HHS Public Access

Author manuscript

Nat Metab. Author manuscript; available in PMC 2022 August 28.

Published in final edited form as:

Nat Metab. 2022 February ; 4(2): 284–299. doi:10.1038/s42255-022-00531-x.

Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in Type 1 Diabetes

Maria Fasolino[1,2,3,4,6,7,*], **Gregory W. Schwartz**[2,3,5,6,7,*], **Abhijeet R. Patil**[1,2,3,4,6,7], **Aanchal Mongia**[2,3,5,6,7], **Maria L. Golson**[1,4,8], **Yue J. Wang**[1,4], **Ashleigh Morgan**[1,4], **Chengyang Liu**[4,9], **Jonathan Schug**[1], **Jinping Liu**[1,4], **Minghui Wu**[1,4], **Daniel Traum**[1,4], **Ayano Kondo**[1,4], **Catherine L. May**[1,4], **Naomi Goldman**[1,2,3,4,6,7], **Wenliang Wang**[1,2,3,4,6,7], **Michael Feldman**[5,7], **Jason H. Moore**[1,7], **Alberto S. Japp**[2,10], **Michael R. Betts**[2,10], **the HPAP Consortium**[**], **Robert B. Faryabi**[2,3,5,6,7,#], **Ali Naji**[2,4,9,#], **Klaus H. Kaestner**[1,3,4,#], **Golnaz Vahedi**[1,2,3,4,6,7,#]

[1]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[2]Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[3]Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[4]Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[5]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[6]Abramson Family Cancer Research Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[#]Corresponding authors: Golnaz Vahedi: vahedi@pennmedicine.upenn.edu, Klaus H. Kaestner: kaestner@pennmedicine.upenn.edu, Ali Naji: Ali.Naji@pennmedicine.upenn.edu, Robert B. Faryabi: faryabi@pennmedicine.upenn.edu.

[*]co-first authors

[**]A list of authors and their affiliations appears at the end of the paper

Author Contributions

C.L. and A.N. procured human pancreatic tissues. A.N. and K.H.K. acquired funding. M. G., Y.J.W., and A.M. generated scRNA-seq libraries. J.S. performed sequencing of scRNA-seq libraries. J. L. performed CyTOF experiments. M.W. and D. T. performed IMC experiments. A. K. assisted in analysis of IMC. C.L.M. performed IHC experiments. N. G. assisted in analysis of CyTOF. W.W. assisted in analysis of scRNA-seq. A.M. performed annotation of IMC and CyTOF. M. Feldman. supervised the HPAP tissue bank. J.H.M., A.S.J., M.R.B. are members of HPAP consortium. K.H.K. supervised M.G., Y.J.W., A.M., J. L., M. W., D. T., A. K., C.L.M. for generation of scRNA-seq, CyTOF, and IMC data. R. B. F. supervised A.M., G.W.S. in analysis of scRNA-seq, CyTOF, and IMC. G.V. supervised M. Fasolino, A.R.P., N.G., and W.W. M. Fasolino., G.W.S., A.R. P. and G.V. performed computational analysis of scRNA-seq, IMC, and CyTOF. M. Fasolino and G.V. wrote the original draft and revised it with comments from all authors. K.H.K. and R.B.F. edited the original and revised manuscript.

Code Availability

Where applicable, scripts used for data processing and analysis have been described in the Supplemental Materials and Methods section and provided on Github https://github.com/GregorySchwartz/multiomics-single-cell-t1d. TooManyCells is a publicly available suite of tools, algorithms, and visualizations (https://github.com/GregorySchwartz/too-many-cells) that was extensively used in this study, and where applicable, the flags used in TooManyCells to generate specific figures are included in the Materials and Methods section.

Competing Interests

M.R.B. has a consulting arrangement with Interius Biotherapeutics. Other authors declare no competing interests.

[7]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[8]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Johns Hopkins University, Baltimore, MD, 21205, USA

[9]Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

[10]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

Abstract

Type 1 Diabetes (T1D) is an autoimmune disease in which immune cells destroy insulin-producing beta cells. The etiology of this complex disease is dependent on the interplay of multiple heterogeneous cell types in the pancreatic environment. Here, we provide a single-cell atlas of pancreatic islets of 24 T1D, autoantibody-positive, and non-diabetic organ donors across multiple quantitative modalities including ~80,000 cells using single-cell transcriptomics, ~7,000,000 cells using cytometry by time-of-flight, and ~1,000,000 cells using *in situ* imaging mass cytometry. We develop an advanced integrative analytical strategy to assess pancreatic islets and identify canonical cell types. We show that a subset of exocrine ductal cells acquires a signature of tolerogenic dendritic cells in an apparent attempt at immune suppression in T1D donors. Our multimodal analyses delineate cell types and processes that may contribute to T1D immunopathogenesis and provide an integrative procedure for exploration and discovery of human pancreas function.

Introduction

Type 1 diabetes (T1D) is an autoimmune disease which occurs as a consequence of the destruction of insulin-producing beta cells in the islets of Langerhans within the pancreas[1,2]. This complex disease is characterized by atypical beta-immune interactions including production of beta cell autoantibodies and the immunological attack on beta cells by cytotoxic $CD8^+$ T cells[3,4].

T1D autoimmunity has been linked to poorly understood genetic and environmental factors. Genome-wide association studies have implicated multiple loci in T1D, with the major histocompatibility complex (MHC) Class II genes as the dominant susceptibility determinant of this disease[5]. However, the precise cellular context through which T1D susceptibility genes cause the destruction of beta cells remains to be discovered. Addressing this question is particularly challenging since the pancreas is a heterogeneous organ, composed of multiple distinct cell types.

Two nontrivial constraints hamper insights into comprehensive identification of the pathogenic cell types in T1D: (1) the inability to safely biopsy the human pancreas of living donors and (2) the significant disease progression and beta cell destruction by the time patients are clinically diagnosed with T1D. Therefore, the majority of T1D studies have been performed on leukocytes from the peripheral blood, which is not the

site of pathogenesis. Of late, the Network for Pancreatic Organ Donors (nPOD)[6] and the Human Pancreas Analysis Program (HPAP)[7] have started collecting pancreatic tissues from hundreds of deceased organ donors diagnosed with T1D. Additionally, given that many T1D patients harbor beta cell autoantibodies (AAbs) in their bloodstream prior to clinical diagnosis, nPOD and HPAP also collect samples from donors with AAbs towards islet proteins but without a medical history of T1D, in hope of elucidating early pathogenic events.

Using these initiatives, we developed a pancreatic islet atlas containing an unprecedented ~80,000 cells using single-cell transcriptomics, ~7,000,000 cells using cytometry by time-of-flight (CyTOF), and ~1,000,000 cells using *in situ* imaging mass cytometry (IMC) in pancreatic tissues of human organ donors collected by HPAP, enabling a resource for extensive exploration and discovery within the pancreatic environment. We also provide an interactive data explorer for simple, direct access to the single-cell transcriptomics data (https://cellxgene.cziscience.com/collections/51544e44-293b-4c2b-8c26-560678423380). Our comprehensive integrative analyses on this unique data set provides cellular and molecular insights into T1D pathogenesis and suggests pancreatic ductal cells may play a role in suppressing $CD4^+$ T cells in pancreatic tissues.

Results

scRNA-seq unravels novel cell states in the human pancreas

To unmask the molecular perturbations occurring in pancreatic tissues during T1D, we constructed 81,313 single-cell RNA-seq (scRNA-seq) libraries from pancreatic islets of 24 human organ donors representing three categories: individuals with T1D (n = 5), those with AAbs toward pancreatic islet proteins but no clinical diagnosis of T1D (‘AAB+’; n = 8), and those with neither AAbs nor a history of T1D (‘Control’; n = 11) (Fig. 1a and Extended Data Fig. 1 a–b, and Tables S1 to S2). The statistics related to reads per cell across donors demonstrated the high-quality of these data sets (Table S3). We filtered outlier cells, removed doublets, and employed the cell type classifier ‘Garnett’[8] (Extended Data Figs. 1c–e, 2a–g, and 3a–e) to cluster 69,645 high-quality cells using ‘TooManyCells’[9] (Fig. 1b–c). The resultant classification was confirmed by both canonical gene marker expression for each cell type and by transferring cluster labels from a previous single-nucleus RNA-seq data set consisting of pancreatic islets[10] to our datasets (Fig. 1c and Extended Data Fig. 4 a–c).

Notably, clustering was clearly driven by cell type, and not by confounding factors such as autoantibody status, age, BMI, phenotypic group, or other factors (Extended Data Figs. 2 d–g, Figs. S1 a–k, and S2 a–l). Additional evidence for the lack of technical noise stems from the observation that cell type clustering was preserved when donors from T1D, AAB+, and Control groups were independently clustered (Fig. S3 a–f).

Considering the reported abnormalities of the exocrine pancreas in T1D[11] and recent efforts indicating the enrichment of sequence polymorphisms associated with T1D within the regulatory elements of exocrine cells[12], we next examined the relationship between pancreatic endocrine and ductal cells. First, we subsetted and reanalyzed the endocrine and

ductal cells to achieve a more granular clustering (Fig. 1d). Upon reclustering, the major cell types – alpha cells, beta cells, delta cells, epsilon cells, PP cells, ductal cells, and acinar cells – were easily discernible (Fig. 1d and Extended Data Fig. 4d). In instances where there were two transcriptionally distinct canonical cell types (i.e. Beta-1 and Beta-2), differential gene expression analysis between populations provided further insights into the underlying molecular differences (Tables S4–S7). For example, cells in the Beta-2 cluster expressed higher levels of stress response genes such as *NPTX2 and GDF15* when compared to those in the Beta-1 cluster. The activation of stress response genes in beta cells in various hyperglycemic states has been reported previously[13–18]. Notably, the comparison of cells in the two ductal clusters revealed that while cells in the Ductal-1 cluster were enriched for transcription factors (TFs) associated with the endocrine cell fate (i.e., *PDX1* and *NKX6-1*), those in the Ductal-2 cluster expressed acinar TFs (i.e., *PTF1A* and *GATA4*).

A substantial number of cells (4,001) were not included in these canonical cell type clusters, but rather formed their own transcriptionally distinct group on the dendrogram. This cluster comprised 5.7% of all profiled cells, with a mixture of cellular classifications and expression of canonical gene markers. We labeled these cells as ‘Hybrid’ cells (Figs. 1d and Extended Data Fig. 4d). Notably, the gene expressed most highly and consistently in Hybrid cells was *INS*, and a comprehensive examination of the cells comprising this cluster ruled out the possibility of them being doublets (Extended Data Figs. 1d and 2 a–e). To further validate the most closely related cell types to these non-canonical cells, we employed a label transfer strategy using a reference pancreatic islet scRNA-seq dataset[19]. We corroborated the assignment of multiple cell types including beta cells and alpha cells to these Hybrid cells (Extended Data Fig. 4 e–f). Cells equivalent to Hybrid cells had been detected earlier[19–21] and were most recently documented in the adult pancreas of mice and humans[20–23]. Nonetheless, we excluded hybrid cells for further analysis to eliminate any cells captured potentially as doublets.

Since immune cell-mediated destruction of viable pancreatic cells is the major pathogenic feature of T1D, we examined the intrapancreatic immune cells profiled by scRNA-seq in detail. First, we subsetted and reclustered the cells constituting the ‘Immune’ cluster from the comprehensive tree (Fig. 1c) and found that this population also contained stellate (*RGS5* high) and Schwann (*PLP1* high) cells along with immune cells (*PTPRC* high) (Fig. S4 a–b). Using the Immunological Genome Project (ImmGen) cell type signatures[24], we further found that the gene signatures of antigen-presenting cells (APCs) such as macrophages, for example *CD68, SPI1, CD14*, and *CD16*, were most frequently expressed in the immune cell subset (Fig. S4 b–c), suggesting that these cell types comprise the majority of the identified immune cells that are collected and cultured along with pancreatic islets.

Studies demonstrating that regulatory elements of immune cells harbor the largest number of risk variants associated with T1D[12,25] imply that immune cells are more susceptible to gene dysregulation compared with other cell types in T1D. To quantify the link between genetic predisposition associated with T1D and cell type-specific gene expression, we used a genetic prioritization model[26,27] and examined the enrichment of sequence variation associated with T1D[12] across our annotated cell types (Fig. S4 d). As a control, we also examined sequence

variation associated with asthma[28,29] and T2D[28,29]. This analysis revealed that immune cells were the top cell type associated with T1D and asthma, which are both immune-mediated disorders. In contrast, beta cells were the top cell type associated with T2D (Bonferroni significance threshold of $P_{S\text{-}LDSC} < 0.05$), in agreement with recent reports demonstrating that risk variants for T2D are enriched in active *cis*-regulatory elements of beta cells[30]. Together, the genetic prioritization model corroborated that gene expression in immune cells is affected by T1D-associated sequence variation.

In addition to successful identification of the major endocrine and exocrine cell types and pancreatic immune cells, we also observed that the overall proportion of these cell types was in accordance with previous work[20,31–35]. Each of the major cell types comprised cells from the three donor groups with varying proportions (Fig. 1 e–g, Extended Data Figs. 4 g–h, Figs. S4 e–g, and S5 a–b). As expected, we found that there was a lower proportion of beta cells in the T1D cohort compared to the AAB+ or Control groups (Fig. S5 a–b). Conversely, both acinar and ductal cells comprised a higher portion in the T1D cohort, reflecting the difficulty of isolating high purity islets from T1D donors. Furthermore, within major cell clusters, there were varying degrees of separation based on donor group, which is to be expected due to likely transcriptomic differences among the three donor states (Figs. 1 e–f and S4a). Notably, Ductal-1 cells clearly separated into distinct T1D-enriched and Control-enriched groups (Fig. 1f). Taken together, our data indicate that transcriptomic differences amongst cell types and not technical biases drive the separation of major cellular clades, and that the donor state further segregates within cell types.

Comparison of endocrine and exocrine cells in AAB+ and T1D donors

We next compared transcriptomic divergence of AAB+ and T1D cells from Controls (Fig. 2a). To perform differential expression analysis between donor groups, we used two complementary analytical strategies: (1) grouping individual cells from different donor groups together (Tables S8–S10) or (2) performing pseudo-bulk analysis for each donor (Tables S11–S13). Plotting the average expression levels of the top 3 differentially expressed genes determined by the first strategy across donor groups confirmed that the predicted differential expression is not driven by one or a few donors (Fig. S6). Since pseudo-bulk methods cause cells from individuals with fewer cells to be more heavily weighted[36], we performed further analysis using genes detected based on the first strategy. Generally, the degree of overlap between dysregulated genes and pathways in AAB+ and T1D states were cell type-dependent (Figs. 2 b–e, S7 a–b, S8 a–b, S9 a–d, S10 a–d, Extended Data Fig 5 a–c). However, some pathways were found to be commonly dysregulated in multiple cell types across T1D and AAB+, including ‘apoptotic signaling’, various protein folding ontologies, various viral-related ontologies, ‘autophagy’, ‘inflammatory pathways’, and ‘stress response’.

We next examined the transcriptional changes in the two populations of annotated beta cells, Beta-1 and Beta-2. A large number of genes was downregulated in T1D (9,512 genes) and AAB+ (3,666 genes) Beta-1 cells compared to Controls, many of which overlapped (2,896 genes, 28%; $p < 2.2e\text{-}16$) between the two donor groups (Figs. 2b and S7a). Notable pathways that were frequently downregulated in Beta-1 cells of AAb+ and T1D donors were

immune/stress response and apoptosis-related (Figs. 2b and S7a). Given that beta cells are destroyed by immune cells in T1D, it is possible that these remaining Beta-1 cells were not targeted by the immune system. It is also possible that these beta cells are able to survive and function after immunological attack by decreasing immune signaling and apoptotic signaling via downregulation of the TP53 pathway (Figs. 2b and S7a), which is notable given that upregulation of the TP53 pathway and an associated increase in susceptibility to apoptosis has been observed in T1D[37,38]. Hence, these results suggest that cells from AAb+ donors in this beta cell population are either spared from destruction or employ similar protective molecular mechanisms to enhance survival and function, which is further supported by the fact that the expression of immune checkpoint protein PDL-1 (CD274) is upregulated in AAB+ Beta-1 cells compared to those from Controls.

The Beta-2 cell population displayed a small proportion of genes (4%; 283 genes; $p < 2.2e-16$) with elevated expression in both T1D and AAB+ cells when compared to Controls (Figs. 2c, and S7b). Additionally, an even smaller number of genes were downregulated in T1D and AAB+ Beta-2 cells when compared to Controls. Nonetheless, several pathways were found to be commonly dysregulated across both donor groups. Two interrelated pathways dysregulated in both T1D and AAB+ Beta-2 cells, namely ‘chaperone-mediated protein folding’ and ‘response to topologically incorrect protein’, suggest a dysregulation of protein folding, an essential function for cellular homeostasis. Additionally, the ‘TNF-alpha/NF-kappa B signaling’ pathway, which has been implicated as an important regulator of autoimmune processes[39], was significantly downregulated across the two donor groups in the Beta-2 cell population (Figs. 2c and S7b). Together, our differential expression analyses extend earlier studies on the pathways triggering beta cell dysfunction and death.

Given the clear segregation of ductal cell populations by donor group, we next examined the transcriptional changes in the two populations of ductal cells, Ductal-1 and Ductal-2. A large number of genes was upregulated in T1D (7,175 genes) and AAB+ (4,371 genes) Ductal-1 cells when compared to Controls, a significant number of which were common between the two donor groups (Figs. 2d and S8a; 2,283 genes; 25%; p-value < 1e-12). Notable induced pathways upregulated in T1D and AAB+ cells are associated with apoptosis, stress, and immune response (Figs. 2d and S8a). In the Ductal-2 cell population, although many upregulated genes were observed in T1D (6,841 genes), there were not nearly as many upregulated genes in AAB+ cells (1,106 genes) when compared to Controls (Figs. 2e and S8b). Furthermore, in the T1D and AAB+ Ductal-2 cell population, there was a modest but significant overlap between upregulated genes (Figs. 2e and S8b; 11%; p-value < 1e-12). Nevertheless, various gene pathways were found to be significantly upregulated across both ductal populations (Figs. 2 d–e and S8 a–b). Taken together, these findings suggest that although AAb+ donors maintain normoglycemia, significant transcriptional dysregulation is occurring in AAB+ endocrine and exocrine cells that is highly similar to that in T1D.

