

An integrated map of cell type-specific gene expression in pancreatic islets

Ruth M Elgamal¹, Parul Kudtarkar², Rebecca L Melton¹, Hannah M Mummey³, Paola Benaglio², Mei-Lin Okino¹, Kyle J Gaulton^{2,#}

1. Biomedical Sciences graduate program, University of California San Diego, La Jolla CA

2. Department of Pediatrics, University of California San Diego, La Jolla CA

3. Bioinformatics and Systems Biology graduate program, University of California San Diego, La Jolla CA

Corresponding author

Contact: kgaulton@health.ucsd.edu

Abstract

Pancreatic islets are comprised of multiple endocrine cell types that produce hormones required for glucose homeostasis, and islet dysfunction is a major factor in the development of type 1 and type 2 diabetes (T1D, T2D). Numerous studies have generated gene expression profiles in individual islet cell types using single cell assays. However, there is no canonical reference of gene expression in islet cell types in both health and disease that is also easily accessible for researchers to access, query, and use in bioinformatics pipelines. Here we present an integrated reference map of islet cell type-specific gene expression from 192,203 cells derived from single cell RNA-seq assays of 65 non-diabetic, T1D autoantibody positive (Aab+), T1D, and T2D donors from the Human Pancreas Analysis Program. We identified 10 endocrine and non-endocrine cell types as well as sub-populations of several cell types, and defined sets of marker genes for each cell type and sub-population. We tested for differential expression within each cell type in T1D Aab+, T1D, and T2D states, and identified 1,701 genes with significant changes in expression in any cell type. Most changes were observed in beta cells in T1D, and, by comparison, there were almost no genes with changes in T1D Aab+. To facilitate user interaction with this reference, we provide the data using several single cell visualization and reference mapping tools as well as open-access analytical pipelines used to create this reference. The results will serve as a valuable resource to investigators studying islet biology and diabetes.

Introduction

The islets of Langerhans in the pancreas are clusters of endocrine cells including alpha, beta, delta, and gamma cell types which each produce hormones that regulate blood glucose levels (1). Dysfunction of beta cells is one of the major pathologies of both type 1 and type 2 diabetes, which collectively affect over 500 million individuals worldwide (2,3). Other cell types in the microenvironment around islets also contribute to the modulation of islet function and diabetes risk such as endothelial and immune cells (4,5). The regulation of gene activity establishes the identity of specific cell types as well as changes in response to environmental stimuli and disease states, and gene activity can be measured by sequence-based expression profiling (6). Understanding the gene expression profiles of islet cell types can therefore provide insight into their function and can also reveal how cells are altered in diabetes.

Single cell technologies enable profiling the expression levels of genes in individual cells, which can then be used to define the gene regulatory profiles of specific cell types (7,8). Numerous studies have assayed gene expression in individual islet cells using single cell techniques (9–13). These studies have defined gene expression profiles of endocrine and non-endocrine cell types in the pancreas, heterogeneous sub-populations of cells representing cellular states within cell types, and changes in disease states including T1D and T2D. A caveat to these studies is that they have been performed using limited sample numbers, and in some cases limited cell numbers, and there has been inconsistency in the results across studies particularly when describing heterogeneity and changes in disease (14,15). In addition, the results of islet single cell studies are not often made easily accessible to researchers, particularly those that are not experts in single cell data analysis, to visualize and query the expression of a gene in each cell type or changes in cell type expression in disease.

The Human Pancreas Analysis Program (HPAP) was developed to comprehensively collect and profile pancreatic islet tissue from human donors to understand the pathogenesis of T1D and T2D (16,17). The data generated by HPAP for each donor includes bulk and single cell RNA-seq and ATAC-seq data as well as histology, genotyping, cellular phenotyping, and other data types. Data generated by HPAP are made freely available to researchers via a web portal PANC-DB (<https://hpap.pmacs.upenn.edu/>) where the raw sequence files from each individual donor can be directly downloaded (16). The rich set of raw islet donor data provided by this resource can then

be used by researchers to create integrated resources which make HPAP accessible to the wider community studying islet biology and diabetes to develop testable hypotheses.

In this study we created a reference map of gene expression in pancreatic islet cell types using single cell RNA-seq data from 65 donors available in HPAP. Using this reference map, we created several additional resources including (i) marker gene lists for every islet cell type and sub-population, (ii) normalized expression levels of genes in each islet cell type, and (iii) changes in gene expression in T1D, T1D autoantibody positive (Aab+), and T2D states in each islet cell type. We host these data in several interactive applications to enable researchers to visualize, query, and analyze this reference. Finally, we provide open-source analytical pipelines used to create the reference map and annotations. These resources are available at www.isletgenomics.org.

Results

Reference map of single cell expression in islets

We downloaded single cell RNA-seq (scRNA-seq) data from 67 donors available in PANC-DB (**Supplementary Table 1**). After pre-filtering of barcodes for each sample based on >500 expressed genes, we excluded two samples which had lower average expressed genes per cell. With the remaining 65 samples, we performed processing and clustering using a custom pipeline. In brief, this pipeline consists of ambient RNA background correction, dimension reduction of log normalized counts, batch correction, Leiden clustering, and post-clustering doublet removal. (**see Methods** for more detail). The resulting map had 192,203 cells which mapped to 14 distinct clusters (**Figure 1A**). In the final map, on average, samples had 2,957 cells with 16,908 unique molecular identifiers (UMI) per cell. Clusters were broadly consistent across samples, and no clusters were preferentially represented by a small number of samples (**Supplementary Figure 1**). We also observed little evidence for residual batch effects in the clusters driven by donor or other variables (**Figure 1B**).

We next annotated the identity of clusters using a curated set of well-established cell type and marker genes (**Supplementary Table 2**). This revealed 10 total cell types including endocrine alpha (*GCG*), beta (*INS*), delta (*SST*), and gamma (*PPY*) cells as well as non-endocrine acinar (*REG1A*), ductal (*CFTR*), endothelial (*PLVAP*), stellate (*PDGFRA*), macrophage (*C1QA/B/C*), and mast cells (*KIT*, *CD69*) (**Figure 1C,D**). We also observed evidence for epsilon cells (*GHRL*)

within the gamma cell cluster (**Figure 1C**), so we labeled this cluster as gamma+epsilon cells. Although islet cell type identity can be annotated using a small number of marker genes, knowledge of a larger set of genes specifically expressed in each cell type can provide potential additional insight into what drives cell identity. We therefore identified genes in each cell type with highly specific expression relative to other cell types in the study (**see Methods**). There were 542 genes with highly selective expression in one or more cell types (**Supplementary Table 3**), including those with no known function in the cell type. For example, in addition to canonical markers *INS* and *IAPP*, we identified beta cell-specific genes with known function in beta cells such as *RBP4* (18), *NPY* (19), *HADH* (20,21), and *C1QL1* (22) as well as those with no known beta cell function to our knowledge such as *SHISAL2B*, which could be targeted in future studies.

