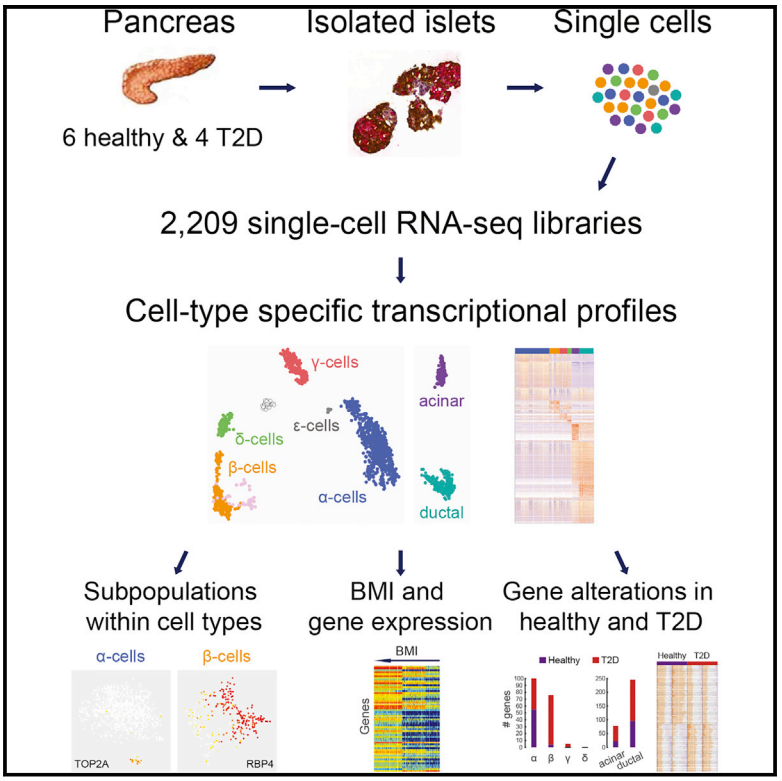
Resource

Cell Metabolism

Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes

Graphical Abstract



Authors

Åsa Segerstolpe, Athanasia Palasantza, Pernilla Eliasson, ..., Maria Kasper, Carina Ämmälä, Rickard Sandberg

Correspondence

rickard.sandberg@ki.se

In Brief

Segerstolpe et al. use single-cell transcriptomics to generate transcriptional profiles of individual pancreatic endocrine and exocrine cells of healthy and type 2 diabetic donors. They revealed cell-type-specific gene expression and novel subpopulations, as well as gene correlations to BMI and gene expression alterations in diabetes.

Highlights

- Single-cell RNA-seq enabled molecular profiling of rare human pancreatic cells
- Subpopulations were identified within endocrine and exocrine cell types
- Genes associated with obesity or diabetes had correlating expression with BMI
- Transcriptional alterations found in type 2 diabetic individuals

Accession Numbers

E-MTAB-5061
E-MTAB-5060

Segerstolpe et al., 2016, Cell Metabolism *24*, 593–607
October 11, 2016
http://dx.doi.org/10.1016/j.cmet.2016.08.020

CellPress

Cell Metabolism
Resource

Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes

Åsa Segerstolpe,[1,2,9] Athanasia Palasantza,[1,9] Pernilla Eliasson,[3] Eva-Marie Andersson,[3] Anne-Christine Andréasson,[3] Xiaoyan Sun,[4] Simone Picelli,[5,10] Alan Sabirsh,[3] Maryam Clausen,[6] Magnus K. Bjursell,[7] David M. Smith,[8] Maria Kasper,[4] Carina Ämmälä,[3] and Rickard Sandberg[1,2,5,11,*]

[1]Department of Cell and Molecular Biology (CMB), Karolinska Institutet, 171 77 Stockholm, Sweden
[2]Integrated Cardio Metabolic Center (ICMC), Karolinska Institutet, 141 57 Huddinge, Sweden
[3]Cardiovascular and Metabolic Diseases (CVMD), Innovative Medicines and Early Development Biotech Unit (iMed), AstraZeneca, 431 83 Mölndal, Sweden
[4]Department of Biosciences and Nutrition and Center for Innovative Medicine, Novum, Karolinska Institutet, 141 83 Huddinge, Sweden
[5]Ludwig Institute for Cancer Research, 171 77 Stockholm, Sweden
[6]Discovery Sciences, Innovative Medicines and Early Development Biotech Unit (iMed)
[7]R&D Information (RDI)
AstraZeneca, 431 83 Mölndal, Sweden
[8]Discovery Sciences, Innovative Medicines and Early Development Biotech Unit (iMed), AstraZeneca, Cambridge Science Park, Milton Road, Cambridge CB4 0WG, UK
[9]Co-first author
[10]Present address: Science for Life Laboratory, 171 65 Stockholm, Sweden
[11]Lead Contact
*Correspondence: rickard.sandberg@ki.se
http://dx.doi.org/10.1016/j.cmet.2016.08.020

SUMMARY

Hormone-secreting cells within pancreatic islets of Langerhans play important roles in metabolic homeostasis and disease. However, their transcriptional characterization is still incomplete. Here, we sequenced the transcriptomes of thousands of human islet cells from healthy and type 2 diabetic donors. We could define specific genetic programs for each individual endocrine and exocrine cell type, even for rare δ, γ, ε, and stellate cells, and revealed subpopulations of α, β, and acinar cells. Intriguingly, δ cells expressed several important receptors, indicating an unrecognized importance of these cells in integrating paracrine and systemic metabolic signals. Genes previously associated with obesity or diabetes were found to correlate with BMI. Finally, comparing healthy and T2D transcriptomes in a cell-type resolved manner uncovered candidates for future functional studies. Altogether, our analyses demonstrate the utility of the generated single-cell gene expression resource.

INTRODUCTION

The pancreas is a vital organ for maintaining metabolic homeostasis, consisting largely of exocrine ductal and acinar cells that produce and deliver digestive enzymes into the gut. Intermingled in the exocrine regions are the islets of Langerhans, comprised of at least five distinct endocrine cell types: α cells (secreting glucagon, GCG), β cells (insulin, INS), γ/PP cells (pancreatic polypeptide, PPY), δ cells (somatostatin, SST), and ε cells (ghrelin, GHRL), together making up less than 2% of pancreas mass. The cell-type composition within human islets of Langerhans is 50%–60% β cells, 30%–45% α cells, less than 10% γ and δ cells, and less than 1% ε cells (Cabrera et al., 2006); however, this composition varies among individuals. The endocrine islets are essential for blood glucose homeostasis and key players in the development of diabetes, which is characterized by loss of functional β cells (Kahn et al., 2006). Type 2 diabetes (T2D) is caused by a combination of increasing INS resistance in peripheral tissues and reduced mass or dysfunction of the β cells.

In order to understand the molecular mechanism governing the function of the pancreas, it has been important to investigate cell-type-specific gene expression in health and disease. Due to the cellular heterogeneity within the islets of Langerhans, it is challenging to interpret whole-islet transcriptome data, and fluorescence-activated cell sorting (FACS)-enriched transcriptome data only exist for a few cell types. In particular, it is hard to distinguish cell-type compositional differences from alterations occurring within specific cell types and address whether subpopulations exist. These issues could be resolved using single-cell transcriptomics (Sandberg, 2014; Stegle et al., 2015). The two studies to date have had too few cells (Li et al., 2016; Wang et al., 2016) to control for inter-individual differences and profile rare cells.

Here, we used single-cell RNA-sequencing (RNA-seq) to generate transcriptional profiles of endocrine and exocrine cell types of the human pancreas in healthy and T2D individuals. We could reveal subpopulations in endocrine and exocrine cell types, identify genes with interesting correlations to BMI in specific cell types, and find alterations in gene expression in T2D.

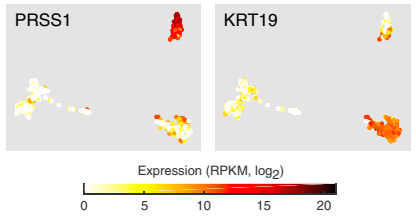
RESULTS

Pancreatic tissue and cultured islets were obtained from six healthy and four T2D donors of varying BMI and age (Figure 1A).