Next, we directly compared T1D to AAB+ cells (Fig. 2f and Tables S8–S13). For both groups of beta cells, genes associated with autophagy, stress response, and immune-related pathways were activated in AAB+ cells compared to T1D cells (Figs. 2 g–h and S7 a–b). Although similar pathways were upregulated in AAB+ Beta-1 and Beta-2 cells, apoptotic and adaptive immune system signaling were only upregulated in Beta-2 AAB+ cells. These

data suggest that this population is undergoing cell death, indicated by the upregulation of adaptive immune cell genes and *BCL10*. In Ductal cell populations, there was a larger number of upregulated genes in T1D (Figs. 2 i–j, and S8 a–b). Notably, apoptotic, metabolic, protein folding, and immune responses were activated in T1D ductal cells in comparison to AAB+ ductal cells (Figs. 2 i–j and S8 a–b). Remarkably, interferon alpha and beta pathways, known to be critical in T1D disease pathogenesis[40–42], were significantly elevated in T1D ductal cells compared to either Control or AAB+ ductal cells (Extended Data Fig. 5d). Our molecular evidence supports more recent findings of exocrine abnormalities in T1D[11,12], positioning these exocrine cells in disease pathogenesis. Taken together, AAB+ cells exhibit significant transcriptional changes like those observed in T1D.

Beta cell gene signature is correlated with the anti-GAD titer

Pancreatic tissues from AAb+ donors collected by HPAP can potentially offer a unique insight into the initial molecular events of T1D pathogenesis. A landmark study following patients from birth determined that ~69% of children with multiple islet autoantibodies progressed to T1D after islet autoantibody seroconversion[43]. Among HPAP donors, only one donor with no history of T1D expressed two islet autoantibodies while the other normoglycemic AAb+ donors were anti-glutamic acid decarboxylase (GAD) autoantibody positive. Considering that the longitudinal children study also revealed that the risk of diabetes in children who had no islet autoantibody was 0.4% in contrast to 14% for children expressing a single islet autoantibody[43], we next focused on the transcriptional landscapes of islets in GAD+ donors and queried for cell types whose transcriptional signature strongly correlated with the GAD titer among the GAD+ subjects. We devised a strategy to determine the number of genes whose expression levels significantly correlated with the GAD titer across GAD+ donors, either positively or negatively. However, we detected only positive correlation of statistical significance between gene expression levels and GAD titers. Strikingly, the top cell type with the largest number of genes (1,473) with significant correlation with the GAD titer in AAb+ donors was Beta-1 cells (Fig. 3a, Table S14). Plotting the average expression levels of cells from each GAD+ donor for these 1,473 genes in Beta-1 cells confirmed this finding (Fig. 3b). To define the identity of genes with an increase in their expression levels correlating with GAD levels, we performed gene-ontology analysis. Our approach highlighted the relevance of endocytosis, lysosome, protein processing in ER and MAPK signaling in Beta-1 cells[13–18] (Fig. 3c). Additional comparison of the cellular clustering of the one AAb+ donor expressing two autoantibodies (IA-2 and ZnT8; AAB+ #5; HPAP043) with GAD+ donors across the AAB+-specific clustering (Fig. S3b) or all donor-type clustering (Fig. 1c) revealed the distinct transcriptional signature of the double autoantibody-expressing AAb+ donor in comparison to single AAb+ GAD+ donors (Figs. 3d and S2l). This analysis also revealed an overall similarity of GAD+ donors, which modestly displayed GAD level-dependent cell co-segregation (Figs. 3d and S2l). Together, our unbiased strategy puts forward Beta-1 cells as the top cell type whose transcriptional outputs correlate with anti-GAD levels, suggesting the dynamic landscape of transcriptome in normoglycemic autoantibody-positive individuals.

MHC Class II expression is enriched in T1D ductal cells

The major genetic susceptibility determinants of T1D have been mapped to the MHC Class II genes[44,45]. We therefore sought to determine which cell types or donor states disproportionately express genes in this pathway. Using our scRNA-seq data, we found that genes associated with MHC Class II activity were enriched in Immune, Endothelial, and Ductal clusters (Extended Data Fig. 6 a–d). The lack of enrichment of the immune cell marker *PTPRC* or other genes associated with immune cells across the endocrine and ductal dendrogram supports the notion that the enrichment of MHC Class II associated genes in ductal cells is not due to immune cell contamination (Extended Data Fig. 4 a,d and Extended Data Fig. 6g). Next, we evaluated the expression of *HLA-DPB1*, an MHC class II gene associated with T1D risk, and *KRT19*, a ductal cell marker, across ductal and endocrine cell types. We identified five clusters with high *HLA-DPB1* and high *KRT19* expression, which accounted for 10.9% of all cells (7,588 cells) (Fig. 4 a–b and Extended Data Fig. 6 e–f). Strikingly, cells from T1D donors disproportionately contributed to this population of MHC Class II-expressing ductal cells (Fig. 4c; p-value < 2.2e-16). This observation is not due to sampling issues pertaining to the difficulty of isolating high purity islets from T1D donors. This conclusion is supported by the fact that even though the Ductal-1 cell population consists of very similar numbers of Control and T1D donor ductal cells (4,217 and 4,154 cells, respectively), there is a marked difference in the percentage of Control versus T1D MHC class II-expressing Ductal-1 cells, at 35% and 91%, respectively (Fig. 4c; p-value < 2.2e-16).

T1D ductal cells assume the transcriptional identity of dendritic cells

Dendritic cells (DCs) are among the major professional antigen-presenting cells expressing MHC Class II proteins with the salient function to ingest antigens and present processed epitopes to T cells, thereby regulating adaptive immune responses by activating or suppressing T cells[46]. Considering that MHC Class II proteins are required for antigen-presentation in dendritic cells, we next evaluated whether there are any other similarities between transcriptional profiles of T1D ductal cells and conventional dendritic cells. Hence, we performed gene-set-enrichment analysis using gene signatures of dendritic cell subtypes, which were recently defined using scRNA-seq profiling in human blood[47]. Remarkably, we found a highly significant enrichment of the DC1 gene signature in Ductal-2 cells of T1D donors, while no other annotated islet cell type revealed such significant and strong enrichment of gene signatures associated with dendritic cell subtypes (Fig. 4d and Extended Data Fig. 7a). DC1 corresponds to the cross-presenting CD141/BDCA-3$^+$ cDC1, which is best marked by CLEC9A[47]. Of note, the enrichment of other dendritic cell subtype gene signatures in T1D ductal cells was not statistically significant (Extended Data Fig. 7a).

To activate T cells, dendritic cells are required to express both MHC Class II proteins and costimulatory proteins CD80 and CD86[46]. In the absence of CD80 and CD86, antigen-presentation by dendritic cells can lead to tolerance and T cell suppression[48]. We found that CD80 and CD86 were not expressed in T1D ductal cells, suggesting a lack of costimulatory signal in these dendritic cell-like ductal cells in T1D donors (Fig. 4e). Additionally, the inhibitory receptor VSIR, which negatively regulates T cell responses[49], showed higher expression in T1D compared with control ductal cells (Extended Data Fig.

7b). Moreover, the ductal cells in T1D expressed high levels of interferon genes such as *ICAM1, ISG20*, and *IRF7* (Figs. 4f and Extended Data Fig. 5d). Hence, our single-cell transcriptional profiling detected an enrichment of ductal cells with transcriptional similarities to tolerogenic DCs. These results imply an unappreciated role for T1D ductal cells potentially acting as decoy receptors in an apparent attempt to deactivate $CD4^+$ T cells by inducing tolerance during immune invasion of the pancreas.

Multimodal confirmation of MHC Class II^+ ductal cells

We next sought to corroborate our transcriptomic-based finding of MHC Class II expression on ductal cells in T1D by employing additional experimental modalities: two high-throughput technologies, CyTOF and IMC, in addition to immunofluorescence experiments. Our integrative approach with CyTOF combined ~ 7,000,000 live, cultured single cells from 12 donors, which had also been profiled by scRNA-seq (4 Control, 4 AAB+, and 4 T1D donors). This additional modality scaled our analytical strategy to millions of cells, measuring the expression levels of 35 proteins (Table S15). Since the strategy we used to annotate cells using scRNA-seq[8] is not applicable to CyTOF measurements, we developed a new machine-learning method to annotate cells based on canonical markers (Extended Data Fig. 8 a–e). Using CyTOF, we identified a population of ductal cells expressing HLA-DR, an MHC Class II protein (Fig. 2a). Notably, we found that cells from T1D donors constituted the largest percentage of this cluster, in agreement with the findings from scRNA-seq (Fig. 5b; p-value < 1e-6). Furthermore, HLA-DR-expressing ductal cells made up a larger percentage of total cells across individual T1D donors compared with Control or AAB+ donors (Fig. 5c, p-value=0.00507). A two-parameter (cytokeratin and HLA-DR) analysis on all single cells analyzed by CyTOF further confirmed the presence of this double-positive population across multiple donors (Fig. 5d and Extended Data Fig. 8 f–g). Notably, these ductal cells did not express CD45, the hallmark of leukocytes (Fig. 5e). The identification of ductal cells with MHC Class II molecules using both scRNA-seq and CyTOF strongly corroborates the increased frequency of this population in T1D.

Having identified a population of ductal cells with MHC Class II molecules enriched in T1D donors by two experimental modalities in our integrative analysis, we next sought to study these ductal cells in pancreatic tissues independent of islet culture by means of anatomical-spatial features in pancreatic tissues by IMC[50]. While measurements with CyTOF and scRNA-seq assays rely on the profiling of dissociated cells, IMC retains spatial information by analyzing tissues fixed directly from the native human pancreas. We again amended our analytical pipeline with an optimized cell annotation approach for the IMC technology. We harnessed the expression levels of 33 proteins quantified by IMC in more than 1 million cells across 143 tissue slides from 19 donors, including 11 individuals not previously assessed by scRNA-seq or CyTOF for an independent validation of our findings (Table S16). This analysis confirmed that MHC Class II-expressing ductal cells were predominately present in T1D donors (Fig. 5 f–h and Extended Data Fig. 9 a–e). MHC Class II-expressing ductal cells were located in all regions of the pancreas (Extended Data Fig. 10a). Remarkably, the frequency of $CD11B^+$ myeloid cells annotated by our analytical strategy in both CyTOF and IMC measurements was highly correlated with the frequency of MHC Class II expressing ductal cells (Extended Data Fig. 10 b–

c). Immunofluorescence staining (IF) in native pancreatic tissues, followed by confocal microscopy, verified the existence of MHC Class II-expressing cells in a Control and a T1D donor (Fig. 5i). We identified MHC Class II-expressing ductal cells in both donors; however, there was a pronounced enrichment of MHC Class II-expressing ductal cells in the T1D pancreas (Fig. 5i). Representative examples of IMC measurements in tissues also confirm this finding (Fig. 6). Finally, cellular neighborhood analysis in pancreatic tissues established that HLADR-expressing ductal cells were surrounded by $CD4^+$ T cells and myeloid cells including $CD11B^+$ dendritic cells (Extended Data Fig. 10 d–f; p-value < 1e-2). Together, our multimodal single-cell measurements from transcriptomics to spatial proteomics in ductal cells suggest that ductal cells are transcriptionally similar to tolerogenic DCs, implying an unappreciated role of these exocrine cells in modulating T cell activity in long-term T1D.

Discussion

Employing three high-throughput single-cell technologies, we provided a comprehensive atlas of millions of cells using integrative multi-modal analyses as a molecular microscope to investigate cellular diversity in the pancreas of T1D, AAb+, and non-diabetic human organ donors. These data, including paired samples across technologies, enable an exploration of the pancreatic environment in both healthy and disease states.

We found that AAb+ donors exhibit similar transcriptional changes as T1D donors in various endocrine and exocrine cells, despite these donors retaining normoglycemia. Remarkably, the unique collection of GAD+ donors in the HPAP database allowed us to delineate Beta-1 cells as the primary cell type whose transcriptional outputs correlate with anti-GAD titers, suggesting the existence of dynamic transcriptional landscape in autoantibody-positive individuals. Although it is impossible to discern at present whether these transcriptional changes are contributing to or are byproducts of disease pathogenesis, the mere discovery of molecular phenotypic changes in pancreatic cells of AAb+ individuals should advance our understanding of early pancreatic perturbations occurring in T1D.

The most striking finding arising from our study is that cells of the exocrine compartment show transcriptional and gene ontological changes in the T1D disease setting. Ductal cells from T1D donors, in contrast with those from non-diabetic or AAb+ donors, express high levels of MHC Class II and interferon pathways, are surrounded by $CD4^+$ T cells and dendritic cells and are transcriptionally similar to tolerogenic dendritic cells. Although our study represents the first report of ductal cells expressing MHC Class II proteins in the T1D context, this finding is in accordance with previous literature documenting an elevation of immune cells in the exocrine pancreas of T1D donors[35,42,51] and regulation of MHC Class II genes by the interferon signaling pathway[52]. Moreover, the expression of MHC Class II proteins in pancreatic ductal adenocarcinomas has been reported[53]. Recent studies also support a role for epithelial cells as facultative, non-professional antigen-presenting cells in the gut and lung[54,55], and expression of MHC Class II proteins in non-lymphoid cells in the pancreas[56] has been shown. We posit that these cells exhibit a tolerogenic response to chronic T cell infiltration in pancreatic tissues and appear to be an ultimately unsuccessful attempt of the pancreas to limit the adaptive T cell response responsible for destroying beta cells. While this interpretation is strongly supported by our multimodal data analysis in

human pancreatic tissues, the limitation of our study relates to lack of functional validation of this hypothesis. Our future efforts utilizing mouse genetics will enable us to further validate the functional relevance of these findings. Together, our study provides a unique resource of millions of cells of the pancreatic environment and unmasks exocrine ductal cells as potential responders to immune infiltration in T1D.

One technical question under intense debate in the scRNA-seq community is how to perform differential expression analysis. Squair *et al.*[57] compared differential expression analysis techniques in scRNA-seq datasets, utilizing bulk RNA-seq data as the ground-truth for measuring false-positives. They concluded that predictions using the pseudo-bulk approach are the most similar to predictions from bulk RNA-seq data. Contradicting Squair *et al.*[57], Zimmerman *et al.*[36] published a study comparing techniques for performing differential expression analysis in scRNA-seq datasets and argued that pseudo-replication is acknowledged as one of the most common statistical mistakes in the scientific literature. Instead, they proposed the use of computationally expensive generalized linear mixed models for the analysis of scRNA-seq data. In summary, the contradictory results of these two studies reveal lack of consensus on alternative differential expression methods. Aware of these challenges in the analysis of scRNA-seq data, we took advantage of multimodal measurements such as IMC, CyTOF, and IHC to assess the reproducibility of our novel findings related to ductal cells in T1D donors across independent experimental assays.

Materials and Methods

Experimental model and subject details

Pancreatic islets were procured by the HPAP consortium (RRID:SCR_016202; https://hpap.pmacs.upenn.edu), part of the Human Islet Research Network (https://hirnetwork.org/), with approval from the University of Florida Institutional Review Board (IRB # 201600029) and the United Network for Organ Sharing (UNOS). A legal representative for each donor provided informed consent prior to organ retrieval. For T1D diagnosis, medical charts were reviewed and C-peptide levels were measured in accordance with the American Diabetes Association guidelines (American Diabetes Association 2009). All donors were screened for autoantibodies prior to organ harvest, and AAb positivity was confirmed post tissue processing and islet isolation.

Organs were processed as previously described[19,35]. Table 1 and 2 summarizes donor information. Pancreatic islets were cultured and dissociated into single cells as previously described (22). Total dissociated cells were used for single cell capture for each of the donors, except AAB+ donor #1 (HPAP019), which was enriched for beta cells.

The C-peptide analysis was performed using a two site immuno-enzymatic assay from Tosoh Bioscience on a Tosoh 2000 auto-analyzer (Tosoh, Biosciences, Inc., South San Francisco, CA). Briefly, the test sample is bound with a monoclonal antibody immobilized on a magnetic solid phase and an enzyme-labeled monoclonal antibody, and then the sample is incubated with a fluorogenic substrate, 4-methylumbelliferyl phosphate (4MUP). The amount of enzyme-labeled monoclonal antibody that binds to the beads is directly proportional to the C-peptide concentration in the test sample. A standard curve is

constructed using calibrator of known concentration, and unknown sample concentrations are calculated using the curve. The C-peptide assay is calibrated against WHO IS 84/510 standard. The assay has a sensitivity level of 0.02 ng/mL. To monitor the assay performance, a set of low, medium, and high C-peptide level quality control samples are analyzed several times per day. The inter-assay coefficients of variability for the low, medium, and high C-peptide controls are 3.2%,1.6%, and 1.8%, respectively. The results of the analyses of the long-term monitoring pools have demonstrated a consistently low variation around the target values, thus ensuring result consistency.

Serum from organ donors is tested for GAD, IA-2, mIAA, and ZnT8A autoantibodies by radioligand-binding assay (RIA) as previously described[58]. Micro IAA (mIAA) and ZnT8A were performed with in-house RBA[59,60] and the assay thresholds (index of 0.010 mIAA and 0.020 for ZnT8A) was set up as 99th percentile of over 100 controls. GAD and IA-2 was performed with NIDDK harmonized standard methods (3) and the upper limits of normal (20 DK units/ml for GAD and 5 DK units/ml for IA-2) was established around the 99th percentile from receiver operating characteristic curves in 500 healthy control subjects and 50 patients with new onset diabetes. In the most recent IASP Workshop, the sensitivity and specificity were 78% and 99% for GAD, 72% and 100% for IA-2, 62% and 99% for mIAA, 74% and 100% for ZnT8A, respectively.

scRNA-seq islet capture, sequencing, and processing

The Single Cell 3’ Reagent Kit v2 or v3 was used for generating scRNA-seq data. 3,000 cells were targeted for recovery per donor. All libraries were validated for quality and size distribution using a BioAnalyzer 2100 (Agilent) and quantified using Kapa (Illumina). For samples prepared using ‘The Single Cell 3’ Reagent Kit v2’, the following chemistry was performed on an Illumina HiSeq4000: Read 1: 26 cycles, i7 Index: 8 cycles, i5 index: 0 cycles, and Read 2: 98 cycles. For samples prepared using ‘The Single Cell 3’ Reagent Kit v3’, the following chemistry was performed on an Illumina HiSeq 4000: Read 1: 28 cycles, i7 Index: 8 cycles, i5 index: 0 cycles, and Read 2: 91 cycles. Cell Ranger (10x Genomics; v3.0.1) was used for bcl2fastq conversion, aligning (using the hg38 reference genome), filtering, counting, cell calling, and aggregating (--normalize=none).

scRNA-seq clustering, doublet removal, & cell type classification

Seurat v3.1.5[32,61] was used for filtering, UMAP generation, and initial clustering. Genes expressed in at least 3 cells were included, as were cells with at least 200 genes. nFeature, nCount, percent.mt, nFeature vs nCount, and percent.mt vs nCount plots were generated to ascertain the lenient filtering criteria of $200 < \text{nFeature} < 8{,}750$, $\text{percent.mt} < 25$, and $\text{nCount} < 125{,}000$. Data was then log normalized, and the top 2,000 variable genes were detected using the “vst” selection method. The data was then linearly transformed (“scaled”), meaning that for each gene, the mean expression across cells is 0 and the variance across cells is 1. Principle component analysis (PCA) was then carried out on the scaled data, using the 2,000 variable genes as input. We employed two approaches to determine the dimensionality of the data, i.e. how many principal components to choose when clustering: (1) a Jackstraw-inspired resampling test that compares the distribution of p-values of each principle component (PC) against a null distribution and (2) an elbow

plot that displays the standard deviation explained by each principal component. Based on these two approaches, 17 PCs with a resolution of 1.2 were used to cluster the cells, and non-linear dimensionality reduction (UMAP) was used with 17 PCs to visualize the dataset.

