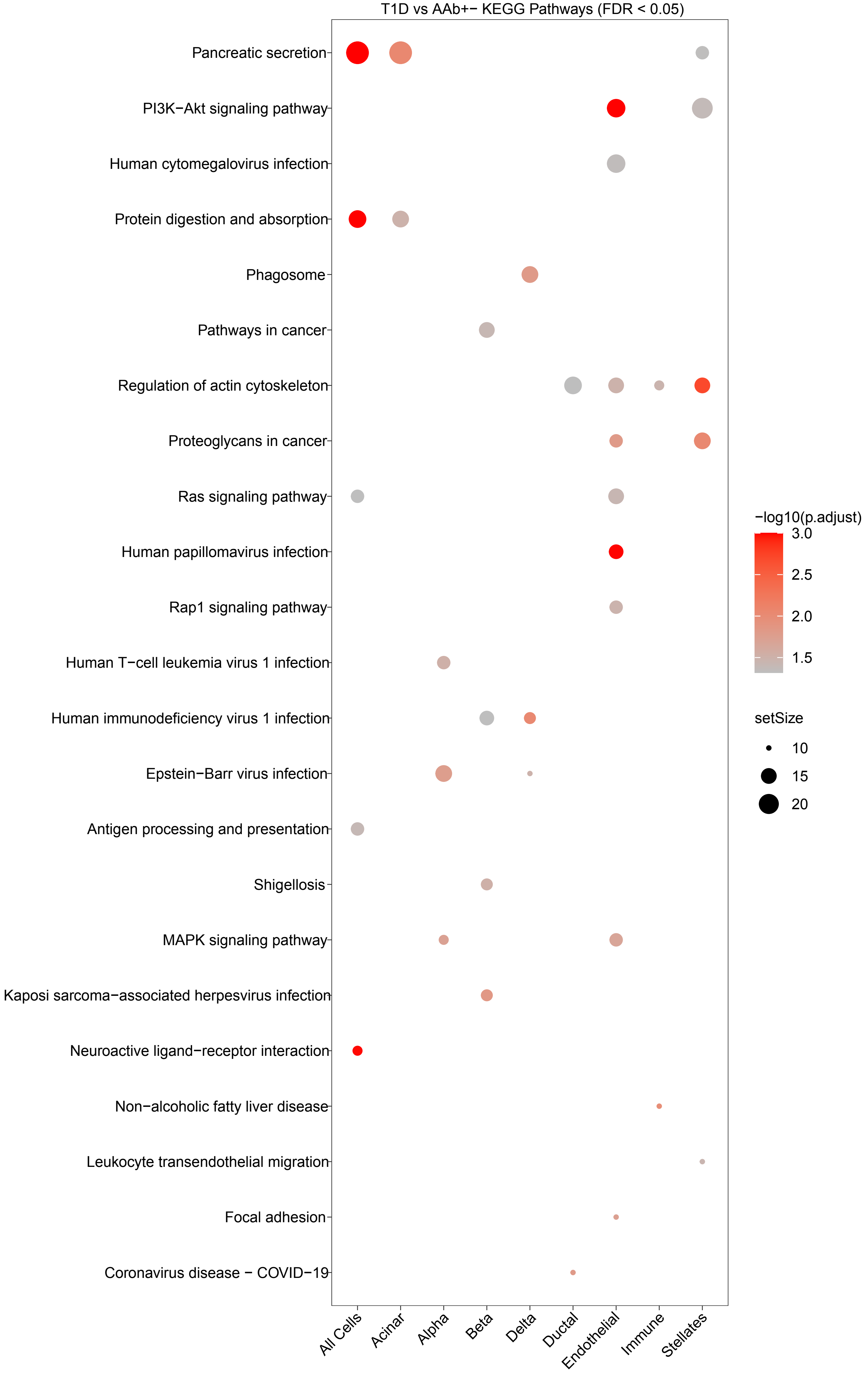
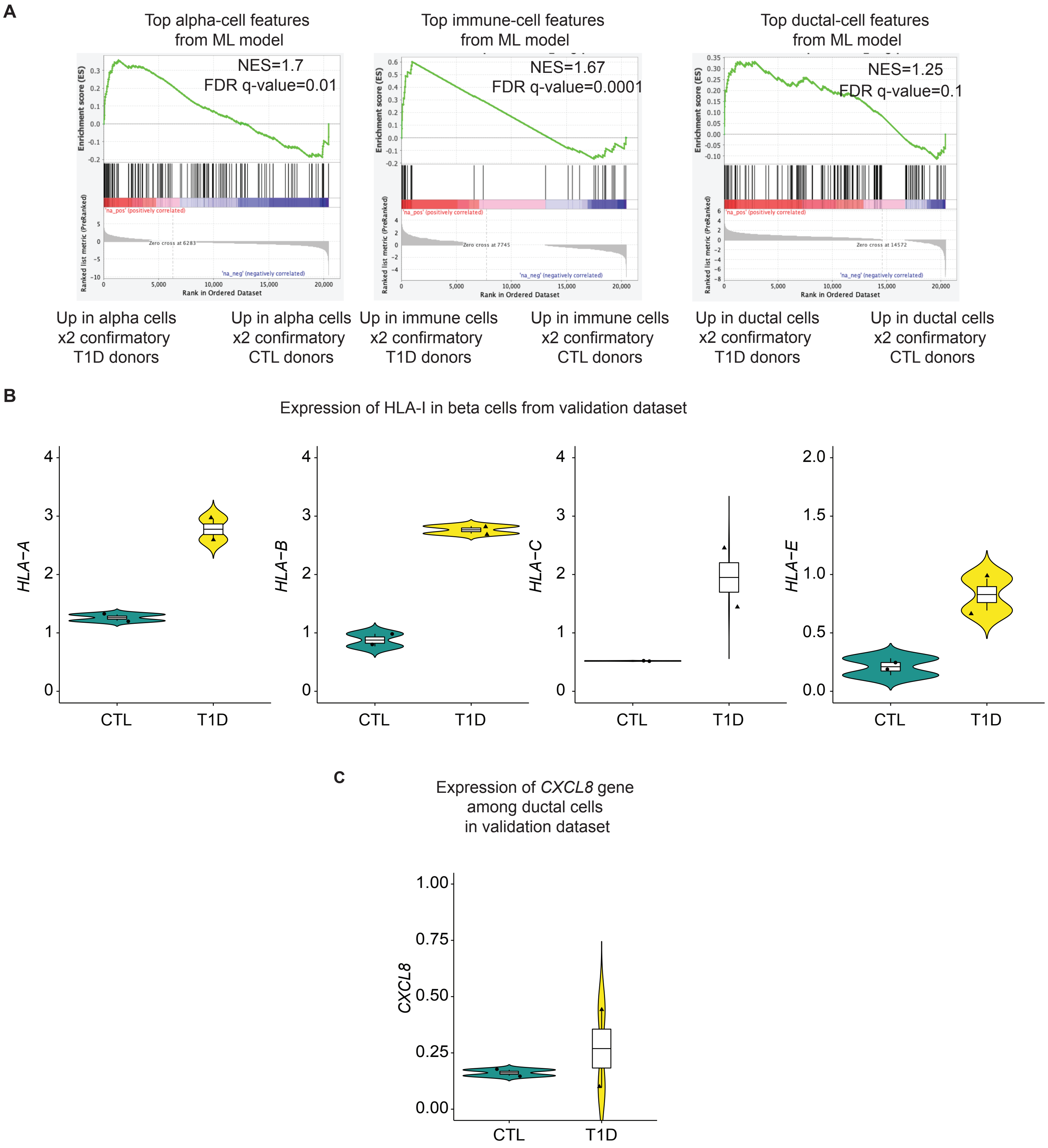


