ARTICLES
https://doi.org/10.1038/s42255-022-00531-x

nature metabolism

Check for updates

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

Maria Fasolino[1,2,3,4,5,6,21], Gregory W. Schwartz[2,3,5,6,7,21], Abhijeet R. Patil[1,2,3,4,5,6], 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[1,2,3,4,5,6], Wenliang Wang[1,2,3,4,5,6], Michael Feldman[6,7], Jason H. Moore[1,6], 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] ✉ and Golnaz Vahedi[1,2,3,4,5,6] ✉

Type 1 diabetes (T1D) is an autoimmune disease in which immune cells destroy insulin-producing beta cells. The aetiology 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 nondiabetic 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 pancreatic function.

T1D is an autoimmune disease that 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 cell and immune cell interactions including production of beta cell autoantibodies (AAbs) 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 (GWAS) have implicated multiple loci in T1D, with the major histocompatibility complex (MHC) class II genes as the dominant susceptibility determinants 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 because 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 perform a biopsy of the human pancreas of living donors and (2) the notable 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 leucocytes 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 individuals with T1D harbour beta cell AAbs in their bloodstream before clinical diagnosis, nPOD and HPAP also collect samples from donors with AAbs to islet proteins but without a medical history of T1D, in the 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 imaging mass cytometry (IMC) in pancreatic tissues of human organ donors collected by the 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

[1]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [2]Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [3]Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [4]Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [5]Abramson Family Cancer Research Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [6]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [7]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [8]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA. [9]Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [10]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [21]These authors contributed equally: Maria Fasolino, Gregory W. Schwartz. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: faryabi@pennmedicine.upenn.edu; ali.naji@pennmedicine.upenn.edu; kaestner@pennmedicine.upenn.edu; vahedi@pennmedicine.upenn.edu

unique dataset provide cellular and molecular insights into T1D pathogenesis and suggest pancreatic ductal cells may play a role in suppressing CD4+ T cells in pancreatic tissues.

Results

Single-cell RNA sequencing 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-sequencing (scRNA-seq) libraries from pancreatic islets of 24 human organ donors representing three categories: individuals with T1D ($n=5$), those with AAbs towards 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, Extended Data Fig. 1a,b and Supplementary Tables 1 and 2). The statistics related to reads per cell across donors demonstrated the high quality of these datasets (Supplementary Table 3). We filtered outlier cells, removed doublets and used the cell-type classifier Garnett[8] (Extended Data Figs. 1–3) to cluster 69,645 high-quality cells using TooManyCells[9] (Fig. 1b,c). The resultant classification was confirmed both by canonical gene marker expression for each cell type and by transferring cluster labels from a previous single-nucleus RNA-seq dataset consisting of pancreatic islets[10] to our datasets (Fig. 1c and Extended Data Fig. 4a–c).

Notably, clustering was clearly driven by cell type, and not by confounding factors such as AAb status, age, body mass index, phenotypic group or other factors (Extended Data Fig. 2d–g, Supplementary Fig. 1a–k and Supplementary Fig. 2a–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 (Supplementary Fig. 3a–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 re-analysed the endocrine and ductal cells to achieve a more granular clustering (Fig. 1d). After reclustering, the major cell types—alpha, beta, delta, epsilon, pancreatic polypeptide (PP), ductal and acinar—were easily discernible (Fig. 1d and Extended Data Fig. 4d). In instances where there were two transcriptionally distinct canonical cell types (that is, Beta-1 and Beta-2), differential gene expression analysis between populations provided further insights into the underlying molecular differences (Supplementary Tables 4–7). 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 hyperglycaemic 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 (that is, *PDX1* and *NKX6-1*), those in the Ductal-2 cluster expressed acinar TFs (that is, *PTF1A* and *GATA4*).

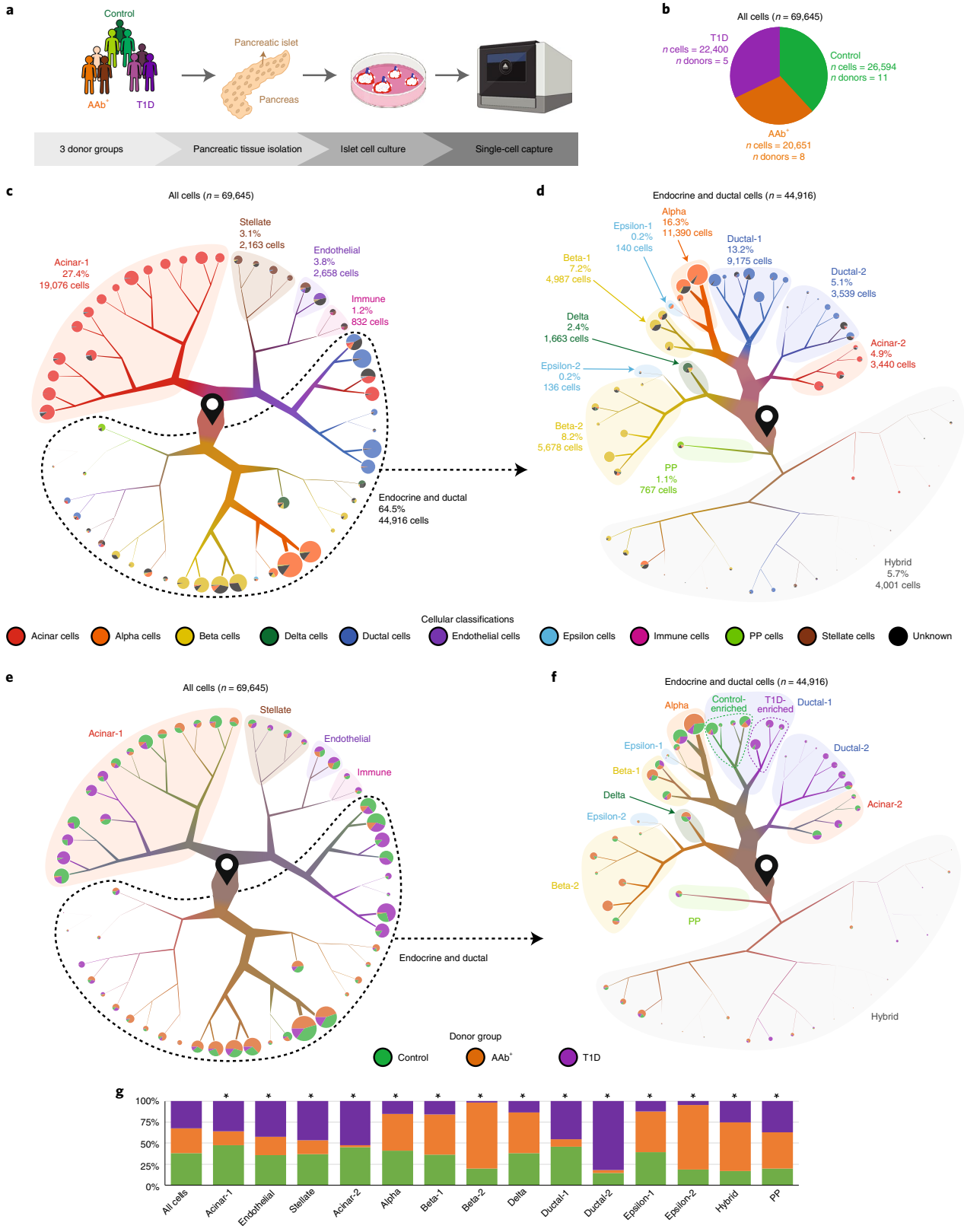
A substantial number of cells (4,001) were not included in these canonical cell-type clusters, but rather formed their own transcriptionally distinct group on the dendrogram. This cluster constituted 5.7% of all profiled cells, with a mixture of cellular classifications and expression of canonical gene markers. We labelled these cells as 'hybrid' cells (Fig. 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 making up this cluster ruled out the possibility of them being doublets (Extended Data Figs. 1d and 2a–e). To further validate the most closely related cell types to these non-canonical cells, we used 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. 4e,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.

Because 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; Supplementary Fig. 4a,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 (Supplementary Fig. 4b,c), suggesting that these cell types constitute the majority of the identified immune cells that are collected and cultured along with pancreatic islets.

Studies demonstrating that regulatory elements of immune cells harbour 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 (Supplementary Fig. 4d). 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

Fig. 1 | Discernment of human pancreatic cell types using single-cell RNA sequencing. a, The transcriptome of single cells from pancreatic islets of three donor types (healthy control donors, AAb+ donors and donors with T1D) was ascertained using the 10x Genomics platform. **b**, Pie chart displaying the proportion of cells constituting 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 colour within the branches indicates the proportion of the cells that are classified by the Garnett cellular classification tool (Supplementary Table 17). 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. Highlighted areas or dashed 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 leucocyte common antigen, high). Graphs represent the percentage of total cells. **d**, Dendrogram visualization and clustering of ductal and endocrine cells. Highlighted regions 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 **c** (**e**) or of endocrine and ductal cells from **d** (**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. *P* values were calculated by the chi-squared test. $^{*}P < 2.2 \times 10^{-16}$.

top cell type associated with T2D (Bonferroni significance threshold of $P_{\text{S-LDSC}} < 0.05$), 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 made up cells from the three donor groups with varying proportions (Fig. 1e–g, Extended Data Fig. 4g,h and Supplementary Figs. 4e and 5a,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 (Supplementary Fig. 5a,b). Conversely, both acinar and ductal cells constituted 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 (Fig. 1e,f and Supplementary Fig. 4a). Notably, Ductal-1 cells clearly separated into distinct T1D-enriched and control-enriched groups (Fig. 1f). Taken together, our data indicate that transcriptomic differences among 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 (Supplementary Tables 8–10) or (2) performing pseudo-bulk analysis for each donor (Supplementary Tables 11–13). Plotting the average expression levels of the top three 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 (Supplementary Fig. 6). As 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 (Fig. 2b–e, Supplementary Figs. 7–10 and Extended Data Fig. 5a–c). However, some pathways were found to be commonly dysregulated in multiple cell types across T1D and AAb$^+$ donors, including 'apoptotic signalling', 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 were 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%; $P < 2.2 \times 10^{-16}$) between the two donor groups (Fig. 2b and Supplementary Fig. 7a). Notable pathways that were frequently downregulated in Beta-1 cells of AAb$^+$ and T1D donors were immune/stress response and apoptosis-related pathways (Fig. 2b and Supplementary Fig. 7a). 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 signalling and apoptotic signalling via downregulation of the TP53 pathway (Fig. 2b and Supplementary Fig. 7a), 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 either are spared from destruction or use 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.

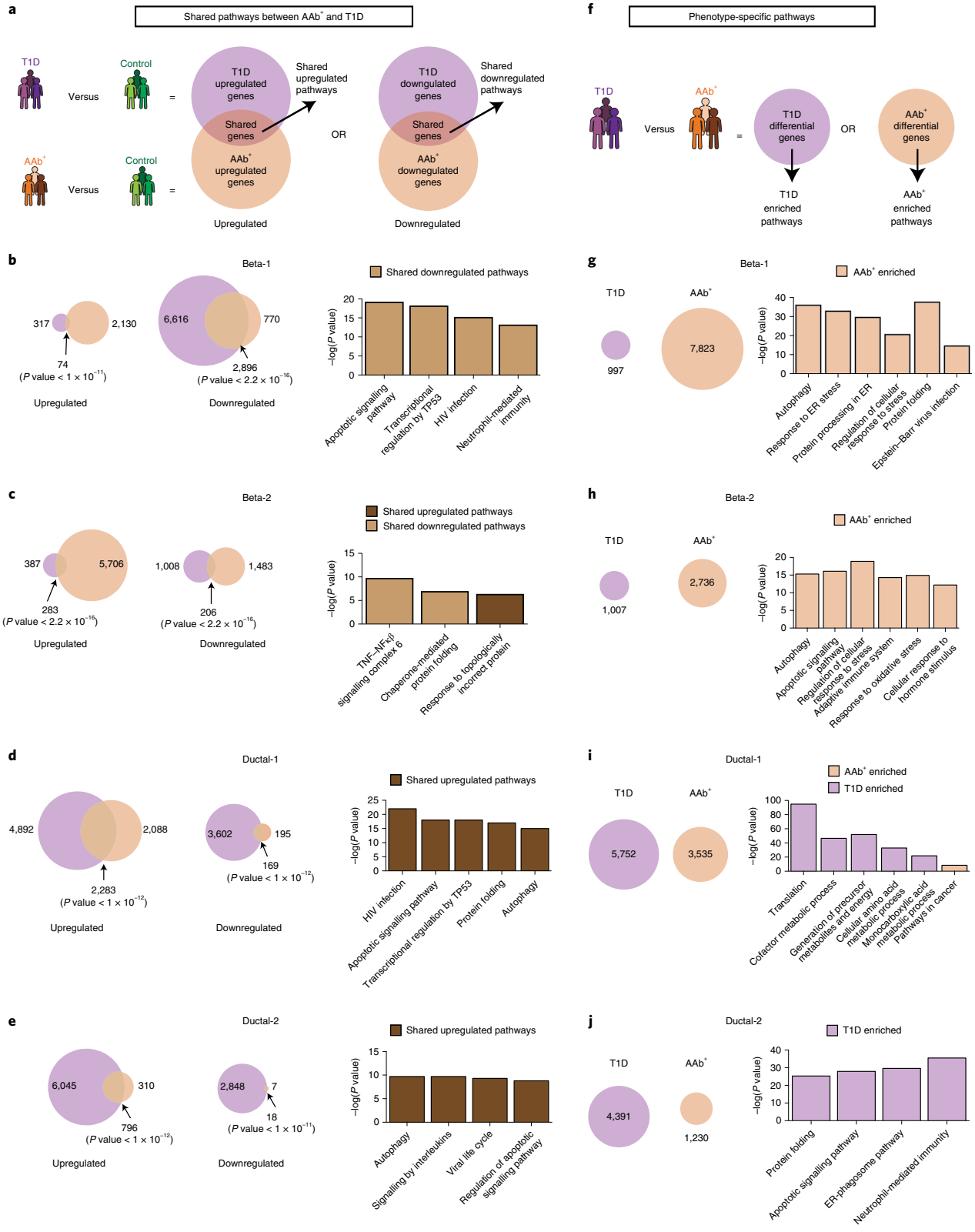
The Beta-2 cell population displayed a small proportion of genes (4%; 283 genes; $P < 2.2 \times 10^{-16}$) with elevated expression in both T1D and AAb$^+$ cells when compared to controls (Fig. 2c and Supplementary Fig. 7b). Additionally, an even smaller number of genes were downregulated in T1D and 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 'tumour necrosis factor/nuclear factor-kappa B signalling' 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 (Fig. 2c and Supplementary Fig. 7b). 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 were 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 (Fig. 2d and Supplementary Fig. 8a; 2,283 genes; 25%; P value $< 1 \times 10^{-12}$). Notable induced pathways upregulated in T1D and AAb$^+$ cells are associated with apoptosis, stress and immune response (Fig. 2d and Supplementary Fig. 8a). 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 (Fig. 2e and Supplementary Fig. 8b). Furthermore, in the T1D and AAb$^+$ Ductal-2 cell population, there was a modest but significant overlap between upregulated genes (Fig. 2e and Supplementary Fig. 8b; 11%; P value $< 1 \times 10^{-12}$). Nevertheless, various gene pathways were found to be significantly upregulated across both ductal populations (Fig. 2d,e and Supplementary Fig. 8a,b). Taken together, these findings suggest