DoubletFinder v2.0[62] was used to demarcate potential doublets in the data as previously described, with the following details: 17 PCs were used for pK identification (no ground-truth) and the following parameters were used when running doubletFinder_v3: PCs = 17, pN = 0.25, pK = 0.0725, nExp = nExp_poi, reuse.pANN = FALSE, and sct = FALSE (Fig. S1d). Scrublet v0.2.1 (*18*) was also used to demarcate potential doublets. We removed all cells that were flagged as doublet by both or either approach.

The raw data for the remaining cells were filtered using the following criteria, which resulted in 69,645 cells remaining: 200 < nFeature < 8,750, percent.mt < 25, and nCount < 100,000. The data were log normalized, the top 2,000 variable genes were detected, the data underwent linear transformation, and PCA was carried out, as described above. Both the Jackstraw-inspired resampling test and an elbow plot of standard deviation explained by each principal component were used to determine the optimal dimensionality of the data, as described above. Based on these two approaches, 26 PCs with a resolution of 1.2 was used to cluster the cells, and UMAP was used with 26 PCs to visualize the 49 clusters detected.

Garnett was used for initial cell classification as previously described[8]. In brief, a cell type marker file (Table S17) with 17 different cell types was compiled using various resources, and this marker file was checked for specificity using the “check_markers” function in Garnett by checking the ambiguity score and the relative number of cells for each cell type. A classifier was then trained using the marker file, with “num_unknown” set to 500, and this classifier was then used to classify cells and cell type assignments were extended to nearby cells, “clustering-extended type” (Louvain clustering) (Fig. S3d). Upon inspection of cluster purity using canonical gene markers of the major pancreatic cell types across the Seurat-generated clusters, we found that the abundant and transcriptionally distinct cell types form generally distinct and unique clusters: beta cells (*INS* high), alpha cells (GCG high), acinar cells (*CPA1* high), ductal cells (*KRT19* high), endothelial cells (*VWF* high) stellate cells (*RSG10* high), and immune cells (*PTPRC*, also known as CD45 or leukocyte common antigen, high) (Fig. S3e). In contrast across the Seurat-generated clusters, the rarer and/or less transcriptionally distinct cell types did not clearly segregate, namely delta cells (*SST* high), PP cells (*PPY* high), and epsilon cells (*GHRL* high).

Integration and label transfer was used to further validate Garnett cell-type assignments as previously described[63]. To label canonical cell types, a previous snRNA-seq data set of adult pancreatic cells (EGAS00001004653) was used as a reference for the “query” datat set presented in this study. First, SCTransform was used to preprocess the data[61]. Briefly, SCTransform uses a generalized linear model (GLM) for each gene with UMI counts as the response variable and sequencing depth as the predictor. To integrate data for UMAP visualization, Seurat integration was used to identify common anchor points between data sets. Seurat uses diagonalized canonical correlation analysis (CCA) followed by L2-normalization and searching for mutual nearest neighbors (MNN). Then, anchors between data sets are compared based on their local neighborhood structure of other anchors

to receive “correction vectors”. These correction vectors are then subtracted from the query gene expression matrix, resulting in an integrated data set[63]. Similarly for label transfer, these anchors between data sets are instead labeled as discrete cell types and similar anchors assign cell labels from the reference cells to the query cells[63]. To assign canonical cell-type labels to Hybrid cells, the same integration and label transfer process was used but with a previous scRNA-seq pancreatic data set as a reference (GSE145126).

We employed the analytical workflow termed ‘TooManyCells’[9], which implements an efficient divisive hierarchical spectral clustering approach along with tree visualizations. We invoked the cellular classifier Garnett, which annotates cell types by training a regression-based classifier from user-provided cell type signatures. Briefly, for the clustering of all cells, the raw data from the 69,645 cells were normalized by total count and gene normalization by median count (TotalMedNorm) followed by term frequency-inverse document frequency (tf-idf) for clustering. For visualization of the comprehensive clustering, the dendrogram was first pruned using the TooManyCells flags ‘--min-distance-search “15”’ and “--smart-cutoff “15”’, followed by pruning using the flag ‘--max-step 6’.

For the clustering of ductal/endocrine cells, data from the ductal/endocrine cell clusters from the comprehensive tree were subsetted and normalized by TotalMedNorm followed by term tf-idf. For visualization of the ductal/endocrine tree, the dendrogram was first pruned using the TooManyCells flags ‘--min-distance-search “7”’ and “--smart-cutoff “7”’ followed by pruning using the flag ‘--max-step 7’. Data from the immune cell cluster from the comprehensive tree were subsetted and normalized by TotalMedNorm followed by tf-idf. For visualization of the immune tree, the dendrogram was first pruned using the TooManyCells flag ‘--max-step 4’. When individual genes were painted across any of the dendrograms, ‘TotalMedNorm’ was employed to normalize gene expression.

Differential Gene Expression, GSEA analysis, and Metascape analysis

Differential genes were found using edgeR through TooManyCells with the normalization “NoneNorm” to invoke edgeR single cell preprocessing, including normalization and filtering. For Metascape analysis[64], less than or equal to 3,000 differential genes (FDR < 0.05 and fold change (FC) > 0.1) were subjected to analysis. The top 20 clusters are displayed and a stringent cut-off of 1e-6 was applied to determine significant gene ontology pathways. For gene-set-enrichment-analysis (GSEA) analysis, GSEA Preranked (4.0.1) was run on a pre-ranked gene list using either user-provided pancreatic gene expression sets or standard hallmark gene signatures provided by the Molecular Signatures Database (MSigDB). Pseudobulk analysis was performed by taking the average of cells within individuals. The differential genes were found using edgeR through multi-sample, multi-group scRNA-seq analysis tool (muscat)[65]. The differential genes were filtered based on the combined threshold of p-value < 0.05 and fold change (FC) > 1.

Hybrid cell co-expression, DE analysis and heatmaps

For the differentially expressed genes (FDR < 0.05 and fold change (FC) > 0.1) between every two sample groups, we calculated the shared and unique genes in each cell type, and visualized it with Venn diagrams. The expression levels of the genes in each cell of the three

groups were extracted from the median normalized count matrix. Then we aggregate the expression levels in each group by taking the average value of the normalized counts. The mean expression values of the three groups were further normalized by the total expression level of that gene. We visualized the normalized expression level of differential genes with heatmaps.

To examine the co-expression of signature genes of some cell types, we normalized the median normalized matrix with log2(N +1). Then we selected the matrix of selected cell types by marker genes. The distribution of the cells from selected cell types by expression level of two marker genes were shown with geom_density_2d_filled() in ggplot2 package of R.

CyTOF data collection, input files, and preprocessing

Flow CyTOF was performed as described previously[50]. Briefly, after isolating the dissociated cells, barcoding was conducted for donors following the manufacturer's protocol (Fluidigm, 101-0804 B1). Following barcoding, metal-conjugated antibody labeling was carried out in 'FoxP3 permeabilization buffer' (eBioscience, 00-8333) with 1% FBS (Hyclone, Cat# 7207) for 12 hours at 4°C at a concentration of up to 3 million cells per 300 µl of antibody cocktail, followed by twice washing with FoxP3 permeabilization buffer. Cells were then incubated with the DNA intercalator Iridium (Fluidigm, 201192A) at a dilution of 1:4,000 in 2% paraformaldehyde (Electron Microscopy Sciences, 15714) in DPBS (Corning, 21-031-CV) at RT for 1 hr. Mass cytometry data were acquired by CyTOF (Fluidigm). Flow CyTOF data analyses of endocrine cell composition was performed using the Cytobank implement (https://www.cytobank.org/).

Normalized FCS files were pre-processed prior to TooManyCells analysis and visualization using FlowJo Version 10.6.1 by gating all events on singlets according to event length and DNA content and then on live cells based on cisplatin exclusion. The Singlet/Live gated population was exported to a CSV file for TooManyCells analysis. Two dimensional plots were visualized for combinations of individual channels.

TooManyCells clustering for CyTOF

TooManyCells was used to generate cell clades of CyTOF data. Cells with less than a total of 1e-16 signal were removed, leaving 6,945,575 cells. Upon inspection of protein levels across a tree with all cells, endocrine and exocrine compartments were further subsetted leading to a refined analysis of 4,521,988 cells. Quantile normalization of the raw counts was used in the clustering step. The resulting tree was pruned by collapsing nodes with less than (7 MAD X median # cells in nodes) cells within them into their parent nodes.

Imaging mass cytometry (IMC) analysis and Cell Segmentation

IMC was performed as described previously[35]. Cell segmentation of all images was performed with the Vis software package (Visiopharm). All image channels were pre-processed with a 3x3 pixel median filter, then cells were segmented by applying a polynomial local linear parameter-based blob filter to the Iridium-193 DNA channel of each image to select objects representing individual nuclei. Identified nuclear objects were

restricted to those greater than 10μm^2, then dilated up to 7 pixels to approximate cell boundaries. Per-cell object mean pixel intensities were then exported for further analysis.

TooManyCells clustering for IMC

TooManyCells was used to generate cell clades of IMC data. Cells with less than a total of 1e^{-16} signal were removed. Upon inspection of protein levels across a tree with all 1,170,001 cells, endocrine and exocrine compartments were further subsetted, leading to the refined analysis of 130,428 cells. The full tree with 1,170,001 cells was used for the assessment of HLA-DR-expressing ductal cells. Quantile normalization of the raw counts was used in the clustering step. The resulting tree was pruned by collapsing nodes containing less than (5 MAD X median # cells in nodes) cells within them into their parent nodes. Subsetting of the tree was done with "--root-cut 3" to focus on node 3 in relevant analyses, with additional pruning of (3 MAD X median # cells in nodes).

Cell-neighborhood analysis for IMC

Three labels were given to cells in the IMC neighborhood analysis: base, neighbor, and distant. Base cells originated from the chosen node, here node 16 in the node 3-focused IMC tree, or node 10 in the complete pruned tree which includes the former node 16. Given the x- and y-coordinates from IMC per cell, each cell's Euclidean distance to a base cell was calculated. If that distance was less than or equal to the chosen value, 20 for the complete pruned tree, the cell was assigned the neighbor label. Otherwise the cell was designated as distant.

Machine-learning method for cell annotation in IMC and CyTOF

To automatically label single cells from proteomic profiles, raw proteomic data along with a signature/marker file (listing unique marker proteins for each cell type) was taken as input. The raw data was normalized with an arcsinh transformation and a cofactor of 200 in case of CyTOF while log transformation followed by unit normalization in case of IMC data. The data was then randomly split into two halves (half donors in one set); cells from 50% of donors are in the training set while the remaining are in the test set. The splitting was done in a stratified fashion based on the disease condition (T1D, AAB+, Control). Semi-supervised learning was employed on the training set (clustering based on proteomics similar cells together) to generate cell labels for the first half of cells based on seeds (cluster centroids) calculated using a handful of labelled cells (0.1-10 percentile cells for each cell type) annotated using markers in the signature file. The annotated training set was used to train an Extreme learning machine (a fast classifier built on a feed forward neural network which does not need training for learning).

IF and confocal microscopy

Tissues were fixed in 10% buffered formalin overnight, washed several times in PBS, then dehydrated through an ethanol and xylenes, then embedded in paraffin and sectioned to 4-8um. Following deparaffinization through xylene and sequential rehydration, slides were subjected to heat antigen retrieval in a pressure cooker with Bulls Eye Decloaking buffer (Biocare). Slides were incubated in primary antibody overnight and secondary antibody

conjugated to peroxidase and then developed using Tyramide Signal Amplification (TSA, Akoya Biosciences). Slides were counter stained with DAPI, and then mounted and imaged on Zeiss LSM800. Primary antibodies used for staining were Mouse anti-CK19 (Santa Cruz sc-6278) and Rabbit anti-HLA-DR (Abcam ab92511).

Statistical analysis of box plots with Control, AAB+, and T1D donor states

The D'Agostino & Pearson omnibus normality test was used to assess whether the data from each group was normally distributed. If any group failed the D'Agostino & Pearson omnibus normality test, the Kruskal-Wallis test was applied. If none of the groups failed the D'Agostino & Pearson omnibus normality test, the one-way ANOVA test was applied.

Statistical analysis of cellular neighborhoods

Differential marker expression significance for neighbors in the IMC analysis was determined using permutation tests. For each marker, the distribution of that marker value for each of the designated n neighbors was compared against 100 distributions derived from n random cells across the entire IMC tree. The resulting p-value was calculated by the ratio of the number of permutations that had a lower median marker value than the observed marker value to the total number of permutations. If this value was < 0.5, the value was subtracted from 1 to switch directionality (number of permutations with a higher median value). To account for the two-tailed test, this value was multiplied by 2 for the final p-value calculation.

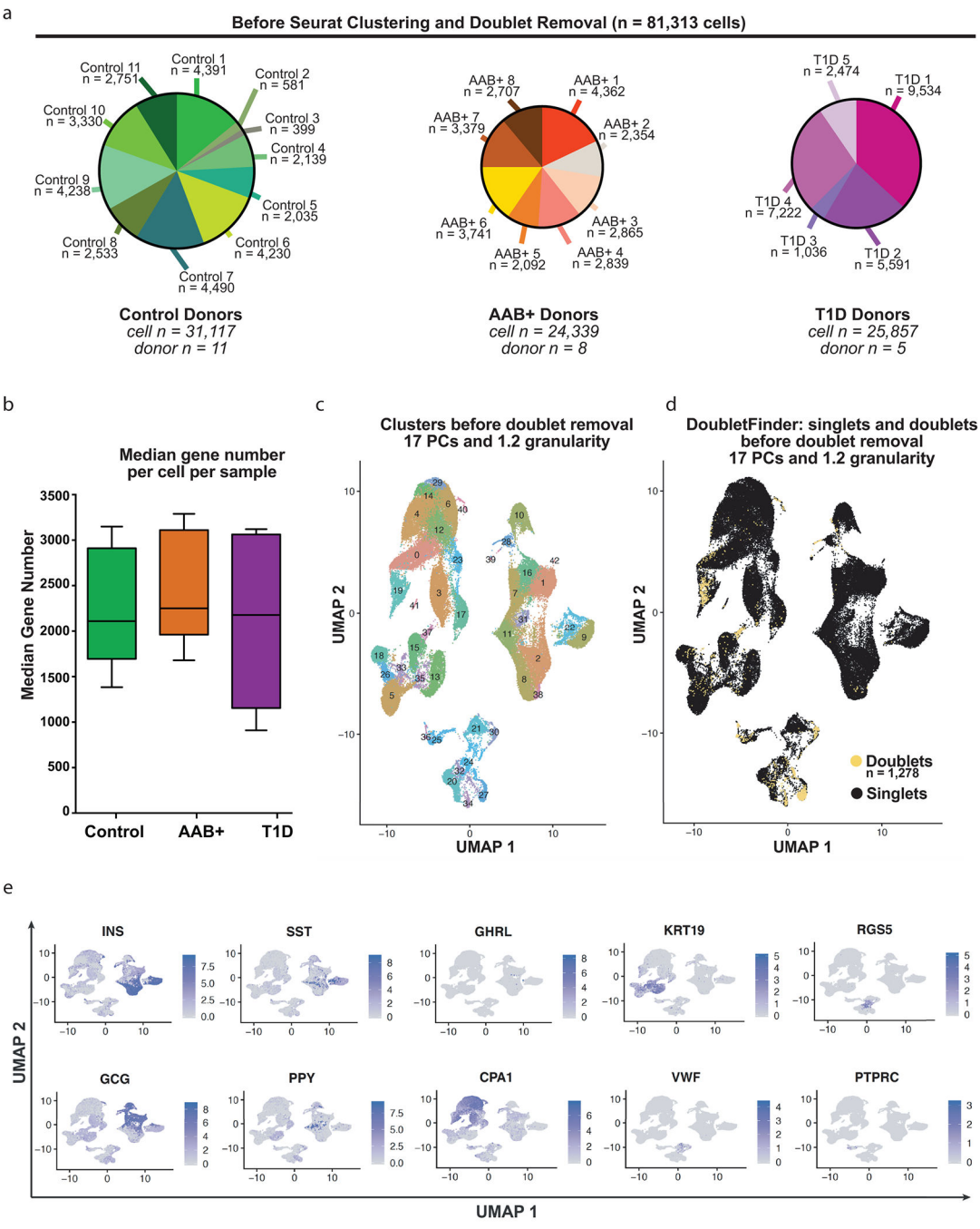
Statistical analysis of gene signatures in GAD+ donors

Pseudobulk counts of GAD+ donors across all cell types were identified using muscat tool[65]. The GAD levels for each GAD+ donors were retrieved from Table 1. To identify the correlation between gene signatures and GAD levels, the Spearman correlation test was conducted in each cell type. The threshold of correlation > 0.9 and p-value < 0.05 were used to determine the significantly correlated genes with GAD levels.

Assessment of common genetic variants associated with T1D

The CELLEX tool takes the scRNA-seq gene expression matrix as input and evaluates multiple metrics such as differential expression T-statistics, gene enrichment score, expression proportion, and normalized specificity index[26]. The average of these metrics is measured as expression specificity. The GWAS trait data and CELLEX estimates are given as input to CELLECT. CELLECT uses the genetic prioritization model (with a threshold of S-LDSC < 0.05) to quantify the association between the common phylogenetic GWAS signal and cell type expression specificity[26].