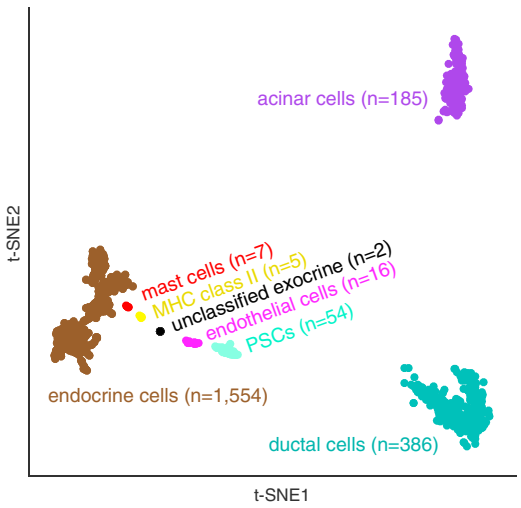
A

Donor	Sex	Age	BMI	HbA1c
H1	male	43	30.8	5.0%
H2	male	25	24.7	5.3%
H3	female	48	35.0	5.7%
H4	male	22	32.9	5.4%
H5	male	27	31.8	N/D
H6	male	23	21.5	5.5%
T2D1	male	57	24.0	7.0%
T2D2	female	37	39.6	7.3%
T2D3	male	52	34.4	7.7%
T2D4	female	55	29.8	7.4%

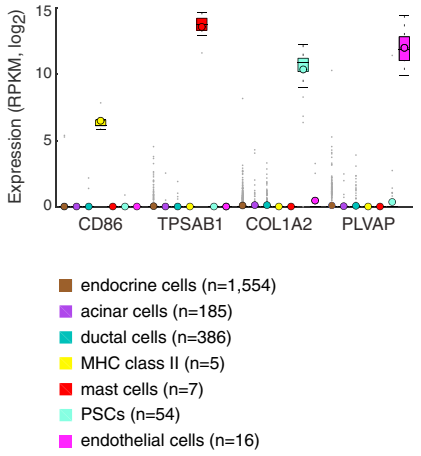
C



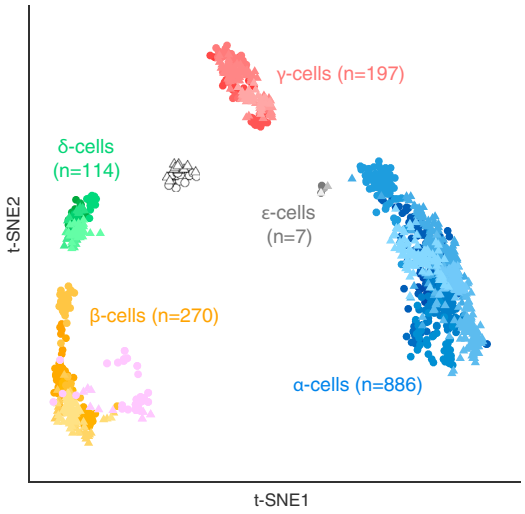
B



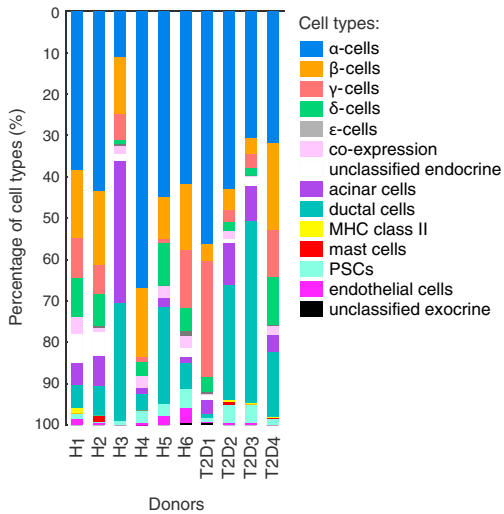
D



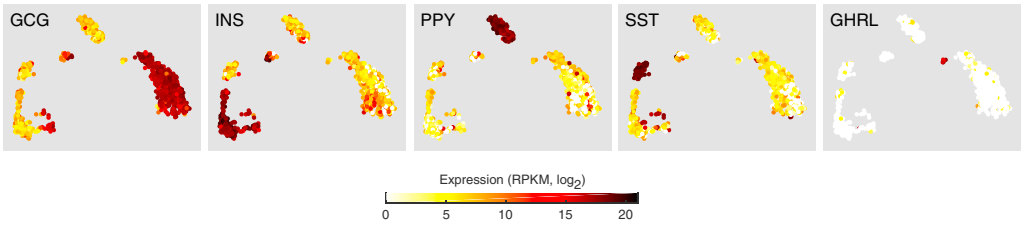
E



G



F



(legend on next page)

Functionality of the islets was confirmed with glucose-stimulated INS secretion (GSIS) (Figure S1A, available online). Islets were dissociated into single-cell suspension and viable individual cells were distributed via FACS into 384-well plates containing lysis buffer (Figure S1B). Next, single-cell cDNA libraries were generated using the Smart-seq2 protocol (Picelli et al., 2014), which enables the capture of full-length poly(A)$^+$ RNAs with higher sensitivity and more even transcriptome coverage than with previous protocols (Picelli et al., 2013; Ramsköld et al., 2012; Stegle et al., 2015). Each single-cell transcriptome was sequenced to ~750,000 reads, sufficient for cell-type classification (Pollen et al., 2014). Gene expression was quantified as reads per kilobase of transcript per million mapped reads (RPKM) using rpkmforgenes (Ramsköld et al., 2009).

Since certain endocrine populations (e.g., ε cells) are present within the islets at low frequencies, we sequenced ~350 cells from each donor. In total we obtained 3,386 cells, and after stringent quality control filtering (see Supplemental Experimental Procedures; Figures S1C–S1H), we retained 2,209 cells. As controls, we also sequenced 16 empty wells for six of the donors (in total 96 wells). The sequencing reads detected in the controls, originating from low levels (<0.5%) of RNA contamination from FACS or library preparation, were below the quality thresholds.

Cell-Type Identification and Expression

We explored the single-cell transcriptome data in an unbiased manner by identifying biological variation in gene expression (Brennecke et al., 2013) (Table S1) and projecting all cells onto two dimensions using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008). The major separation of cells reflected gene expression differences between exocrine and endocrine cell types (Figure 1B), with two clusters corresponding to exocrine acinar (n = 185) and ductal (n = 386) cells based on expression of their respective markers, *PRSS1* and *KRT19* (Figure 1C), whereas the third cluster contained the endocrine cells (n = 1,554). The five smaller observed clusters (Figure 1B) corresponded to pancreatic stellate cells (PSCs; n = 54, high expression of collagen genes, matrix metalloproteinases, *TIMP1*, *FN1*, *POSTN*, and *ACTA2*) (Liu and Du, 2015), endothelial cells (n = 16, expressing *PLVAP*, *CD31/PECAM1*, *VWF*, and *ANGPT2*), mast cells (n = 7, expressing tryptase genes *TPSB2*, *TPSD1*, and *TPSAB1*), antigen-presenting MHC class II cells (n = 5, high expression of *CD74*, *CD86*, *HLA-DPA1*, *HLA-DPB1*, and *HLA-DRA*) and two cells of unknown origin ("unclassified exocrine," high expression of, e.g., *RCAN1*, *SPP1*, and *NOV*) (Figure 1D; Table S1). The cell clusters were obtained without using knowledge of cell types or prior purification of cell populations.

To resolve the various endocrine cell types, we performed a similar analysis only on the endocrine cells, which separated them into discrete clusters (Figure 1E) with distinct hormone expression and allowed their classification into α, β, γ, δ, and ε cells (Figure 1F). The endocrine cells clustered together irrespectively of inter-individual differences and T2D status (Figure 1E), demonstrating that cell-type-specific expressions are associated with the dominating transcriptome patterns. A group of cells characterized by a high expression of multiple hormones and an elevated number of expressed genes was detected among β cells ("co-expression," n = 39; Figures 1E and S2A). Although such cells have been reported (Blodgett et al., 2015), this group likely corresponds to cell doublets and therefore was excluded. Indeed, image analyses identified 0.3% cell doublets during single-cell sorting (Supplemental Experimental Procedures), and hormone co-expression was found in ~0.9%–1.8% of the remaining endocrine cells after removal of the potential cell doublets (Figures S2C and S2D). A second cluster of cells (unclassified endocrine, n = 41; Figure 1E, white group) characterized by low number of genes expressed (Figure S2B) was also excluded from the analyses.

Variations in Cell-Type Composition

The composition of cells profiled from each donor varied substantially (Figures 1G and S3A; Table S1), both in terms of exocrine and endocrine cells and within the endocrine cell types. To determine if the observed composition of cell types corresponded to cell types present in the non-dissociated tissue and islet preparations, we performed immunohistochemistry staining with GCG and INS on the pancreatic tissues and purified islets (Figure S3B) (eight donors), and FACS analyses on dissociated islets using antibodies targeting GCG, INS, and SST (two donors) (Figure S3C). Analyses of the estimated fractions of α, β, and δ cells in the tissue and islets revealed that α cells were more abundant and β cells less numerous in the single-cell data than in the corresponding tissue and islet preparations (Figure S3D). This could reflect lower survival of β cells during single-cell dissociation and FACS. For γ and δ cells, the numbers were in general agreement except for an increase of γ cells in the single-cell data in one T2D donor (Figure 1G). We concluded that single-cell dissociation and FACS introduce a systematic bias in cell-type proportions. In order to compare single-cell to

Figure 1. Single-Cell Transcriptome Analyses of Human Pancreas

(A) Table of donor information (HbA1c, glycated hemoglobin).

(B) Projection of all cells (n = 2,209) onto two dimensions using t-SNE based on the expression values (log_2RPKM) of the 1,000 genes with highest biological variation across cells.