Fig. 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, Venn diagrams indicate the numbers of upregulated and downregulated genes, as well as overlapping genes, across the two donor states, for each cell type. Right, bar graphs displaying notable gene ontologies that are shared across disease states for upregulated and downregulated genes. The *P* values presented are the results of hypergeometric cumulative distribution function 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, circles indicate the numbers of genes that are 'T1D enriched' or 'AAb enriched', for each cell type. Right, bar graphs displaying notable gene ontologies that are enriched for each donor state.

that although AAb+ donors maintain normoglycaemia, 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 Supplementary Tables 8–13). For both groups of beta cells, genes associated with autophagy, stress response and immune-related



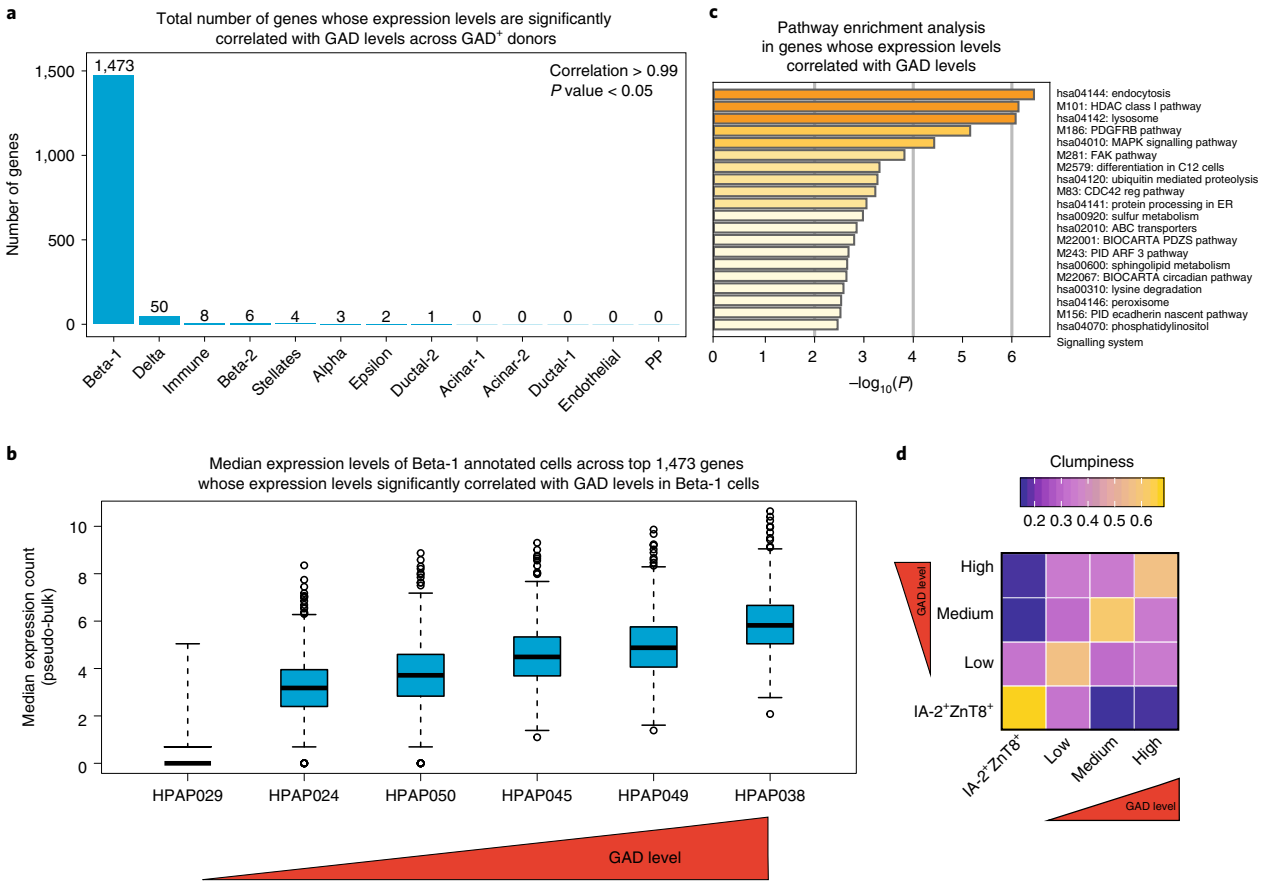


Fig. 3 | The gene signature of Beta-1 cells in GAD$^+$ donors is correlated with donors' anti-GAD AAb titres. a, Transcriptional outputs of Beta-1 cells positively correlate with the anti-GAD AAb titre in AAb$^+$ donors. In every annotated cell type, we searched for genes whose expression level correlated with anti-GAD AAb levels in normoglycaemic 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 titres corroborated our query. The total number of cells was 6,904. Box-and-whisker plots of the given values show the lower 25th percentile (Q1), interquartile range (IQR), the median (Q2), the upper 75th percentile (Q3) and the minimum (minimum value in the data, Q1 − 1.5 × IQR and maximum (maximum value in the data, Q3 + 1.5 × IQR) values. The dots represent 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 the ER and MapK signalling pathway in Beta-1 cells. **d**, Comparison of the cell clustering of the one normoglycaemic AAb$^+$ donor expressing two AAbs (IA-2 and ZnT8) with GAD$^+$ donors using clumpiness revealed the distinct transcriptional signature of the double AAb-expressing AAb donor and the single AAb-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.

pathways were activated in AAb$^+$ cells compared to T1D cells (Fig. 2g,h and Supplementary Fig. 7a,b). Although similar pathways were upregulated in AAb$^+$ Beta-1 and Beta-2 cells, apoptotic and adaptive immune system signalling 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 (Fig. 2i,j and Supplementary Fig. 8a,b). Notably, apoptotic, metabolic, protein folding and immune responses were activated in T1D ductal cells in comparison to AAb$^+$ ductal cells (Fig. 2i,j and Supplementary Fig. 8a,b). Remarkably, interferon-α and interferon-β 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-glutamic acid decarboxylase titre. 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 AAbs progressed to T1D after islet AAb seroconversion[43]. Among HPAP donors, only one donor with no history of T1D expressed two islet AAbs, while the other normoglycaemic AAb$^+$ donors were anti-glutamic acid decarboxylase (GAD) AAb positive. Considering that the longitudinal study of children also revealed that the risk of diabetes in children who had no islet AAbs was 0.4%

in contrast to 14% for children expressing a single islet AAb[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 titre among the GAD$^+$ donors. We devised a strategy to determine the number of genes whose expression levels significantly correlated with the GAD titre across GAD$^+$ donors, either positively or negatively. However, we detected only a positive correlation of statistical significance between gene expression levels and GAD titres. Strikingly, the top cell type with the largest number of genes (1,473) that were significantly correlated with the GAD titre in AAb$^+$ donors was Beta-1 cells (Fig. 3a and Supplementary Table 14). 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 endoplasmic reticulum (ER) and MAP kinase signalling in Beta-1 cells[13–18] (Fig. 3c). Additional comparison of the cellular clustering of the one AAb$^+$ donor expressing two AAbs (IA-2 and ZnT8; AAb$^+$ no. 5; HPAP043) with GAD$^+$ donors across the AAb$^+$-specific clustering (Supplementary Fig. 3b) or all donor-type clustering (Fig. 1c) revealed the distinct transcriptional signature of the double AAb-expressing AAb$^+$ donor in comparison to single AAb$^+$ GAD$^+$ donors (Fig. 3d and Supplementary Fig. 2l). This analysis also revealed an overall similarity of GAD$^+$ donors, which modestly displayed GAD-level-dependent cell co-segregation (Fig. 3d and Supplementary Fig. 2l). 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 normoglycaemic AAb$^+$ individuals.

Major histocompatibility complex 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. 6a–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. 4a,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. 4a,b and Extended Data Fig. 6e,f). Strikingly, cells from T1D donors disproportionately contributed to this population of MHC class II-expressing ductal cells (Fig. 4c; P value $< 2.2 \times 10^{-16}$). 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 $< 2.2 \times 10^{-16}$).

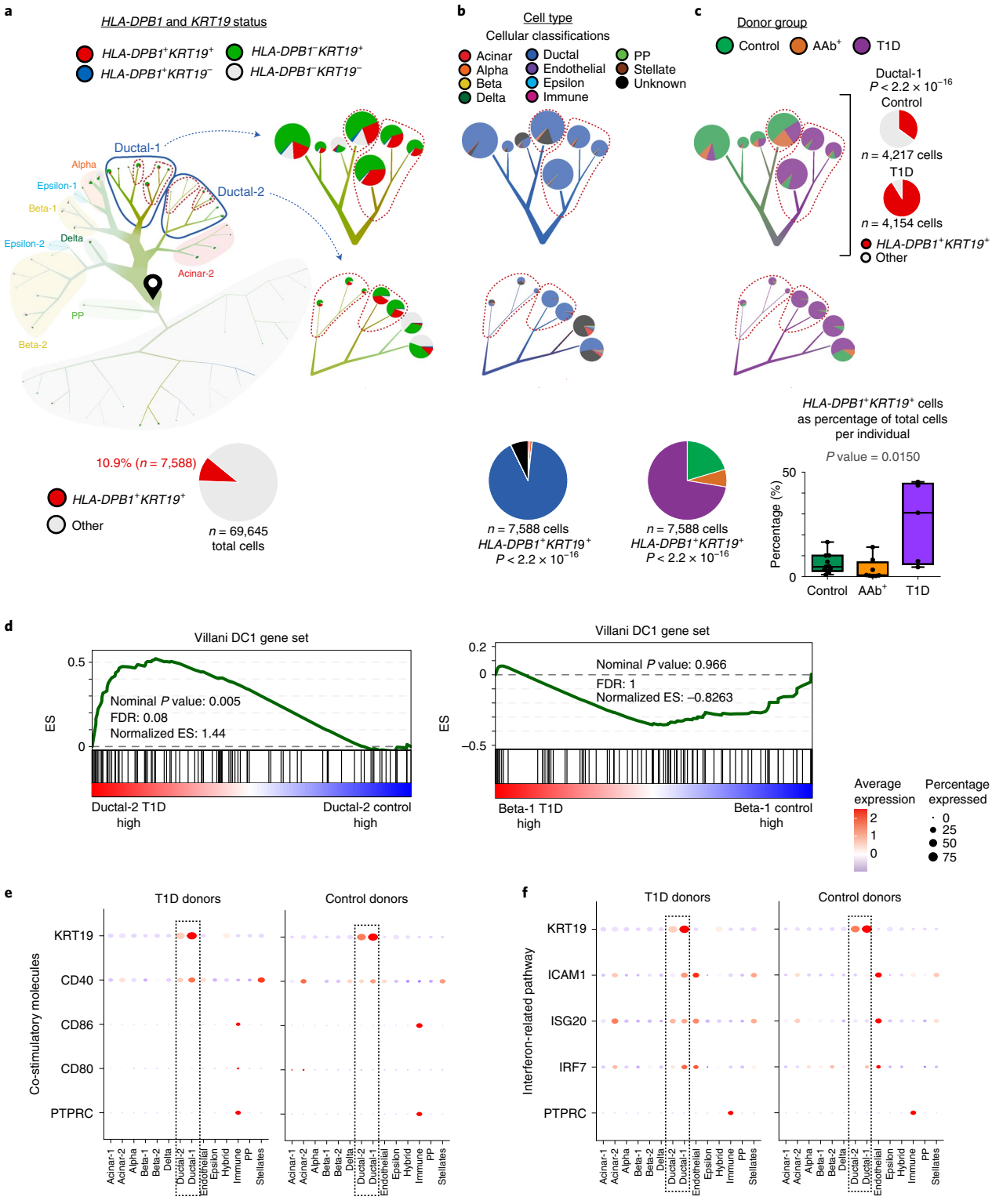
T1D ductal cells assume the transcriptional identity of dendritic cells. Dendritic cells (DCs) are among the major professional APCs expressing MHC class II proteins with the salient function to ingest antigens and present processed epitopes to T cells, thereby regulating adaptive immune responses by activating or suppressing T cells[46]. Considering that MHC class II proteins are required for antigen presentation in DCs, we next evaluated whether there are any other similarities between transcriptional profiles of T1D ductal cells and conventional DCs. Hence, we performed gene-set-enrichment analysis (GSEA) using gene signatures of DC 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 DC 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 DC subtype gene signatures in T1D ductal cells was not statistically significant (Extended Data Fig. 7a).