Extended Data



Extended Data Fig.1: Cell numbers and clustering before complete filtering

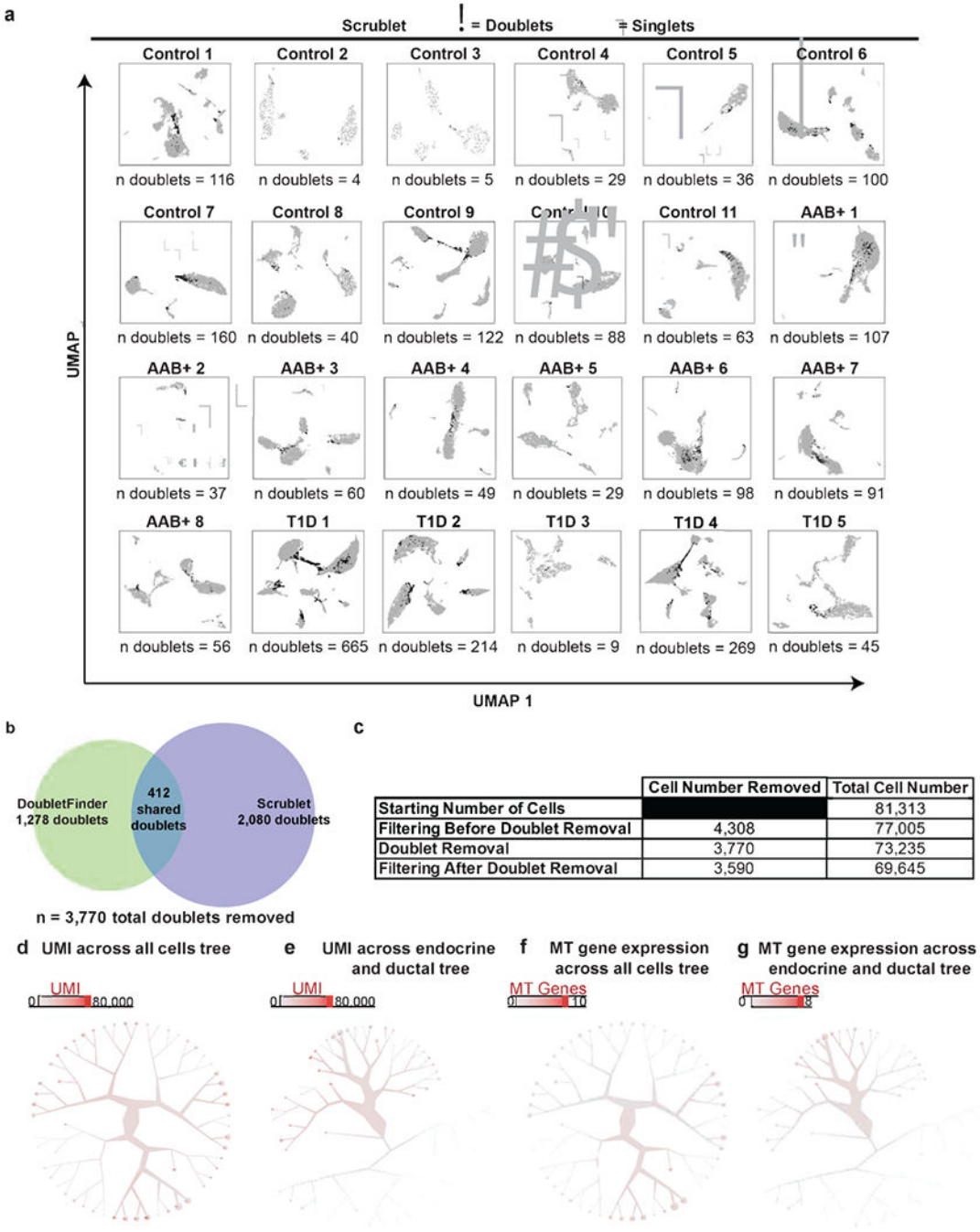
a) Pie chart displaying the cell numbers and proportions of each individual donor per donor type.

b) Box plot displaying the average gene number per cell per donor type.

c) UMAP visualization of cell clusters for all cells.

d) Doublets and singlets, as identified using DoubletFinder, across cell clusters visualized by UMAP.

e) UMAP visualization of the normalized gene expression counts of each canonical gene marker of each major cell type.

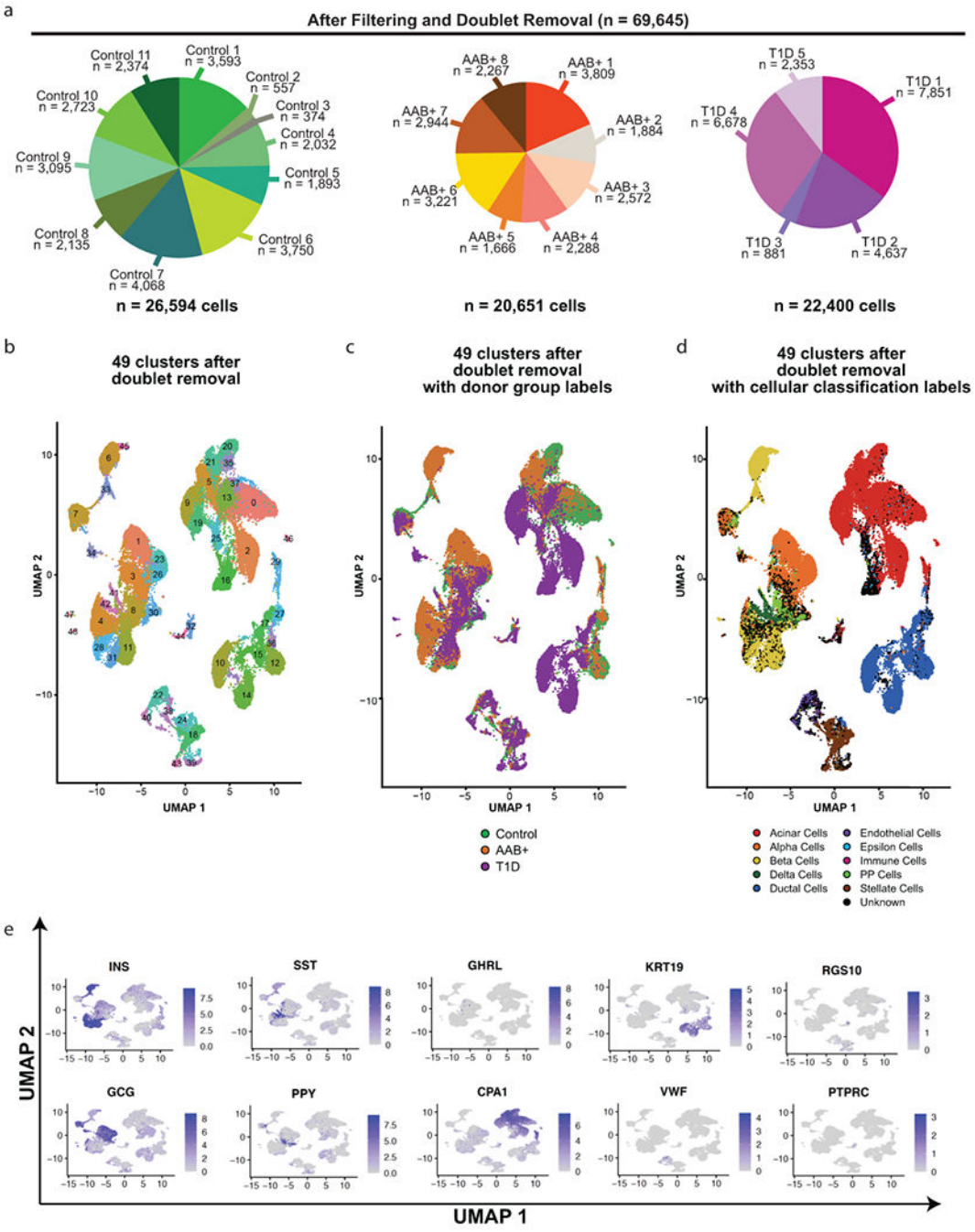


	Cell Number Removed	Total Cell Number
Starting Number of Cells		81,313
Filtering Before Doublet Removal	4,308	77,005
Doublet Removal	3,770	73,235
Filtering After Doublet Removal	3,590	69,645

Extended Data Fig.2: Doublet removal and UMI counts

a) Doublets and singlets, as identified using Scrublet, across cell clusters visualized by UMAP per individual.

b) Venn diagram indicating the number of cells deemed doublets by DoubletFinder and Scrublet, as well as cells that were commonly identified by both approaches.
c) Table indicating the number of cells removed and the resulting total cell number for each step of filtering.
d) Unique molecular identifier (UMI) counts per cell projected across the dendrogram visualization and clustering of all cells from Figure 1c. Pie charts at the end of the branches display the breakdown of UMI counts per cell within that terminal cluster. Cells begin at the start pin symbol, and from there are partitioned based on similarities and differences in gene expression.
e) UMI counts per cell projected across the dendrogram visualization and clustering of ductal and endocrine cells from Figure 1d. Pie charts at the end of the branches display the breakdown of UMI counts per cell within that terminal cluster. Cells begin at the start pin symbol, and from there are partitioned based on similarities and differences in gene expression.
f) Expression of genes associated with mitochondrial function projected across the dendrogram visualization and clustering of all cells from Figure 1c.
g) Expression of genes associated with mitochondrial function projected across the dendrogram visualization and clustering of ductal and endocrine cells from Figure 1d.



Extended Data Fig.3: Cell numbers and clustering after complete filtering

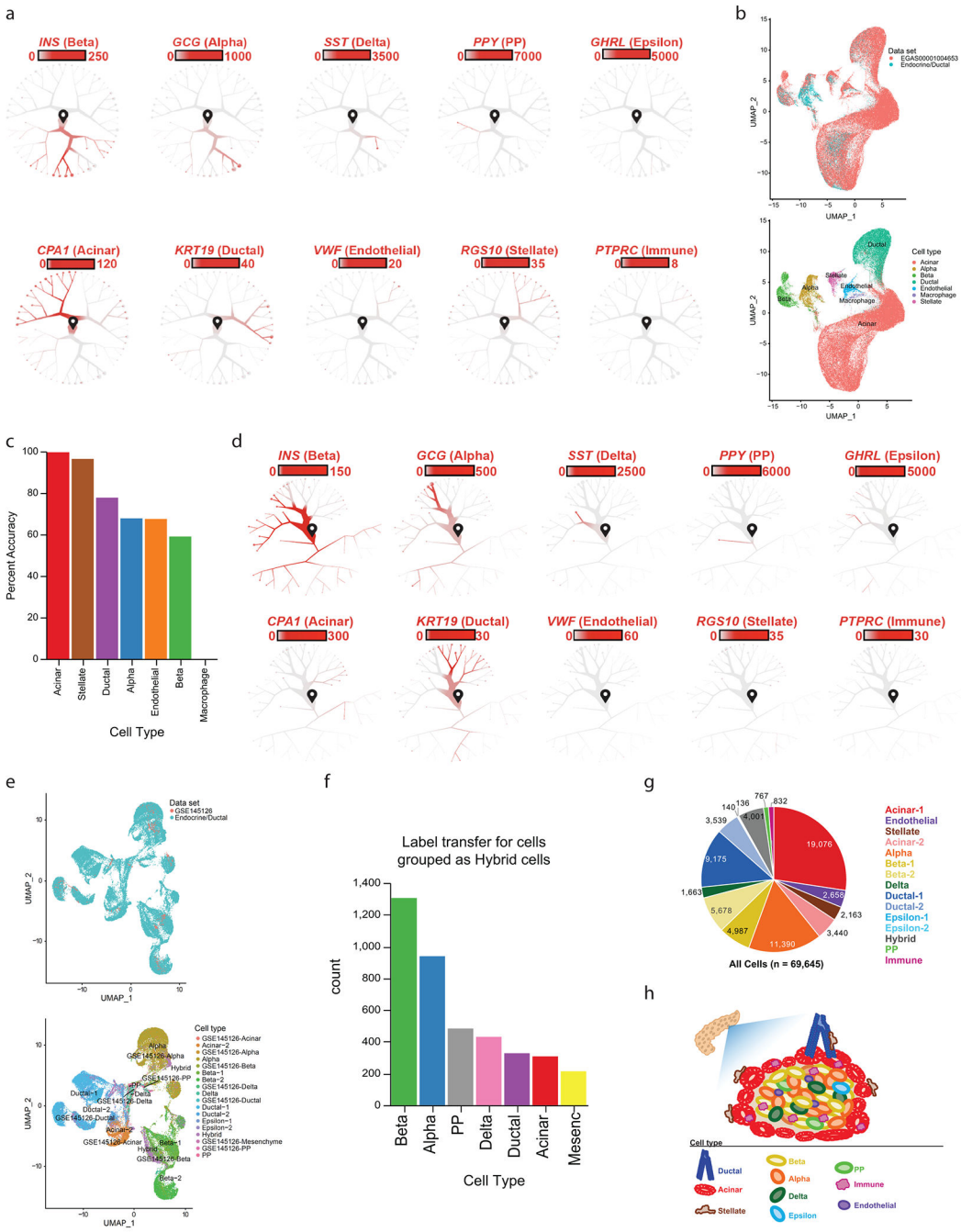
a) Pie chart displaying the cell numbers/proportions of each individual donor per donor type.

b) UMAP visualization of cell clusters for all cells.

c) UMAP visualization donor groups across clusters for all cells.

d) UMAP visualization of Garnett cellular classifications across clusters for all cells.

e) UMAP visualization of the normalized gene expression counts of each canonical gene marker of each major cell type.



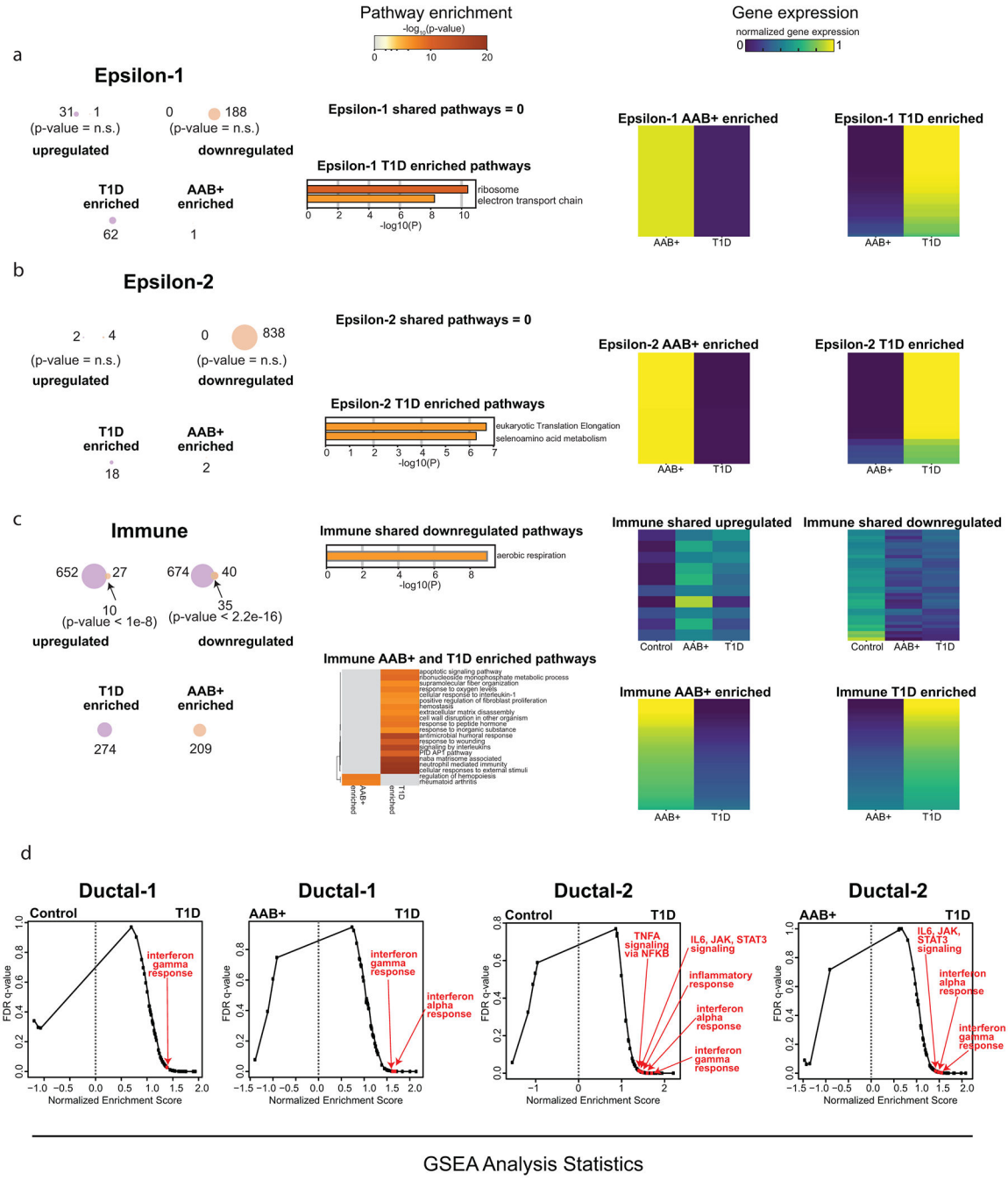
Extended Data Fig.4: Marker gene expression confirms canonical cell types

a) Dendrograms highlighting the expression of each canonical gene marker of each major cell type across the dendrogram of all cells in Figure 1c.

b) The classification of our scRNA-seq data was confirmed by a label transfer strategy using a previous single-nucleus RNA-seq data set in pancreatic islets[10].

c) Bar plot demonstrates percentages of agreement between previous annotation and our strategy using a label-transfer strategy.