Figure S4. Significant KEGG pathways (FDR < 0.05) based on ranked genes obtained from T1D vs. AAb+ comparison across unannotated all cells and annotated different cell types, Related to Figure 3.

Figure S5. Significant KEGG pathways (FDR < 0.05) based on ranked genes obtained from AAb+ vs. CTL comparison across unannotated all cells and annotated different cell types, Related to Figure 3.



Top alpha-cell features from ML model

Top immune-cell features from ML model

Top ductal-cell features from ML model

Expression of HLA-I in beta cells from validation dataset

Expression of *CXCL8* gene among ductal cells in validation dataset

Figure S6. Expression of top features in independent validation dataset, related to Figure 4.
(A). Top features from ML model compared with features from validation dataset.
(B). Average expression of HLA-I genes among beta cells in the independent validation dataset of two T1D and CTL donors.
(C). Average expression of CXCL8 gene among ductal cells in the independent validation dataset.
(D). Bar charts with error bars showing the cell type distribution across all conditions

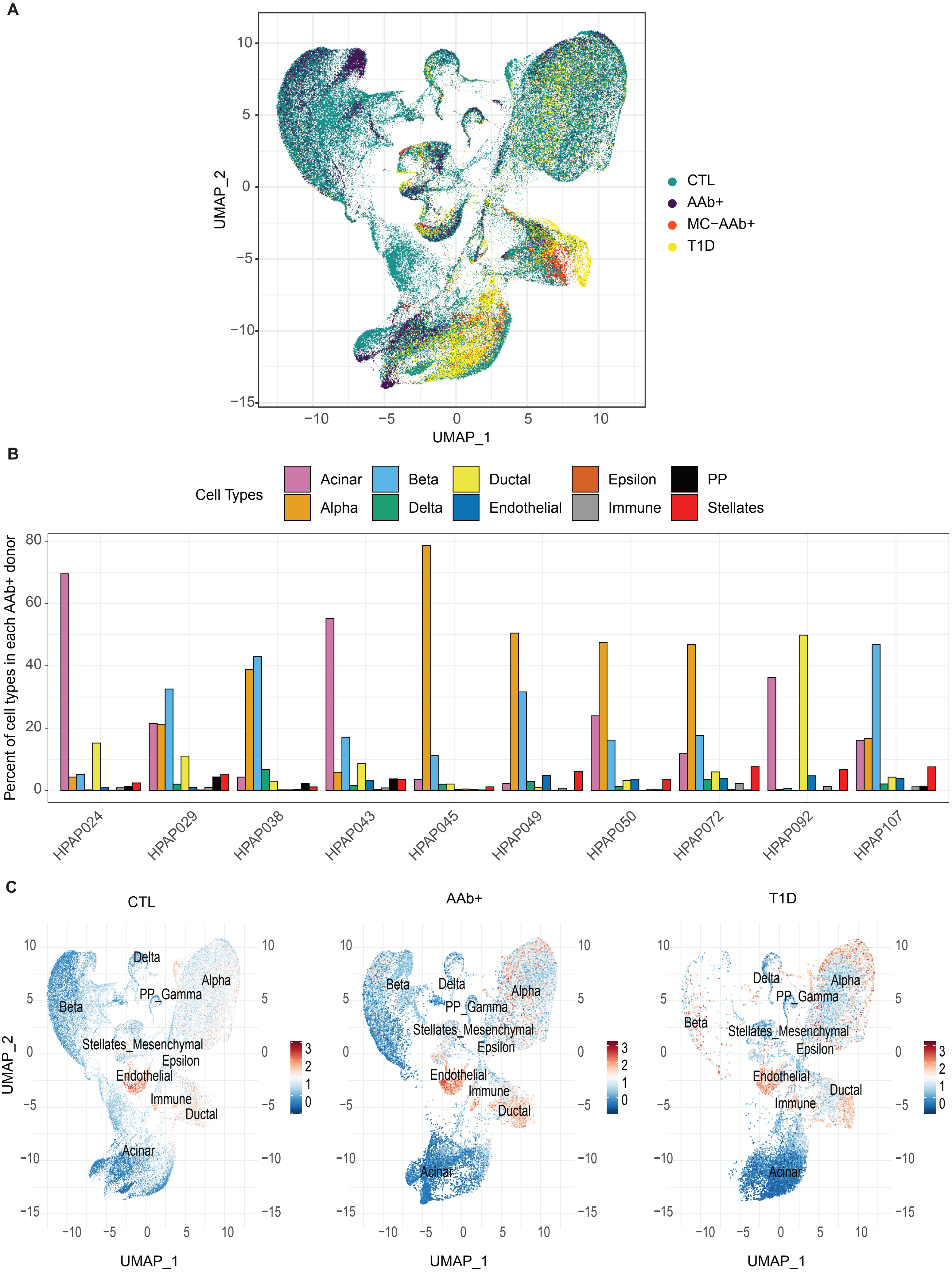


Figure S7. UMAPs and grouped charts showing AAb+ donors cell type information, related to Figure 5.
(A). UMAP highlighting the cells from misclassified AAb+ donors such as HPAP092 and HPAP107.
(B). Grouped bar charts showing the percent-wise cell type distribution in each of the AAb+ donors.
(C). UMAPs showing the module gene score for HLA-I genes across different conditions.