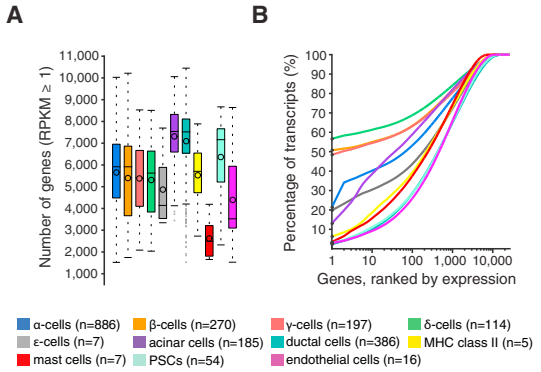
(C) Expression (log_2RPKM) of exocrine marker genes (*PRSS1* for acinar and *KRT19* for ductal cells) overlaid onto the 2D t-SNE as shown in (B).

(D) Boxplots displaying the expression levels in the seven obtained clusters of marker genes for MHC class II antigen-presenting cells (*CD86*), mast cells (*TPSAB1*), pancreatic stellate cells (*COL1A2*), and endothelial cells (*PLVAP*). Median and mean are shown as a line and circle, respectively. Edges of each box indicate the 25th and 75th percentiles. Bars extend to extreme data points and outliers are plotted as gray dots.

(E) Two-dimensional t-SNE projection of the endocrine cells (n = 1,554) based on the expression values (log_2RPKM) of the 500 genes with highest biological variation across endocrine cells. The obtained clusters were assigned to the endocrine cell types based on the hormone expression levels in (F). Colors correspond to cell types and shadings indicate donors. Healthy and T2D cells are marked with circles and triangles, respectively.

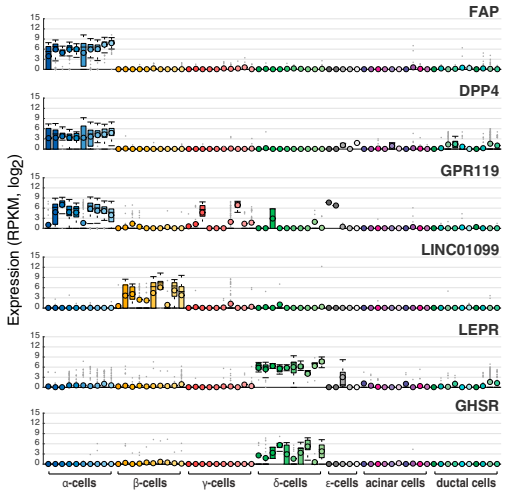
(F) t-SNE representation of cells as in (E) illustrating the expression of the five endocrine hormones: *GCG*, *INS*, *PPY*, *SST*, and *GHRL*. Color scale is according to log_2RPKM values, with white and red colors corresponding to minimum (zero) and maximum (log_2RPKM = 21) expression, respectively.

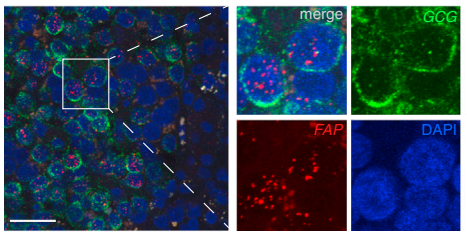
(G) Bar graphs showing the percentage of cells classified into cell types, per donor.

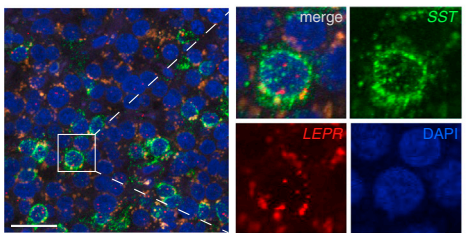


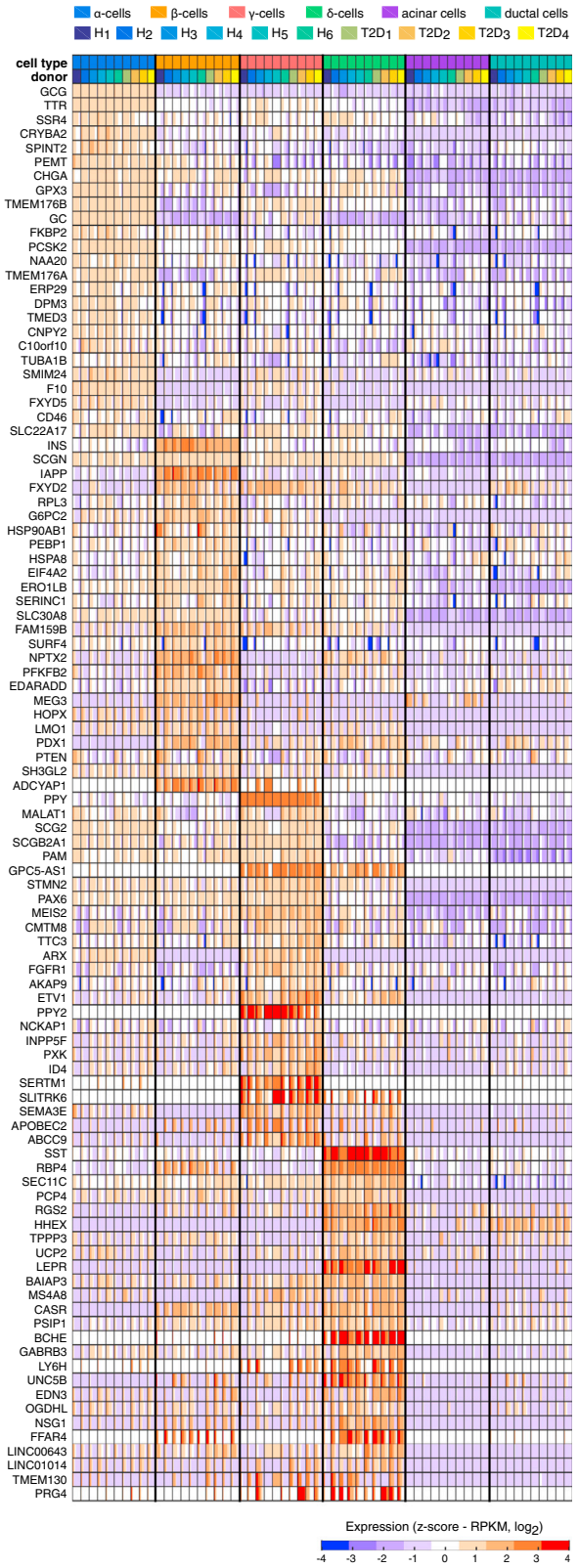
C

α-cells	β-cells	γ-cells	δ-cells	acinar cells	ductal cells
174	94	64	55	332	408









(legend on next page)

whole-islet RNA-seq, we correlated the average gene expression in single cells with the expression in whole-islet sequencing from the same donors. This analysis demonstrated that single-cell data give a reliable view of the whole islet (Spearman's ρ, 0.86–0.92; Figure S4A) and that with only tens of single cells, correlations start to saturate (Figure S4B).

Gene Expression Signatures of Endocrine Cells

The average number of genes expressed per cell type was roughly 5,500 for endocrine and 7,000 for exocrine cells (Figure 2A), with slight variation using higher expression threshold (Figure S1H). Strikingly, hormone expression accounted for a large fraction of the transcriptome in endocrine cells, reflecting their functional dedication to hormone secretion, with *INS*, *PPY*, and *SST* transcripts alone accounting for ~50% of the total cellular transcripts in the β, γ, and δ cells, respectively (Figure 2B). In α and ε cells, the expression of *GCG* and *GHRL*, respectively, accounted for ~20% of their transcriptomes (Figure 2B; Table S1).

Next, we investigated the specific gene expression in the six major cell types (α, β, γ, δ, acinar, and ductal cells) to shed light on their respective functions, using one-way ANOVA on the five healthy male donors (to exclude confounding effects of sex and disease). This analysis revealed several hundred genes that were expressed at significantly higher levels in α, β, γ, δ, or exocrine cell types (Table S2; Figures 2C, 2D, and 3C). Most of the previous transcriptome studies of the pancreas have investigated whole islets or used FACS to sort for either α or β cells followed by microarray analysis or RNA-seq. For α or β cells, several studies in mouse (Benner et al., 2014; Ku et al., 2012) or human islets (Blodgett et al., 2015; Bramswig et al., 2013; Dorrell et al., 2011; Nica et al., 2013) have found cell-type-specific transcripts with 2-fold or higher expression levels than in other cell types. The genes we identified as being α and β cell enriched overlapped largely with the previous studies. For example, in α cells we detected known markers such as *GCG*, *LOXL4*, *DPP4*, *GC*, and *FAP* (Figure 2E) (Dorrell et al., 2011). Importantly, RNA in situ hybridization verified significant co-expression of *FAP* and *GCG* in α cells (Figure 2F; Table S3). We also observed cell-type enriched expression of genes that have not previously been linked to specific endocrine cell types. *GPR119* has been identified as a potential target for the treatment of diabetes, since agonists have been shown to enhance nutrient-stimulated INS and GLP-1 release and also to increase β cell mass in mice in vivo (Chu et al., 2008; Gao et al., 2011). We could assign the expression of *GPR119* primarily to α cells in humans (Figure 2E), in contrast to previous reports (Chu et al., 2008; Odori et al., 2013).