To activate T cells, DCs are required to express both MHC class II proteins and co-stimulatory proteins CD80 and CD86 (ref. [46]). In the absence of CD80 and CD86, antigen presentation by DCs 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 co-stimulatory signal in these DC-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 ductal cells compared with control ductal cells (Extended Data Fig. 7b). Moreover, the ductal cells in T1D expressed high levels of interferon genes including *ICAM1, ISG20* and *IRF7* (Fig. 4f and Extended Data Fig. 5d). Hence, our single-cell transcriptional profiling detected an enrichment of ductal cells with transcriptional

Fig. 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 coexpression of *HLA-DPB1* and *KRT19* transcripts in individual cells by scRNA-seq across the ductal and endocrine dendrogram from Fig. 1d. Left bottom, pie chart demonstrating *HLA-DPB1*$^+$*KRT19*$^+$ cells as a percentage of the total cells. Right, magnified view of the clusters of cells with a 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 neighbouring 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 a high percentage (25% or greater) of *HLA-DPB1*$^+$*KRT19*$^+$ cells (outlined in red dashed lines) and neighbouring clusters of cells. Bottom, pie chart displaying the relative proportion of cellular classification status of *HLA-DPB1*$^+$*KRT19*$^+$ cells. The P value was generated by the chi-squared test. **c**, Top left, dendrogram visualization of donor groups across the magnified clusters of cells with a high percentage (25% or greater) of *HLA-DPB1*$^+$*KRT19*$^+$ cells (outlined in red) as well as neighbouring clusters of cells. Top right, pie charts displaying the relative proportion of *HLA-DPB1*$^+$*KRT19*$^+$ cells in control (top) or T1D (bottom) Ductal-1 cells. The P value was generated by Fisher's exact test. Bottom left, pie charts displaying the relative proportions of *HLA-DPB1*$^+$*KRT19*$^+$ cells by donor group. The P value was generated by the chi-squared test. Bottom right, boxplots 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, with the line in the middle representing the median. Whiskers extend from the minimum to the maximum values, all data points are shown, and the P value reflects the result of the Kruskal–Wallis test. **d**, T1D ductal cells are transcriptionally similar to tolerogenic DCs. GSEA was performed using gene signatures of DC subtypes, which were recently defined using scRNA-seq in human blood[47]. The DC1 gene signature was enriched in Ductal-2 cells but not Beta-1 cells of T1D donors. **e**, The co-stimulatory proteins *CD80* and *CD86* were not expressed in T1D ductal cells. **f**, Ductal cells of T1D donors express interferon-associated genes including *ISG20*, *ICAM1* and *IRF7* compared with ductal cells of control donors. ES, enrichment score.

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 major histocompatibility complex class II+ ductal cells. We next sought to corroborate our transcriptomic-based finding of MHC class II expression on ductal cells in T1D by using 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 (Supplementary Table 15). As 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. 8a–e). Using CyTOF, we identified a population of ductal cells expressing HLA-DR, an MHC class II protein encoded by the human leucocyte antigen (HLA) complex (Fig. 5a). 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 $< 1 \times 10^{-6}$). 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 analysed by CyTOF further confirmed the presence of this double-positive population across multiple donors (Fig. 5d and Extended Data Fig. 8f,g). Notably, these ductal cells did not express CD45, the hallmark of leucocytes (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 and 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 analysing 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 (Supplementary Table 16). This analysis confirmed that MHC class II-expressing ductal cells were predominately present in T1D donors (Fig. 5f–h and Extended Data Fig. 9a–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. 10b,c). Immunofluorescence staining 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 confirmed this finding (Fig. 6). Finally, cellular neighbourhood analysis in pancreatic tissues established that HLA-DR-expressing ductal cells were surrounded by CD4$^+$ T cells and myeloid cells including CD11B$^+$ DCs (Extended Data Fig. 10d–f; P value $< 1 \times 10^{-2}$). 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.

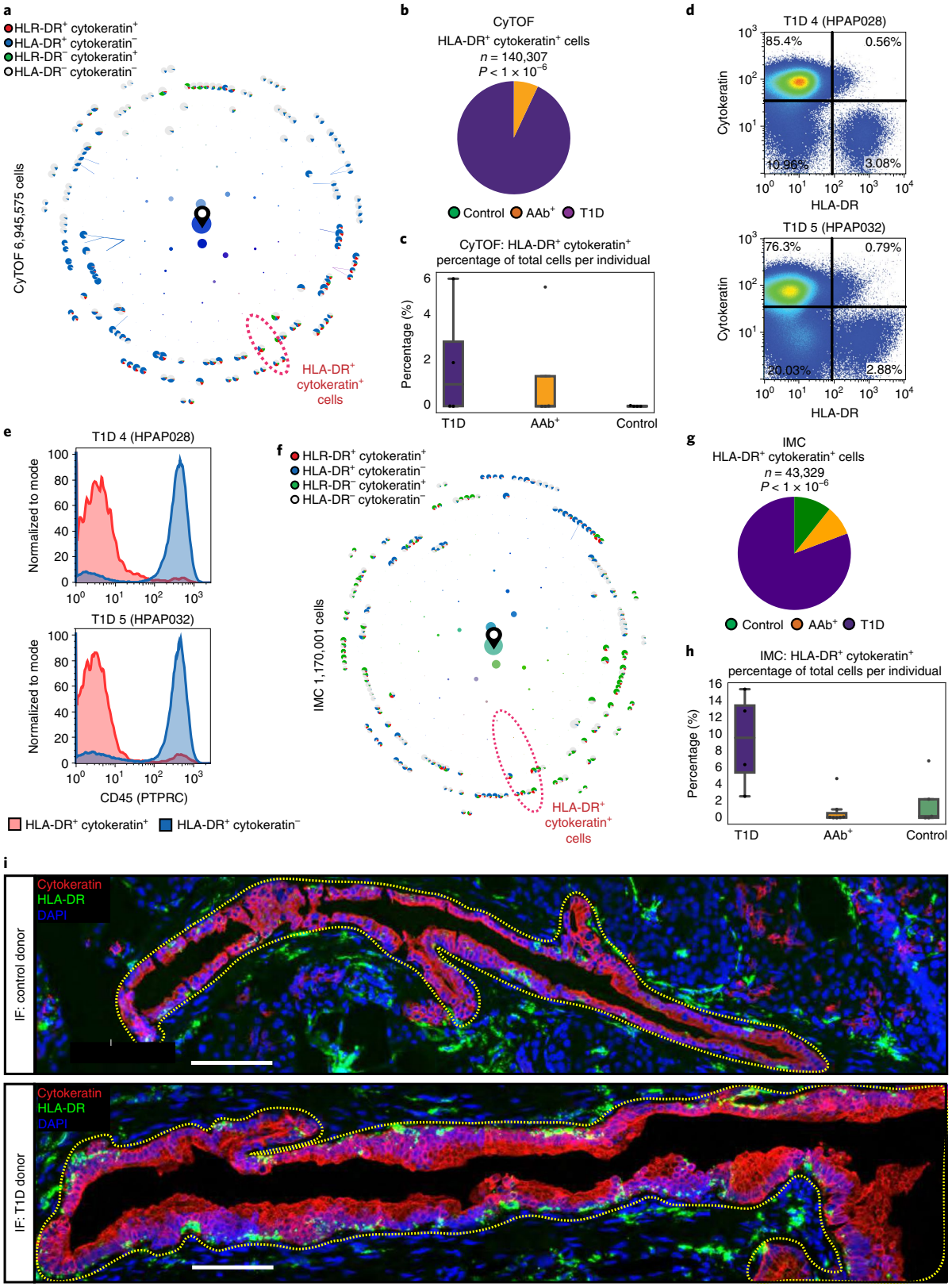
Discussion

Using three high-throughput single-cell technologies, we provided a comprehensive atlas of millions of cells using integrative multimodal analyses as a molecular microscope to investigate cellular diversity in the pancreas of T1D, AAb$^+$ and nondiabetic 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 normoglycaemia. 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 titres, suggesting the existence of dynamic transcriptional landscape in AAb$^+$ individuals. Although it is impossible to discern at present whether these transcriptional changes are contributing to or are by-products of disease pathogenesis, the mere discovery of molecular phenotypic changes in the 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 nondiabetic or AAb$^+$ donors, express high levels of MHC class II and interferon pathways, are surrounded by CD4$^+$ T cells and DCs and are transcriptionally similar to tolerogenic DCs. Although, to our knowledge, 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 signalling pathway[52]. Moreover, the

Fig. 5 | Three single-cell resolution protein-based approaches corroborate the existence of MHC class II-expressing ductal cells in T1D. a, Dendrogram visualization of coexpression of HLA-DR and cytokeratin protein coexpression in single cells analysed with 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**, Boxplots displaying the 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: 4 AAb$^+$, 4 control and 4 T1D). The box-and-whisker plot shows 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 IQR. **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 the coexpression of HLA-DR and cytokeratin proteins in single cells analysed by 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 chi-squared test. The P value shows 0.000 by both chi-square function from scipy.stats (Python) for the observed frequency array (34,983, 3,711, 4,635):(T1D, AAb$^+$, control) and cannot provide an exact P value. **h**, Boxplots displaying HLA-DR$^+$ cytokeratin$^+$ cells as a percentage of total cells per individual across donor groups from the IMC data (P value = 1×10^{-16}; number of donors: 7 AAb$^+$, 5 control, 4 T1D). The P value was obtained from a one-way analysis of variance (ANOVA) test. **i**, Representative confocal microscopy images from the pancreas of a T1D donor (top) and a control donor (bottom) displaying HLA-DR$^+$ cytokeratin$^+$ labelled by immunofluorescence (IF; control, $n = 3$; T1D, $n = 2$).


a
HLR-DR+ cytokeratin+
HLA-DR+ cytokeratin−
HLR-DR− cytokeratin+
HLA-DR− cytokeratin−
CyTOF 6,945,575 cells
HLA-DR+ cytokeratin+ cells
b
CyTOF
HLA-DR+ cytokeratin+ cells
n = 140,307
P < 1 × 10−6
Control AAb+ T1D
c
CyTOF: HLA-DR+ cytokeratin+ percentage of total cells per individual
Percentage (%)
T1D AAb+ Control
d
T1D 4 (HPAP028)
85.4%
0.56%
10.96%
3.08%
Cytokeratin
HLA-DR
T1D 5 (HPAP032)
76.3%
0.79%
20.03%
2.88%
e
T1D 4 (HPAP028)
T1D 5 (HPAP032)
Normalized to mode
CD45 (PTPRC)
HLA-DR+ cytokeratin+
HLA-DR+ cytokeratin−
f
HLR-DR+ cytokeratin+
HLA-DR+ cytokeratin−
HLR-DR− cytokeratin+
HLA-DR− cytokeratin−
IMC 1,170,001 cells
HLA-DR+ cytokeratin+ cells
g
IMC
HLA-DR+ cytokeratin+ cells
n = 43,329
P < 1 × 10−6
Control AAb+ T1D
h
IMC: HLA-DR+ cytokeratin+ percentage of total cells per individual
Percentage (%)
T1D AAb+ Control
i
IF: control donor
IF: T1D donor
Cytokeratin
HLA-DR
DAPI