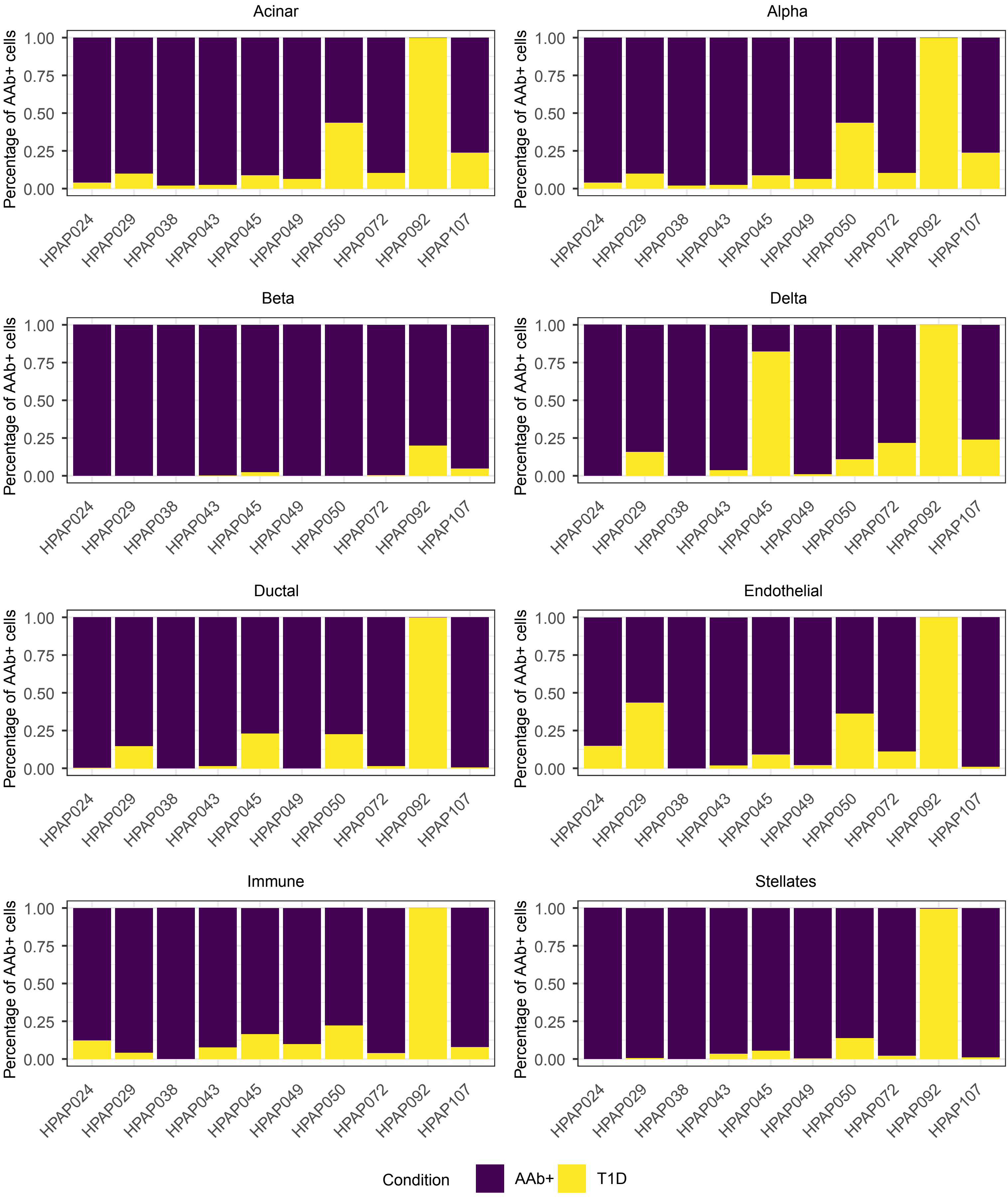


Figure S8. Stacked bar chart showing the donor wise percentage of AAb+ cells classified as T1D or AAb+ across annotated cell types on SCT (SCTransform) assay, related to Figure 5.