In β cells, we found elevated expression levels of, e.g., *INS*, *IAPP*, *ADCYAP1*, *PDX1*, *MAFA*, *NKX6-1*, and *MEG3* (Figures 2D and 3B). Additionally, β cell-specific protein expression of both PDX1 and NKX6-1 was verified using FACS analysis, where the PDX1+/NKX6-1+ cell population coincided with the expression of β cell marker C-peptide protein (Figure S5A). We also detected long non-coding RNAs, e.g., *LINC01099*, with expression restricted mainly to β cells (Figure 2E).

Transcriptional Profiling of Rare Endocrine Cell Types

The transcriptomes of γ, δ, and ε cells are largely unknown in human. We noted that many receptors were highest or exclusively expressed in δ cells (e.g., *UNC5B*, *GABRB3*, *GABRG2*, *CASR*, *FFAR4/GPR120*, and *KCNJ2*) (Figure 2D; Table S2). We also identified pronounced expression of the leptin receptor (*LEPR*) in δ cells (Figure 2E). Previous studies in murine and human cells have reported LEPR protein expression in α, β, and δ cells (Kieffer et al., 1996, 1997; Tudurí et al., 2009). RNA in situ hybridization validated significant co-expression of *LEPR* and *SST* in δ cells (Figure 2G; Table S3). Expression of the GHRL receptor (*GHSR*) has been reported on mouse α cells (Date et al., 2002), rat INS-1 β cell line (Wierup et al., 2004), and to be partly localized to β cells in human islets (Granata et al., 2007). Intriguingly, we uncovered that the *GHSR* was specifically expressed in δ cells (Figure 2E).

We also identified genes specifically expressed in γ cells, e.g., *SERTM1*, *SPOCK1*, *ABCC9*, and *SLITRK6* (Figure 2D; Table S2), none of which, to the best of our knowledge, have been assigned to pancreatic or γ cell functions. Among the 2,209 cellular transcriptomes captured, seven were from ε cells. Analysis of these cells allowed us to identify elevated expression of 39 genes, including the expression of *GHRL* and *GHRLOS*. The ε cells uniquely expressed several interesting receptors like *NPY1R*, *OPRK1*, *PTGER4*, and *ASGR1*, and processing enzymes such as *PCSK6* (Figure 3A; Table S2).

Transcription Factor Expression in Endocrine Cells

We could confirm the expression of many transcription factors in endocrine cells, e.g., *ISL1* and *NEUROD1*, and restricted

Figure 2. Characterization of Endocrine Cell Transcriptomes

(A) Boxplots showing the number of genes detected in each cell type (expression threshold, RPKM ≥ 1). Median and mean are shown as a line and circle, respectively. Edges of each box indicate the 25th and 75th percentiles. Bars extend to extreme data points and outliers are plotted as gray dots.

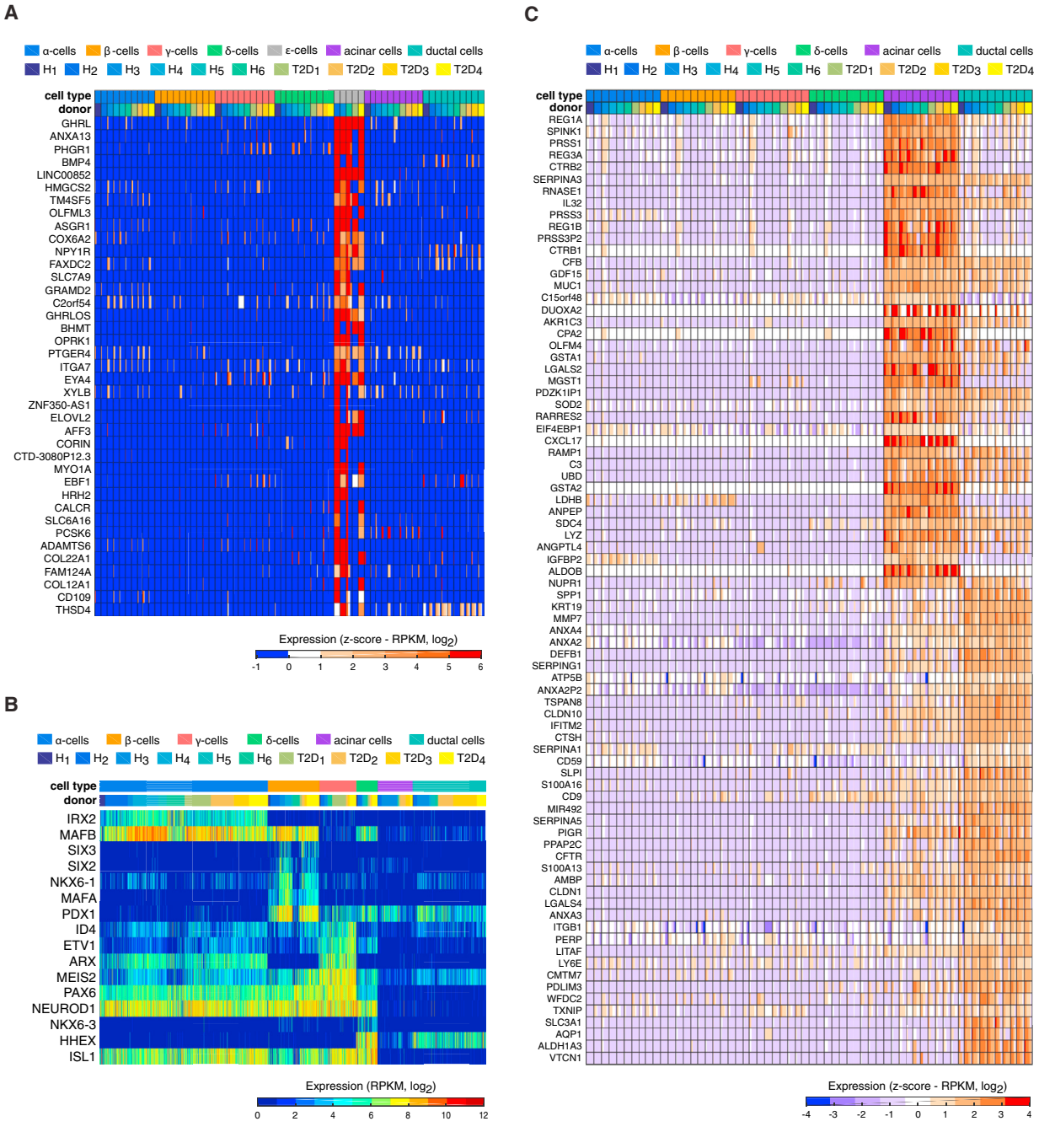
(B) Percentage of all mRNAs in respect to the total transcriptome in each cell type (Table S1), with genes ranked according to the expression magnitude in descending order (x axis).

(C) Table with the number of significantly enriched genes in α, β, γ, δ, acinar, and ductal cells.

(D) Heatmap with expression distributions for the top 25 enriched genes in each of the four endocrine cell types (α, β, γ, and δ cells). The genes were selected based on the magnitude of expression range among the four endocrine cell types. The expression profiles of the ten donors are shown separately and for each endocrine and exocrine cell type, with labels indicating cell type and donor (top). Colors correspond to standardized $\log_2$ expression values, where each cell in the heatmap contains the distribution of values across the cells for each cell type and donor.

(E) Boxplots with the expression levels of selected cell-type enriched genes identified in the differential expression analysis: *FAP*, *DPP4*, and *GPR119* for α cells; *LINC01099* for β cells; and *LEPR* and *GHSR* for δ cells. Gene expression is shown for α, β, γ, δ, ε, acinar, and ductal cells, with color shadings representing different donors. Median and mean are shown as a line and circle, respectively. Edges of each box indicate the 25th and 75th percentiles. Bars extend to extreme data points and outliers are plotted as gray dots.

(F and G) Single-molecule RNA FISH on pancreatic tissue section. (F, left) A representative islet (donor H5) co-stained with *GCG* (green), *FAP* (red), and DAPI (blue). (G, left) Islet (donor H5) co-stained with *SST* (green), *LEPR* (red), and DAPI (blue). (F and G, right) Zoom-in on merged or individual channels. Quantification and additional images in Table S3. Scale bar represents 25 μm.



(legend on next page)

expression of *MAFA* in β cells and *IRX2* in α cells (Figure 3B). However, many transcription factors were expressed in a subset of endocrine or exocrine cells, including *ARX* expression in α and γ cells, in line with the increase in γ and α cell numbers after forced *ARX* expression in mouse embryonic islets (Collombat et al., 2007). We found *HHEX* expression in both δ and ductal cells while *MAFB* expression was restricted to α, β, and δ cells (Figure 3B). We also detected several enrichments of factors in particular cell types. For example, *NKX6-3* was enriched in δ cells and *SIX2* and *SIX3* were restricted to β cells. In γ cells, we observed significantly elevated expression of *MEIS2*, *ETV1*, and *ID4* (Figure 3B), whereas *AFF3* was specifically expressed in ε cells (Figure 3A).

Cell-Type-Specific Expression in Other Pancreatic Cell Types

The transcriptomes of exocrine cell types had higher numbers of significantly enriched genes, 332 and 408 genes, respectively, for acinar and ductal cells (Figures 2C and 3C; Table S2). Additionally, we detected a large number of genes that were expressed at higher levels in both exocrine cell types but were absent in endocrine cells (data not shown). The pancreatic stellate cells and endothelial cells also expressed a large number of specific genes (Figure 3D; Table S2).