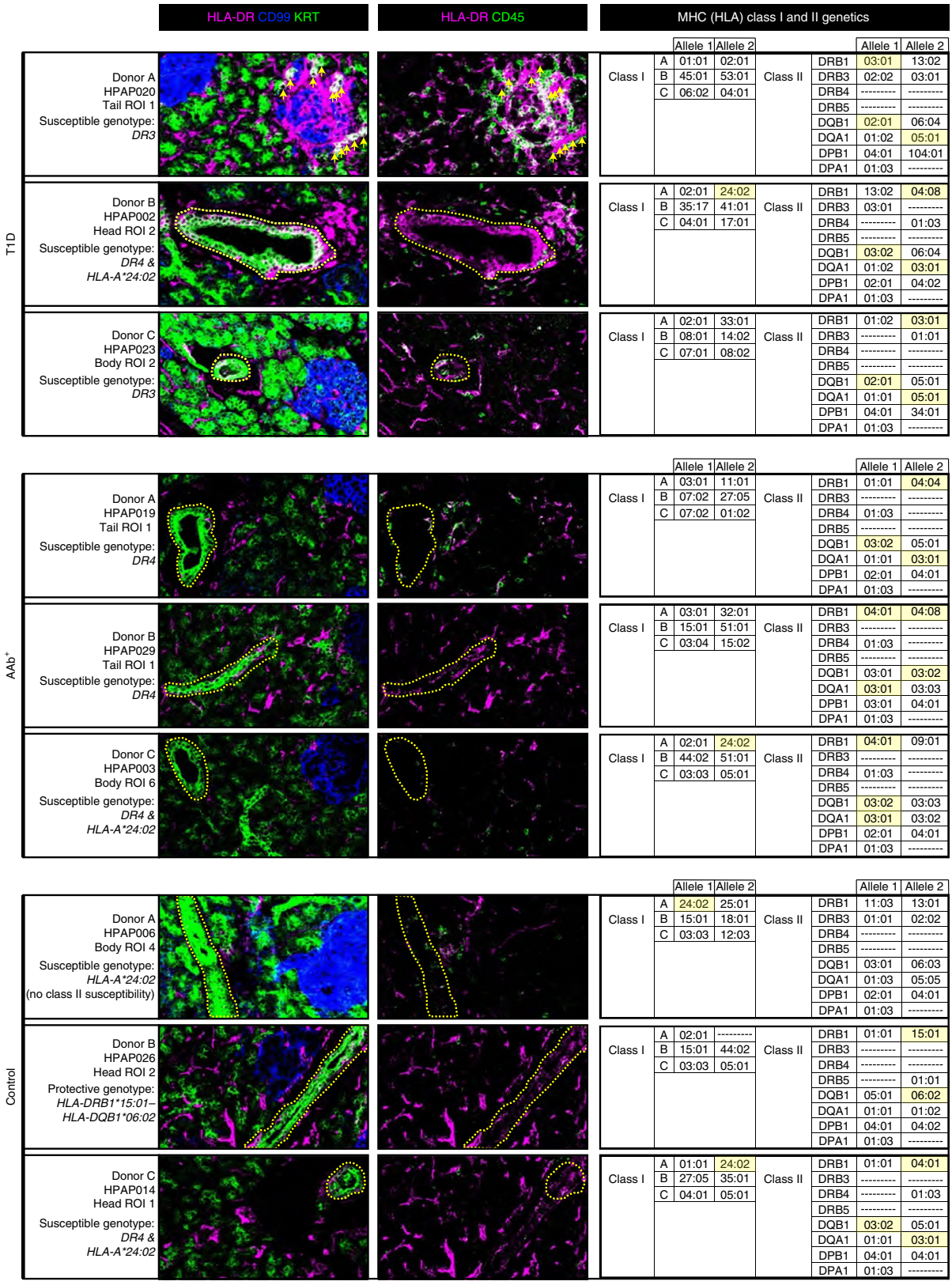

HLA-DR CD99 KRT
HLA-DR CD45
MHC (HLA) class I and II genetics
T1D
AAb+
Control
Donor A
HPAP020
Tail ROI 1
Susceptible genotype:
DR3
Donor B
HPAP002
Head ROI 2
Susceptible genotype:
DR4 &
HLA-A*24:02
Donor C
HPAP023
Body ROI 2
Susceptible genotype:
DR3
Donor A
HPAP019
Tail ROI 1
Susceptible genotype:
DR4
Donor B
HPAP029
Tail ROI 1
Susceptible genotype:
DR4
Donor C
HPAP003
Body ROI 6
Susceptible genotype:
DR4 &
HLA-A*24:02
Donor A
HPAP006
Body ROI 4
Susceptible genotype:
HLA-A*24:02
(no class II susceptibility)
Donor B
HPAP026
Head ROI 2
Protective genotype:
HLA-DRB1*15:01–
HLA-DQB1*06:02
Donor C
HPAP014
Head ROI 1
Susceptible genotype:
DR4 &
HLA-A*24:02
Class I
Class II
Allele 1
Allele 2
A 01:01 02:01
B 45:01 53:01
C 06:02 04:01
DRB1 03:01 13:02
DRB3 02:02 03:01
DRB4 --------- ---------
DRB5 --------- ---------
DQB1 02:01 06:04
DQA1 01:02 05:01
DPB1 04:01 104:01
DPA1 01:03 ---------
A 02:01 24:02
B 35:17 41:01
C 04:01 17:01
DRB1 13:02 04:08
DRB3 03:01 ---------
DRB4 --------- 01:03
DRB5 --------- ---------
DQB1 03:02 06:04
DQA1 01:02 03:01
DPB1 02:01 04:02
DPA1 01:03 ---------
A 02:01 33:01
B 08:01 14:02
C 07:01 08:02
DRB1 01:02 03:01
DRB3 --------- 01:01
DRB4 --------- ---------
DRB5 --------- ---------
DQB1 02:01 05:01
DQA1 01:01 05:01
DPB1 04:01 34:01
DPA1 01:03 ---------
A 03:01 11:01
B 07:02 27:05
C 07:02 01:02
DRB1 01:01 04:04
DRB3 --------- ---------
DRB4 01:03 ---------
DRB5 --------- ---------
DQB1 03:02 05:01
DQA1 01:01 03:01
DPB1 02:01 04:01
DPA1 01:03 ---------
A 03:01 32:01
B 15:01 51:01
C 03:04 15:02
DRB1 04:01 04:08
DRB3 --------- ---------
DRB4 01:03 ---------
DRB5 --------- ---------
DQB1 03:01 03:02
DQA1 03:01 03:03
DPB1 03:01 04:01
DPA1 01:03 ---------
A 02:01 24:02
B 44:02 51:01
C 03:03 05:01
DRB1 04:01 09:01
DRB3 --------- ---------
DRB4 01:03 ---------
DRB5 --------- ---------
DQB1 03:02 03:03
DQA1 03:01 03:02
DPB1 02:01 04:01
DPA1 01:03 ---------
A 24:02 25:01
B 15:01 18:01
C 03:03 12:03
DRB1 11:03 13:01
DRB3 01:01 02:02
DRB4 --------- ---------
DRB5 --------- ---------
DQB1 03:01 06:03
DQA1 01:03 05:05
DPB1 02:01 04:01
DPA1 01:03 ---------
A 02:01 ---------
B 15:01 44:02
C 03:03 05:01
DRB1 01:01 15:01
DRB3 --------- ---------
DRB4 --------- ---------
DRB5 --------- 01:01
DQB1 05:01 06:02
DQA1 01:01 01:02
DPB1 04:01 04:02
DPA1 01:03 ---------
A 01:01 24:02
B 27:05 35:01
C 04:01 05:01
DRB1 01:01 04:01
DRB3 --------- ---------
DRB4 --------- 01:03
DRB5 --------- ---------
DQB1 03:02 05:01
DQA1 01:01 03:01
DPB1 04:01 04:01
DPA1 01:03 ---------

Fig. 6 | Representative examples of IMC measurement corroborate that MHC class II-positive ductal cells are present in pancreatic tissues. Left, 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 at 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 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 (refs. [45,66,67]). Notably, HLA-DR$^+$ ductal cells were found across all HLA genotypes, including both susceptible and protective genotypes.

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, nonprofessional APCs 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 immunohistochemistry to assess the reproducibility of our findings related to ductal cells in T1D donors across independent experimental assays.

Methods

Experimental model and donor 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 no. 201600029) and the United Network for Organ Sharing. A legal representative for each donor provided informed consent before 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 AAbs before organ collection, and AAb positivity was confirmed after tissue processing and islet isolation.

Organs were processed as previously described[19,35]. Supplementary Tables 1 and 2 summarize 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). Briefly, the test sample was bound with a monoclonal antibody immobilized on a magnetic solid phase and an enzyme-labelled monoclonal antibody, and then the sample was incubated with a fluorogenic substrate, 4-methylumbelliferyl phosphate (4MUP). The amount of enzyme-labelled monoclonal antibody that binds to the beads is directly proportional to the C-peptide concentration in the test sample. A standard curve was constructed using calibrator of known concentration, and unknown sample concentrations are calculated using the curve. The C-peptide assay was calibrated against WHO IS 84/510 standard. The assay had a sensitivity level of 0.02 ng ml^{-1}. To monitor the assay performance, a set of low, medium and high C-peptide-level quality-control samples were analysed several times per day. The inter-assay coefficients of variability for the low, medium and high C-peptide controls were 3.2%, 1.6% and 1.8%, respectively. The results of the analyses of the long-term monitoring pools demonstrated a consistently low variation around the target values, thus ensuring result consistency.

Serum from organ donors was tested for GAD, IA-2, mIAA and ZnT8A AAbs by radioligand-binding assay as previously described[58]. Micro IAA (mIAA) and ZnT8A were performed with in-house radiobinding assay[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. Measuring GAD and IA-2 levels was performed with NIDDK harmonized standard methods (3) and the upper limits of normal (20 DK units per ml for GAD and 5 DK units per ml for IA-2) was established around the 99th percentile from receiver operating characteristic curves in 500 healthy controls and 50 individuals 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 and 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. A total of 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 HiSeq 4000: 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 and cell-type classification. Seurat (v3.1.5)[32,61] was used for filtering, uniform manifold approximation and projection (UMAP) generation and initial clustering. Genes expressed in at least three cells were included, as were cells with at least 200 genes. nFeature, nCount, percent.mt, nFeature versus nCount and percent.mt versus nCount plots were generated to ascertain the lenient filtering criteria of 200 < nFeature < 8,750, percent.mt < 25 and nCount < 125,000. Data were then log normalized, and the top 2,000 variable genes were detected using the 'vst' selection method. Data were then linearly transformed ('scaled'), meaning that for each gene, the mean expression across cells was 0 and the variance across cells was 1. Principal-component analysis (PCA) was then carried out on the scaled data, using the 2,000 variable genes as input. We used two approaches to determine the dimensionality of the data, that is, how many principal components (PCs) to choose when clustering: (1) a Jackstraw-inspired resampling test that compares the distribution of *P* values of each PC against a null distribution and (2) an elbow plot that displays the standard deviation explained by each PC. Based on these two approaches, 17 PCs with a resolution of 1.2 were used to cluster the cells, and nonlinear 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 (Supplementary Fig. 1d). 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 PC 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 (Supplementary Table 17) 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, that is, the 'clustering-extended type' (Louvain clustering; Supplementary Fig. 3d). After 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 leucocyte common antigen, high; Supplementary Fig. 3e). 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 dataset of adult pancreatic cells (EGAS00001004653) was used as a reference for the 'query' dataset presented in this study. First, SCTransform was used to preprocess the data[61]. Briefly, SCTransform uses a generalized linear model for each gene with unique molecular identifier count 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 datasets. Seurat uses diagonalized canonical correlation analysis followed by L2-normalization and searching for mutual nearest neighbours. Then, anchors between datasets are compared based on their local neighbourhood structure of other anchors to receive 'correction vectors'. These correction vectors are then subtracted from the query gene expression matrix, resulting in an integrated dataset[63]. Similarly for label transfer, these anchors between datasets are instead labelled 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 dataset as a reference (GSE145126).

We used 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 the 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 used to normalize gene expression.

Differential gene expression, gene-set enrichment analysis 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 (false discovery rate (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 1×10^{-6} was applied to determine significant gene ontology pathways. For GSEA, 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. Pseudo-bulk analysis was performed by taking the average of cells within individuals. The differential genes were found using edgeR through a 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 FC > 1.

Hybrid cell coexpression, differential expression analysis and heat maps. For the differentially expressed genes (FDR < 0.05 and FC > 0.1) between every two sample groups, we calculated the shared and unique genes in each cell type, and visualized the data using 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 aggregated 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 heat maps.

To examine the coexpression of signature genes of some cell types, we normalized the median normalized matrix with $\log_2(N+1)$. 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 the ggplot2 package of R.

Cytometry time-of-flight 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 labelling was carried out in 'FoxP3 permeabilization buffer' (eBioscience, 00-8333) with 1% FBS (Hyclone, 7207) for 12 h at 4 °C at a concentration of up to 3 million cells per 300 µl of antibody cocktail, followed by washing twice 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 room temperature for 1 h. Mass cytometry data were acquired by CyTOF (Fluidigm). Flow CyTOF data analyses of endocrine cell composition were performed using the Cytobank implement (https://www.cytobank.org/).

Normalized FCS files were pre-processed before TooManyCells analysis and visualization using FlowJo (v10.6.1) 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 cytometry by time of flight. TooManyCells was used to generate cell clades of CyTOF data. Cells with less than a total of 1×10^{-16} signal were removed, leaving 6,945,575 cells. Following 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 median absolute deviation (MAD) × median number of cells in nodes) cells within them into their parent nodes.

Imaging mass cytometry 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 3 × 3-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 10 µm^2, then dilated up to seven pixels to approximate cell boundaries. Per-cell object mean pixel intensities were then exported for further analysis.

TooManyCells clustering for imaging mass cytometry. TooManyCells was used to generate cell clades of IMC data. Cells with less than a total of 1×10^{-16} signal were removed. After 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 × median number of cells in nodes) cells within them into their parent nodes. Subsetting of the tree was performed with '--root-cut 3' to focus on node 3 in relevant analyses, with additional pruning of (3 MAD × median number of cells in nodes).

Cell-neighbourhood analysis for imaging mass cytometry. Three labels were given to cells in the IMC neighbourhood analysis: base, neighbour 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, that is, 20 for the complete pruned tree, the cell was assigned the neighbour label. Otherwise, the cell was designated as distant.

Machine-learning method for cell annotation in imaging mass cytometry and cytometry by time of flight. 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) were taken as input. The raw data were

normalized with an arcsinh transformation and a cofactor of 200 in the case of CyTOF, while log transformation followed by unit normalization in the case of IMC data. The data were then randomly split into two halves (half donors in one set); cells from 50% of donors were in the training set, while the remaining were in the test set. The splitting was done in a stratified fashion based on the disease condition (T1D, AAb$^+$ and control). Semi-supervised learning was used 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 that does not need training for learning).

Immunofluorescence and confocal microscopy. Tissues were fixed in 10% buffered formalin overnight, washed several times in PBS, then dehydrated using ethanol and xylene, then embedded in paraffin and sectioned to 4–8 µm. 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 (Akoya Biosciences). Slides were counterstained with DAPI, and then mounted and imaged on a 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 boxplots with control, AAb$^+$ and T1D donor states. The D'Agostino–Pearson omnibus normality test was used to assess whether the data from each group were 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, a one-way ANOVA was applied.

Statistical analysis of cellular neighbourhoods. Differential marker expression significance for neighbours in the IMC analysis was determined using permutation tests. For each marker, the distribution of the observed marker value for each of the designated n neighbours 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. Pseudo-bulk counts of GAD$^+$ donors across all cell types were identified using the muscat tool[65]. The GAD levels for each GAD$^+$ donor were retrieved from Supplementary 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 type 1 diabetes. 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].

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Gene Expression Omnibus 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/.

Code availability

Where applicable, scripts used for data processing and analysis are available in the Supplemental Information and Methods and provided on GitHub at 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 Methods.

Received: 27 April 2021; Accepted: 14 January 2022;
Published online: 28 February 2022

References

1. Powers, A. C. Type 1 diabetes mellitus: much progress, many opportunities. *J. Clin. Invest.* https://doi.org/10.1172/JCI142242 (2021).
2. Michels, A. W., Redondo, M. J. & Atkinson, M. A. The pathogenesis, natural history, and treatment of type 1 diabetes: time (thankfully) does not stand still. *Lancet Diabetes Endocrinol.* https://doi.org/10.1016/S2213-8587(21)00344-2 (2021).
3. Boldison, J. & Wong, F. S. Immune and pancreatic beta cell interactions in type 1 diabetes. *Trends Endocrinol. Metab.* **27**, 856–867 (2016).
4. Unanue, E. R. & Wan, X. The immunoreactive platform of the pancreatic islets influences the development of autoreactivity. *Diabetes* **68**, 1544–1551 (2019).
5. Barrett, J. C. 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 (2009).
6. Campbell-Thompson, M. et al. Network for Pancreatic Organ Donors with Diabetes: developing a tissue biobank for type 1 diabetes. *Diabetes Metab. Res. Rev.* **28**, 608–617 (2012).
7. Kaestner, K. H., Powers, A. C., Naji, A., Consortium, H. & Atkinson, M. A. NIH initiative to improve understanding of the pancreas, islet and autoimmunity in type 1 diabetes: the Human Pancreas Analysis Program. *Diabetes* **68**, 1394–1402 (2019).
8. Pliner, H. A., Shendure, J. & Trapnell, C. Supervised classification enables rapid annotation of cell atlases. *Nat. Methods* **16**, 983–986 (2019).
9. Schwartz, G. W. et al. TooManyCells identifies and visualizes relationships of single-cell clades. *Nat. Methods*, https://doi.org/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 (2021).
11. Campbell-Thompson, M., Rodriguez-Calvo, T. & Battaglia, M. Abnormalities of the exocrine pancreas in type 1 diabetes. *Curr. Diab. Rep.* **15**, 79 (2015).
12. Chiou, J. et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398–402 (2021).
13. Tersey, S. A. 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 (2012).
14. Marhfour, I. et al. Expression of endoplasmic reticulum stress markers in the islets of patients with type 1 diabetes. *Diabetologia* **55**, 2417–2420 (2012).
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 (2013).
16. Nakayasu, E. S. et al. Comprehensive proteomics analysis of stressed human islets identifies GDF15 as a target for type 1 diabetes intervention. *Cell Metab.* **31**, 363–374 (2020).
17. Unsicker, K., Spittau, B. & Krieglstein, K. The multiple facets of the TGF-β family cytokine growth/differentiation factor-15/macrophage inhibitory cytokine-1. *Cytokine Growth Factor Rev.* **24**, 373–384 (2013).
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 (2010).
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 (2020).
20. Wang, Y. J. et al. Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* **65**, 3028–3038 (2016).
21. Teo, A. K. K. et al. Single-cell analyses of human islet cells reveal de-differentiation signatures. *Cell Death Discov.* **4**, 14 (2018).
22. Chiang, M. K. & Melton, D. A. Single-cell transcript analysis of pancreas development. *Dev. Cell* **4**, 383–393 (2003).
23. Riedel, M. J. et al. Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas. *Diabetologia* **55**, 372–381 (2012).
24. Heng, T. S. & Painter, M. W., Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
25. Vahedi, G. et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558–562 (2015).
26. Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *Elife* https://doi.org/10.7554/eLife.55851 (2020).
27. Ludwig, M. Q. et al. A genetic map of the mouse dorsal vagal complex and its role in obesity. *Nat. Metab.* **3**, 530–545 (2021).
28. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–908 (2018).

29. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
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 (2020).
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 (2016).
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 (2018).
33. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394 (2016).
34. Segerstolpe, A. et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* **24**, 593–607 (2016).
35. Wang, Y. J. 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 (2019).
36. Zimmerman, K. D., Espeland, M. A. & Langefeld, C. D. A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* **12**, 738 (2021).
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 (2009).
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.* **99**, 675–683 (2021).
39. Christen, U. et al. A dual role for TNF 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 (2001).
40. Lombardi, A., Tsomos, E., Hammerstad, S. S. & Tomer, Y. Interferon-α: the key trigger of type 1 diabetes. *J. Autoimmun.* **94**, 7–15 (2018).
41. Marro, B. S. et al. Progression of type 1 diabetes from the prediabetic stage is controlled by interferon-α signaling. *Proc. Natl Acad. Sci. USA* **114**, 3708–3713 (2017).
42. Osum, K. C. et al. Interferon-γ drives programmed death-ligand 1 expression on islet beta cells to limit T cell function during autoimmune diabetes. *Sci. Rep.* **8**, 8295 (2018).
43. Ziegler, A. G. et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA* **309**, 2473–2479 (2013).
44. Noble, J. A. & Valdes, A. M. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr. Diab. Rep.* **11**, 533–542 (2011).
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 (2007).
46. Wculek, S. K. et al. Dendritic cells in cancer immunology and immunotherapy. *Nat. Rev. Immunol.* **20**, 7–24 (2020).
47. Villani, A. C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes and progenitors. *Science* https://doi.org/10.1126/science.aah4573 (2017).
48. Appleman, L. J. & Boussiotis, V. A. T cell anergy and co-stimulation. *Immunol. Rev.* **192**, 161–180 (2003).
49. Wang, L. et al. VISTA, a novel mouse Ig superfamily ligand that negatively regulates T cell responses. *J. Exp. Med.* **208**, 577–592 (2011).
50. Wang, Y. J. et al. Single-cell mass cytometry analysis of the human endocrine pancreas. *Cell Metab.* **24**, 616–626 (2016).
51. Rodriguez-Calvo, T., Ekwall, O., Amirian, N., Zapardiel-Gonzalo, J. & von Herrath, M. G. Increased immune cell infiltration of the exocrine pancreas: a possible contribution to the pathogenesis of type 1 diabetes. *Diabetes* **63**, 3880–3890 (2014).
52. Steimle, V., Siegrist, C. A., Mottet, A., Lisowska-Grospierre, B. & Mach, B. Regulation of MHC class II expression by interferon-γ mediated by the transactivator gene *CIITA*. *Science* **265**, 106–109 (1994).
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 (2020).
54. Wosen, J. E., Mukhopadhyay, D., Macaubas, C. & Mellins, E. D. Epithelial MHC class II expression and its role in antigen presentation in the gastrointestinal and respiratory tracts. *Front Immunol.* **9**, 2144 (2018).
55. Kondo, A. et al. Highly multiplexed image analysis of intestinal tissue sections in patients with inflammatory bowel disease. *Gastroenterology* **161**, 1940–1952 (2021).
56. Russell, M. A. 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 (2019).
57. Squair, J. W. et al. Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
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 (2010).
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. USA* **97**, 1701–1706 (2000).
60. Wenzlau, J. M. et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc. Natl Acad. Sci. USA* **104**, 17040–17045 (2007).
61. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
62. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA-sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).
63. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
64. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
65. Crowell, H. L. et al. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).
66. Howson, J. M. et al. Evidence of gene–gene interaction and age-at-diagnosis effects in type 1 diabetes. *Diabetes* **61**, 3012–3017 (2012).
67. Howson, J. M., Walker, N. M., Clayton, D. & Todd, J. A., Type 1 Diabetes Genetics Consortium. 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**, 31–45 (2009).

Acknowledgements

We thank our colleagues for helpful discussions, particularly: A. Chandra, E. L. Prak, B. Stanger, M. Silverman, G. Beatty, K. Z., M. Lazar, R. Vonderheide, A. Minn and E. J. Wherry. Some schematics throughout the paper were created with Biorender.com. We thank A. Georgescu for confocal microscopy and the University of Pennsylvania Diabetes Research Center for the use of the Functional Genomics Core (P30-DK19525). This work was supported by National Institute of Health grants UC4 DK112217 and U01DK112217 (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.

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 measurements. C.L.M. performed immunohistochemistry experiments. N.G. assisted in analysis of CyTOF data. W.W. assisted in analysis of scRNA-seq data. A.M. performed annotation of IMC and CyTOF data. M. Feldman supervised the HPAP tissue bank. J.H.M., A.S.J. and M.R.B. are members of the HPAP consortium. K.H.K. supervised M.G., Y.J.W., A.M., J. L., M.W., D.T., A.K. and C.L.M. for generation of scRNA-seq, CyTOF and IMC data. R.B.F. supervised A.M. and G.W.S. in analysis of scRNA-seq, CyTOF and IMC data. 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 data. 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.

Competing interests

M.R.B. has a consulting arrangement with Interius BioTherapeutics. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s42255-022-00531-x.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42255-022-00531-x.

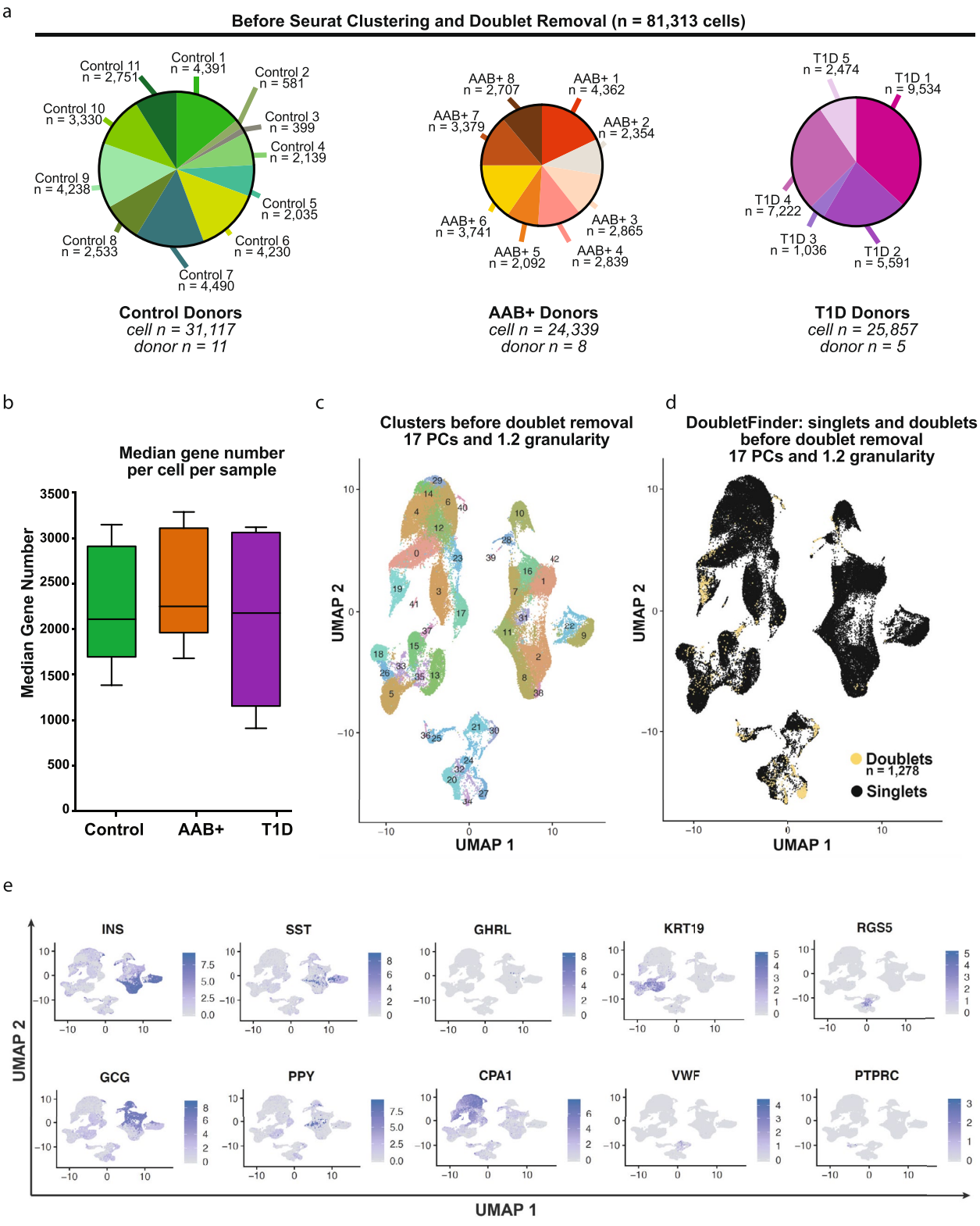
Correspondence and requests for materials should be addressed to Robert B. Faryabi, Ali Naji, Klaus H. Kaestner or Golnaz Vahedi.

Peer review information *Nature Metabolism* thanks Tune Pers, Raghavendra Mirmira and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Isabella Samuelson.

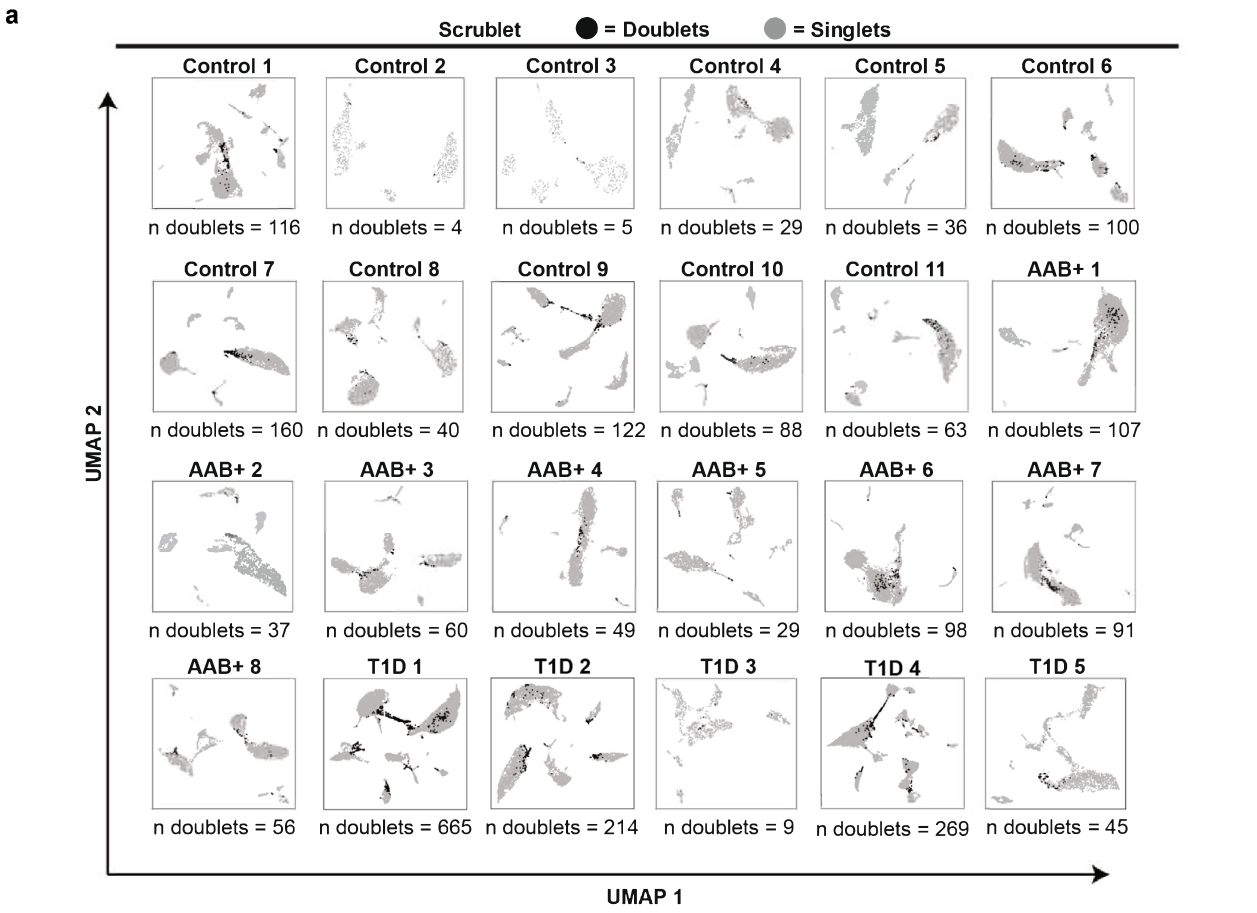
the HPAP Consortium

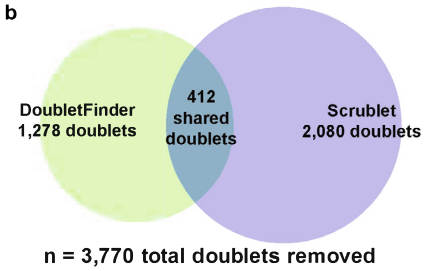
Maria Fasolino[11,12,13,14,15,16], Gregory W. Schwartz[12,13,15,16,17], Abhijeet R. Patil[11,12,13,14,15,16], Aanchal Mongia[12,13,15,16,17], Maria L. Golson[11,14,18], Yue J. Wang[11,14], Ashleigh Morgan[11,14], Chengyang Liu[14,19], Jonathan Schug[11], Jinping Liu[11,14], Minghui Wu[11,14], Daniel Traum[11,14], Ayano Kondo[11,14], Catherine L. May[11,14], Naomi Goldman[11,12,13,14,15,16], Wenliang Wang[11,12,13,14,15,16], Michael Feldman[16,17], Jason H. Moore[11,16], Alberto S. Japp[10,12], Michael R. Betts[12,20], Robert B. Faryabi[12,13,15,16,17], Ali Naji[12,14,19], Klaus H. Kaestner[11,13,14] and Golnaz Vahedi[11,12,13,14,15,16]

[11]Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [12]Institute for Immunology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [13]Epigenetics Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [14]Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [15]Abramson Family Cancer Research Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [16]Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [17]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [18]Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA. [19]Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [20]Department of Microbiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.