We also identified several cell types with multiple distinct clusters (**Figure 1A**). In most cases, these clusters represent previously described cell sub-types or states; for example, we identified quiescent and activated states of stellate cells (23), a MUC5B+ sub-population of ductal cells (23), and a sub-population of ‘cycling’ alpha cells (24) (**Figure 1A,C**). We identified 141 genes highly specific to a cell sub-type when compared to other cells of that same cell type (**Supplementary Table 4, see Methods**). We also identified a cluster comprised of both alpha and beta cells, which did not appear to be doublets, as the cells expressing insulin and glucagon were largely distinct (**Supplementary Figure 2**). These cells may represent cellular states of alpha and beta cells, as has been observed in other studies (15,24,25); however, due to the ambiguity over what cell populations these cells exactly represent, we excluded the cluster from downstream analyses.

We next compared the proportions of each islet cell type across samples (**Figure 1E**). As the purity of the islet preparations in PANC-DB varies dramatically, we assessed the proportion of different endocrine cell types as a function of the total number of endocrine cells per sample. Among non-diabetic samples, there was substantial variability in the proportion of different islet cell types; for example, the proportion of beta cells in islets ranged from .19 to .71. When considering disease states, we observed decreased proportion of beta cells in islets in T1D compared to ND control as expected (avg. ND=.40, avg. T1D=.18; Wilcox $P=4.1 \times 10^{-4}$). For beta cells, we observed a slight decrease in T2D (avg. ND=.40, avg. T2D=.365; Wilcox $P=.36$) and increase in T1D Aab+ (avg. ND=.40, avg. T2D=.49; Wilcox $P=.24$), although these estimates were not significant. While we also observed increased alpha cell proportion in T1D and T2D, this is likely explained by the relative decrease in beta cells.

We finally annotated the repertoire of genes expressed in each islet cell type in the non-diabetic state. For each cell type, we aggregated reads from all cells in non-diabetic samples, calculated normalized transcripts-per-million (TPM) expression levels from the 'pseudo'-bulk counts for each gene, and defined genes expressed in a cell type (average TPM>1). There were between 7.6k-11.7k genes expressed per cell type in non-diabetic samples, with acinar cells (7.6k) and beta cells (8.2k) and having the smallest number of expressed genes. We next calculated gene expression levels for samples separately for each disease state. There were between 8.9k-11.9k genes expressed per cell type in T1D, 7.6k-11.7k genes expressed per cell type in T2D, and 7.4k-11.5k per cell type in T1D Aab+. Interestingly, for almost every cell type, we observed a larger number of expressed genes in both T1D and T2D compared to non-diabetic samples.

Changes in islet cell type-specific gene expression in T1D and T2D

Identifying genes with changes in cell type activity in disease and pre-disease states can provide insight into disease pathogenesis. We therefore next determined changes in cell type gene expression in T2D (n=17), T1D (n=10), and T1D autoantibody positive (Aab+) non-diabetic (n=11) compared to non-diabetic control (n=27). We tested for differential expression of genes in a cell type from 'pseudo'-bulk profiles across samples accounting for donor and technical characteristics as well as ambient background RNA signal (**see Methods**). In total, across all conditions, we identified 1,701 genes with significant changes in expression (FDR<.10) in at least one cell type (**Figure 2A**).

For T2D, there were 84 genes with significant changes in any cell type compared to non-diabetic (ND) (**Supplementary Table 5**). Most of these genes were significant in beta cells (n=79), although we also observed a few differential genes in acinar, alpha, and endothelial cells. In beta cells, the genes with the largest increases in expression in T2D included *TSHR* which is the receptor for thyroid stimulating hormone, *SLC4A4* which is a bicarbonate co-transporter, and *TNFRSF11B* which is a cytokine receptor for tumor necrosis factor (TNF) family proteins. Genes with up-regulated expression in beta cells were enriched (FDR<.20) for biological processes related to protein modification, kinase activity, and ion binding, and by comparison down-regulated genes were enriched for processes related to protein transport and Golgi apparatus.

For T1D, we identified 1,808 genes with significant changes in expression compared to control (**Supplementary Table 6**). We observed the largest number of gene expression changes in beta

cells (n=1,305), although there were also significant changes in cycling alpha cells (n=260), alpha cells (n=35), acinar cells (n=14), mast cells (n=22) and multiple other cell types. In beta cells, the genes with the largest increases in expression in T1D included multiple MHC class I genes (e.g. *HLA-A*, *HLA-B*) as well as genes involved in MHC processing and presentation such as *B2M* and *CD74* (**Figure 2B**). Genes up-regulated in beta cells were broadly enriched (FDR<.20) for biological processes related to MHC antigen processing and presentation ($P=2.21 \times 10^{-17}$), cytokine signaling ($P=5.77 \times 10^{-14}$) and differentiation ($P=7.30 \times 10^{-15}$), whereas down-regulated genes were enriched for processes related to stimulus response ($P=5.19 \times 10^{-18}$), response to ER stress ($P=5.89 \times 10^{-9}$), mitochondrial organization ($P=1.02 \times 10^{-8}$), response to metal ions ($P=2.67 \times 10^{-8}$), and apoptotic signaling ($P=4.15 \times 10^{-8}$) (**Figure 2C**). In addition to MHC class I genes, we observed up-regulation of genes mediating cytokine signaling responses such as *IRF1/2*, *STAT1/4*, and *NFKB1*, genes involved in cell survival such as *BCL6*, and the beta cell autoantigen *GAD1* (**Figure 2D**).

Finally, among individuals positive for T1D autoantibodies (Aab+) we identified very few genes with significant changes in expression compared to ND donors with no autoantibodies in any cell type (**Figure 2A**). While there was some evidence for up-regulation of MHC class I genes, we observed no clear change in expression of cytokine responsive factors such as IRFs, STATs and *NFKB1* (**Figure 2D**). By comparison, a recent study using largely the same set of samples from HPAP reported substantial changes in expression in beta cells in T1D Aab+ (28). Given the dramatic differences in the changes in expression between studies, more careful consideration of analytical approaches as well as larger sample sizes will be needed to reconcile these differences.

Reference map and resource availability

The integrated map of gene expression in islet cell types generated by this study can be used to understand gene activity in physiological and disease states. In addition, this map can be used as part of bioinformatics pipelines, for example to perform reference mapping of new single cell RNA-seq datasets. We therefore provide several resources and interactive applications to facilitate the wide use of this integrated map in a variety of downstream analyses. These resources are all available at: <http://www.isletgenomics.org>.

First, we provide the islet cell type expression map in two interactive single cell browsers, CELLxGENE (29) and ShinyCell (30), which enable visualizing patterns across individual cells for

example cell type identity, variables such as donor or library, technical factors such as number of features or percent mitochondrial reads, or the expression level of selected genes (**Supplementary Figure 3**). In addition, we provide the expression map within Azimuth (31), which can be used for rapid on-the-fly reference mapping of new datasets (**Supplementary Figure 4**).