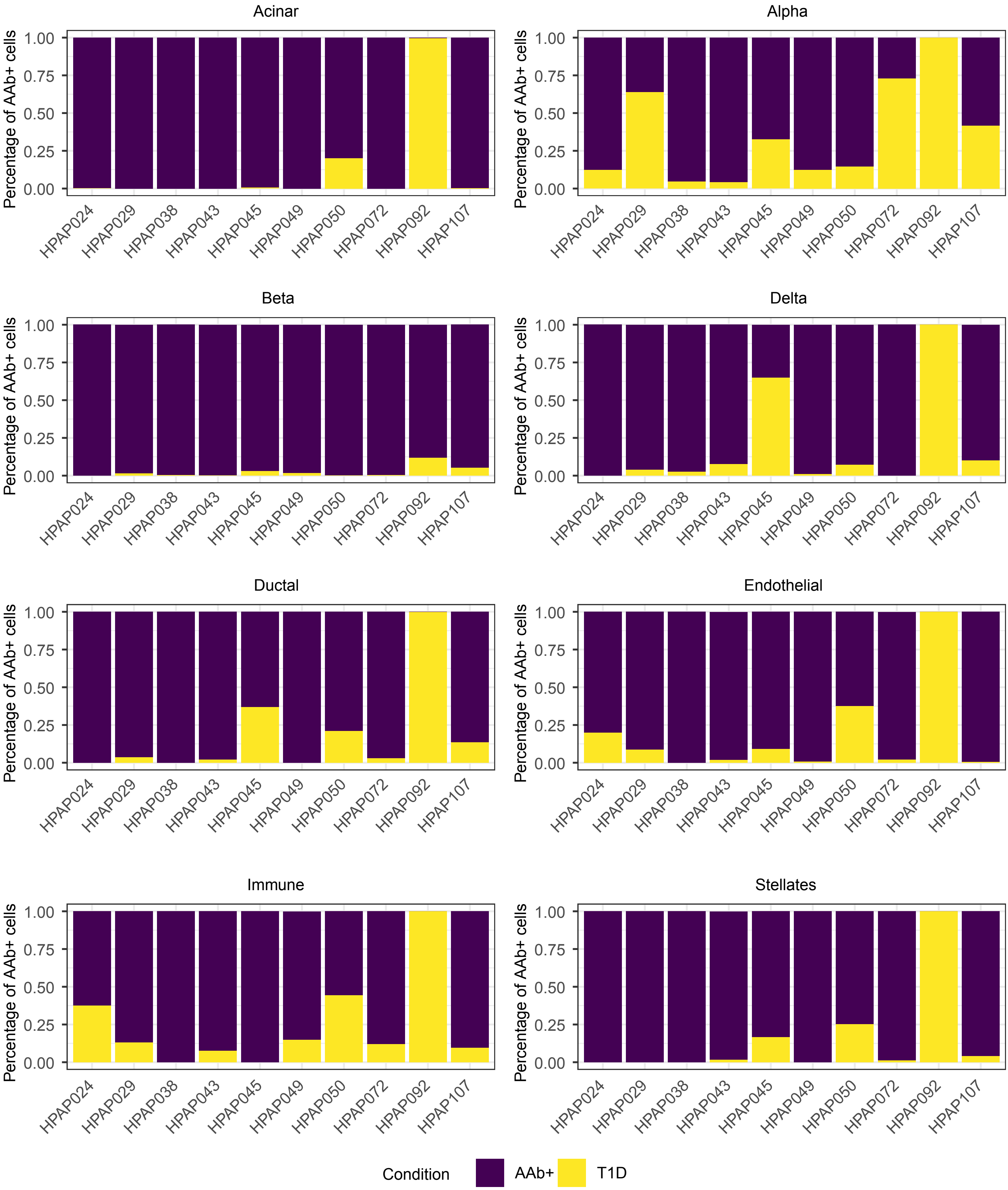


Figure S9. Stacked bar chart showing the donor wise percentage of AAb+ cells classified as T1D or AAb+ across annotated cell types on RNA (LogNormalize) assay, related to Figure 5.