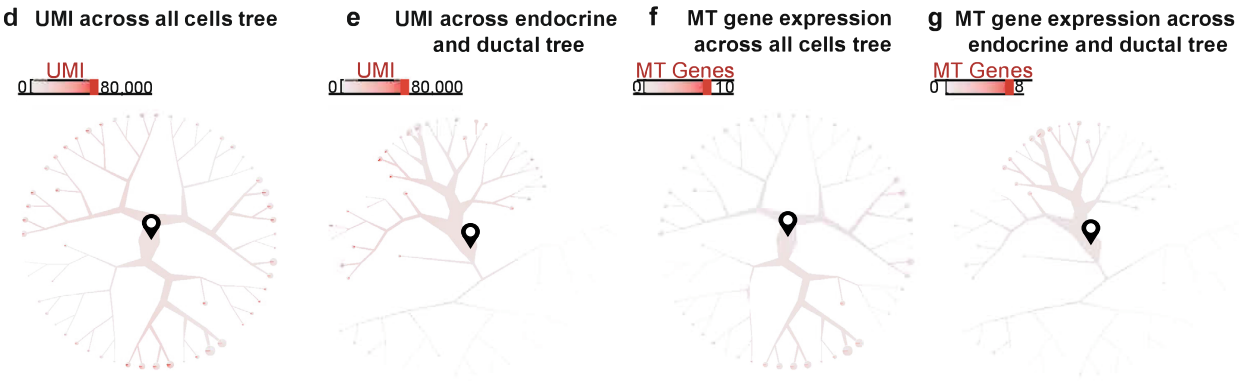
Extended Data Fig. 1 | Cell numbers and clustering before complete filtering. a) Pie chart displaying the cell numbers and proportions of each individual donor per donor type. b) Box plot displaying the average gene number per cell per donor type. c) UMAP visualization of cell clusters for all cells. 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.





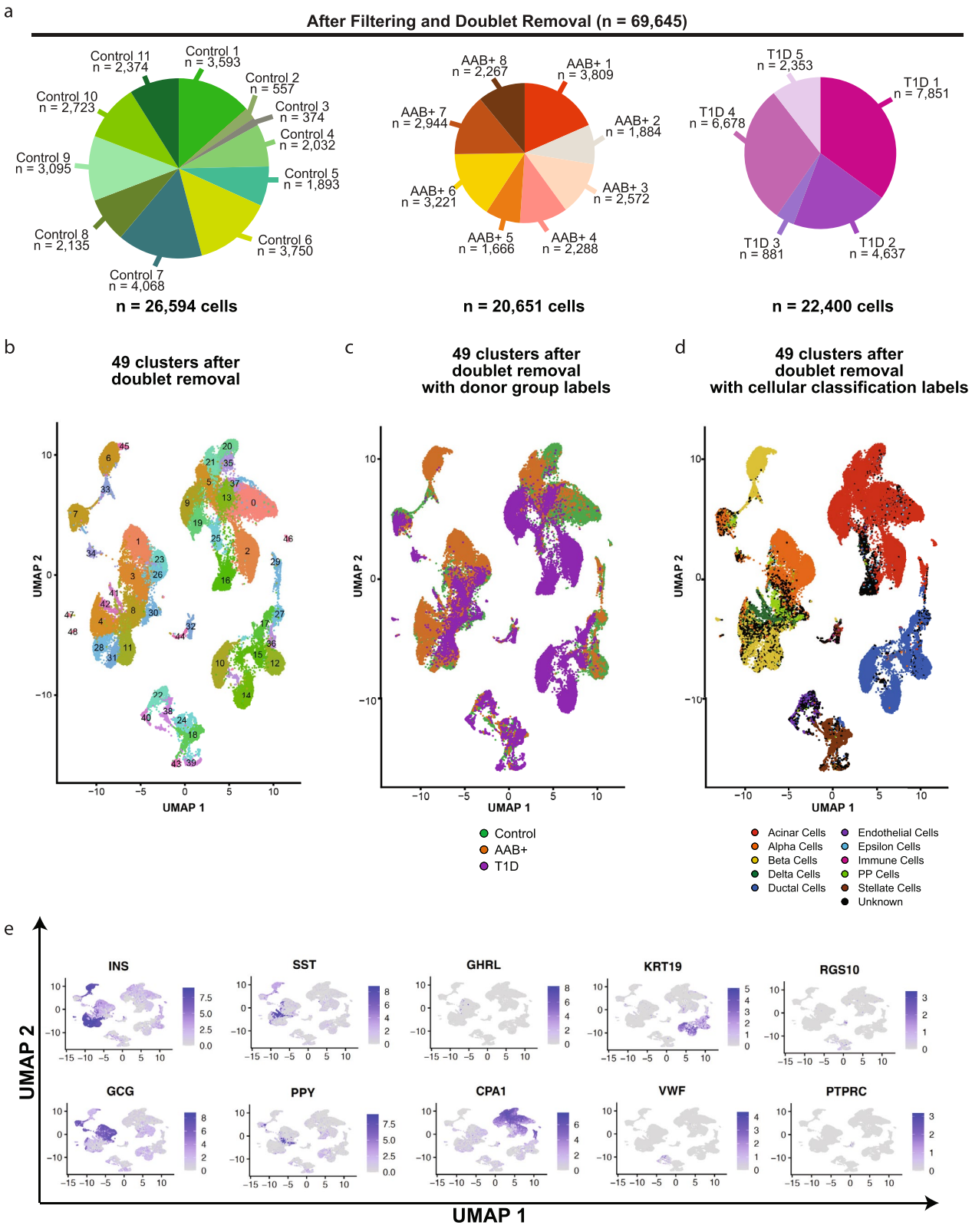
c

	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



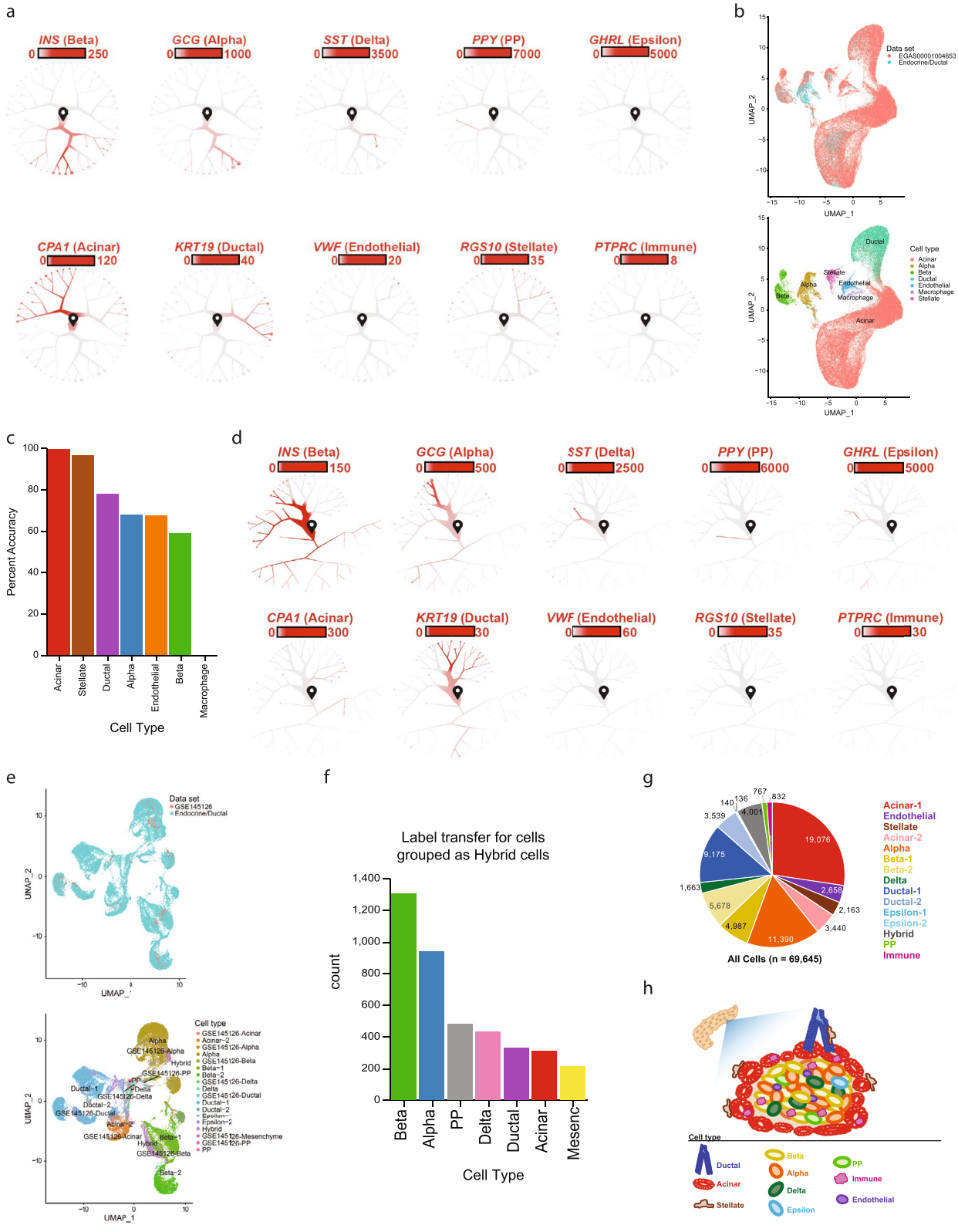
Extended Data Fig. 2 | See next page for caption.

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 Fig. 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 Fig. 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 Fig. 1c. g) Expression of genes associated with mitochondrial function projected across the dendrogram visualization and clustering of ductal and endocrine cells from Fig. 1d.



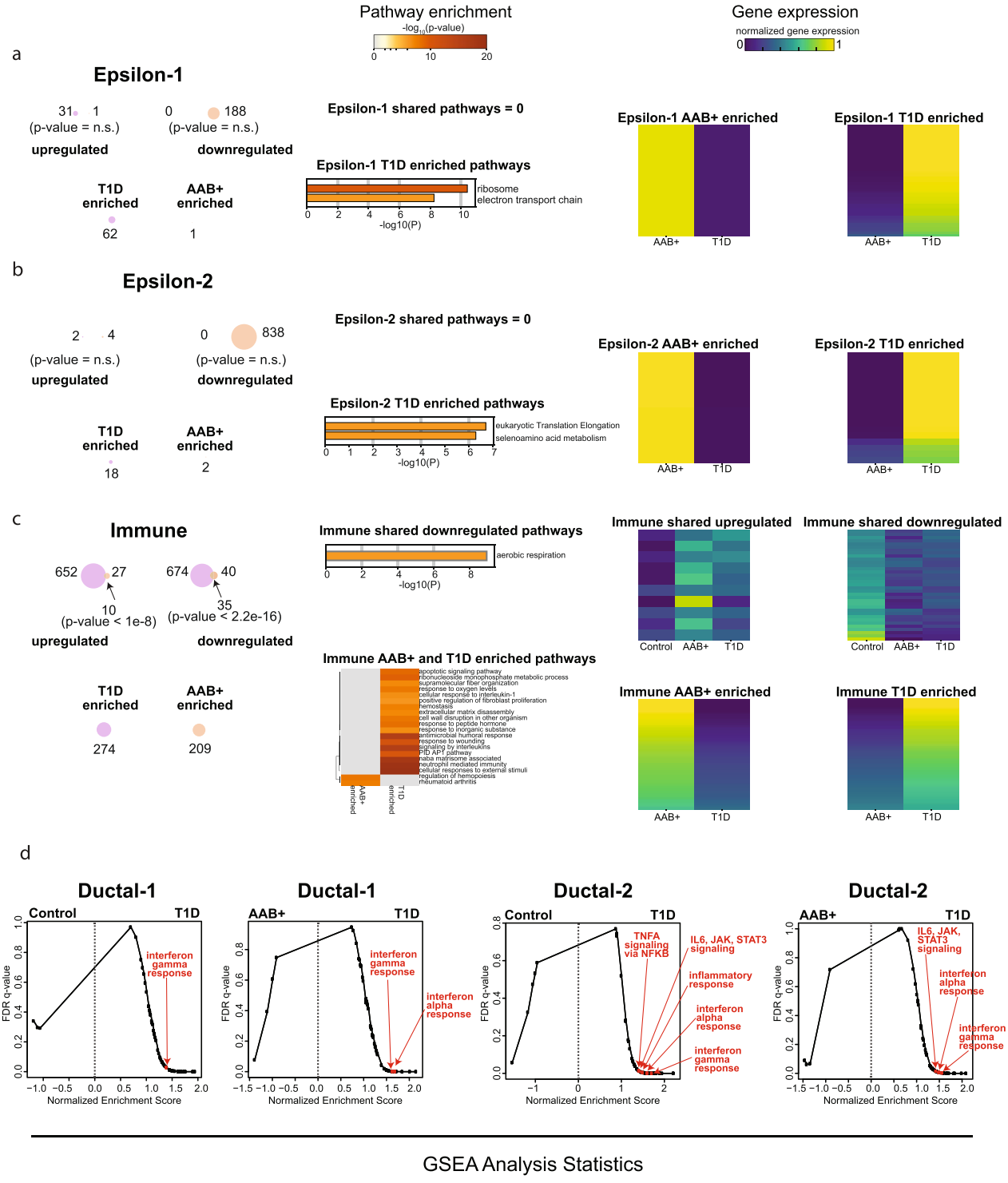
Extended Data Fig. 3 | See next page for caption.