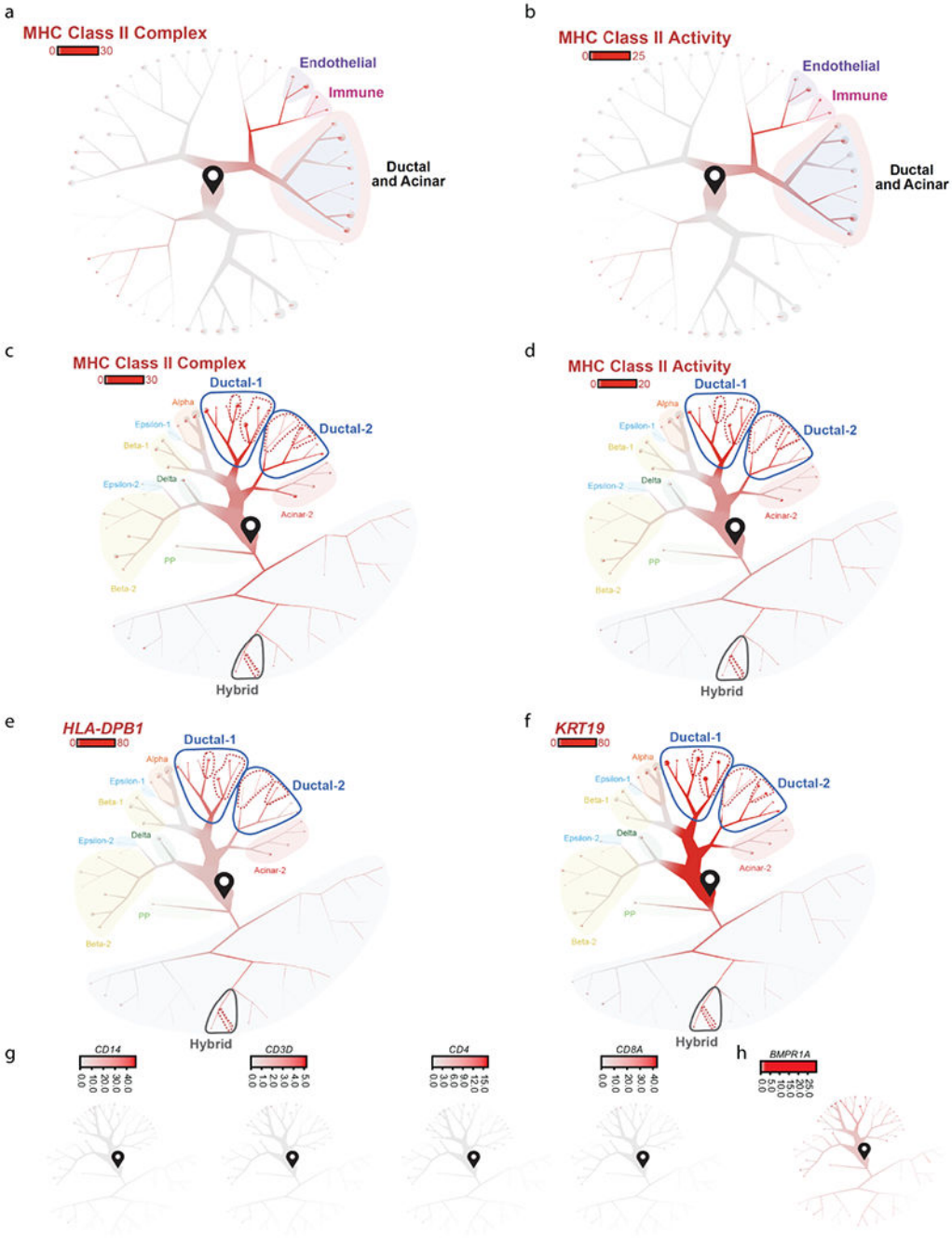
d) Dendrograms highlighting the expression of each canonical gene marker of each major cell type across the dendrogram of ductal and endocrine cells in Figure 1d.
e) To further validate the most closely related cell types to Hybrid cells, we used a label transfer strategy to a previous pancreatic islet scRNA-seq data set[19]. In concordance with Garnett and canonical gene markers, we corroborated the assignment of beta, alpha, and PP cells to these Hybrid cells.
f) Bar plot demonstrates annotation results of label transfer for cells grouped as Hybrid cells.
g) Pie chart displaying the cell numbers/proportions of each cell type defined in Figure 1, c and d.
h) Schematic of the human pancreatic islet anatomy and major cell types.



Extended Data Fig.5: Gene and gene ontology pathways that are shared and different across disease states in Epsilon-1, Epsilon-2, and Immune cells

(a-c) (Left) For each cell type, Venn diagrams indicate the numbers of upregulated and downregulated genes, as well as overlapping genes, across the two disease states. Circles indicate the numbers of genes that are 'T1D enriched' or 'AAB enriched'. p-values presented are the results of hypergeometric CDF tests (one-tailed test for overrepresentation). (Middle) For each cell type, displayed are gene ontology pathways that are shared across T1D and AAB+ cells when compared to Control cells (top) or pathways that are differently enriched in T1D cells vs AAB+ cells (bottom). The top 20 clusters are

displayed and a stringent cut-off of 1e-6 was applied to determine significant gene ontology pathways. (Right) Heatmaps displaying the degree of gene expression changes of genes (rows) that are shared (top) or differential (bottom) across AAB+ and T1D disease states. (d) GSEA analysis plots of FDR q-value vs Normalized Enrichment Score. For both ductal populations, Ductal-1 and Ductal-2, T1D cells were compared to AAB+ or Control cells to determine differentially enriched gene sets. Demarcated in red and labeled are signatures of interest.



Extended Data Fig.6: Corroboration of HLA-DR+ Ductal cells

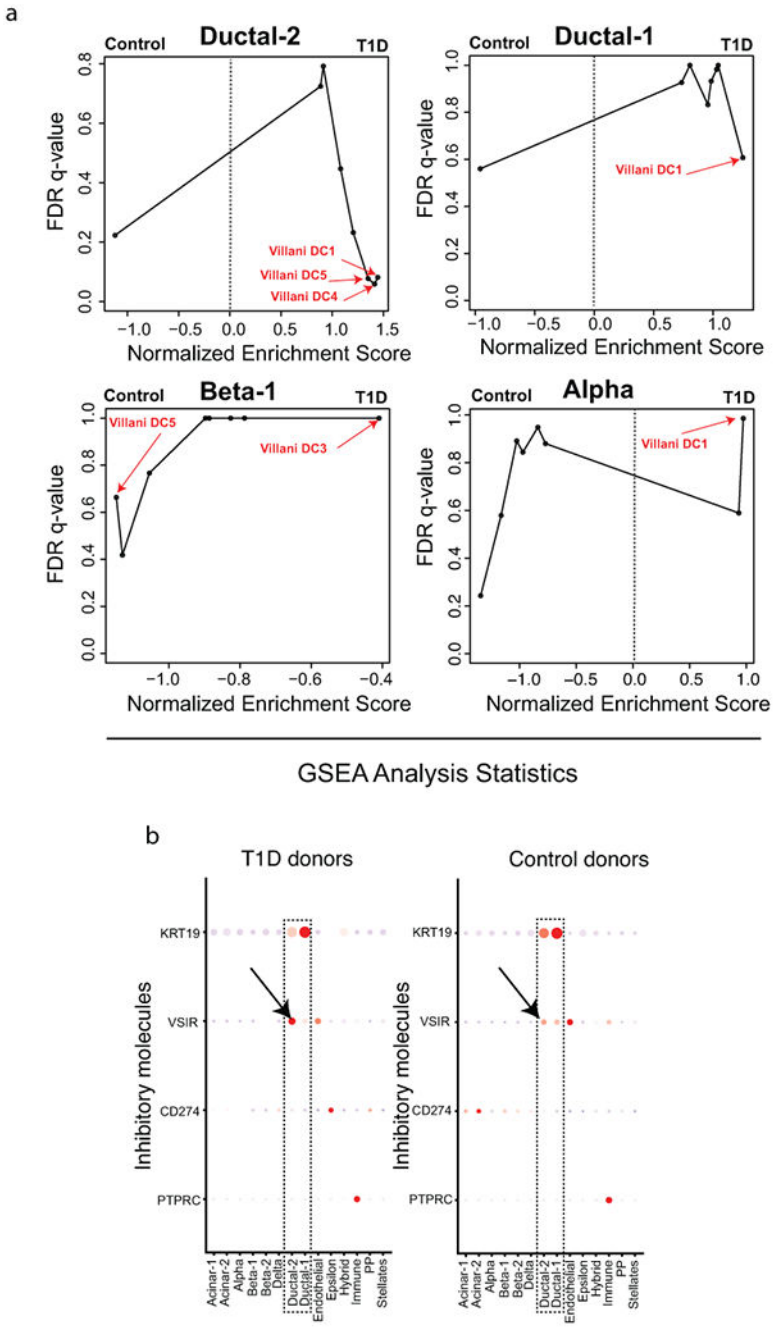
(a-b) Dendrograms highlighting the expression of the MHC class II complex (a) or MHC class II activity (b) across the dendrogram of all cells in Figure 1C. Scale bars represent normalized transcript numbers (mean across all MHC class II complex genes (a) or MHC class II activity genes (b)).

(c-d) Dendrograms highlighting the expression of the MHC class II complex (c) or MHC class II activity (d) across the dendrogram of ductal and endocrine cells in Figure 1D. Scale bars represent normalized transcript numbers (mean across all MHC class II complex genes (c) or MHC class II activity genes (d)).

(e-f) Dendrograms highlighting the expression of the *HLA-DPB1* (E) or *KRT19* (f) across the dendrogram of ductal and endocrine cells in Figure 1D. Scale bars represent normalized transcript numbers.

g) Dendrograms highlighting the expression of the immune-related genes across the dendrogram of ductal and endocrine cells in Figure 1D. Scale bars represent normalized transcript numbers.

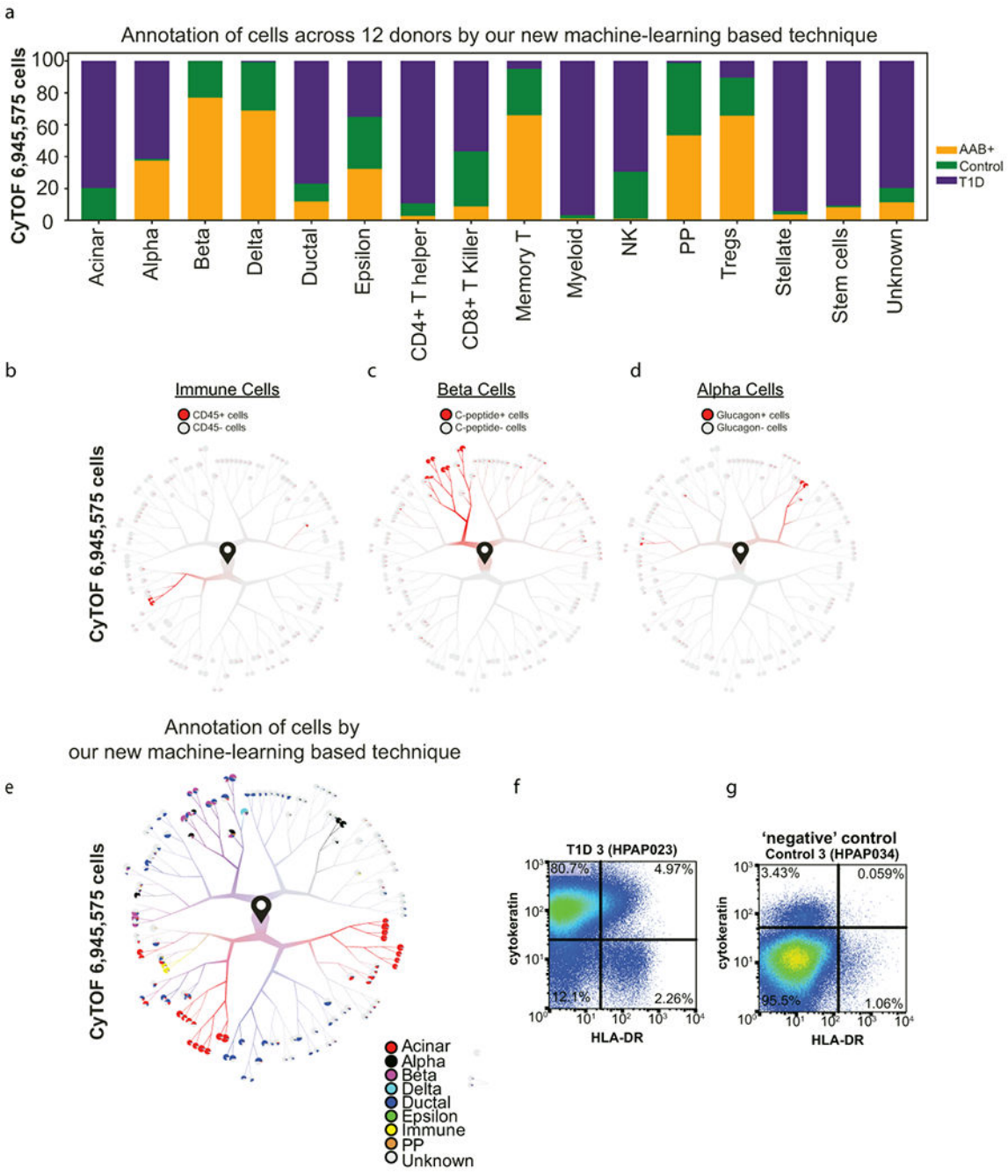
(h) Dendrograms highlighting the expression of the *BMPR1A* across the dendrogram of ductal and endocrine cells in Figure 1D. Scale bars represent normalized transcript numbers.



Extended Data Fig.7: GSEA analysis across annotated cells types for dendritic cells gene sets.

a) DC1 gene signature is significantly enriched within Ductal-2 cells of T1D donors. Integrated GSEA analysis for dendritic cells gene sets from Villani et al[47] across ranked lists of differentially expressed genes between T1D and control donors.

b) Expression analysis of the inhibitory marker VSIR in dendritic cells demonstrates the high level of this gene in T1D ductal cells compared with control ductal cells.



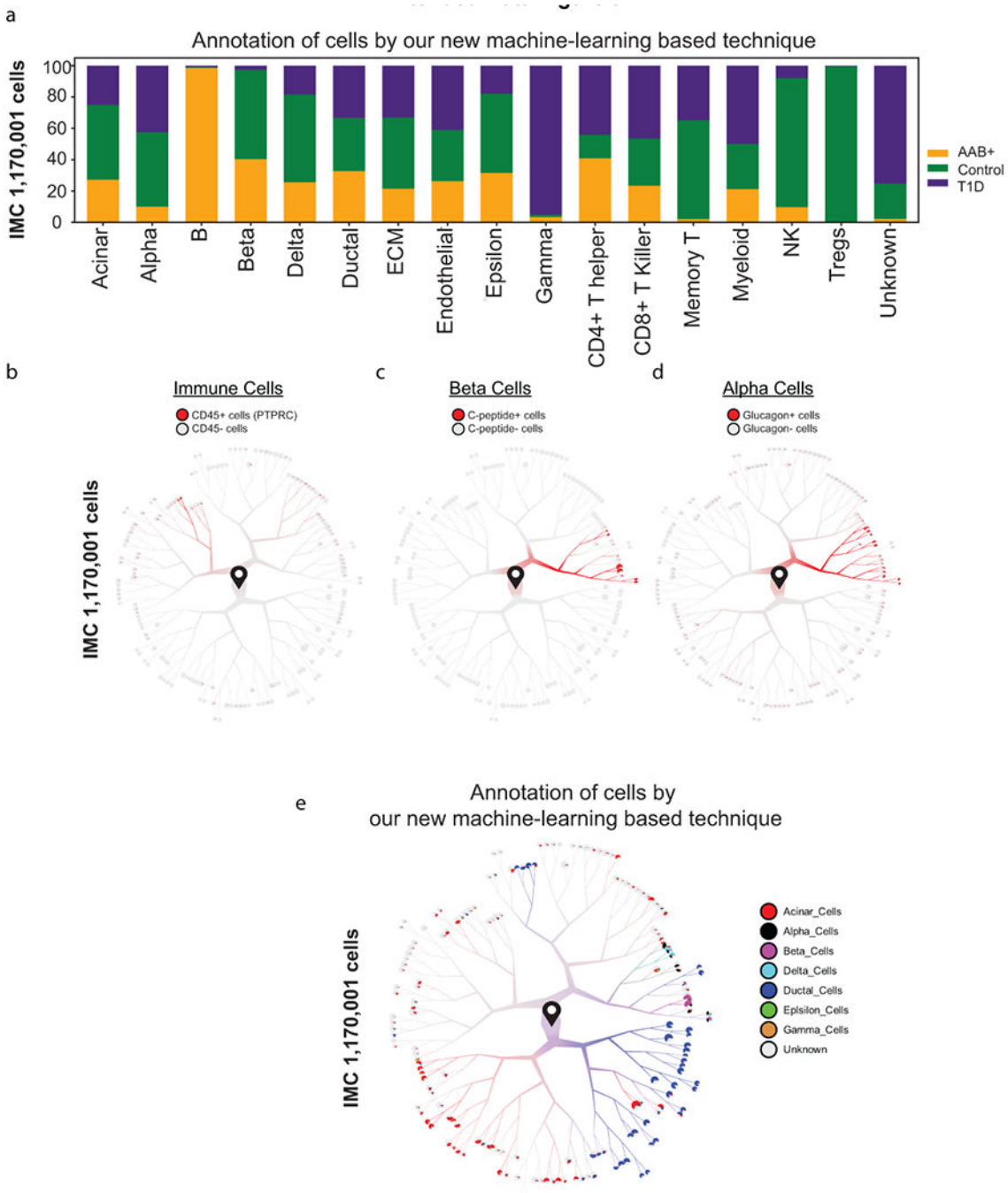
Extended Data Fig.8: CyTOF validation of canonical cell types

a) Bar graph displaying the proportion of cells for all major pancreatic cell types from each donor group where cell annotations were obtained by our new machine-learning based strategy using CyTOF measurements across 12 donors.

b) Dendrogram visualization of the immune cell cluster, CD45 positive (+) cells, as determined by the analysis of the flow cytometry by time-of-flight (CyTOF) data.

c) Dendrogram visualization of the beta cell cluster, C-peptide positive (+) cells, as determined by the analysis of the CyTOF data.

d) Dendrogram visualization of the alpha cell cluster, Glucagon positive (+) cells, as determined by the analysis of the CyTOF data.
e) Major cell types projected on TooManyCells tree based on our machine-learning based annotation using CyTOF data (n=6,945,575 cells).
f) Two-parameter CyTOF analysis of HLA-DR and cytokeratin protein expression in single cells from T1D donor #3 (HPAP023).
g) Two parameter CyTOF analysis of HLA-DR and cytokeratin protein expression in single cells from Control donor #3 (HPAP034), a donor with a very low percentage of HLA-DR+ ductal cells as determined by unbiased analysis of CyTOF data with TooManyCells.