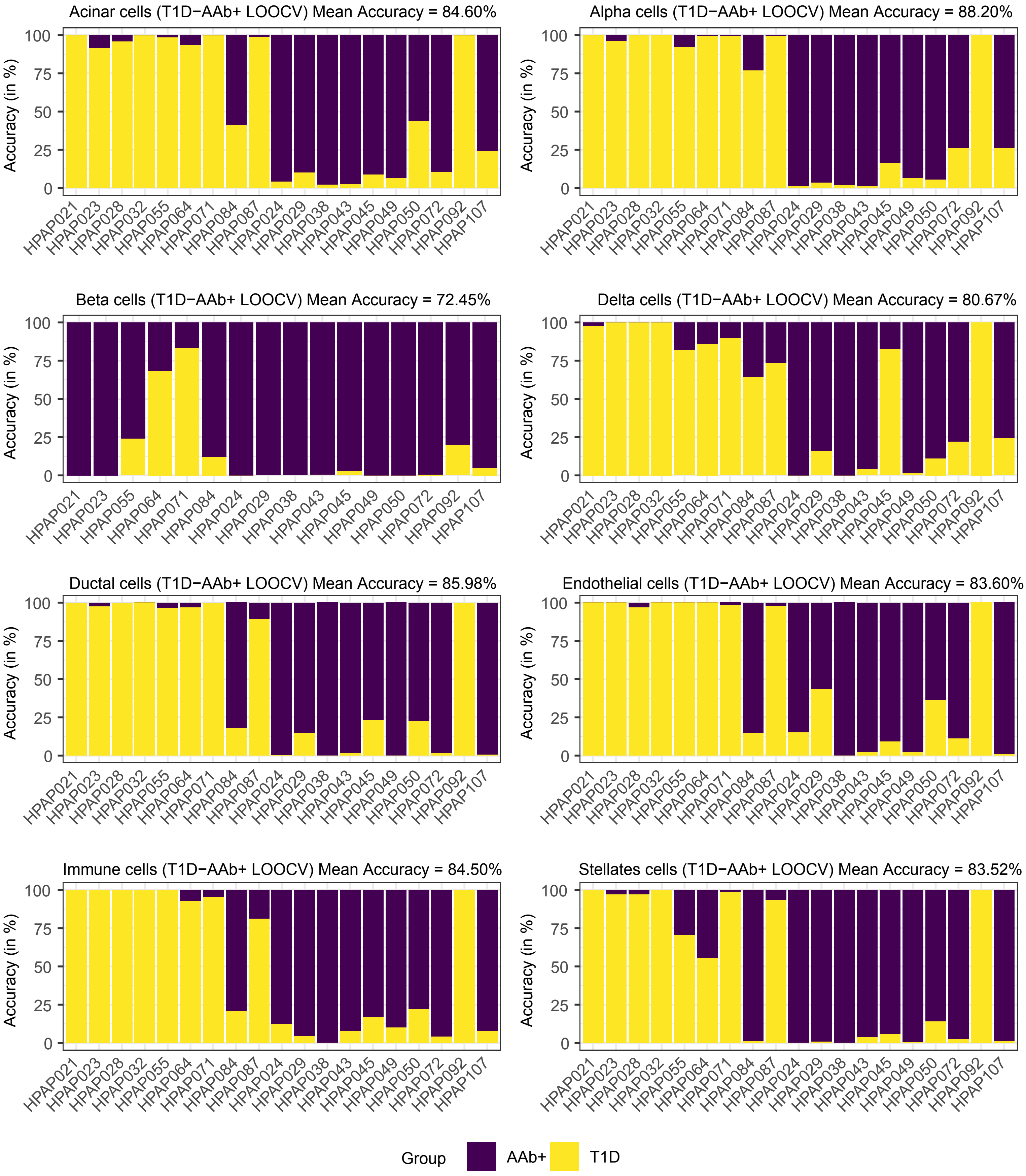


Figure S10. Stacked bar chart showing the donor wise classification accuracy based on LOOCV approach for T1D vs. AAb+ classifier across annotated cell types on SCT (SCTransform) assay, related to Figure 5.