Second, we provide annotations of activity in each islet cell type including marker genes, normalized gene expression levels, and changes in gene expression in T1D, T1D Aab+ and T2D. We developed several interactive applications that enable users to select specific genes to view expression levels in each cell type, as well as to visualize changes in cell type expression in different disease states (**Supplementary Figure 5**).

Finally, the analytical pipelines used for data processing and clustering, defining expressed genes in each cell type, and defining differentially expressed genes are provided open access.

Discussion

Maps of gene expression levels in individual cell types within a heterogeneous tissue are valuable tools for hypothesis generation to understand cell type function and identity, gene activity, and changes in disease. In addition, these maps can be used for reference mapping of single cell RNA-seq datasets to facilitate annotation of cell identity and perform integrated analyses (31–33). While repositories such as PANC-DB provide access to a rich resource of raw sequence data and phenotypic information on human islet donors generated by HPAP (16), drawing insight from these data is a major challenge to researchers without single cell data analysis expertise. Our study provides an integrated map of gene expression profiles in islet cell types and changes in disease derived from the single cell RNA-seq experiments in HPAP, which will help enable downstream analyses and hypothesis generation for many non-single cell expert investigators.

There are several areas where the map can be further improved in future iterations. First, we were unable to separate a population of epsilon cells, likely due both to the rarity of epsilon cells and the sparsity of single cell RNA-seq profiles. By comparison, several studies profiling islets using different single cell technology resolved small epsilon cell populations (9,34). We also did not identify other rare cell types in the pancreas such as Schwann cells or lymphoid cell types. The samples profiled by HPAP are purified islets where other cell types outside of the islet

microenvironment are preferentially removed, and therefore the profiles of non-islet cell types are under-represented. Furthermore, the repertoire of discrete states that exist within each cell type, as well as any sub-types for example with distinct spatial localizations, remains to be resolved. Continued profiling of donors and cells from both purified islets and whole pancreas will help to define profiles for all pancreatic cell type and sub-types. Finally, even after accounting for ambient background RNA there is still residual expression of genes in off-target cell types, particularly for highly expressed genes in common cell types, where improvements in background correction methods for single cell RNA-seq are needed to estimate islet cell type-specific expression more accurately.

Genes with significant changes in cell type-specific expression in T1D and T2D provide insight into disease pathogenesis, for example up-regulation of MHC class I genes and cytokine signaling processes in T1D and differentiation related processes in T2D. There were few changes in gene expression in T1D Aab+ where we identified almost no significant genes. By comparison, a recent study using largely the same set of HPAP samples found marked changes in beta cell expression profiles in T1D Aab+ (28). Given substantial heterogeneity between donors and in disease processes, much larger sample sizes will be needed to determine the true extent to which gene profiles change in different disease and physiological states. Larger sample sizes will also enable finer grained partitioning of samples to understand gene regulatory differences between phenotypic sub-groupings. For example, many Aab+ samples in HPAP are single GAD+ (28), yet there is great diversity in the number and type of autoantibodies that individuals can have with different rates of progression to T1D (35,36). Similarly, profiling islets from impaired glucose tolerant donors will help understand changes that occur during progression to T2D (37).

In summary, our map of islet cell type-specific expression and associated resources of cell type-specific gene activity in physiological and disease states provided by this study will be a valuable reference to the islet and diabetes research community.

Methods

Data Availability

Organ procurement and processing was performed by the Human Pancreas Analysis Program (HPAP) as previously described (16). Single cell RNA-sequencing (scRNA-seq) data from isolated and dissociated pancreatic islets were made publicly available by HPAP and raw fastq

files for experiments from 67 donors (10 T1D, 17 T2D, 29 ND, 11 ND but AAB+) were downloaded from the PANC-DB data portal. Cell Ranger 6.0.1 (10x Genomics) software was used to perform alignment to the GRCh38 human reference genome and generate count matrices.

Preliminary filtering

Barcodes were filtered for a minimum of 500 expressed genes per cell and less than 15% mitochondrial reads. Two samples (HPAP-027 and HPAP-093) were removed since the mean number of expressed genes per cell after this filtering step was markedly lower than for other samples (<1000).

Ambient RNA correction

Ambient RNA removal was performed to account for extracellular RNA contamination that may get trapped in a droplet during library generation. SoupX 1.6.1 (38) was used on raw feature barcode matrices for ambient RNA removal on the remaining 65 samples using the automated contamination fraction estimation method. Raw count values for each sample were corrected using the SoupX contamination estimates and the round to integer feature, ensuring resulting counts remain integers for use in downstream analyses (39).

Data processing and clustering

The SoupX-corrected count matrices were merged and log normalized with a scale factor of 1000. The variance stabilizing transformation (vst) method was used to find the 2000 most variable features. Data was scaled and principal component analysis was performed with 20 principal components using Seurat 4.2.0 (31). Harmony 0.1.1 (40) was used for batch correction using donor, 10x Genomics assay chemistry (10x 3' v2 or 10 3' v3), and tissue source (nPOD or UPenn) as covariates. Uniform manifold approximation and projection (UMAP) and neighbors were calculated using the reduction from Harmony. Clustering was performed in Seurat 4.2.0 using the Leiden algorithm at a resolution of 0.5.

Post-clustering doublet removal

Scrublet 0.2.3 (41) was used to identify doublets with the default parameters (expected doublet rate of 6%, minimum counts of 2, minimum cells of 3, minimum gene variability percentile of 85%, and 30 principal components). For each sample, RNA count matrices were extracted, saved in MatrixMarket format, and input into Scrublet with default parameters. There were 4,382 barcodes flagged as doublets, and we removed these barcodes from the merged Seurat object and re-

performed Harmony integration and clustering (resolution 0.3) with the remaining barcodes as described above. We further curated a set of cell type-specific marker genes, and clusters that contained marker genes for two or more different cell types were further sub-clustered using the Leiden algorithm at resolutions of 0.15-0.25. Any sub-clusters expressing multiple cell type marker genes were presumed to be residual doublets, and we manually removed these sub-clusters representing a total of 13,036 barcodes. We then re-performed Harmony integration and clustering using the final set of barcodes as described above.

Cell type specific marker genes

While there are well established marker genes for many pancreatic islet cell types, we performed an unbiased analysis of cluster-specific marker genes using the FindAllMarkers function in Seurat 4.2.0 (31). For each cluster, a Wilcoxon rank sum test was run on log-normalized counts in comparison to the remaining clusters. Genes were considered markers if they were expressed in more than 25% of cells in the cell type and less than 25% of other cells, had log2 fold change threshold greater than 1, and had adjusted P-value less than .05. For cell types with multiple clusters, we used the FindMarkers function to identify genes with cluster-specific activity relative to other cells in the same cell type.