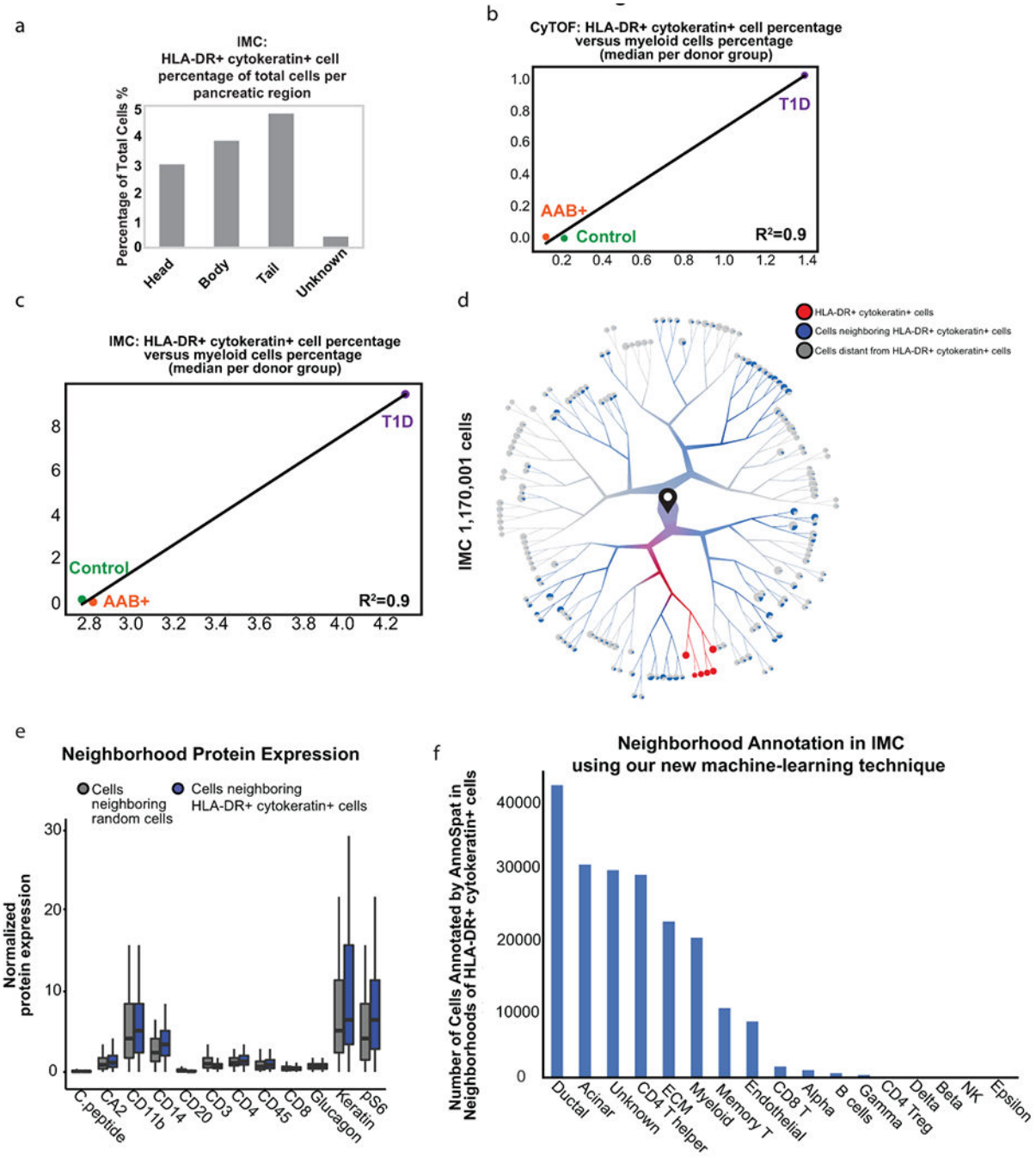


Extended Data Fig.9: IMC validation of HLA-DR+ ductal cells

a) Bar graph displaying the proportion of cells for all major pancreatic cell types from each donor group where cell annotations were obtained by our machine-learning-based strategy using IMC measurements. Further manual inspection of CD19 and FOXP3 staining used for annotating B and Tregs indicated low quality of these markers across tissue slides.

b) Dendrogram visualization of the immune cell cluster, CD45 positive (+) cells, as determined by the analysis of the imaging mass cytometry (IMC) data analysis.

c) Dendrogram visualization of the beta cell cluster, C-peptide positive (+) cells, as determined by the analysis of the IMC data analysis.
d) Dendrogram visualization of the alpha cell cluster, Glucagon positive (+) cells, as determined by the analysis of the IMC data analysis.
e) Major cell types projected on TooManyCells tree as they were annotated by our machine-learning based strategy using IMC data (n=1,170,001 cells).



Extended Data Fig.10: Cellular neighborhood analysis in IMC data demonstrates the enrichment of CD4+ T cells surrounding HLA-DR+ ductal cells

a) Bar plot displaying the proportion of HLA-DR+ cytokeratin+ cells from each pancreatic region determined by IMC.
b-c) HLA-DR+ cytokeratin+ cells versus percentage of myeloid cells. For each donor group, the median of percentage of each annotated immune subtype and the median HLA-DR+ ductal cell percentage of total cells across all individual donors per donor group was computed. Only myeloid cells demonstrated significant correlation with respect to the number of HLA-DR+ cytokeratin+ cells across donor groups.
d) Dendrogram visualization of the clusters of HLA-DR+ cytokeratin+ cells (red), cells neighboring HLA-DR+ cytokeratin+ (blue), and cells distant from HLA-DR+ cytokeratin+ cells (grey) as determined by leveraging the spatial architecture provided by IMC data.
e) Boxplots showing the normalized protein expression of different canonical markers in cells neighboring HLA-DR+ cytokeratin+ cells (blue) versus cells neighboring random cells (grey). The number of random cells evaluated was equal to the number of HLA-DR+ cytokeratin+ cells. Differential marker expression significance for neighbors in the IMC analysis was determined using permutation tests. For each marker, the distribution of that marker value for each of the designated n neighbors was compared against 100 distributions derived from n random cells across the entire IMC tree. * indicates p-value < 0.01. Total number of cells in both blue and gray groups is 195,633. Box-and-whisker plots (centre, median; box limits, upper (75th) and lower (25th) percentiles; whiskers, 1.5 × interquartile range; points, outliers).
f) $CD4^+$ T cells are the number one immune subtypes enriched at the neighborhood of HLA-DR+ cytokeratin+ cells. Annotation of neighbors of HLA-DR+ cytokeratin+ cells was performed our machine-learning based strategy.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank our colleagues for helpful discussions, particularly: Aditi Chandra, Eline Luning Prak, Ben Stanger, Michael Silverman, Gregory Beatty, Kenneth Zaret, Mitchell Lazar, Robert Vonderheide, Andy Minn, and E. John Wherry. We thank Andrei Georgescu for confocal microscopy and the University of Pennsylvania Diabetes Research Center (DRC) for the use of the Functional Genomics Core (P30-DK19525). This work was supported by NIH grants UC4 DK112217 and U01DK11221702A1 (to A.N., K.K., M.B., J.M., M.F., R.B.F., and G.V.), R01CA230800 and Susan G. Komen CCR185472448 (to R.B.F.), and R01HL145754, U01DK127768, U01DA052715, the Burroughs Wellcome Fund, the Chan Zuckerberg Initiative, W. W. Smith Charitable Trust, the Penn Epigenetics Institute and the Sloan Foundation awards to G.V.

Data Availability

The GEO accession number associated with this paper is GSE148073. Additional data are publicly available at https://hpap.pmacs.upenn.edu/. Furthermore, a user-friendly web portal for exploration of the scRNA-seq data is available at https://cellxgene.cziscience.com/e/37b21763-7f0f-41ae-9001-60bad6e2841d.cxg/

HPAP Consortium Authors:

Maria Fasolino[1,2,3,4,6,7*], Gregory W. Schwartz[2,3,5,6,7*], Abhijeet R. Patil[1,2,3,4,6,7], Aanchal Mongia[2,3,5,6,7], Maria L. Golson[1,4,8], Yue J. Wang[1,4], Ashleigh Morgan[1,4], Chengyang Liu[4,9], Jonathan Schug[1], Jinping Liu[1,4], Minghui Wu[1,4], Daniel Traum[1,4], Ayano Kondo[1,4], Catherine L. May[1,4], Naomi Goldman[1,2,3,4,6,7], Wenliang Wang[1,2,3,4,6,7], Michael Feldman[5,7], Jason H. Moore[1,7], Alberto S. Japp[2,10], Michael R. Betts[2,10], Robert B. Faryabi[2,3,5,6,7#], Ali Naji[2,4,9#], Klaus H. Kaestner[1,3,4#], Golnaz Vahedi[1,2,3,4,6,7#]

Affiliations:

[1]Department of Genetics, [2]Institute for Immunology, [3]Epigenetics Institute, [4]Institute for Diabetes, Obesity and Metabolism, [5]Department of Pathology and Laboratory Medicine, [6]Abramson Family Cancer Research Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA; [7]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA; [8]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Johns Hopkins University, Baltimore, MD, 21205, USA, [9]Department of Surgery, [10]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.

References

1. Powers AC Type 1 diabetes mellitus: much progress, many opportunities. J Clin Invest 131, doi:10.1172/JCI142242 (2021).
2. Michels AW, Redondo MJ & Atkinson MA The pathogenesis, natural history, and treatment of type 1 diabetes: time (thankfully) does not stand still. Lancet Diabetes Endocrinol, doi:10.1016/S2213-8587(21)00344-2 (2021).
3. Boldison J & Wong FS Immune and Pancreatic beta Cell Interactions in Type 1 Diabetes. Trends Endocrinol Metab 27, 856–867, doi:10.1016/j.tem.2016.08.007 (2016). [PubMed: 27659143]
4. Unanue ER & Wan X The Immunoreactive Platform of the Pancreatic Islets Influences the Development of Autoreactivity. Diabetes 68, 1544–1551, doi:10.2337/dbi18-0048 (2019). [PubMed: 31331989]
5. Barrett JC et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 41, 703–707, doi:10.1038/ng.381 (2009). [PubMed: 19430480]
6. Campbell-Thompson M et al. Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. Diabetes Metab Res Rev 28, 608–617, doi :10.1002/dmrr.2316 (2012). [PubMed: 22585677]
7. Kaestner KH, Powers AC, Naji A, Consortium H & Atkinson MA NIH Initiative to Improve Understanding of the Pancreas, Islet, and Autoimmunity in Type 1 Diabetes: The Human Pancreas Analysis Program (HPAP). Diabetes 68, 1394–1402, doi:10.2337/db19-0058 (2019). [PubMed: 31127054]
8. Pliner HA, Shendure J & Trapnell C Supervised classification enables rapid annotation of cell atlases. Nat Methods 16, 983–986, doi:10.1038/s41592-019-0535-3 (2019). [PubMed: 31501545]
9. Schwartz GW et al. TooManyCells identifies and visualizes relationships of single-cell clades. Nat Methods, doi:10.1038/s41592-020-0748-5 (2020).
10. Tosti L et al. Single-Nucleus and In Situ RNA-Sequencing Reveal Cell Topographies in the Human Pancreas. Gastroenterology 160, 1330–1344 e1311, doi:10.1053/j.gastro.2020.11.010 (2021). [PubMed: 33212097]

11. Campbell-Thompson M, Rodriguez-Calvo T & Battaglia M Abnormalities of the Exocrine Pancreas in Type 1 Diabetes. Curr Diab Rep 15, 79, doi:10.1007/s11892-015-0653-y (2015). [PubMed: 26318606]
12. Chiou J et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature 594, 398–402, doi:10.1038/s41586-021-03552-w (2021). [PubMed: 34012112]
13. Tersey SA et al. Islet beta-cell endoplasmic reticulum stress precedes the onset of type 1 diabetes in the nonobese diabetic mouse model. Diabetes 61, 818–827, doi:10.2337/db11-1293 (2012). [PubMed: 22442300]
14. Marhfour I et al. Expression of endoplasmic reticulum stress markers in the islets of patients with type 1 diabetes. Diabetologia 55, 2417–2420, doi:10.1007/s00125-012-2604-3 (2012). [PubMed: 22699564]
15. Engin F et al. Restoration of the unfolded protein response in pancreatic beta cells protects mice against type 1 diabetes. Sci Transl Med 5, 211ra156, doi:10.1126/scitranslmed.3006534 (2013).
16. Nakayasu ES et al. Comprehensive Proteomics Analysis of Stressed Human Islets Identifies GDF15 as a Target for Type 1 Diabetes Intervention. Cell Metab 31, 363–374 e366, doi:10.1016/j.cmet.2019.12.005 (2020). [PubMed: 31928885]
17. Unsicker K, Spittau B & Krieglstein K The multiple facets of the TGF-beta family cytokine growth/differentiation factor-15/macrophage inhibitory cytokine-1. Cytokine Growth Factor Rev 24, 373–384, doi:10.1016/j.cytogfr.2013.05.003 (2013). [PubMed: 23787157]
18. Carstensen M et al. Macrophage inhibitory cytokine-1 is increased in individuals before type 2 diabetes diagnosis but is not an independent predictor of type 2 diabetes: the Whitehall II study. Eur J Endocrinol 162, 913–917, doi:10.1530/EJE-09-1066 (2010). [PubMed: 20167682]
19. Avrahami D et al. Single-cell transcriptomics of human islet ontogeny defines the molecular basis of beta-cell dedifferentiation in T2D. Mol Metab 42, 101057, doi:10.1016/j.molmet.2020.101057 (2020). [PubMed: 32739450]
20. Wang YJ et al. Single-Cell Transcriptomics of the Human Endocrine Pancreas. Diabetes 65, 3028–3038, doi:10.2337/db16-0405 (2016). [PubMed: 27364731]
21. Teo AKK et al. Single-cell analyses of human islet cells reveal de-differentiation signatures. Cell Death Discov 4, 14, doi:10.1038/s41420-017-0014-5 (2018).
22. Chiang MK & Melton DA Single-cell transcript analysis of pancreas development. Dev Cell 4, 383–393, doi:10.1016/s1534-5807(03)00035-2 (2003). [PubMed: 12636919]
23. Riedel MJ et al. Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas. Diabetologia 55, 372–381, doi:10.1007/s00125-011-2344-9 (2012). [PubMed: 22038519]
24. Heng TS, Painter MW & Immunological Genome Project, C. The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol 9, 1091–1094, doi:10.1038/ni1008-1091 (2008). [PubMed: 18800157]
25. Vahedi G et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. Nature 520, 558–562, doi:10.1038/nature14154 (2015). [PubMed: 25686607]
26. Timshel PN, Thompson JJ & Pers TH Genetic mapping of etiologic brain cell types for obesity. Elife 9, doi:10.7554/eLife.55851 (2020).
27. Ludwig MQ et al. A genetic map of the mouse dorsal vagal complex and its role in obesity. Nat Metab 3, 530–545, doi:10.1038/s42255-021-00363-1 (2021). [PubMed: 33767443]
28. Loh PR, Kichaev G, Gazal S, Schoech AP & Price AL Mixed-model association for biobank-scale datasets. Nat Genet 50, 906–908, doi:10.1038/s41588-018-0144-6 (2018). [PubMed: 29892013]
29. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209, doi:10.1038/s41586-018-0579-z (2018). [PubMed: 30305743]
30. Rai V et al. Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. Mol Metab 32, 109–121, doi:10.1016/j.molmet.2019.12.006 (2020). [PubMed: 32029221]
31. Baron M et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst 3, 346–360 e344, doi:10.1016/j.cels.2016.08.011 (2016). [PubMed: 27667365]

32. Butler A, Hoffman P, Smibert P, Papalexi E & Satija R Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411–420, doi:10.1038/nbt.4096 (2018). [PubMed: 29608179]
33. Muraro MJ et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst 3, 385–394 e383, doi:10.1016/j.cels.2016.09.002 (2016). [PubMed: 27693023]
34. Segerstolpe A et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab 24, 593–607, doi:10.1016/j.cmet.2016.08.020 (2016). [PubMed: 27667667]
35. Wang YJ et al. Multiplexed In Situ Imaging Mass Cytometry Analysis of the Human Endocrine Pancreas and Immune System in Type 1 Diabetes. Cell Metab 29, 769–783 e764, doi:10.1016/j.cmet.2019.01.003 (2019). [PubMed: 30713110]
36. Zimmerman KD, Espeland MA & Langefeld CD A practical solution to pseudoreplication bias in single-cell studies. Nat Commun 12, 738, doi:10.1038/s41467-021-21038-1 (2021). [PubMed: 33531494]
37. Jailwala P et al. Apoptosis of CD4+ CD25(high) T cells in type 1 diabetes may be partially mediated by IL-2 deprivation. PLoS One 4, e6527, doi: 10.1371/journal.pone.0006527 (2009). [PubMed: 19654878]
38. Slominski B, Skrzypkowska M, Ryba-Stanislawowska M, Mysliwiec M & Trzonkowski P Associations of TP53 codon 72 polymorphism with complications and comorbidities in patients with type 1 diabetes. J Mol Med (Berl) 99, 675–683, doi: 10.1007/s00109-020-02035-1 (2021). [PubMed: 33495869]
39. Christen U et al. A dual role for TNF-alpha in type 1 diabetes: islet-specific expression abrogates the ongoing autoimmune process when induced late but not early during pathogenesis. J Immunol 166, 7023–7032, doi:10.4049/jimmunol.166.12.7023 (2001). [PubMed: 11390446]
40. Lombardi A, Tsomos E, Hammerstad SS & Tomer Y Interferon alpha: The key trigger of type 1 diabetes. J Autoimmun 94, 7–15, doi:10.1016/j.jaut.2018.08.003 (2018). [PubMed: 30115527]
41. Marro BS et al. Progression of type 1 diabetes from the prediabetic stage is controlled by interferon-alpha signaling. Proc Natl Acad Sci U S A 114, 3708–3713, doi:10.1073/pnas.1700878114 (2017). [PubMed: 28325871]
42. Osum KC et al. Interferon-gamma drives programmed death-ligand 1 expression on islet beta cells to limit T cell function during autoimmune diabetes. Sci Rep 8, 8295, doi :10.1038/s41598-018-26471-9 (2018). [PubMed: 29844327]
43. Ziegler AG et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA 309, 2473–2479, doi:10.1001/jama.2013.6285 (2013). [PubMed: 23780460]
44. Noble JA & Valdes AM Genetics of the HLA region in the prediction of type 1 diabetes. Curr Diab Rep 11, 533–542, doi:10.1007/s11892-011-0223-x (2011). [PubMed: 21912932]
45. Nejentsev S et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature 450, 887–892, doi:10.1038/nature06406 (2007). [PubMed: 18004301]
46. Wculek SK et al. Dendritic cells in cancer immunology and immunotherapy. Nat Rev Immunol 20, 7–24, doi:10.1038/s41577-019-0210-z (2020). [PubMed: 31467405]
47. Villani AC et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, doi:ARTN eaah4573 10.1126/science.aah4573 (2017).
48. Appleman LJ & Boussiotis VA T cell anergy and costimulation. Immunol Rev 192, 161–180, doi:10.1034/j.1600-065x.2003.00009.x (2003). [PubMed: 12670403]
49. Wang L et al. VISTA, a novel mouse Ig superfamily ligand that negatively regulates T cell responses. J Exp Med 208, 577–592, doi:10.1084/jem.20100619 (2011). [PubMed: 21383057]
50. Wang YJ et al. Single-Cell Mass Cytometry Analysis of the Human Endocrine Pancreas. Cell Metab 24, 616–626, doi:10.1016/j.cmet.2016.09.007 (2016). [PubMed: 27732837]
51. Rodriguez-Calvo T, Ekwall O, Amirian N, Zapardiel-Gonzalo J & von Herrath MG Increased immune cell infiltration of the exocrine pancreas: a possible contribution to the pathogenesis of type 1 diabetes. Diabetes 63, 3880–3890, doi:10.2337/db14-0549 (2014). [PubMed: 24947367]
52. Steimle V, Siegrist CA, Mottet A, Lisowska-Grospierre B & Mach B Regulation of MHC class II expression by interferon-gamma mediated by the transactivator gene CIITA. Science 265, 106–109, doi:10.1126/science.8016643 (1994). [PubMed: 8016643]