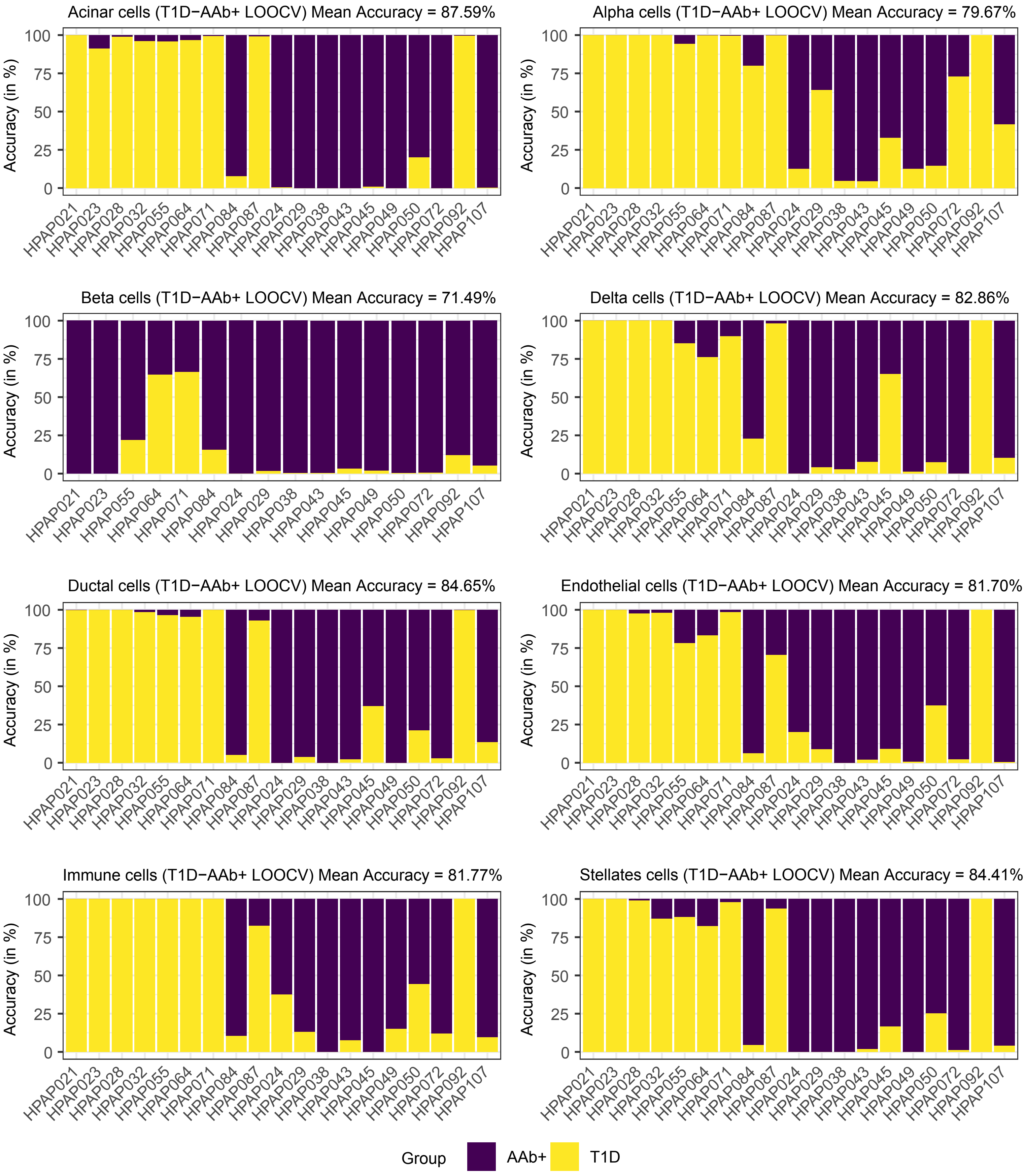
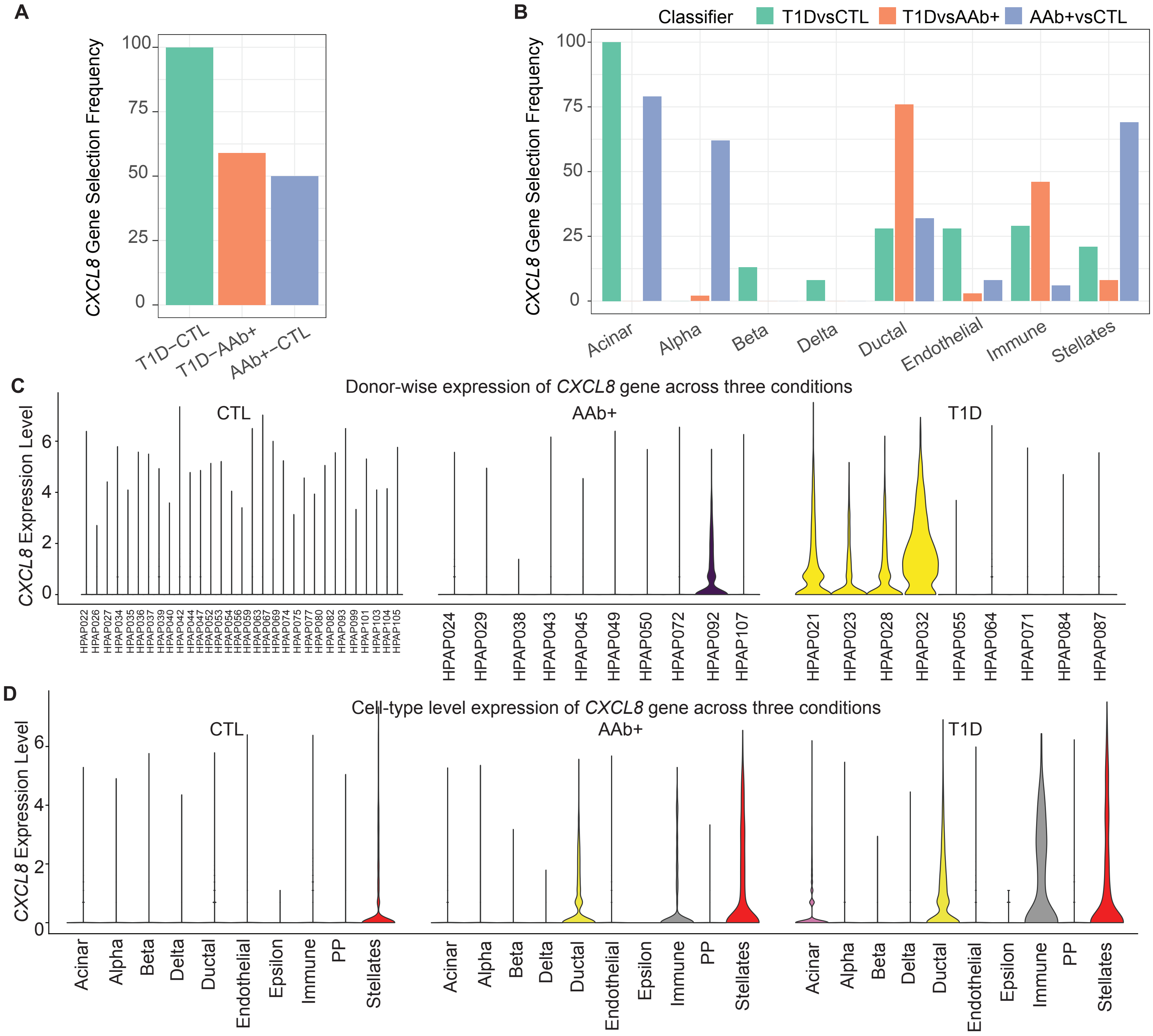


Figure S11. Stacked bar chart showing the donor wise classification accuracy based on LOOCV approach for T1D vs. AAb+ classifier across annotated cell types on RNA (LogNormalize) assay, related to Figure 5.