Using our cell-type-specific expression data, we assessed to what extent cell-type-specific gene expression can be identified from whole-islet sequencing (Taneera et al., 2012) by simply correlating gene expression with cell-type-defining hormone or marker gene expression. Our results indicate that only few of the cell-type enriched genes would be captured, as our strongest cell-type enriched genes were either seldom significant or had low ranks (Figure S5B).

The Identification of Subpopulations and Cellular States

In order to explore heterogeneity within cell populations, we investigated the clustering of cells within each cell type. After assigning cells to cell types, the second strongest determinant of the cellular transcriptomes was inter-individual differences (Figure S6).

Clustering analysis after the removal of donor effect (see Supplemental Experimental Procedures) uncovered subpopulations of α, β, and acinar cells. For γ, δ, and ductal cells, no robust separation was observed (Figure S6). Twelve α cells belonging to multiple donors (both healthy and T2D) grouped separately from all other α cells (Figures 4A and S6A). *GCG* expression was similar in both clusters; however, differential expression analysis showed a mild downregulation of several α enriched genes (*C10orf10*, *PEMT*, *PLCE1*, *ARRDC4*, *CRYBA2*, *LOXL4*, *RGS4*, and *SMIM24*) and a high expression of proliferation-associated genes (e.g., *TOP2A*, *MKI67*, *CENPF*, *BIRC5*, and *CDK1*) in these twelve cells (Figure 4A). Therefore, the rare α cells constituted proliferating α cells that were distinguished by a signature of 439 significantly upregulated genes (Table S4). Importantly, this group was not detectable before adjusting for donor effects (Figure S6A). Searching for proliferative gene expression signatures across the complete single-cell dataset revealed additional acinar (n = 8) and ductal (n = 2) cells with expression of proliferative markers (Figure S7). The gene expression signatures of these rare proliferating cells might provide clues to the mechanisms of self-renewal of pancreatic tissue.

Sub-clustering of β cells revealed five clusters of cells with combinatorial expression of *RBP4*, *FFAR4/GPR120*, *ID1*, *ID2*, and *ID3* (Figure 4B; Table S4). Cells of the five clusters expressed *INS* at similar levels (Figure 4B). Interestingly, clusters 1 and 5 expressed *RBP4* (also expressed in δ cells; Figure 2D), an adipokine primarily expressed in the liver and adipocytes (Klöting et al., 2007). Increased circulating levels of RBP4 are prominent in obese and T2D individuals and correlate with INS resistance in the periphery (Yang et al., 2005). The same *RBP4*-expressing β cells (clusters 1 and 5) also express *FFAR4.* FFAR4 agonists have been shown to induce INS release in mouse pancreatic islets (Moran et al., 2014). β cell clusters 1, 2, and 3 were characterized by expression of *ID1* and *ID3*, regulators of basic helix-loop-helix (bHLH) transcription factors (Ling et al., 2014). To the best of our knowledge, these subgroups of β cells have never been described before.

Finally, we identified two clusters of acinar cells (Figure 4C). Cells belonging to cluster 1 were characterized by elevated expression of inflammatory related genes (Figure 4C; Table S4), including components of the MHC class II molecule, e.g., *CD74*, *HLA-DMA*, *HLA-DRA*, and *HLA-DRB1* (Figure 4C). Cluster 1 also expressed pro-inflammatory chemokines (e.g., *CXCL1*, *CXCL6*, *CCL2*, *CCL20*, and *CX3CL1*), cytokines (*IL17C* and *IL18*), and transcription factors and immune regulators (*STAT1*, *NFKBIA*, *NFKBIZ*, *HIF1A*, *SOX4*, and *ONECUT2*). Cells of cluster 2 expressed higher levels of key acinar genes encoding secretory digestive enzymes (e.g., *CEL*, *CELA2A*, and *AMY*) and important transcription factors that regulate expression of digestive enzymatic genes including *PTF1A* and *RBPJL* (Figure 4C).

Gene Expression Alterations with Increased Obesity

In order to investigate gene expression signatures relating to obesity, we correlated the expression of each gene with donor BMI for the five healthy males, separately per cell type (see Supplemental Experimental Procedures; Table S5). Several of the genes correlating with BMI have earlier been implicated in

Figure 3. Cell-Type-Specific Expression

(A) Heatmap showing ε cell enriched gene expression (rows) across the seven cell types for the ten donors (columns). Labels indicating the cell type and donor are shown on the top. Colors in the heatmap correspond to standardized $\log_2$ expression values, where each cell in the heatmap contains the distribution of values across the cells for each cell type and donor.

(B) Heatmap with expression levels of transcription factors (rows) across the cells from the six cell types for the ten donors (columns). Blue and red colors correspond to minimum (zero) and maximum ($\log_2$RPKM = 12) expression, respectively.

(C) Heatmap with expression of significantly enriched genes (rows) in acinar and ductal cells across the six cell types for the ten donors (columns). The genes were selected based on the magnitude of expression range between the two exocrine cell types. The heatmap was generated as in (A).

(D) Heatmap showing the differentially expressed (DE) genes (columns) between the pancreatic stellate (PSCs) and the endothelial cells (rows). The results were obtained using single-cell differential expression (SCDE), and only the genes with a $\log_2$ fold change in expression of at least seven between the two groups are shown. The colors in the heatmap correspond to standardized $\log_2$ expression values, where each cell in the heatmap contains the distribution of values across the cells of each cell type.

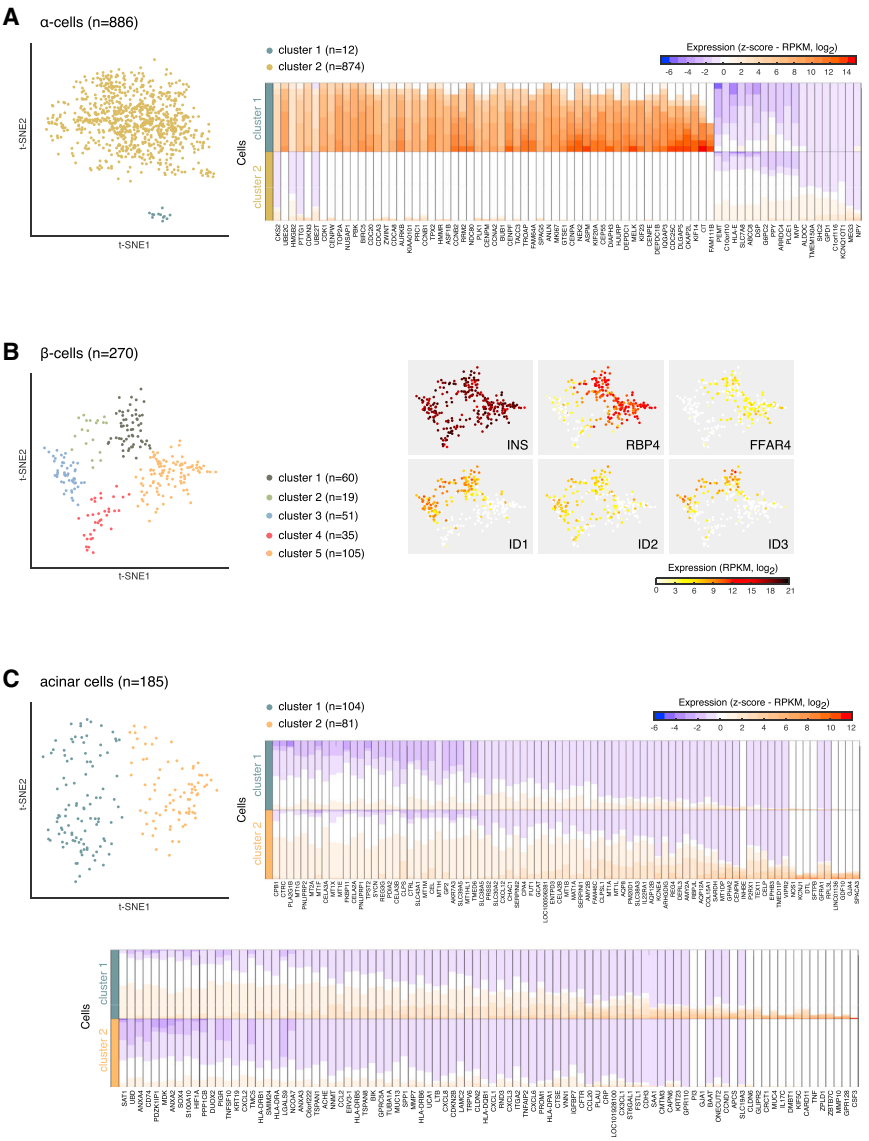


Figure 4. Uncovering Subtypes of Endocrine and Exocrine Cells

(A) (Left) Two-dimensional t-SNE representation of all α cells (n = 886, 10 donors) using donor-normalized expression values of the 500 most variable genes in α cells. Colored according to cluster assignments. (Right) Heatmap illustrating the top DE genes (columns) between the two α cell clusters (rows). The colors in the heatmap correspond to standardized $\log_2$ expression values, where each cell in heatmap contains the distribution of values across the cells in each cluster.