53. Moncada R et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. Nat Biotechnol 38, 333–342, doi:10.1038/s41587-019-0392-8 (2020). [PubMed: 31932730]
54. Wosen JE, Mukhopadhyay D, Macaubas C & Mellins ED Epithelial MHC Class II Expression and Its Role in Antigen Presentation in the Gastrointestinal and Respiratory Tracts. Front Immunol 9, 2144, doi:10.3389/fimmu.2018.02144 (2018). [PubMed: 30319613]
55. Kondo A et al. Highly Multiplexed Image Analysis of Intestinal Tissue Sections in Patients With Inflammatory Bowel Disease. Gastroenterology 161, 1940–1952, doi:10.1053/j.gastro.2021.08.055 (2021). [PubMed: 34529988]
56. Russell MA et al. HLA Class II Antigen Processing and Presentation Pathway Components Demonstrated by Transcriptome and Protein Analyses of Islet beta-Cells From Donors With Type 1 Diabetes. Diabetes 68, 988–1001, doi:10.2337/db18-0686 (2019). [PubMed: 30833470]
57. Squair JW et al. Confronting false discoveries in single-cell differential expression. Nat Commun 12, 5692, doi:10.1038/s41467-021-25960-2 (2021). [PubMed: 34584091]
58. Bonifacio E et al. Harmonization of glutamic acid decarboxylase and islet antigen-2 autoantibody assays for national institute of diabetes and digestive and kidney diseases consortia. J Clin Endocrinol Metab 95, 3360–3367, doi:10.1210/jc.2010-0293 (2010). [PubMed: 20444913]
59. Yu L et al. Early expression of antiinsulin autoantibodies of humans and the NOD mouse: evidence for early determination of subsequent diabetes. Proc Natl Acad Sci U S A 97, 1701–1706, doi:10.1073/pnas.040556697 (2000). [PubMed: 10677521]
60. Wenzlau JM et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proc Natl Acad Sci U S A 104, 17040–17045, doi:10.1073/pnas.0705894104 (2007). [PubMed: 17942684]
61. Hafemeister C & Satija R Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol 20, 296, doi:10.1186/s13059-019-1874-1 (2019). [PubMed: 31870423]
62. McGinnis CS, Murrow LM & Gartner ZJ DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst 8, 329–337 e324, doi:10.1016/j.cels.2019.03.003 (2019). [PubMed: 30954475]
63. Stuart T et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019). [PubMed: 31178118]
64. Zhou Y et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 10, 1523, doi:10.1038/s41467-019-09234-6 (2019). [PubMed: 30944313]
65. Crowell HL et al. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. Nat Commun 11, 6077, doi:10.1038/s41467-020-19894-4 (2020). [PubMed: 33257685]
66. Howson JM et al. Evidence of gene-gene interaction and age-at-diagnosis effects in type 1 diabetes. Diabetes 61, 3012–3017, doi:10.2337/db11-1694 (2012). [PubMed: 22891215]
67. Howson JM, Walker NM, Clayton D, Todd JA & Type 1 Diabetes Genetics, C. Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. Diabetes Obes Metab 11 Suppl 1, 31–45, doi:10.1111/j.1463-1326.2008.01001.x (2009). [PubMed: 19143813]

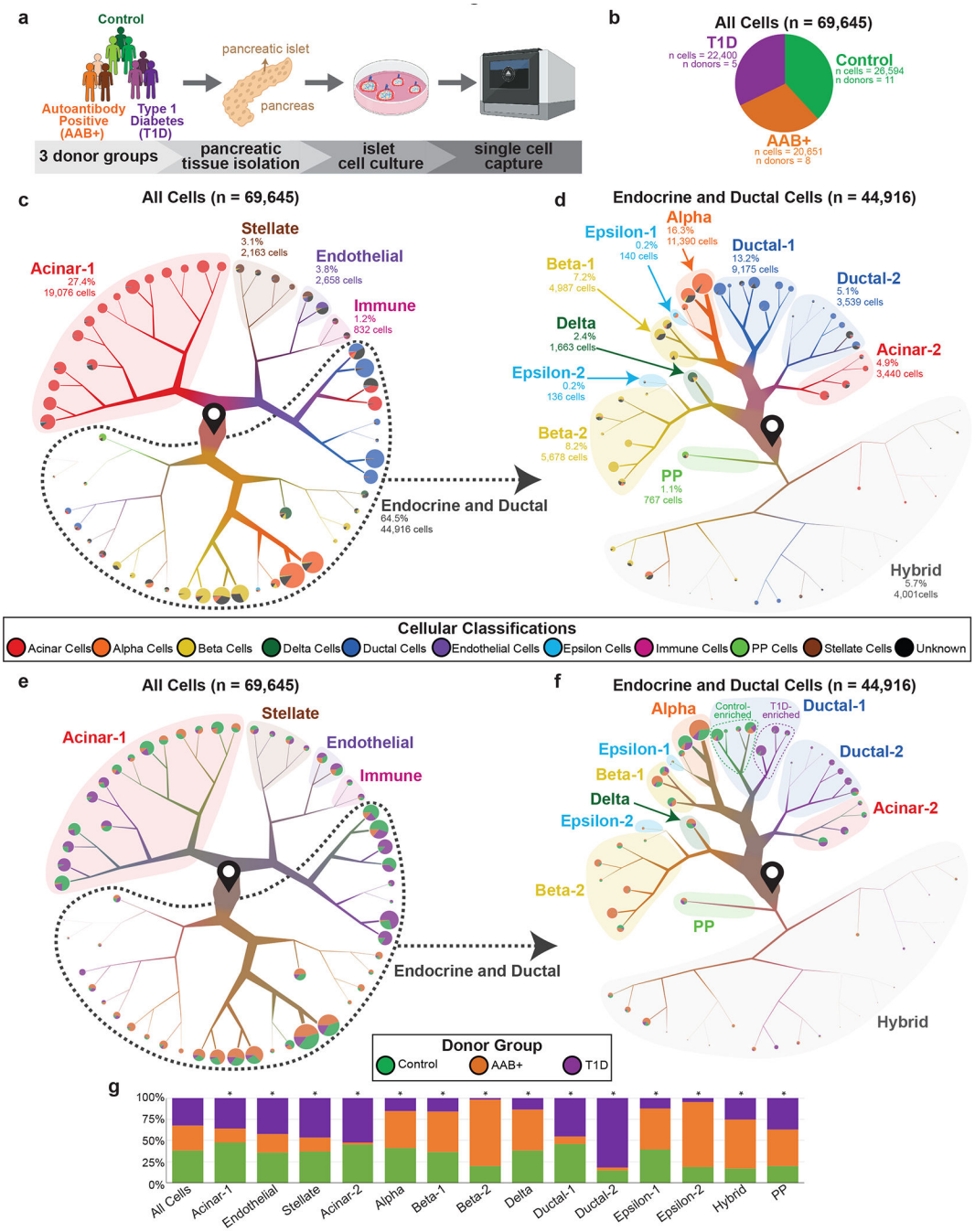


Figure 1: Discernment of human pancreatic cell types using single-cell RNA-seq

a) The transcriptome of single cells from pancreatic islets of 3 donor types (healthy Control donors, autoantibody positive (AAb+) donors, and donors with Type 1 diabetes (T1D)) was ascertained using the 10x Genomics platform.

b) Pie chart displaying the proportion of cells comprised by each donor group.

c) TooManyCells dendrogram visualization and clustering of all cells. Cells begin at the start pin symbol, and are then partitioned based on transcriptional similarities and differences. The color within the branches indicates the proportion of the cells that are classified by

the Garnett cellular classification tool (Table S17). Each bifurcation denotes significant transcriptional differences between the two cell groups. Pie charts at the end of the branches display the breakdown of Garnett cellular classification of cells within that terminal cluster. Highlighting or dotted lines surrounding particular clusters of cells with labels define cell types based on Garnett cellular classifications and canonical gene expression. Branch thickness and pie-chart size is proportional to cell number. Branch length is not indicative of any factor, but is merely a means by which to display cells within a defined space. Beta cells (*INS* high), alpha cells (*GCG* high), delta cells (*SST* high), PP cells (*PPY* high), epsilon cells (*GHRL* high), acinar cells (*CPA1* high), ductal cells (*KRT19* high), endothelial cells (*VWF* high), stellate cells (*RSG10* high), and immune cells (*PTPRC*, also known as CD45 or leukocyte common antigen, high). Percentages provided represent the percentage of total cells.

d) Dendrogram visualization and clustering of ductal and endocrine cells. Highlighting or dotted lines surrounding particular clusters of cells with labels define cell types based on Garnett cellular classifications and canonical gene expression.

(e-f) Group donor type projected across the dendrogram visualization and clustering of all cells from Figure 1C (e) or of endocrine and ductal cells from Figure 1D (f). Pie charts at the end of the branches display the breakdown of donor type within that terminal cluster.

(g) Bar graph displaying the proportion of cells from each donor group for all major pancreatic cell types. The p-values are calculated by the Chi-squared test

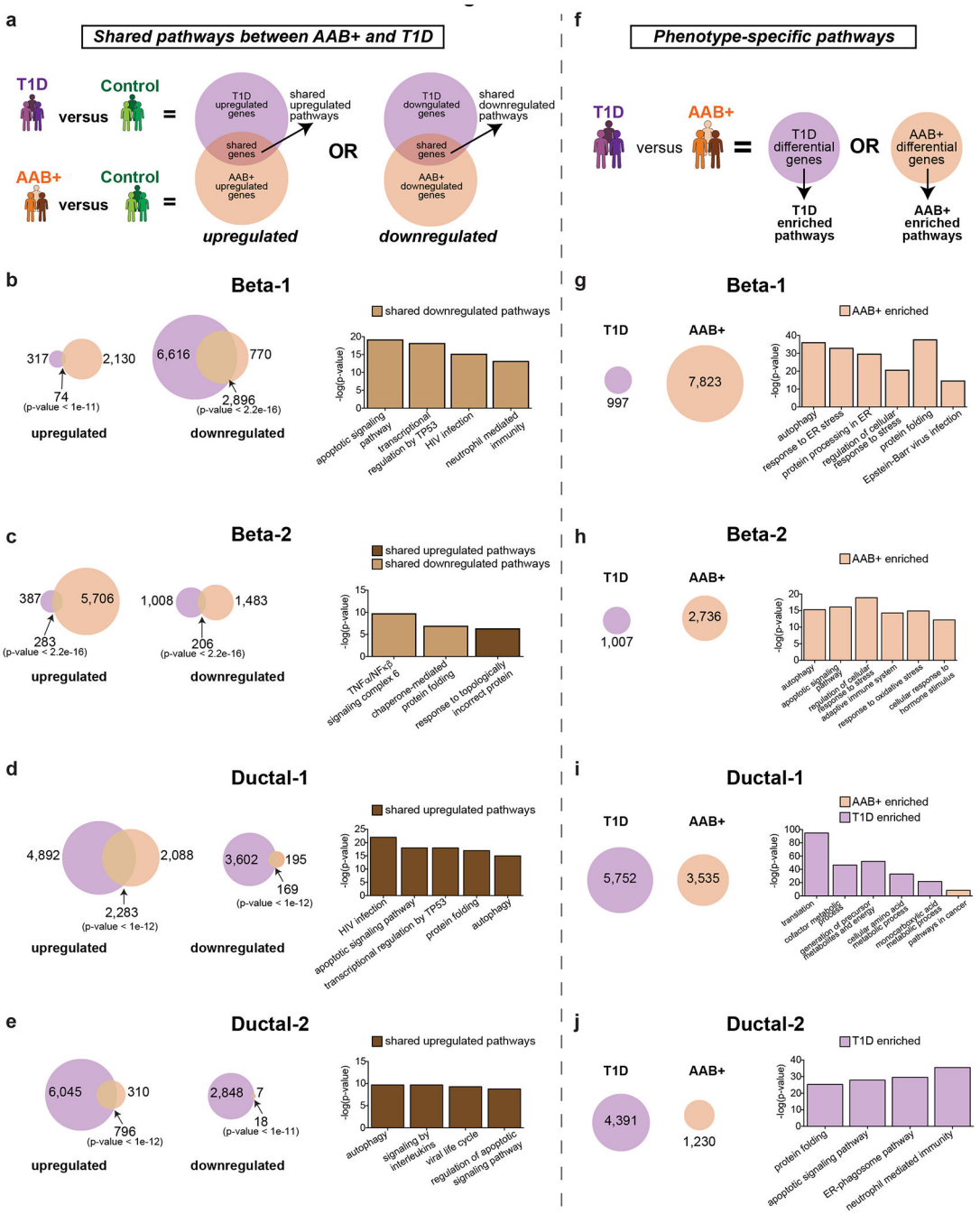


Figure 2: AAb+ and T1D donors have both common and distinct transcriptomic changes in endocrine and exocrine cell types

a) For each cell type, two pairwise differential comparisons were carried out: (1) T1D versus Control (referred to as 'T1D upregulated' (T1D/Control) or 'T1D downregulated' (Control/T1D)) and (2) AAB+ versus Control (referred to as 'AAB+ upregulated' (AAB+/Control) or 'AAB+ downregulated' (Control/AAB+)). T1D upregulated genes were then compared to AAB+ upregulated genes to find commonly upregulated genes, and subsequently commonly upregulated gene ontologies and pathways, across these two donor groups; this exact same approach was carried out for downregulated genes as well.

(b-e) (Left) For each cell type, Venn diagrams indicate the numbers of upregulated and downregulated genes, as well as overlapping genes, across the two donor states. (Right) Bar graph displaying notable gene ontologies that are shared across disease states for upregulated and downregulated genes. The p-values presented are the results of hypergeometric CDF tests (one-tailed test for overrepresentation).
(f) Transcriptional differences between cells from T1D and AAb+ donors were determined by directly comparing T1D to AAB+ cells to generate lists of differentially expressed genes that are enriched in T1D cells or AAB+ cells, and enriched gene ontology pathways were discovered from these differential gene lists.
(g-j) (Left) For each cell type, circles indicate the numbers of genes that are ‘T31D enriched’ or ‘AAB enriched’. (Right) Bar graph displaying notable gene ontologies that are enriched for each donor state.

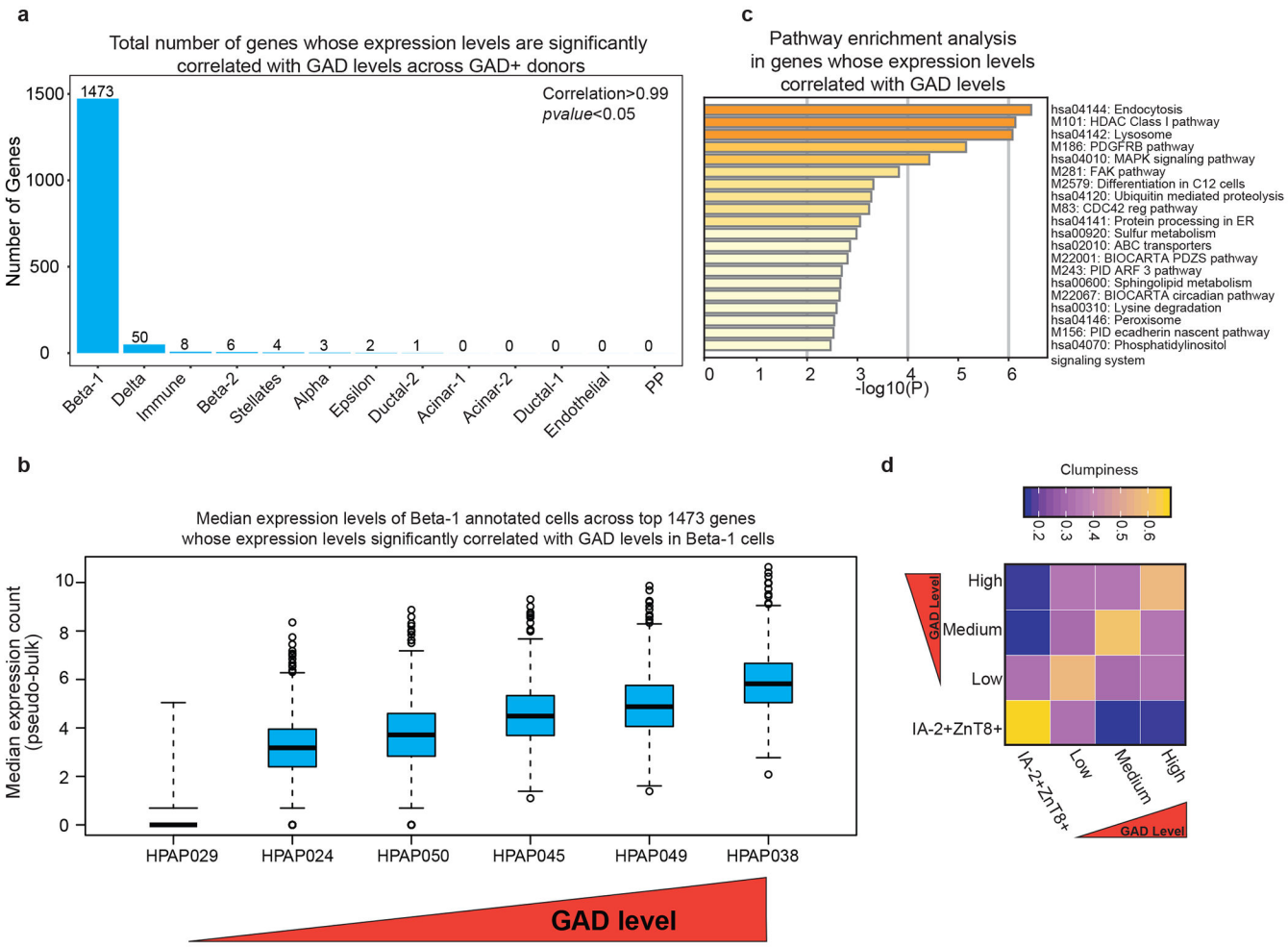


Figure 3: The gene signature of Beta-1 cells in GAD+ donors is correlated with donors' anti-GAD autoantibody titers.

a) Transcriptional outputs of Beta-1 cells positively correlate with the anti-GAD AAb titer in AAb+ donors. In every annotated cell type, we searched for genes whose expression level correlated with anti-GAD AAb levels in normoglycemic GAD+ donors (R^2>0.99 and p-value<0.05).

b) Plotting the average expression levels of cells from each GAD+ donor for the top 1,473 genes in Beta-1 cells with statistically significant correlation with the GAD titers corroborated our query. The total number of cells is 6,904. The plot shows a box-and-whisker plot of the given values. Lower 25th percentile (Q1), Interquartile range (IQR), Median (Q2), Upper 75th percentile (Q3). Minimum (Minimum value in the data, Q1-1.5 * IQR), Maximum (Maximum value in the data, Q3+1.5*IQR). The dots represents potential outliers.

c) A gene-ontology analysis in 1,473 genes related to Beta-1 cells using metaScape highlighted the relevance of endocytosis, protein processing in ER and MapK signaling pathway in Beta-1 cells.

d) Comparison of the cell clustering of the one normoglycemic AAb+ donor expressing two autoantibodies (IA-2 and ZnT8) with GAD+ donors using clumpiness revealed the

distinct transcriptional signature of the double autoantibody-expressing autoantibody donor and the single autoantibody-expressing GAD+ donors. Clumpiness is a measure for finding the level of aggregation between labels distributed among the leaves of a hierarchical tree and extensively measures the relationships between metadata. Here, each leaf of the dendrogram contains a collection of labels (different AAB donor group). The more the labels group together within the dendrogram, the higher the clumpiness value. This analysis also demonstrates the overall similarity of GAD+ donors, which modestly displayed GAD level-dependent cell co-segregation.