Cell type gene expression profiles

We aggregated reads from cells in each cell type and created 'pseudo'-bulk counts from contamination-corrected RNA counts. We calculated transcripts per million (TPMs) for each donor in each cell type using GENCODE v38 GRCh38.p13 (42) gene size annotations. Differential gene expression analyses were performed using DESeq 1.34.0 (43), comparing cell type profiles between ND controls and T1D Aab+, T1D, and T2D. Sex, scaled age, scaled BMI, 10x kit chemistry, and tissue procurement source were included as covariates. Genes were only tested for a cell type if at least half of the samples per tested condition had at least 5 counts. Multiple test correction was performed using the Benjamini-Hochberg false detection rate correction at an alpha of 10%. We performed gene set enrichment analyses (GSEA) (44) for up- and down-regulated genes in beta cells using Gene Ontology, KEGG pathway, and REACTOME terms. For conditions with less than 100 genes differentially expressed at FDR<.10, we used an uncorrected p-value threshold of .001 as input to GSEA. From the results of GSEA we considered terms significant at FDR<.20 and which also contained less than 1000 genes.

Data availability

The raw sequence data is available on the PANC-DB website. Processed files and derived annotations generated by this study are available at isletgenomics.org. Custom code is available at <https://github.com/Gaulton-Lab/HPAP-scRNA-seq>.

Author contributions

K.J.G. conceived of the study and obtained funding. K.J.G. and R.M.E. wrote the manuscript. R.M.E. performed all single cell data analyses. R.M.E., R.L.M., H.M.M., P.B., and M.L.O. developed analytical pipelines for single cell data processing and analysis. P.K. developed and implemented tools for data visualization and interaction.

Acknowledgements

This work was supported by the National Institutes of Health grant numbers DK-105554, DK-114650, and DK-120429 to KJG. This manuscript used data acquired from the Human Pancreas Analysis Program (HPAP-RRID:SCR_016202) Database, a Human Islet Research Network (RRID:SCR_014393) consortium (UC4-DK-112217, U01-DK-123594, UC4-DK-112232, and U01-DK-123716).

References

1. Da Silva Xavier G. The Cells of the Islets of Langerhans. *J Clin Med*. 2018 Mar 12;7(3):54.
2. Mobasser M, Shirmohammadi M, Amiri T, Vahed N, Hosseini Fard H, Ghojzadeh M. Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *Health Promot Perspect*. 2020;10(2):98–115.
3. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*. 2019 Nov;157:107843.
4. Aamodt KI, Powers AC. Signals in the pancreatic islet microenvironment influence β -cell proliferation. *Diabetes Obes Metab*. 2017 Sep;19 Suppl 1(Suppl 1):124–36.
5. Brissova M, Aamodt K, Brahmachary P, Prasad N, Hong JY, Dai C, et al. Islet microenvironment, modulated by vascular endothelial growth factor-A signaling, promotes β cell regeneration. *Cell Metab*. 2014 Mar 4;19(3):498–511.
6. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009 Jan;10(1):57–63.

7. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* 2017 Aug 18;9(1):75.
8. Kanter I, Kalisky T. Single cell transcriptomics: methods and applications. *Front Oncol.* 2015;5:53.
9. Segerstolpe Å, Palasantza A, Eliasson P, Andersson EM, Andréasson AC, Sun X, et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* 2016 Oct 11;24(4):593–607.
10. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* 2016 Oct 26;3(4):346-360.e4.
11. Camunas-Soler J, Dai XQ, Hang Y, Bautista A, Lyon J, Suzuki K, et al. Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. *Cell Metab.* 2020 May 5;31(5):1017-1031.e4.
12. Li J, Klughammer J, Farlik M, Penz T, Spittler A, Barbieux C, et al. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* 2016 Feb;17(2):178–87.
13. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, Sivakamasundari V, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* 2017 Feb;27(2):208–22.
14. Wang YJ, Kaestner KH. Single-Cell RNA-Seq of the Pancreatic Islets--a Promise Not yet Fulfilled? *Cell Metab.* 2019 Mar 5;29(3):539–44.
15. Mawla AM, Huising MO. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. *Diabetes.* 2019 Jul;68(7):1380–93.
16. Kaestner KH, Powers AC, Naji A, HPAP Consortium, Atkinson MA. NIH Initiative to Improve Understanding of the Pancreas, Islet, and Autoimmunity in Type 1 Diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes.* 2019 Jul;68(7):1394–402.
17. Shapira SN, Naji A, Atkinson MA, Powers AC, Kaestner KH. Understanding islet dysfunction in type 2 diabetes through multidimensional pancreatic phenotyping: The Human Pancreas Analysis Program. *Cell Metab.* 2022 Dec 6;34(12):1906–13.
18. Huang R, Bai X, Li X, Zhao L, Xia M. Retinol Binding Protein 4 Impairs Pancreatic Beta-Cell Function, Leading to the Development of Type 2 Diabetes. *Diabetes.* 2018 Jul 1;67(Supplement_1):1826-P.
19. Yang CH, Onda DA, Oakhill JS, Scott JW, Galic S, Loh K. Regulation of Pancreatic β -Cell Function by the NPY System. *Endocrinology.* 2021 Aug 1;162(8):bqab070.
20. Martens GA, Vervoort A, Van de Casteele M, Stangé G, Hellemans K, Van Thi HV, et al. Specificity in beta cell expression of L-3-hydroxyacyl-CoA dehydrogenase, short chain, and

- potential role in down-regulating insulin release. *J Biol Chem*. 2007 Jul 20;282(29):21134–44.
21. Hardy OT, Hohmeier HE, Becker TC, Manduchi E, Doliba NM, Gupta RK, et al. Functional genomics of the beta-cell: short-chain 3-hydroxyacyl-coenzyme A dehydrogenase regulates insulin secretion independent of K⁺ currents. *Mol Endocrinol*. 2007 Mar;21(3):765–73.
 22. Gupta R, Nguyen DC, Schaid MD, Lei X, Balamurugan AN, Wong GW, et al. Complement 1q-like-3 protein inhibits insulin secretion from pancreatic β -cells via the cell adhesion G protein-coupled receptor BAI3. *J Biol Chem*. 2018 Nov 23;293(47):18086–98.
 23. Tosti L, Hang Y, Debnath O, Tiesmeyer S, Trefzer T, Steiger K, et al. Single-Nucleus and In Situ RNA-Sequencing Reveal Cell Topographies in the Human Pancreas. *Gastroenterology*. 2021 Mar;160(4):1330-1344.e11.
 24. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, et al. Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *Nat Genet*. 2021 Apr;53(4):455–66.
 25. Xin Y, Dominguez Gutierrez G, Okamoto H, Kim J, Lee AH, Adler C, et al. Pseudotime Ordering of Single Human β -Cells Reveals States of Insulin Production and Unfolded Protein Response. *Diabetes*. 2018;67(9):1783–94.
 26. Albanus RD, Tang X, Taylor HJ, Manickam N, Erdos M, Narisu N, et al. Single-cell gene expression and chromatin accessibility profiling of human pancreatic islets at basal and stimulatory conditions nominates mechanisms of type 1 diabetes genetic risk [Internet]. *Genomics*; 2022 Nov [cited 2023 Jan 22]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2022.11.12.516291>
 27. Lu TTH, Heyne S, Dror E, Casas E, Leonhardt L, Boenke T, et al. The Polycomb-Dependent Epigenome Controls β Cell Dysfunction, Dedifferentiation, and Diabetes. *Cell Metab*. 2018 Jun 5;27(6):1294-1308.e7.
 28. Fasolino M, Schwartz GW, Patil AR, Mongia A, Golson ML, Wang YJ, et al. Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. *Nat Metab*. 2022 Feb;4(2):284–99.
 29. Cellxgene [Internet]. Available from: <https://chanzuckerberg.github.io/cellxgene/>
 30. Ouyang JF, Kamaraj US, Cao EY, Rackham OJL. ShinyCell: simple and sharable visualization of single-cell gene expression data. Mathelier A, editor. *Bioinformatics*. 2021 Oct 11;37(19):3374–6.
 31. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021 Jun;184(13):3573-3587.e29.
 32. Kang JB, Nathan A, Millard N, Rumker L, Moody DB, Korsunsky I, et al. Efficient and precise single-cell reference atlas mapping with Symphony [Internet]. *Bioinformatics*; 2020 Nov [cited 2021 Mar 13]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.11.18.389189>

33. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Büttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. *Nat Biotechnol*. 2022 Jan;40(1):121–30.
34. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst*. 2016 Oct 26;3(4):385-394.e3.
35. Jacobsen LM, Bocchino L, Evans-Molina C, DiMeglio L, Goland R, Wilson DM, et al. The risk of progression to type 1 diabetes is highly variable in individuals with multiple autoantibodies following screening. *Diabetologia*. 2020 Mar;63(3):588–96.
36. Yu L, Zhao Z, Steck AK. T1D Autoantibodies: room for improvement? *Curr Opin Endocrinol Diabetes Obes*. 2017 Aug;24(4):285–91.
37. Wang G, Chiou J, Zeng C, Miller M, Matta I, Han JY, et al. Integration of single-cell multiomic measurements across disease states with genetics identifies mechanisms of beta cell dysfunction in type 2 diabetes [Internet]. *Systems Biology*; 2023 Jan [cited 2023 Jan 26]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2022.12.31.522386>
38. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience*. 2020 Dec 26;9(12):giaa151.
39. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018 Jun;36(5):411–20.
40. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16(12):1289–96.
41. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems*. 2019 Apr;8(4):281-291.e9.
42. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012 Sep;22(9):1760–74.
43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
44. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005 Oct 25;102(43):15545–50.

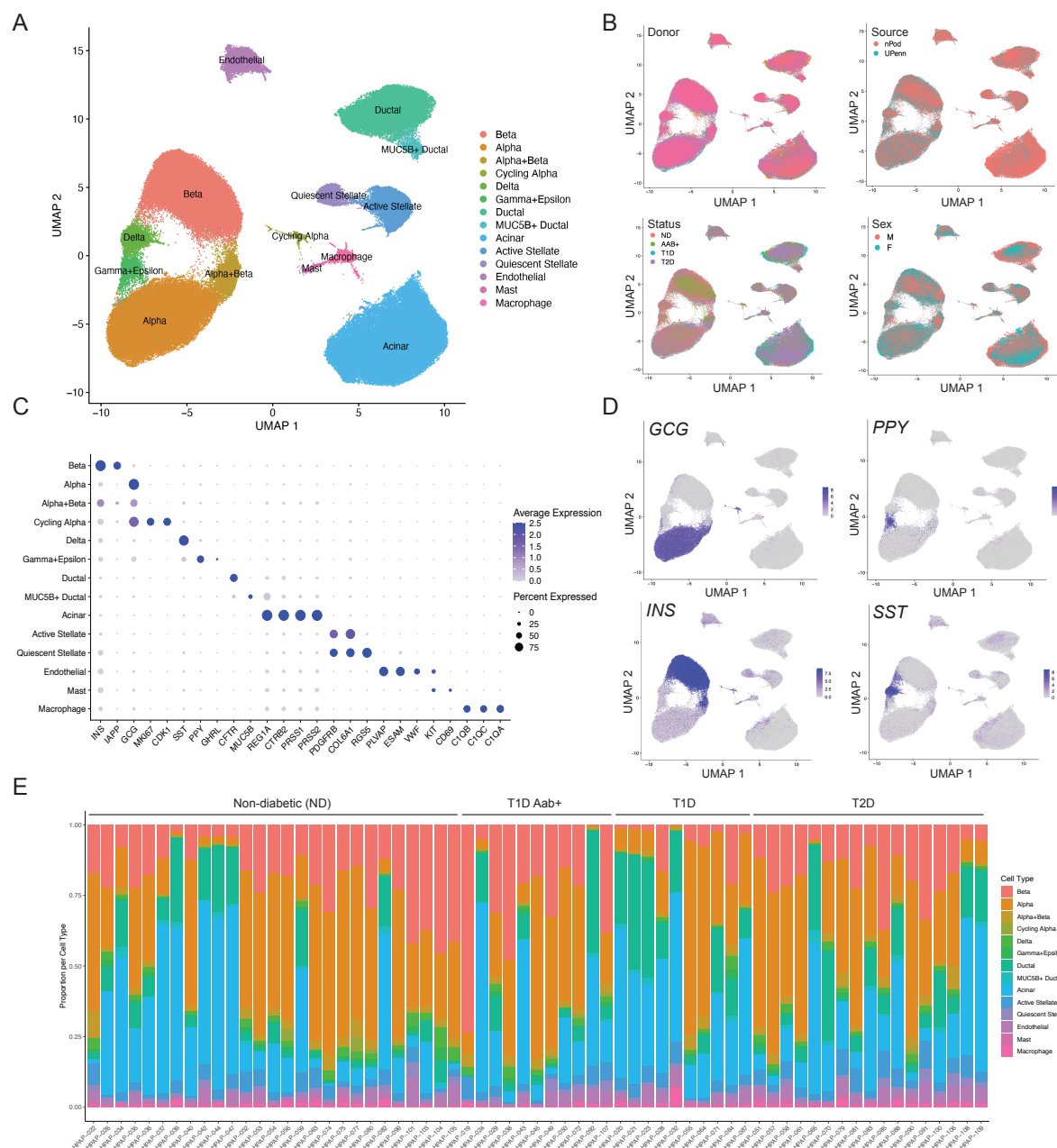


Figure 1. Map of gene expression in pancreatic islet cell types. (A) UMAP plot showing clustering of single cell RNA-seq profiles from 192,203 cells from HPAP donors in the PANC-DB website. Clusters are labeled based on cell type and sub-type identity using known marker genes. (B) Cells labeled based on variables such as donor, sex, disease status, and tissue source. (C) Dot plot showing normalized expression level and percent expressing cells for selected marker genes in each cluster. (D) Cells labeled with expression level of islet cell type hormones insulin (*INS*), glucagon (*GCG*), somatostatin (*SST*), and pancreatic polypeptide (*PPY*). (E) Proportion of cells generated from each donor from each cell type, grouped by disease state.