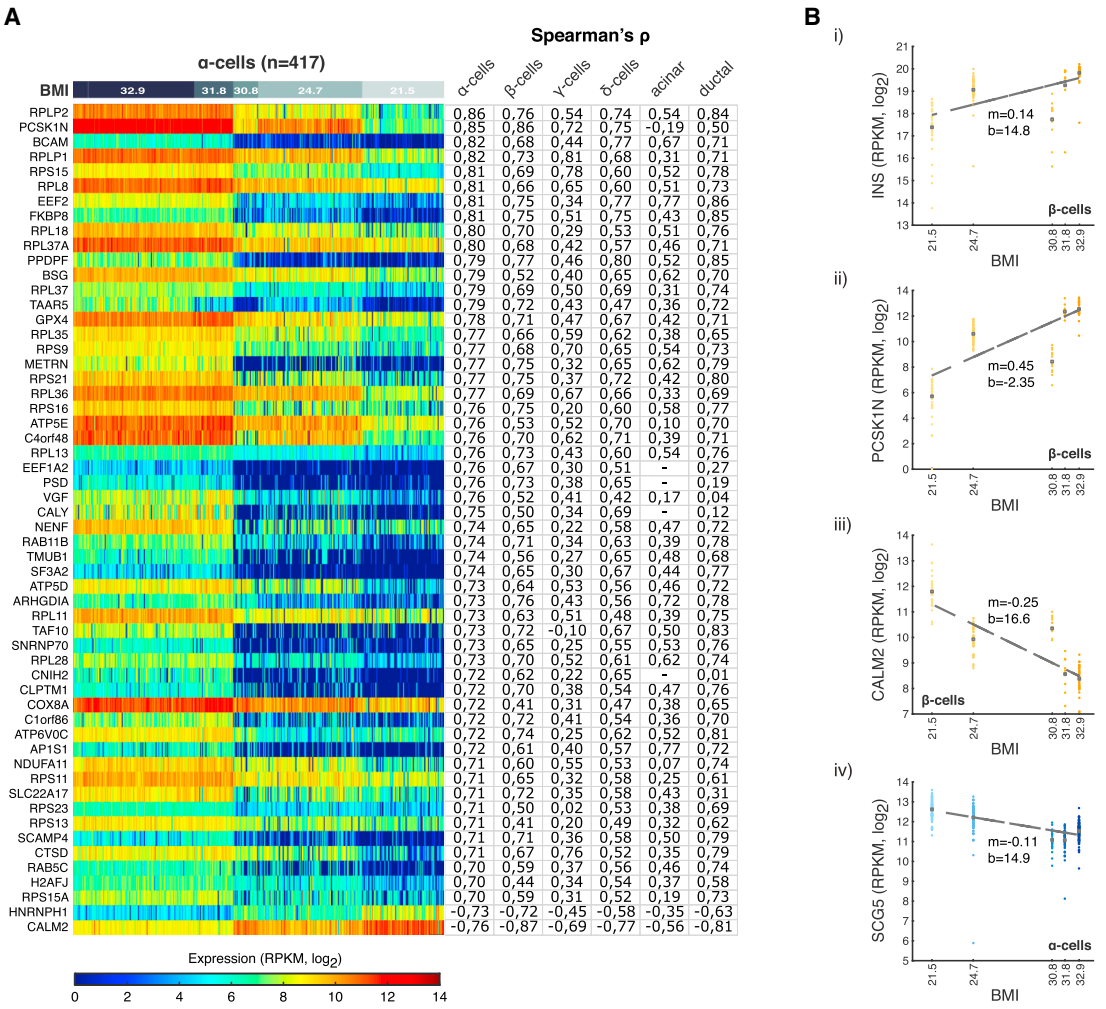
(B) (Left) Two-dimensional t-SNE representation of β cells (n = 270, 10 donors) using donor-normalized expression values of the 50 most variable genes in β cells. (Right) t-SNE representation of cells, colored according to expression ($\log_2$RPKM) of DE genes among the five clusters.

(C) (Left) Two-dimensional t-SNE projection of acinar cells (n = 185, 10 donors) using donor-normalized expression values of the 100 most variable genes in acinar cells. (Right) Heatmaps illustrating the top DE genes (columns) per cell cluster (rows). The heatmap was generated as in (A).

obesity or diabetes progression. Expression of the PCSK1 inhibitor (*PCSK1N*) showed a strong positive correlation with BMI in all endocrine cell types (Figures 5A and 5B). PCSK1 converts prohormones to active hormones with substrates such as INS, SST, and enkephalins. Altered function of PCSK1 is associated with obesity both in mouse and human (Nead et al., 2015). In addition, transgenic mice overexpressing *PCSK1N* have an obese phenotype (Wei et al., 2004). Analysis also revealed that *SCG5* was negatively correlated with BMI in all cell types (Figure 5B). SCG5 is a molecular chaperone of PCSK2, the enzyme responsible for cleavage of pro-GCG to GCG and also involved in pro-IAPP to IAPP processing (Braks and Martens, 1994; Wang et al., 2001), implicating an α and β cell-related dysfunction in GCG and IAPP processing, respectively, correlating with increasing BMI. Interestingly, *PPDPF* expression in α, β, δ, and ductal cells was positively correlated with BMI (Figure 5A). PPDPF regulates exocrine pancreas development in zebrafish, where overexpression leads to relative expansion of the exocrine pancreas (Jiang et al., 2008). Finally, we noted that *CALM2* was negatively correlated with BMI in endocrine (α, β, γ, and δ) and ductal cells (Figures 5A and 5B), which is of interest given the role of calmodulins in calcium sensing and signaling, e.g., in β cells and INS release (Gilon et al., 2014). Naturally, previous bulk analyses should, in principle, be able to identify genes that are correlating to BMI in all (or the most abundant) cell types, whereas those specific to particular cell types would not be easily revealed without single-cell resolution. This is apparent when comparing correlations obtained with all cells to correlations based on separate cell types (Figure 5C). For example, positive correlations of *INS*, *PPY*, and *SST* in β, γ, and δ cells were stronger in cell-type resolved analyses. Interestingly, several genes with previous links to obesity and diabetes were correlating with BMI in the cell-type resolved analyses, but were not correlating in the bulk analyses (Figure 5C).

Gene Expression Alterations in T2D

Contrasting the transcriptomes from healthy and T2D individuals in each cell type independently identified significant (adjusted p value $\leq$ 0.01) alterations in gene expression in endocrine and exocrine cell types (Figure 6A; Table S6). INS deficiency in T2D can result from decreased β cell mass and reduced INS production in remaining β cells. Three out of four T2D donors had lower numbers of β cells (Figure S3A), and we detected significantly lower *INS* mRNA levels in T2D β cells (Figure 6B). The most significant difference in β cells was a downregulation of *FXYD2* in T2D individuals (Figure 6B). *FXYD2* encodes the



(legend on next page)

gamma subunit of an Na,K-ATPase. Mice lacking *FXYD2* are more glucose tolerant, have a substantial pancreatic β cell hyperplasia, and have increased plasma INS levels (Arystarkhova et al., 2013). In line with these experiments, overexpression of FXYD2 in renal cells reduced the rate of cell division (Wetzel et al., 2004). Our data also revealed several important genes that were upregulated in T2D β cells, e.g., *GPD2* and *LEPROTL1* (Figure 6B). GPD2 is one of two key enzymes of the NADH shuttles into mitochondria and aids the subsequent ATP production and glucose-induced activation of mitochondrial metabolism and INS secretion in β cells (Eto et al., 1999). Moreover, *LEPROTL1* is also known as endospanin-2. Overexpression of the LEPROTL1 protein decreases the surface localization of the leptin receptor (Séron et al., 2011) and the growth hormone receptor (*GHR*) in cells (Touvier et al., 2009), thereby inhibiting the hormonal actions. Endospanin-1, a paralog of *LEPROTL1* with similar activity (Séron et al., 2011), was recently shown to be upregulated in obese mice, and silencing of the gene restored central leptin responsiveness in these mice, leading to body weight loss (Vauthier et al., 2014).

We also observed transcriptional alterations in α cells of T2D individuals. RGS4 has been shown to be a negative regulator of GSIS in mouse MIN6 insulinoma cells via its inhibitory actions on the M(3) muscarinic receptor (*CHRM3*) (Ruiz de Azua et al., 2010). The expression of *RGS4* and *CHRM3* in pancreas has not been established before and we detected them significantly enriched in α and γ cells, respectively (Table S2). We found a significantly decreased expression of *WFS1* in T2D α cells (Figure 6B; Table S6), in agreement with a recent study (Taneera et al., 2012).

Finally, we used gene set enrichment analysis (GSEA) to investigate whether transcriptional alterations in T2D were enriched for specific functional categories or previously defined gene signatures. The analysis revealed that genes responsible for energy metabolism in mitochondria and protein synthesis were significantly downregulated in most cell types in T2D individuals (Figure 6C; Table S7). In contrast, among the genes found upregulated in T2D, we observed, for example, enrichment for apoptosis, diabetic nephropathy, and cytokine signaling (Figure 6C).