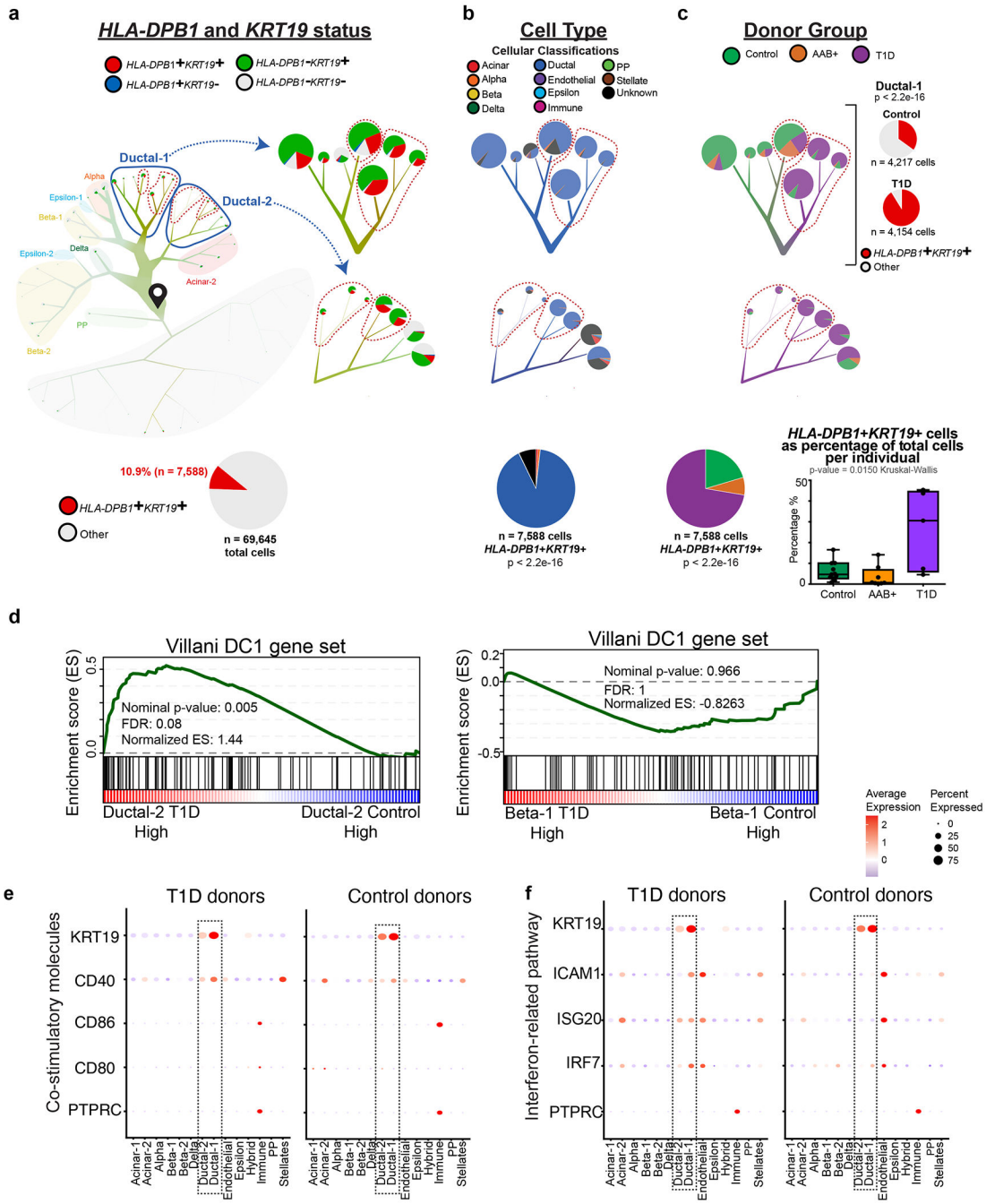


Figure 4: Single-cell RNA-seq profiling enables the identification of MHC Class II-expressing ductal cells with transcriptional similarities to dendritic cells in T1D

a) (Left Top) Dendrogram visualization of co-expression of *HLA-DPB1* and *KRT19* gene transcripts in individual cells by scRNA-seq across the ductal and endocrine dendrogram from Figure 1D. (Left Bottom) Pie chart demonstrating *HLA-DPB1+KRT19+* cells as percentage of total cells. (Right) Magnified view of the clusters of cells with high percentage (25% or greater) of *HLA-DPB1+KRT19+* cells with *HLA-DPB1* and *KRT19* status displayed across these clusters (outlined in red dashed lines) and neighboring clusters of

cells. Cells begin at the start pin symbol and from there are partitioned based on similarities and differences in gene expression.

b) (Top) Dendrogram visualization of cellular classification status across the magnified clusters of cells with high percentage (25% or greater) of *HLA-DPB1+KRT19+* cells (outlined in red dash lines) and neighboring clusters of cells. (Bottom) Pie chart displaying the relative proportion of cellular classification status of *HLA-DPB1+KRT19+* cells. The p-value presented is the result of the Chi-squared test.

c) (Top Left) Dendrogram visualization of donor group across the magnified clusters of cells with a high percentage (25% or greater) of *HLA-DPB1+KRT19+* cells (outlined in red) as well as neighboring clusters of cells. (Top Right) Pie chart displaying the relative proportion of *HLA-DPB1+KRT19+* cells in Control (top) or T1D (bottom) Ductal-1 cells. The p-value presented is the result of the Fisher exact test. (Bottom Left) Pie chart displaying the relative proportion of donor group of *HLA-DPB1+KRT19+* cells. The p-value presented is the result of the Chi-squared test. (Bottom Right) Box plots displaying the *HLA-DPB1+KRT19+* cell percentage of total cells per individual across donor groups. 24 total donors: 11 controls, 8 AAB+,and 5 T1D. A box-and-whisker plot is depicted with the box extending from the 25th to 75th percentiles, the line in the middle representing the median, whiskers extending from the minimum to the maximum, and all data points shown.

d) T1D ductal cells are transcriptionally similar to tolerogenic dendritic cells. Gene-set enrichment analysis was performed using gene signatures of DC subtypes, which were recently defined using scRNA-seq in human blood[47]. The DC1 dendritic cell gene signature was enriched in Ductal-2 cells but not Beta-1 cells of T1D donors.

e) The co-stimulatory proteins CD80 or CD86 are not expressed in T1D ductal cells.

f) Ductal cells of T1D donors express interferon-associated genes including *ISG20*, *ICAM1*, and *IRF7* compared with those of control donors.

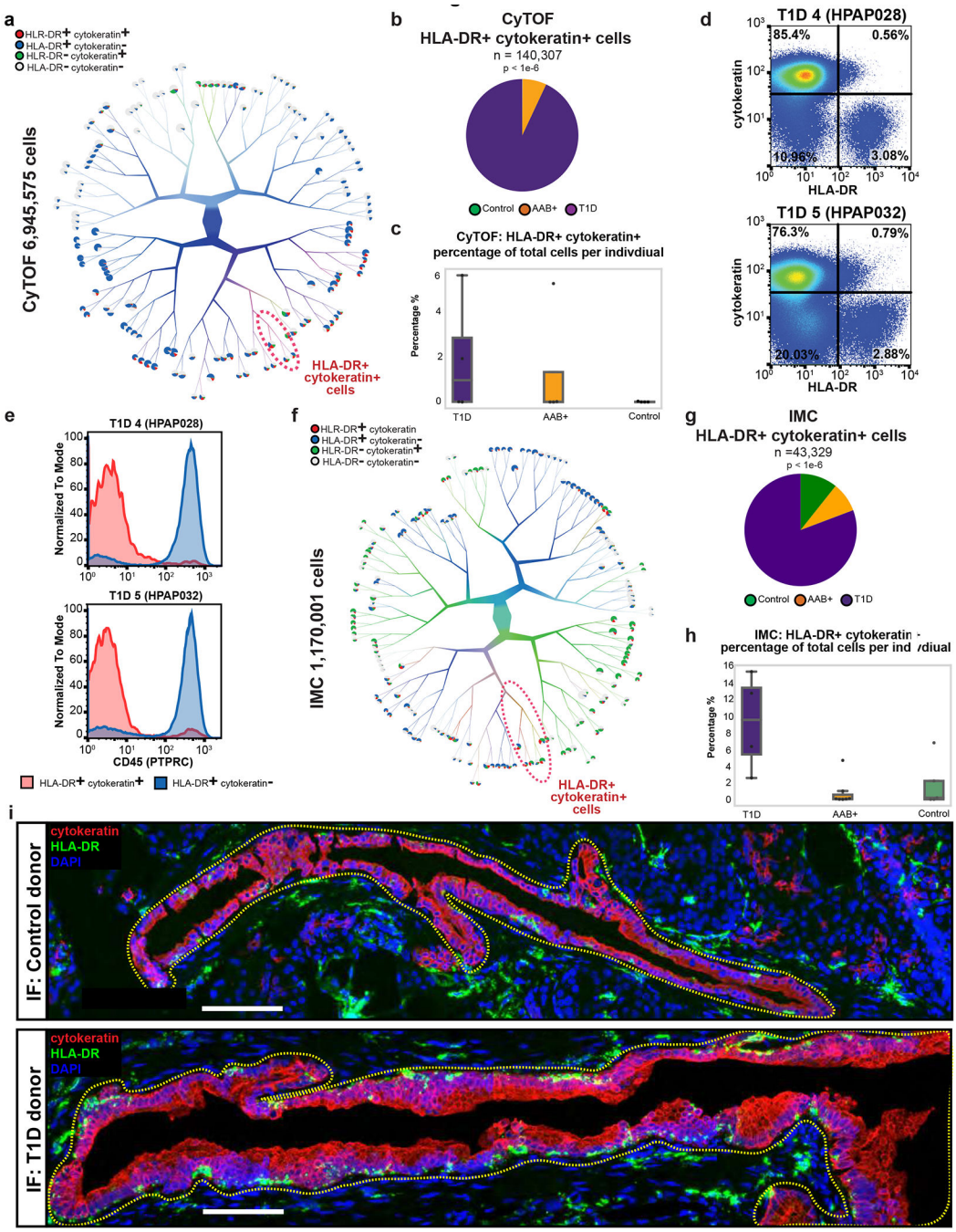


Figure 5: Three single-cell resolution protein-based approaches corroborate the existence of MHC Class II-expressing ductal cells in T1D

a) Dendrogram visualization of co-expression of HLA-DR and cytokeratin protein coexpression in single cells analyzed with flow cytometry by time-of-flight (CyTOF). b) Pie chart displaying HLA-DR+ cytokeratin+ cells and the relative proportions of each donor group from the CyTOF data. The p-value was calculated by the Chi-squared test. c)Box plots displaying HLA-DR+ cytokeratin+ cell percentage of total cells per individual across donor groups derived from the CyTOF data (p-value=0.00507). Number of donors: AAB+ : 4; Control : 4; T1D : 4. Figure depicts box-and-whisker plot showing the quartiles,

minimum non-outlier calculated by (Q1 – 1.5*IQR), 25th percentile/lower quartile Q1, 50th percentile/median Q2, 75th percentile/upper quartile Q3, maximum non-outlier calculated by (Q3 + 1.5*IQR) of the variable (hybrid percentage of total cells per individual) while the whiskers extend to show the rest of the distribution, except for points that are determined to be “outliers” (dots outside whiskers) using a method that is a function of the inter-quartile range.

d) Two-parameter CyTOF analysis of HLA-DR and cytokeratin protein expression in single cells from T1D donor #4 (HPAP028) and T1D donor #5 (HPAP032).

e) CD45 (PTPRC) expression levels in HLA-DR+ cytokeratin+ and HLA-DR+ cytokeratin- single cells.

f) Dendrogram visualization of co-expression of HLA-DR and cytokeratin proteins in single cells analyzed by imaging mass cytometry (IMC).

g) Pie chart displaying HLA-DR+ cytokeratin+ cells and the relative proportions of each donor group from the IMC data. The p-value was calculated by the Chi-squared test. The p-value shows 0.000 by both Chi-square function from scipy.stats (python) for the observed frequency array [34983,3711,4635] : [T1D,AAB+,Control], and cannot provide exact p-value.

h) Box plots displaying HLA-DR+ cytokeratin+ cells as a percentage of total cells per individual across donor groups from the IMC data (p-value=1e-16). Number of donors: AAB+ : 7; Control : 5; T1D : 4. p-value is obtained from a one-way ANOVA test.

i) Representative confocal microscopy image from the pancreas of a T1D donor (top) and Control donor (bottom) displaying HLA-DR+ cytokeratin+ labeled by immunofluorescence (IF). Control (n=3) and T1D (n=2).

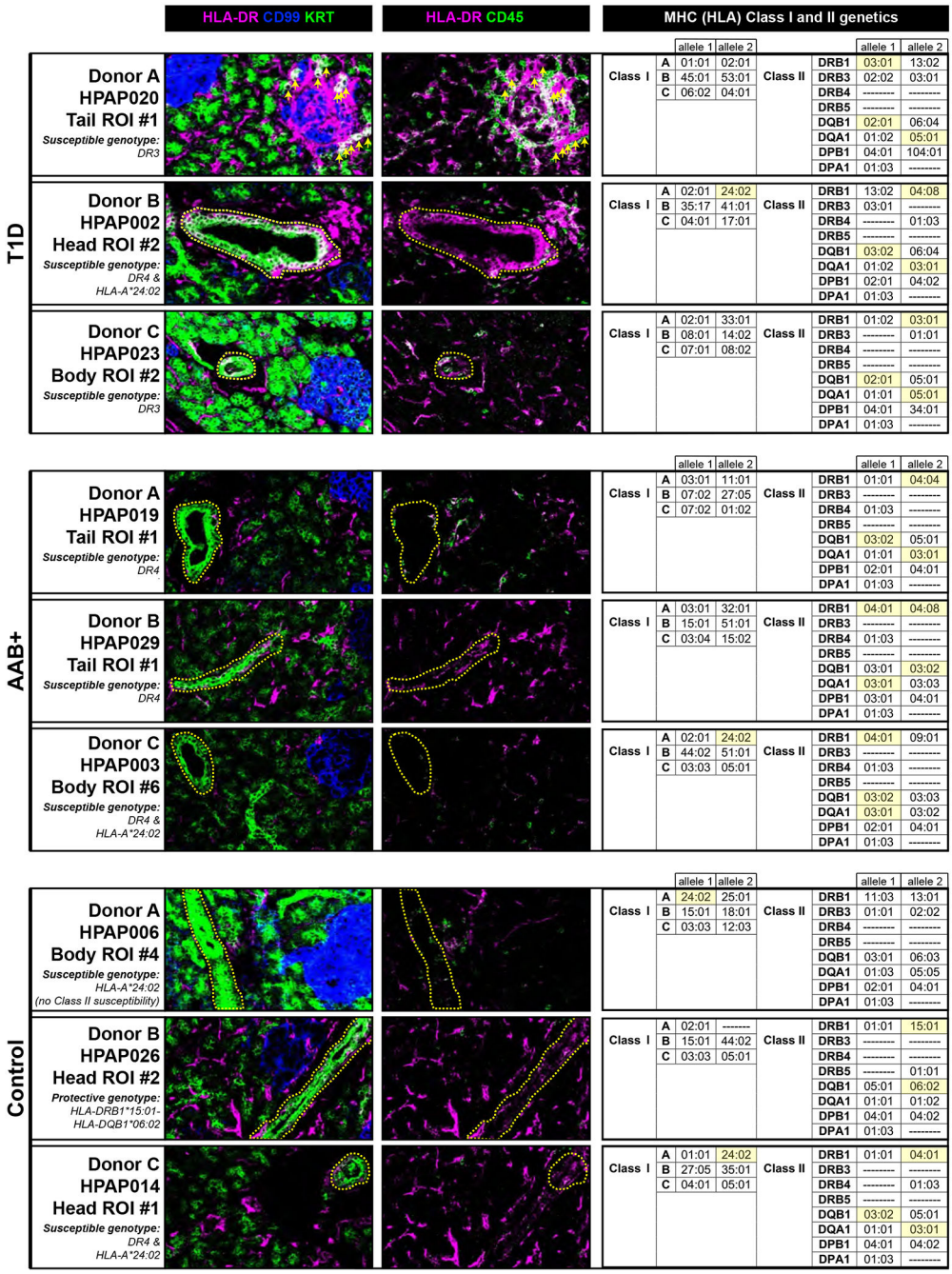


Figure 6: Representative examples of IMC measurement corroborates that MHC Class II positive ductal cells are present in pancreatic tissues.

(Left) Imaging mass cytometry (IMC) in a region of interest (ROI) in pancreatic tissue from three representative individual donors for each donor group type (T1D, AAB+, and Control). HLA-DR is a general marker of MHC Class II (HLA-DR) expression, CD99 is a general islet marker, KRT (pan-keratin) is a ductal cell marker, and CD45 (PTPRC) is a general immune cell marker. Notably, HLA-DR+ ductal cells were primarily located in large ductal structures (outlined in yellow). The images presented here are publicly available at https://www.pancreatlas.org/datasets/508.

(Right) HLA typing performed by next-generation sequencing. Comprehensive clinical information about each donor is provided in PANC-DB: https://hpap.pmacs.upenn.edu/. Highlighted in yellow are the particular HLA alleles contributing to the susceptible or protective genotypes, which are abbreviated for each donor on the left side of the figure as follows. The four susceptible genotypes assessed were (1) HLA-DRB1*03:01-HLA-DQA1*05:01-HLA-DQB1*02:01 (abbreviated as ‘DR3’, referring to the haplotype bearing the DRB1*03 allele); (2) HLA-DRB1*04:01/02/04/05/08-HLA-DQA1*03:01-HLA-DQB1*03:02/04 (or HLA-DQB1*02) (abbreviated as ‘DR4’, referring to the haplotype bearing the DRB1*04 allele); (3) HLA-A*24:02; and (4) HLA-B*39:06. The two protective genotypes assessed were (1) HLA-DRB1*15:01-HLA-DQB1*06:02 and (2) HLA-DRB1*07:01-HLA-DQB1*03:03[45,66,67]. Notably, HLA-DR+ ductal cells were found across all HLA genotypes, including both susceptible and protective genotypes.