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.



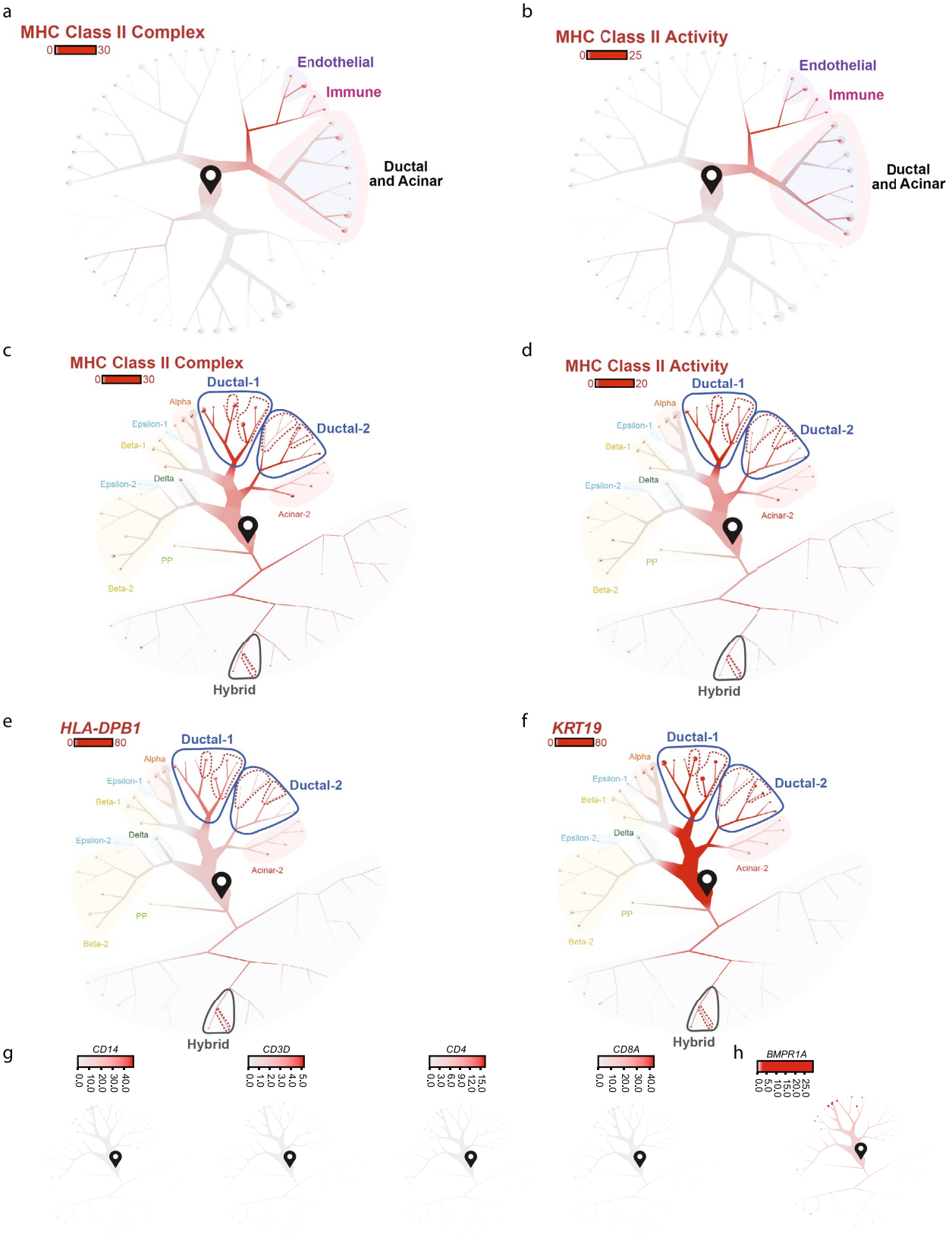
Extended Data Fig. 4 | See next page for caption.

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 Fig. 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 Fig. 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 Fig. 1, c and d. h) Schematic of the human pancreatic islet anatomy and major cell types.



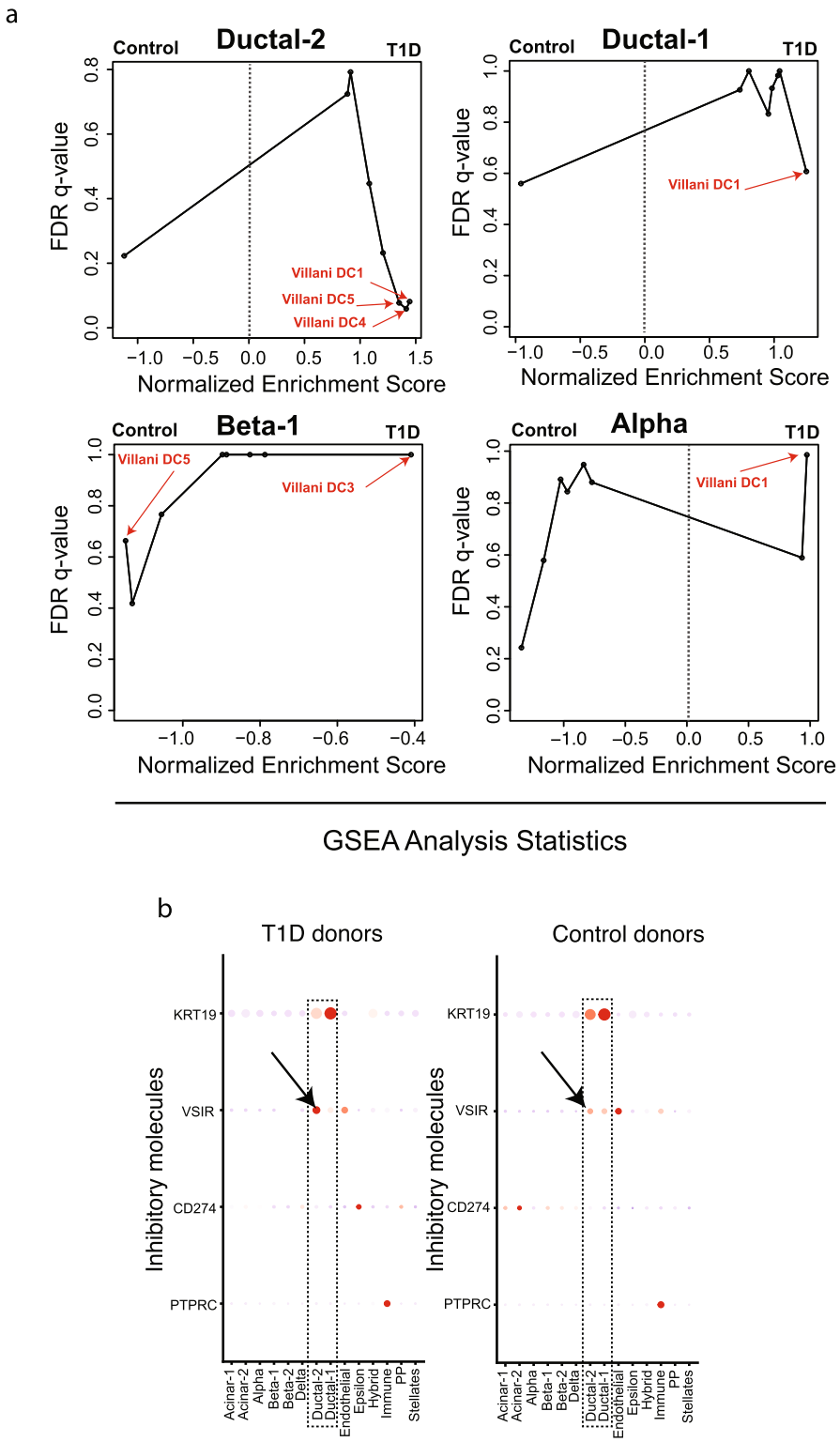
Extended Data Fig. 5 | See next page for caption.

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.

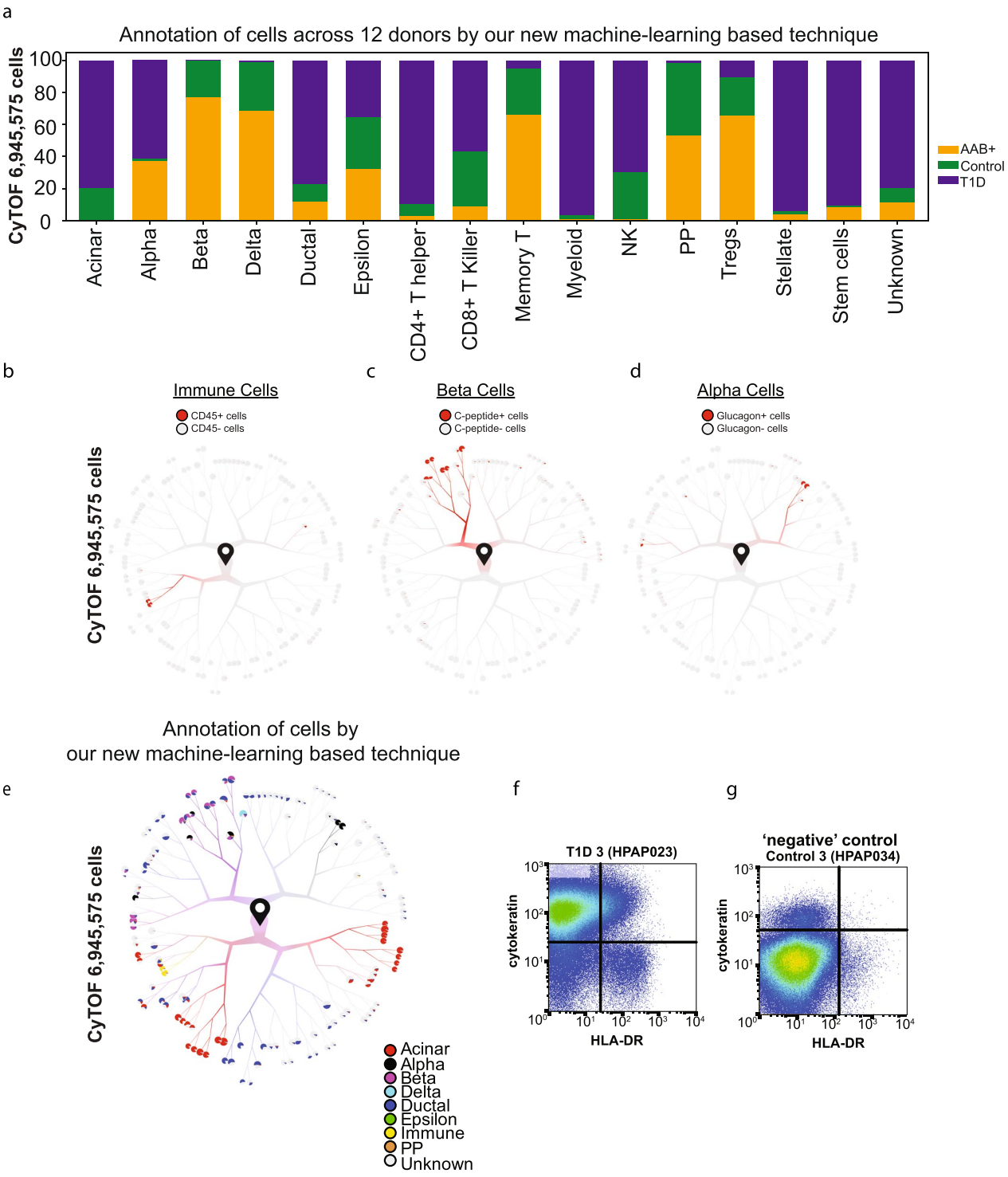


Extended Data Fig. 6 | See next page for caption.

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

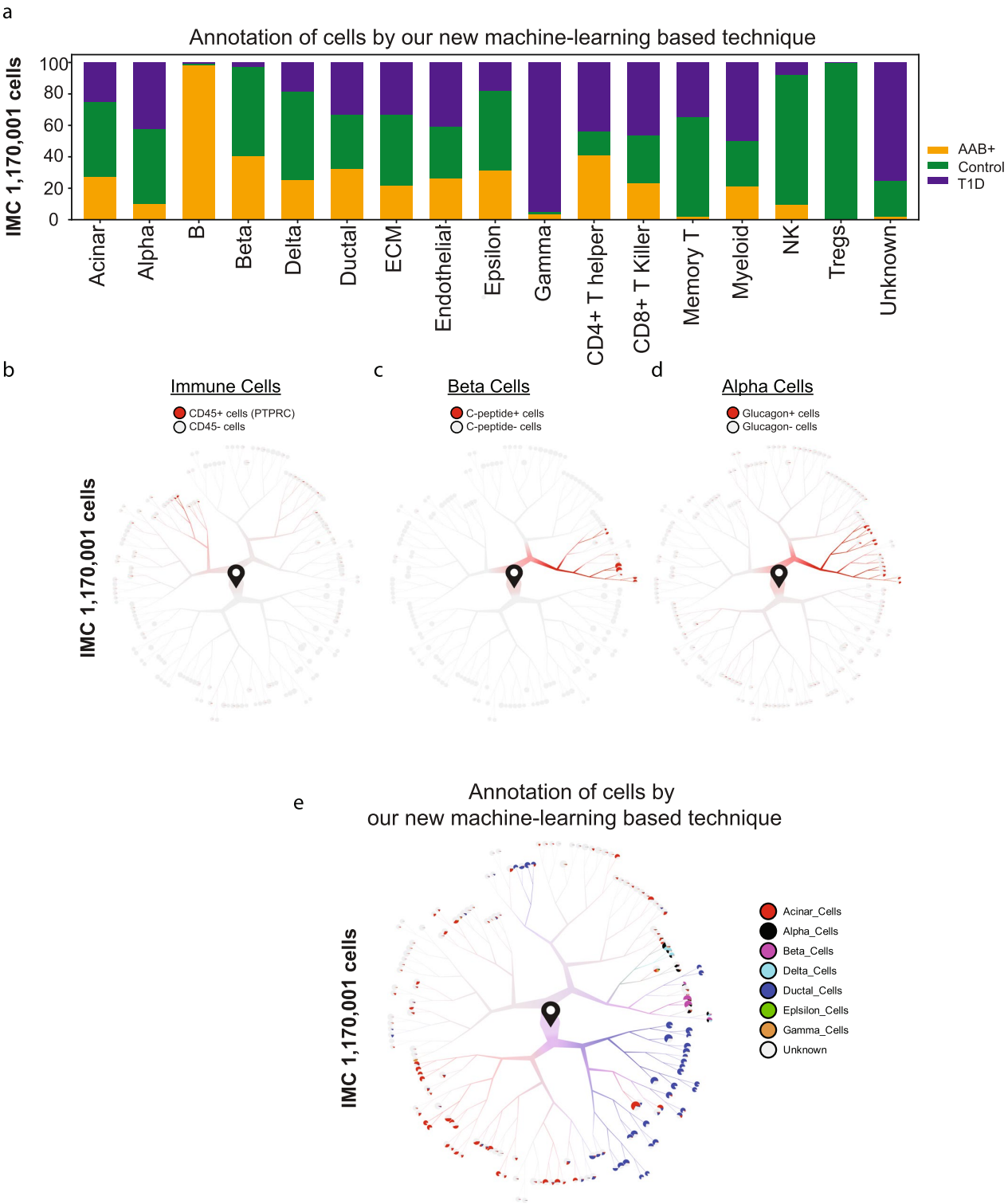


Extended Data Fig. 7 | GSEA analysis across annotated cells types for dendritic cells gene sets. a) DC1 gene signature is significantly enriched within Ductal-2 cells of T1D donors. Integrated GSEA analysis for dendritic cells gene sets from Villani et al[47] across ranked lists of differentially expressed 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.