Functional Validation of GLP1R

GCG-like peptide receptor 1 (*GLP1R*), expressed in β and δ cells (Figure 7A), is of interest for targeted diabetic treatment due to the ability of GLP1R agonists to induce GSIS in β cells (Nadkarni et al., 2014). Moreover, GLP1 analogs (e.g., exenatide and liraglutide) are commonly used anti-diabetic drugs (Montanya, 2012). To validate the functionality of the isolated human islets, we measured both the INS response to exenatide treatment and the GLP1R protein expression. Immunohistochemistry detected GLP1R solely in the endocrine islets and predominantly in β cell enriched regions (Figure 7B). Additionally, $GLP1R^+$ cellular fractions (FACS) were significantly enriched for β and δ cells (Figure 7C). Finally, we observed a significant increase in INS release from islets treated with 10 nM exenatide in high glucose conditions, regardless of whether the islets originated from healthy or T2D donors (Figure 7D). We conclude that the islets are functional in terms of GLP1/GLP1R signaling and that GLP1R protein expression was in agreement with the single-cell RNA-seq data.

DISCUSSION

With the development of single-cell RNA-seq, high-throughput profiling of transcriptomes across cell types, states, and subpopulations has become feasible (Sandberg, 2014). Here, we generated a unique resource of 2,209 full-length, single-cell transcriptomes from cultured pancreatic islets of ten donors. Our data are available through a user-friendly and searchable web portal (http://sandberg.cmb.ki.se/pancreas). Unbiased analyses of the data allowed the identification of cell types, demonstrating that cell-type-specific gene expression programs were the strongest determinant of the cellular transcriptomes. Importantly, due to the large number of profiled cells, we could simultaneously define the transcriptional signatures of both abundant and rare cell types in the pancreas, including δ, γ, ε, stellate, immune, and endothelial cells.

Despite the few numbers of scattered δ and ε cells within the islets of Langerhans, their transcriptomes hinted that they may have important, previously unrecognized roles in islet homeostasis. With multiple receptors, δ cells could sense and react to a variety of hormonal signals. First, they express the GHRL receptor, which enables them to receive input locally from GHRL-producing ε cells and/or systemically from GHRL release from the intestine during digestion.

In our transcriptional profile of δ cells, we also uncovered that they uniquely express the leptin receptor. Because leptin is released predominantly from adipocytes (Fasshauer and Blüher, 2015), the initial pancreatic response to leptin levels could be mediated through δ cells. In this respect, it is also tempting to speculate whether the scattered positions of δ cells within human islets of Langerhans enable them to integrate both paracrine and endocrine factors before using SST and other factors (van der Meulen et al., 2015) to signal to their immediate environment.

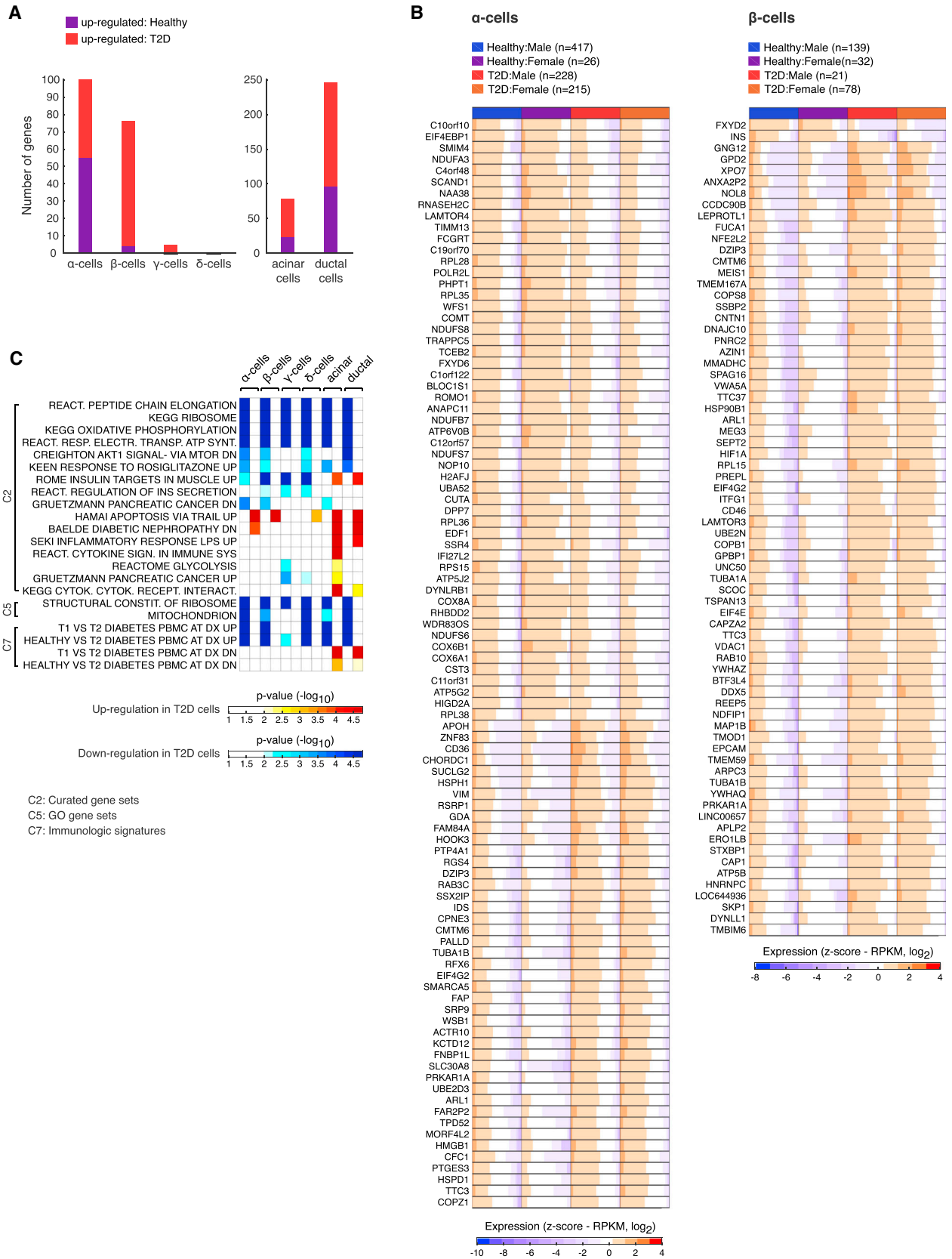
After assigning cells to cell types, it was apparent that cells within cell types grouped according to donor (Figure S6). Only

Figure 5. Gene Expression Correlates to Donor Physiological Characteristics

(A) Heatmap of gene expression associated with BMI in α cells (n = 417, healthy male donors). Genes (rows) with a positive or negative correlation of at least 0.7 in magnitude are ranked in descending order of coefficient. Cells (columns) are ordered in descending order of BMI. Blue and red colors correspond to minimum (zero) and maximum (log_2RPKM = 14) expression, respectively. The corresponding Spearman's correlation coefficients computed for the genes in each cell type (based on the cells from healthy male donors) are displayed on the right of the heatmap.

(B) Scatterplots showing the expression levels of four genes with a robust correlation with BMI. The fitted curve in each scatter shows the linear regression between the expression and BMI, where m is the slope and b the y-intercept.

(C) Scatterplots of Spearman's correlations of genes toward BMI computed based on overall expression in donors (x axis) or for the specific indicated cell type (y axis). Colored are the genes with an absolute correlation coefficient of at least 0.6 within a particular cell type that is also at least 10% higher in respect to the correlation computed using all cells.



(legend on next page)