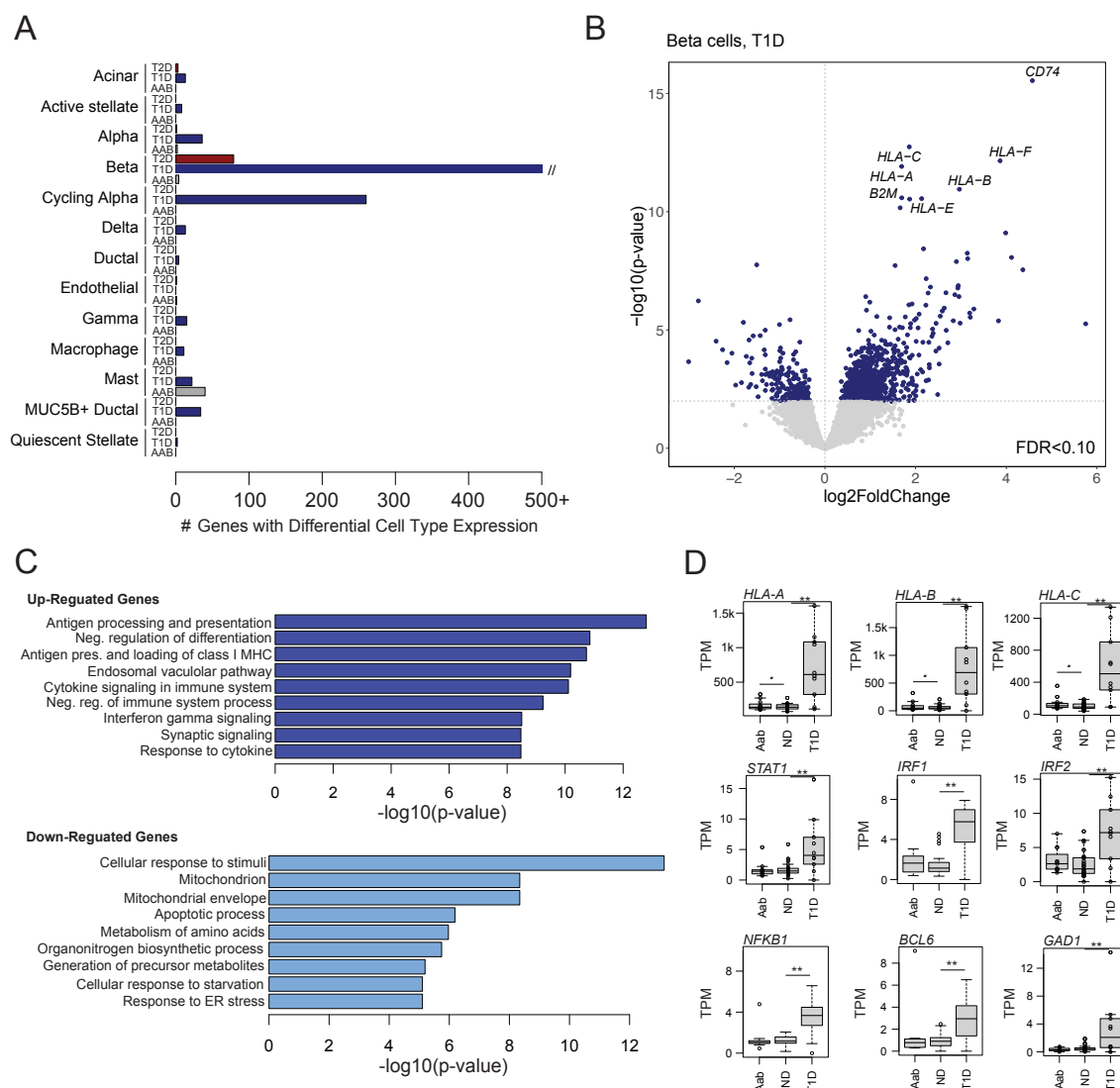
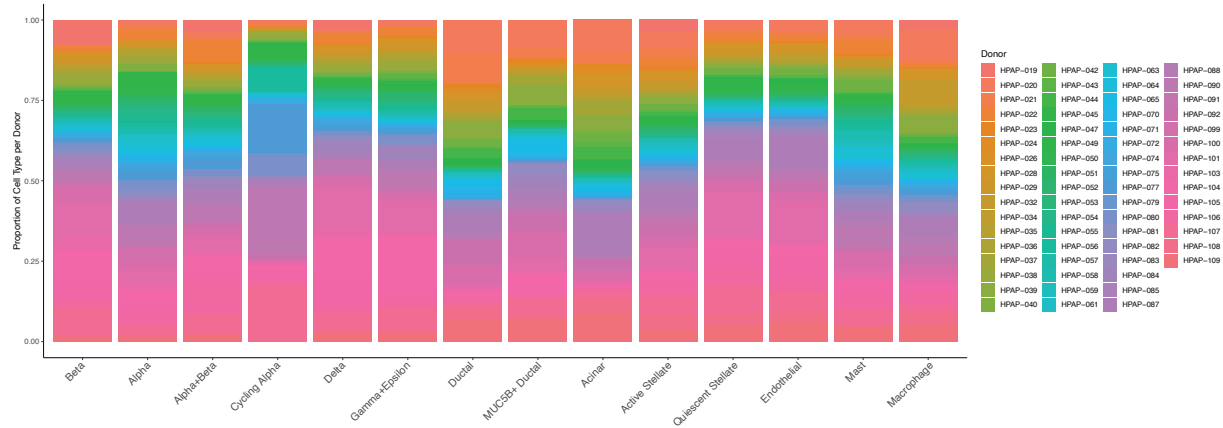
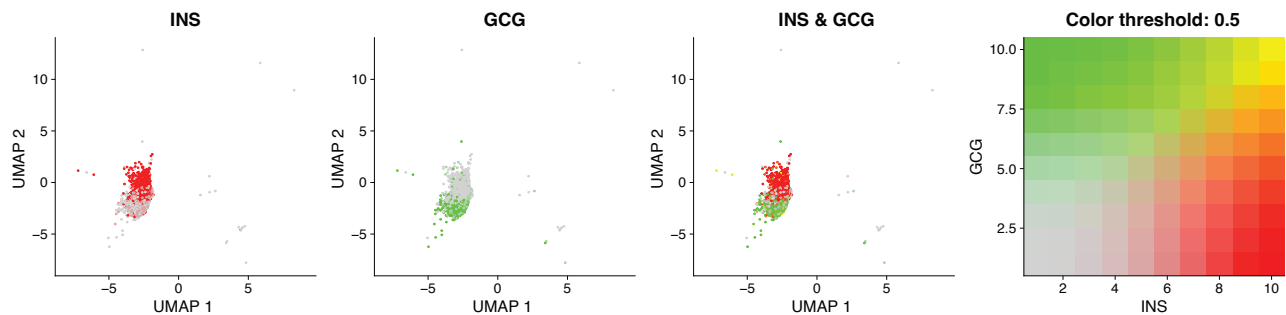