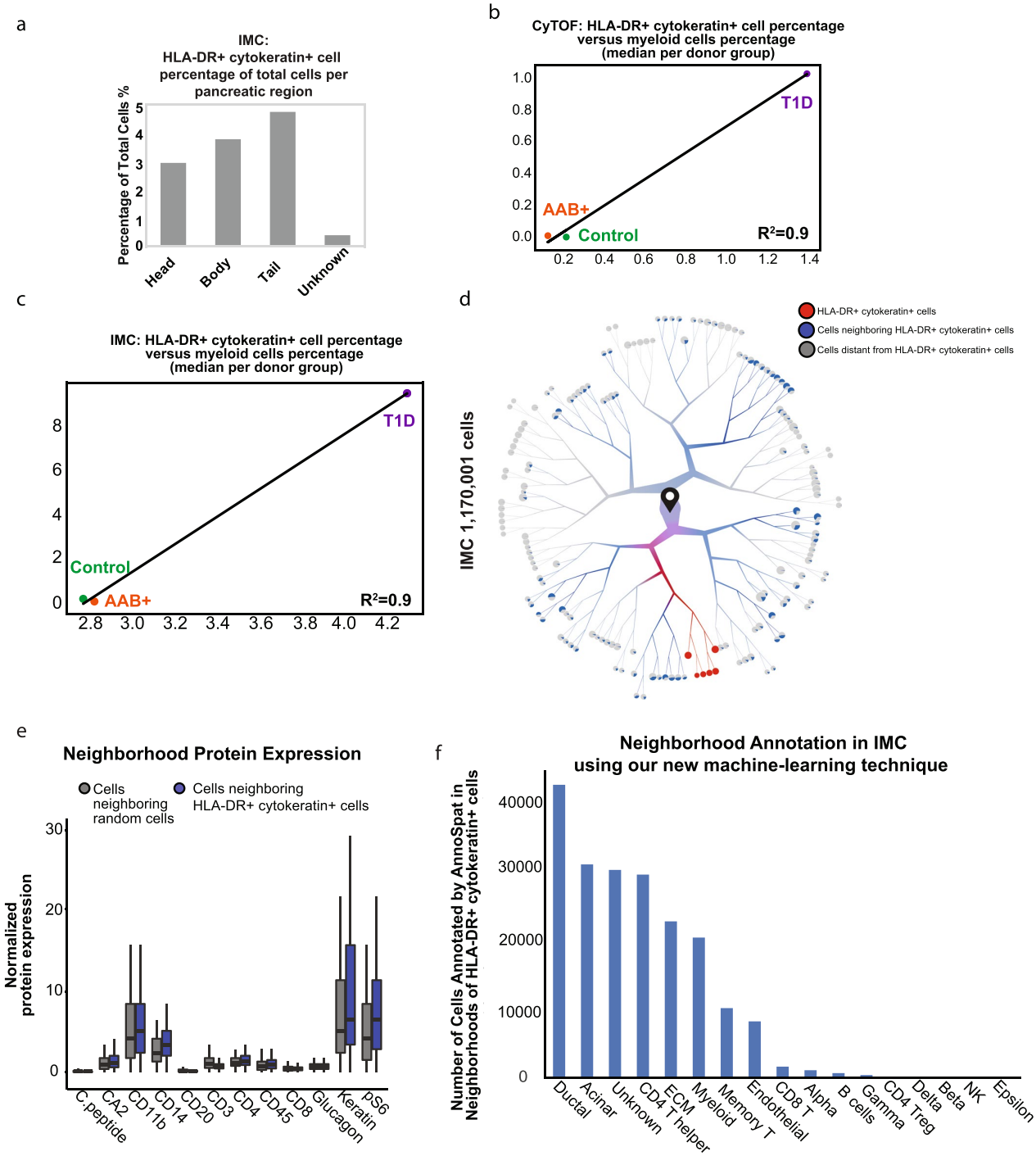


Extended Data Fig. 8 | See next page for caption.

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 | See next page for caption.

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.

nature research

Corresponding author(s): Golnaz Vahedi, Klaus H. Kaestner, Ali Naji, Robert B. Faryabi

Last updated by author(s): January 20 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

scRNA-seq
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 (bcl2fastq v2.20.0.422) conversion, aligning (using the hg38 reference genome), filtering, counting, cell calling, and aggregating (--normalize=none) (fig. S1, a and b).

CyTOF
Mass cytometry data were acquired on CyTOF (Fluidigm).

IMC
From Wang et al. 2016:
"Antibody Labeling and Image Acquisition
Four to eight mm FFPE sections were stained with an antibody cocktail (Table S1) containing all antibodies. Briefly, tissue sections were de-paraffinized with xylene and carried through sequential rehydration from 100% Ethanol to 70% Ethanol before being transferred to PBS. Heat-induced antigen retrieval was performed in a decloaking chamber (Biocare Medical) at 95oC for 30 min in Tris/EDTA buffer (10mM Tris, 1mM EDTA, pH9.2). Slides were cooled to room temperature (RT) and were subsequently blocked with PBS+3%BSA for 1h at RT. Meanwhile, the antibody cocktail was prepared in PBS+1%BSA buffer, with the appropriate dilution for each of the antibodies (Table S1). Each slide was incubated with 100 ml of the antibody cocktail overnight at 4oC. The next day, slides were washed 3 times with PBS and labeled with 1:400 dilution of Intercalator-Ir (Fluidigm 201192B) in PBS for 30 min at RT. Slides were briefly washed with H2O three times and air dried for at least 30 min before IMC acquisition. The IMC was purchased from Fluidigm (Fluidigm, Hyperion Imaging System). All IMC operation was performed following Fluidigm's operation"

IHC
Slides were imaged on Zeiss LSM800

Data analysis

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.

scRNA-seq clustering, doublet removal, & cell type classification
Seurat v3.1.5 (19, 50) was used for filtering, UMAP generation, and initial clustering. Genes expressed in at least 3 cells were kept, 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 > nFeature < 8,750, percent.mt < 25, and 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 (fig. S1, c and e).
Two independent methods were used to detect and remove doublets. DoubletFinder v2.0 (51) 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 (fig. S2a). Given that a small percentage of cells were demarcated as doublets by both methods, we removed all cells that were flagged as doublet by both or either approach, leading to the removal of 3,770 cells (fig. S2, b and c).
Following doublet removal, the raw data for the remaining 73,235 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 (fig. S2c and fig. S3a). 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 (fig.S3, b and c).
Garnett was used for initial cell classification as previously described (9). In brief, a cell type marker file (table S53) with 17 different cell types was compiled using various resources (18-22, 28), 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, 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, 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) (fig. S3e).
To overcome the apparent limitation in grouping the major canonical cell types, we employed the analytical workflow termed 'TooManyCells' (10), which implements an efficient divisive hierarchical spectral clustering approach along with tree visualizations to preserve the relationships among cell clusters at varying resolutions. In conjunction with 'TooManyCells' clustering, we invoked the cellular classifier Garnett (9), which annotates cell types by training a regression-based classifier from user-provided cell type signatures (table S53; (34)). 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'. TooManyCells enabled distinct clustering of all major known cell types associated with pancreatic islets, as confirmed by Garnett cell labels and the inspection of canonical gene markers for each cell type (Fig. 1C and fig. S4a).
The raw data from different cell types were then subsetted from the comprehensive clustering in Figure 1C in order to cluster cells on a cell-type basis. 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'. For the clustering of immune cells, 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 (http://metascape.org/gp/index.html#/main/step1; (52)), 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) (53) was run on a pre-ranked gene list using either user-provided pancreatic gene expression sets (22, 28) or standard hallmark gene signatures provided by the Molecular Signatures Database (MSigDB) (54).

CyTOF
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 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 (fig S15 a to c), endocrine and exocrine compartments were further subsetted leading to a refined analysis of 4,521,988 cells (Fig. 4A). 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.

IMC
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 10 m2, then dilated up to 7 pixels to approximate cell boundaries. Per-cell object mean pixel intensities were then exported for further analysis.

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 (Fig 6F). Quantile normalization of the raw counts was used in the clustering step. The resulting tree was pruned by collapsing nodes with 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). 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, here 20 for the complete pruned tree, the cell was assigned the neighbor label. Otherwise the cell was designated as distant.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Accession codes will be available before publication; https://hpap.pmacs.upenn.edu/. 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/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All high quality scRNA-seq libraries collected to date by HPAP were used in this study, which is comprised 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).
Data exclusions	Since all high quality scRNA-seq libraries collected to date by HPAP were used in this study, no samples were excluded.
Replication	These are human donor samples on a hard to harvest (and in some instances rare) population of cells, and therefore, replication is not plausible.
Randomization	Randomization was not carried out. The contribution of covariates on our findings was assessed (please see figure S5 and S6). Covariates were NOT found to affect cell clustering, and therefore, not affect our findings: autoantibody type, age, amylase levels, BMI, cold ischemic time, collection period, c-peptide levels, culture days, CyTOF purity, DCD or DBD, group (T1D, AAb^{+}, or control), HbA1c%, individual, lipase levels, ancestry, sex, viability, or warm ischemic time.
Blinding	Blinding was not carried out because the major approach used in this study, RNA-seq, is generally considered to be unbiased

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: Mouse anti-CK19 (Santa Cruz sc-6278), Rabbit anti-HLA-DR (Abcam ab92511)

Validation:

Adapted from the IMC paper in which this approach was described, Y. J. Wang 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 (2019): "20 of the 33 antibodies used in the current panel were directly purchased from FluidigmR (https://www.fluidigm.com). For the remaining 13 antibodies, carrier-free antibodies were purchased from different vendors and tested by immunofluorescent staining on FFPE sections. Based on anticipated epitope abundance, either pancreas or spleen sections were used for antibody testing. The staining results of antibodies were assessed by board-certified pathologists. Antibodies producing the expected results were conjugated with lanthanide metals using the Maxpar X8 metal conjugation kit following manufacturer's protocol (Fluidigm 201300). Post-conjugation, antibody specificities were again tested using immunofluorescent staining, followed by titration in the IMC platform. The antibody characterization procedure and results are shown in Figures S1A and S1B and Table S1. More details on antibody quality assessment can be found in Tables S2 and S3."

Adapted from the IMC paper in which this approach was described, Y.J. Wang et al. Single-Cell Mass Cytometry Analysis of the Human Endocrine Pancreas. Cell Metab. 24(4):616-626 (2016): "Antibodies were validated by the following three methods. (1) Immunofluorescent labeling (Figures S1A–S1E): the labeling efficiency of each antibody was tested by direct labeling of human islet cells in suspension. Briefly, human pancreatic cells were dissociated and labeled with the test antibodies following the same procedure as when performing mass cytometry. Subsequently, Cy3-conjugated secondary antibodies were applied. Cells were cytospun onto microscope slides and imaged under a fluorescent scope. Only those antibodies displaying the expected staining patterns were used in downstream experiments. (2) Flow cytometry (Figure S1F): cells were dissociated and labeled as in mass cytometry experiment, followed by secondary antibody staining in the Cy3 channel. Cellular events were subsequently acquired by a BD LSRII following a standard flow cytometry protocol. (3) Stimulation followed by CyTOF2 sample acquisition (Figure S1G): human islets were (a) incubated in retinoic acid at 50 nM final concentration for 72 hr; (b) serum starved overnight, followed by stimulation with 1.25 ng/ml Leptin for 4 hr; or (c) serum starved for 48 hr, followed by stimulation with Prolactin at 200 ng/ml for 30 min. After stimulation, cells were dissociated and processed following a normal mass cytometry sample preparation protocol. Antibodies that passed initial quality control were titrated in CyTOF with 1:100, 1:200, 1:500, 1:1,000, and 1:10,000 dilutions. Unlabeled cells were used as a negative control. Cytokeratin 19 Antibody (A53-B/A2): sc-6278. was validated by Western blot analysis of Cytokeratin 19 expression in MCF7 (A), MIA PaCa-2 (B) and MDA-MB-231 (C) whole cell lysates."

Human research participants

Policy information about studies involving human research participants

Population characteristics: All donors were screened for autoantibodies prior to organ harvest, and AAb positivity was confirmed again post tissue processing and islet isolation. The contribution of covariates on our findings was assessed (please see figure S5 and S6). Covariates were NOT found to affect cell clustering, and therefore, not affect our findings: autoantibody type, age, amylase levels, BMI, cold ischemic time, collection period, c-peptide levels, culture days, CyTOF purity, DCD or DBD, group (T1D, AAb $^{+}$, or control), HbA1c%, individual, lipase levels, ancestry, sex, viability, or warm ischemic time. Extensive data regarding population characteristics of the human research participant is provided in Table S1.

Recruitment: 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 again post tissue processing and islet isolation.

Ethics oversight: Pancreatic islets were procured by the HPAP consortium under 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).

Note that full information on the approval of the study protocol must also be provided in the manuscript.