E

Average expression and Percent Expression of *CXCL8* across all cell types in three conditions

Cell Types	Average Expression of *CXCL8* in CTL	Percent Expressed in CTL	Average Expression of *CXCL8* in AAb+	Percent Expressed in AAb+	Average Expression of *CXCL8* in T1D	Percent Expressed in T1D
Acinar	0.67	13.54	0.56	11.64	3.41	30.24
Alpha	0.45	4.53	0.07	1.79	0.31	6.64
Beta	0.24	3.81	0.01	0.56	0.05	1.93
Delta	0.13	1.78	0.02	1.23	0.48	7.06
Ductal	1.35	22.32	3.56	41.79	7.92	55.03
Endothelial	2.92	9.23	2.46	14.58	3.1	18.49
Epsilon	0.02	1.25	0	0	0.21	16.28
Immune	4.23	24.92	6.1	26.46	19.69	50.66
PP Gamma	0.45	4.21	0.19	4.06	5.22	14.37
Stellates	18.57	30.37	16.55	42.14	35.91	45.61

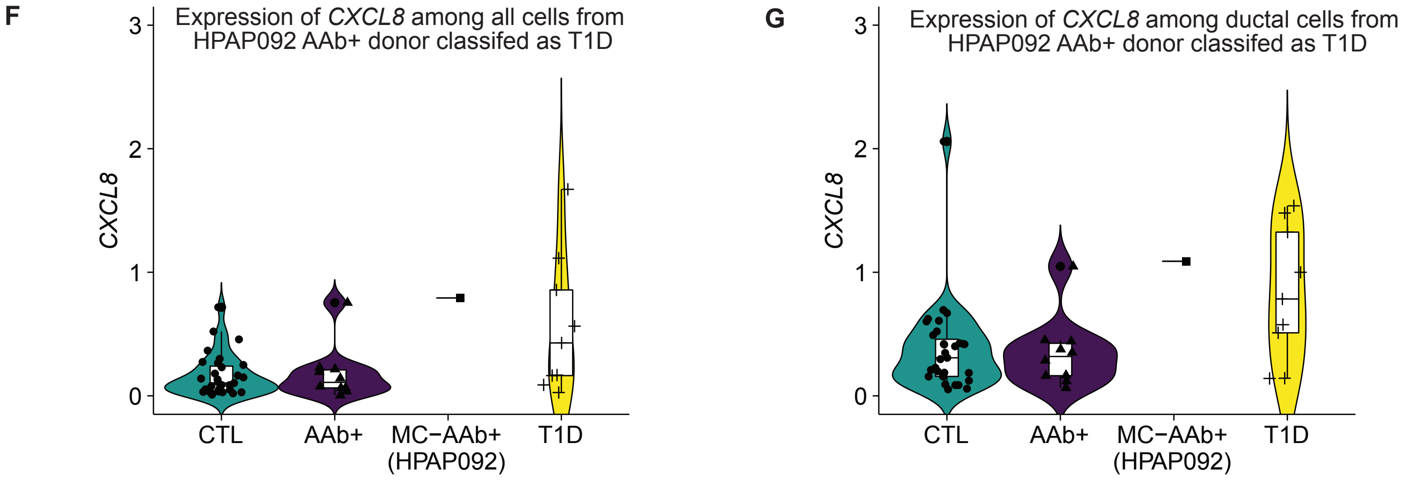


Figure S12. Expression of CXCL8 gene in cell types across healthy, AAb+, and T1D donors, related to Figure 6.
(A). *CXCL8* gene selection frequency across different pairwise classifiers across unannotated (all cells) cells.
(B). *CXCL8* gene selection frequency across different pairwise classifiers across annotated cells.
(C). Donor-wise expression of *CXCL8* gene across three conditions.
(D). Cell-type level expression of *CXCL8* gene across three conditions.
(E). Average expression and Percent Expression of *CXCL8* across all cell types in three conditions.
(F). Expression of *CXCL8* among all cells from HPAP092 AAb+ donor classified as T1D.
(G). Expression of *CXCL8* among ductal cells from HPAP092 AAb+ donor classified as T1D