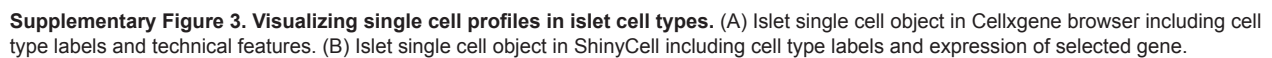
Figure 2. Changes in pancreatic islet cell type expression in disease states. (A) Number of genes in each cell type with significant changes (FDR<.10) in expression in T1D, T2D, and T1D autoantibody positive (AAB) compared to non-diabetic (ND) donors using DESeq2. The x-axis is capped at 500 genes for legibility, although the number of genes significant in beta cells in T1D is much higher. (B) Differential gene expression in beta cells in T1D. Genes with up-regulated expression are on the right side of the dashed line, and genes with down-regulated expression are on the left side. The hard line represents the p-value threshold corresponding to FDR<.10. Genes with the most significant increase in expression are labelled. (C) Gene sets enriched among genes with up-regulated and down-regulated beta cell expression in T1D. (D) Normalized beta cell expression level of HLA class I and selected stress response genes across ND, AAB and T1D donors. **FDR<.10, *uncorrected p<.10

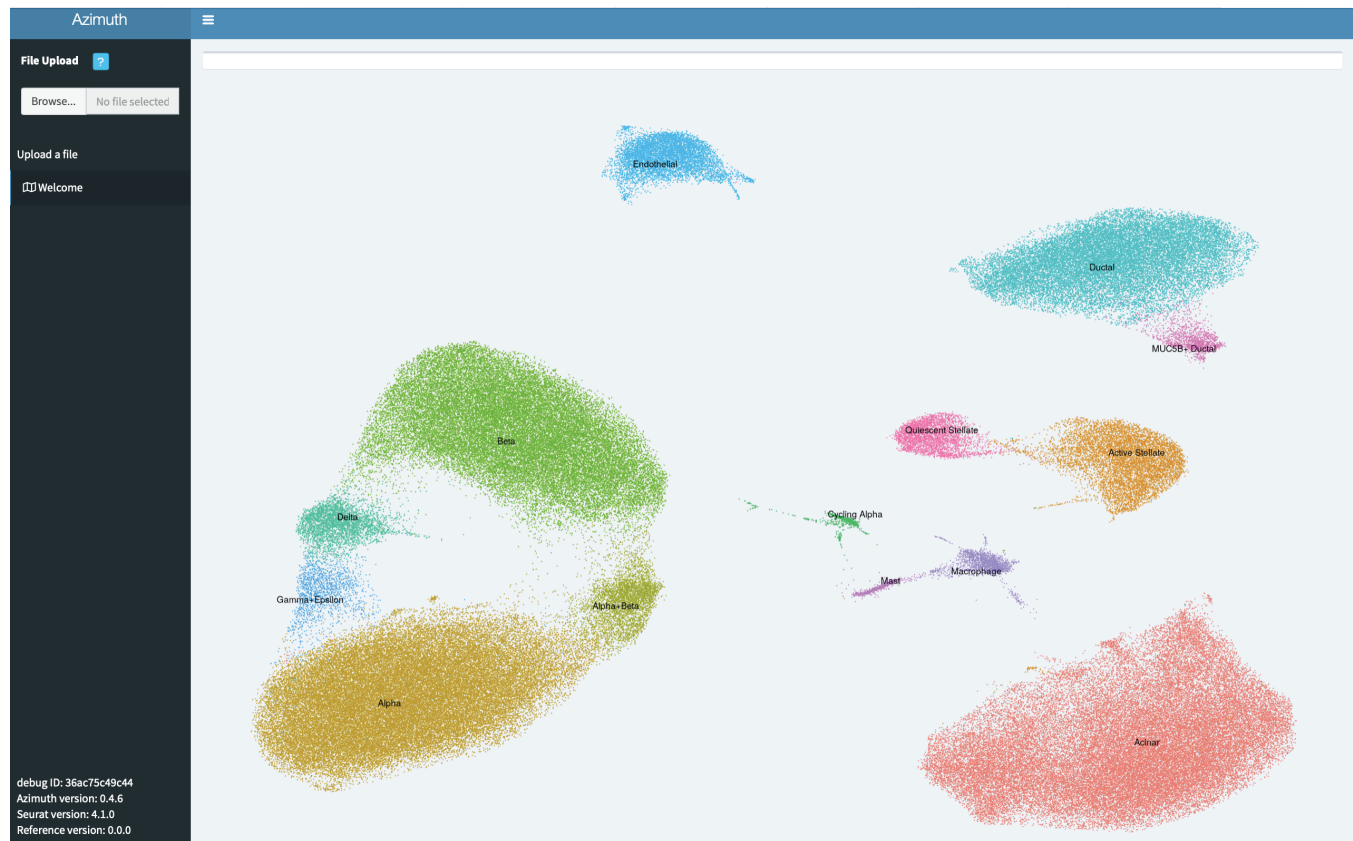


Supplementary Figure 1. Sample representation per cluster. Proportion of cells in each cluster from all donor samples in HPAP. No clusters have representation from only a small number of donors.



Supplementary Figure 2. Marker gene expression in alpha+beta cluster. Expression level of insulin and glucagon separately, as well as overlaid together, in the alpha+beta cluster. The insulin and glucagon expressing cells within this cluster appear largely distinct.





Supplementary Figure 4. Reference mapping of islet single cell gene expression. Screenshot of islet cell type expression in the Azimuth application which enables users to upload their own data and perform annotation of cell type identity.

A

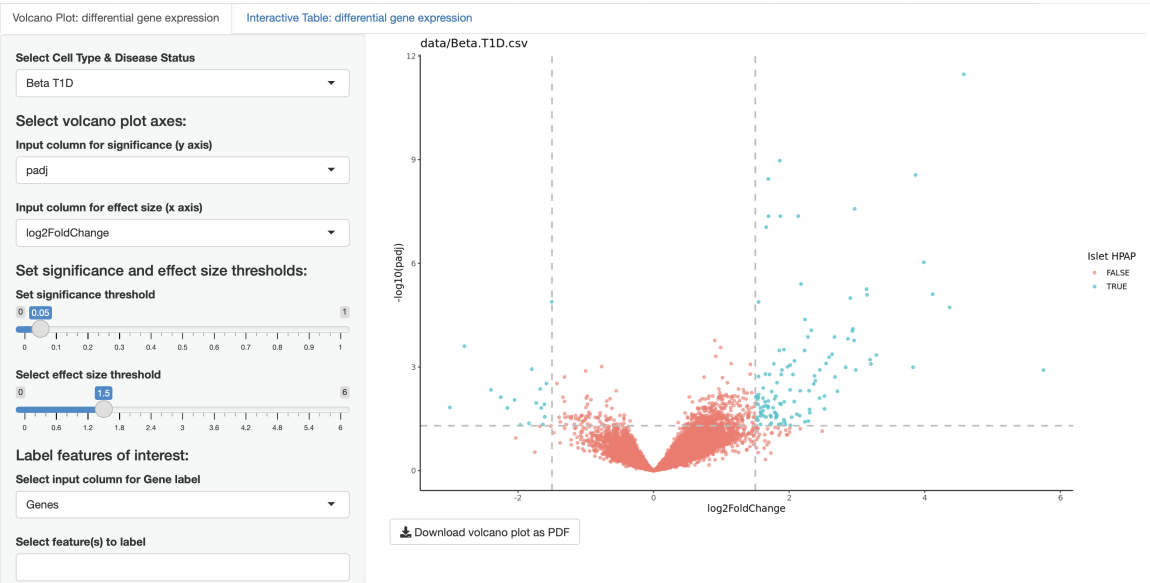
Gene expression by diabetes status Gene expression by diabetes status & celltype Gene expression by samples Gene expression by samples & celltype Gene expression

Show 25 entries

Gene	Acinar	Alpha	Alpha.Beta	Beta	Delta	Ductal	Endothelial	Gamma.Epsilon	Macrophage	Mast	Stellate
ATBG	3.212641e+00	4.675087e+01	54.81347390	2.847356e+01	33.68548633	1.096773e+00	6.28970261	45.46132081	4.548228e+01	42.3763157	4.586925e+01
ATBG-AS1	2.320169e-01	3.400946e+00	1.36038210	1.841105e+00	2.05322227	1.244174e-01	1.04563459	3.54291173	5.364167e+00	1.3527215	4.622447e+00
ATCF	4.828836e-01	4.863233e+00	1.07596670	4.115999e-01	1.95889522	1.630135e-01	0.21819035	3.46709175	3.037363e-01	0.1312187	1.402195e-01
A2M	1.808825e-01	3.933402e-01	3.06446338	6.079455e-02	0.01745652	1.525629e-01	90.65869108	0.05574615	1.025682e+02	5.0701892	1.095499e+02
A2M-AS1	1.879183e-01	2.791529e-01	0.10145575	1.444395e-01	0.39866154	3.725872e-01	1.19519589	0.28253513	2.292314e+00	6.9657428	6.839012e-01
A2ML1	2.405591e-03	1.549727e-02	0.00000000	3.989762e-03	0.00000000	1.719600e-03	0.02021480	0.00000000	7.533070e-03	0.0000000	1.494020e-02
A2ML1-AS1	4.543127e-03	4.980638e-02	0.00000000	1.541975e-02	0.10896336	3.179456e-02	0.05003053	0.29445625	0.000000e+00	0.0000000	2.637038e-02
A2ML1-AS2	3.574426e-04	1.167073e-02	0.00000000	0.000000e+00	0.00000000	1.992948e-02	0.00000000	0.00000000	0.000000e+00	0.0000000	0.000000e+00
ATGALT2	2.697417e-04	2.966884e-03	0.00000000	1.668849e-03	0.00000000	0.000000e+00	0.05488237	0.00000000	1.835194e-01	0.0000000	5.348360e-02
ATGALT	1.476048e-01	8.363954e-01	0.50686921	1.355326e-01	0.66745200	1.195134e-01	83.85668240	2.18426302	3.063457e+00	0.6775548	2.593378e+01
ATGNT	0.000000e+00	3.551597e-03	0.00000000	3.067596e-04	0.00000000	5.401533e-02	0.00000000	0.00000000	0.000000e+00	0.0000000	3.154041e-03
AAAS	1.809350e+00	4.000987e+00	1.59150034	1.854512e+00	2.93225258	5.206393e+00	4.89089147	3.01089727	3.971746e+00	1.9164218	4.704595e+00
AACS	1.748550e+00	1.262717e+01	8.36942443	5.985727e+00	7.38081529	6.528281e+00	3.45588917	8.81303437	3.334672e+00	6.2701363	3.830804e+00
ADAC	2.330674e-01	2.674009e-01	0.31363282	1.063461e-01	0.32911268	7.674094e-01	0.19566886	0.12962847	1.380664e+00	0.0000000	1.672720e+00
ADACL2	3.856332e-04	8.978832e-04	0.00000000	0.000000e+00	0.00000000	0.000000e+00	0.00000000	0.00000000	0.000000e+00	0.0000000	0.000000e+00
ADACL2-AS1	0.000000e+00	0.000000e+00	0.00000000	0.000000e+00	0.00000000	5.780649e-02	0.00000000	0.00000000	0.000000e+00	0.0000000	0.000000e+00
ADACL3	0.000000e+00	0.000000e+00	0.00000000	0.000000e+00	0.00000000	5.175550e-04	0.00000000	0.00000000	0.000000e+00	0.0000000	0.000000e+00
ADACL4	0.000000e+00	1.496356e-03	0.00000000	5.380547e-04	0.00000000	5.772612e-04	0.00000000	0.00000000	0.000000e+00	0.0000000	1.002968e-02
ADAT	5.883759e-01	2.095760e+00	1.08894384	1.807596e+00	1.37009128	2.053139e+00	0.23176832	1.67013473	2.308009e+01	0.2230563	2.011294e+00
AGAB	4.113689e+00	1.496549e+01	7.39879284	4.955785e+00	9.81552314	1.873571e+01	9.82559641	12.67439143	6.064798e+00	8.7762598	8.197081e+00
AK1	1.341733e+00	9.094585e+00	3.84183234	2.523636e+00	4.91200539	8.038964e+00	8.98594722	6.87635383	8.912870e+00	9.1263572	8.205656e+00
AMOC	9.782857e+00	1.571561e+01	9.78155343	9.108193e+00	9.44062464	5.850733e+01	84.94739670	15.74072116	1.896133e+01	32.3489074	3.893165e+01
AMP	1.108992e+01	3.348245e+01	16.21611908	1.142300e+01	22.98702357	4.695322e+01	55.04142572	30.01772298	2.625547e+01	25.1772350	5.982888e+01
ANAT	2.825549e-02	8.765268e-02	0.07663297	2.345003e-02	0.09784104	1.336035e-01	0.37817407	0.06044421	7.863072e-03	0.0000000	1.398239e-01
AAR2	2.170895e+00	5.136582e+00	2.78785243	1.350566e+00	2.53867584	8.567738e+00	12.99747687	4.45703714	5.783500e+00	4.8341910	1.009426e+01

Gene Acinar Alpha Alpha.Beta Beta Delta Ductal Endothelial Gamma.Epsilon Macrophage Mast Stellate

B



Supplementary Figure 5. Querying and visualizing islet cell type-specific expression. (A) Interactive table showing normalized expression levels of genes in each identified cell type. (B) Interactive visualization of differential cell type expression in disease states.