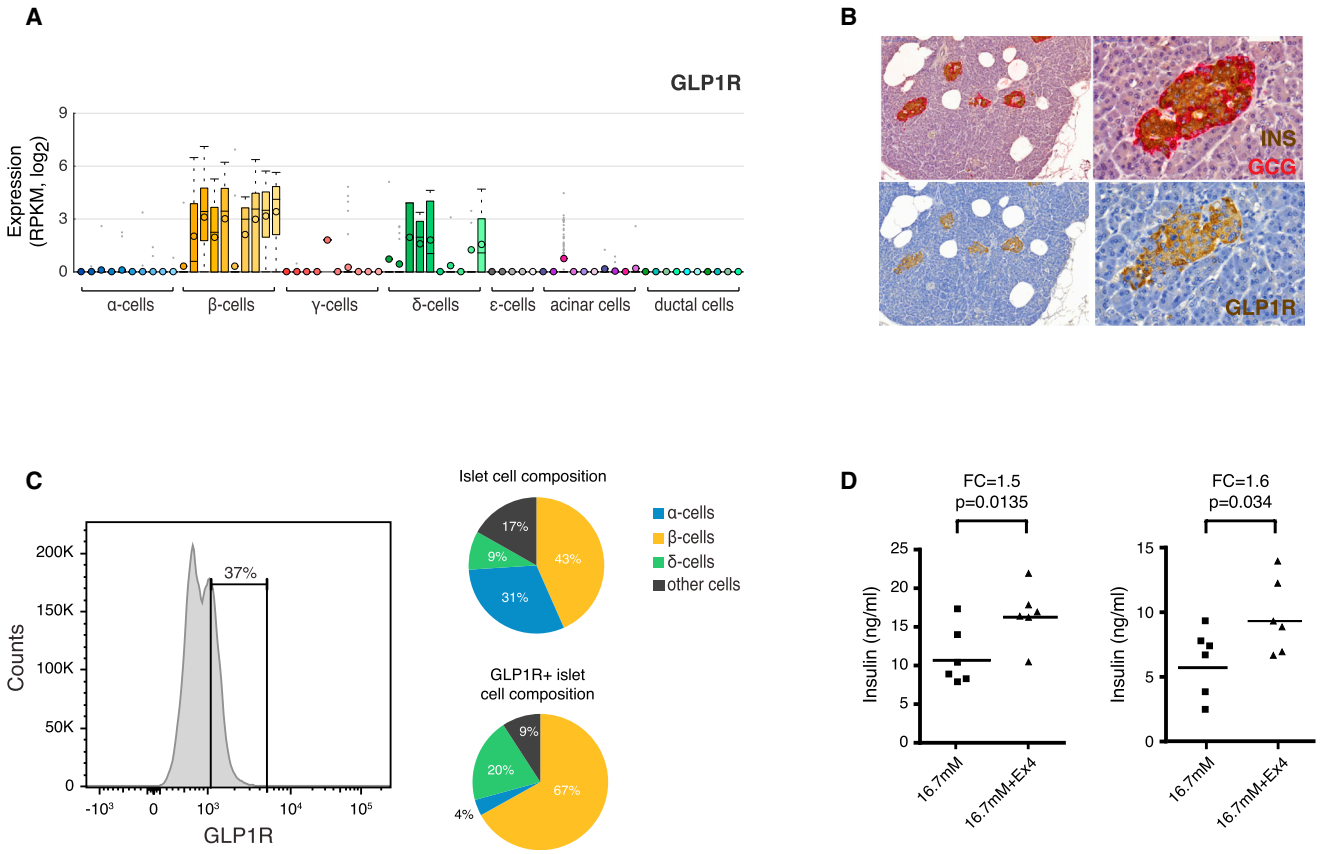


Figure 7. Functional Analysis of GLP1R

(A) Boxplot showing expression of *GLP1R* in α, β, γ, δ, ε, acinar, and ductal cells, with color shadings representing different donors. Median and mean are shown as a line and circle, respectively. Edges of each box indicate the 25th and 75th percentiles. Bars extend to extreme data points and outliers are plotted as gray dots.

(B) Immunohistochemistry of pancreatic tissue for INS (brown) and GCG (red) (top left and zoom-in on top right), or GLP1R (brown) (lower left and zoom-in on lower right).

(C) (Left) FACS analysis of islet cells stained with fluorescently conjugated GLP1R antagonist. Gate shows fraction of the total islet cells positive for GLP1R. (Right) Pie charts with the percentage of α, β, δ, and other cells in FACS analyses of total islet cells (upper) or the GLP1R+ cellular fraction (lower), labeled with INS, GCG, and SST antibodies.

(D) Dot plots showing GSIS of human islets from one healthy (left) and one T2D (right) donor treated with 16.7 mM glucose ± 10 nM Exenatide (Ex4).

after correcting for donor differences were we able to identify subpopulations and cellular states. The large number of profiled cells allowed us to investigate the heterogeneity within each cell type. Interestingly, we unraveled distinct sub-clusters of endocrine and exocrine cells, including proliferating α cells and subsets of β cells with combinatorial expression of transcriptional regulators. Moreover, we identified a group of acinar cells expressing MHC class II genes (*CD74*, *HLA-DMA*, *HLA-DRA*, *HLA-DRB1*, and *HLA-DRB5*). MHC class II molecules are predominantly expressed by professional antigen-presenting cells. In addition, peripheral cells in close contact with environmental cues for immunological reactions (e.g., intestinal epithelial cells, airway epithelial cells, and keratinocytes) have been shown to express MHC class II molecules (Kambayashi and Laufer, 2014). It is possible that the antigen-presenting acinar cells are localized to pancreatic areas in closer contact with environmental cues.

For metabolic diseases, it is expected that single-cell transcriptome analyses can be paradigm changing. The single-cell resolution allows for the parallel identification of disease-associated transcriptional alterations in each cell type independently (Sandberg, 2014). Moreover, although islets or tissues in health

Figure 6. Altered Gene Expression in Cells from T2D Individuals

(A) Bar graphs showing the number of DE genes between cells from healthy and T2D donors per cell type.

(B) Heatmaps with the DE genes between healthy and T2D α (left) and β cells (right). Labels indicating the disease status and sex of the cells are shown on the top. Colors in the heatmap correspond to standardized $\log_2$ expression values, where each cell in the heatmap contains the distribution of values across the cells of each group.

(C) Heatmap of the enriched gene sets (using GSEA) within the α, β, γ, δ, acinar, and ductal cell types for healthy and T2D groups. The heatmap is colored according to the adjusted p values ($-\log_{10}$), with red and blue colors corresponding to enrichments among genes up- and downregulated in T2D, respectively. Sets with no significant enrichment are indicated with white color.

and disease vary in cellular composition, single-cell transcriptomics circumvents these challenges by first assigning cells to cell types, which then allows comparisons in a cell-type resolved manner. Our analyses revealed potential compensatory actions for the lower β cell mass in T2D, including the downregulation of *FXYD2* and upregulation of *GPD2*. It is intriguing that mice lacking FXYD2 had a pronounced pancreas phenotype with increased proliferation of β cells (Arystarkhova et al., 2013). The downregulation of *FXYD2* could indicate an effort to stimulate β cell proliferation also in humans, but whether the function of FXYD2 is shared in mice and men needs to be established. Moreover, elevated levels of GPD2 in T2D β cells could increase INS secretion via increase of NADH shuttles into mitochondria. Finally, aberrations in mitochondrial respiratory functions were also prominent in the functional gene enrichment results for several cell types in the pancreas from the T2D donors.

In this study, we generated a resource of single-cell transcriptomes from healthy and T2D donors that should be of great value to the research community. Our computational analyses demonstrated the power of cell-type resolved analyses and revealed cell-type-specific gene expression programs, subpopulations, and transcriptional alterations in T2D. Similar single-cell analyses of other metabolic tissues and pathologies will significantly advance our understanding of heterogeneity within metabolic tissues in health and disease.

EXPERIMENTAL PROCEDURES

Single-Cell RNA-Seq of Pancreatic Islets

Human tissue and primary islets were purchased from Prodo Laboratories Inc., providing islets isolated from donor pancreases obtained from deceased individuals with research consent from Organ Procurement Organizations (OPOs). The use and storage of human islets and tissue samples were performed in compliance with the Declaration of Helsinki, ICH/Good Clinical Practice and was approved by the independent Regional Ethics Committee. Human islet samples (85%–95% pure) were cultured for 4 days in complete Prodo Islet Media Standard PIM(S) to recover after arrival. Islets were dissociated and distributed by FACS into 384-well plates. Single-cell RNA-seq libraries were produced with the Smart-seq2 protocol as previously described (Picelli et al., 2014). Sequencing was carried out on an Illumina HiSeq 2000 generating 43 bp single-end reads.

RNA-Seq of Whole Islets

RNA was isolated from whole islets from seven donors (healthy donors H3, H4, and H6 and all T2D donors) using QIAGEN RNeasy Microkit with on-column DNase digestion (QIAGEN) and processed with Illumina TruSeq Stranded mRNA Library Prep Kit. Indexed libraries were pooled and sequenced (paired-end 75 bp) on an Illumina NextSeq 500.

Single-Molecule mRNA Fluorescence In Situ Hybridization

mRNAs were visualized by single-molecule fluorescence in situ hybridization (smFISH) using the RNAscope Fluorescent Multiplex Kit (Advanced Cell Diagnostics, Inc.) on formalin-fixed, paraffin-embedded (FFPE) sections of pancreas from four healthy donors. The following RNAscope probes were used: Hs-GCG (ACD556741), hs-SST (ACD310591), hs-FAP (ACD411971), and hs-LEPR (ACD406371). Images were acquired on a Nikon A1R confocal microscope. See Table S3 for detailed information about how each cell was counted together with the statistics per slide and donor.

Processing of RNA-Seq Data

Sequence reads were aligned toward the human genome (hg19 assembly) using STAR (v2.3.0e), and uniquely aligned reads within RefSeq gene annotations were used to quantify gene expression as RPKMs using rpkmforgenes (Ramsköld et al., 2009).

Cell-Type Classification

Cell-type classification was defined in a two-dimensional t-SNE space that was constructed using a set of genes with highest biological variation. The assignment of the formed clusters to cell types was performed based on the expression levels of hormones or other known marker genes. The cell-type classification of all cells analyzed is included in Table S1.

Differential Expression Analysis

Differential expression analysis between cell types (α, β, γ, δ, acinar, and ductal cells) was performed using one-way ANOVA with $\log_2$-transformed expression data from the five healthy male donors.

Analysis of Heterogeneity in the Cell Types

To investigate heterogeneity within cell types, single-cell expression data of each cell type independently was ranked based on biological variability, and using this as input, the cells were projected onto two dimensions with the t-SNE. In all cases, the obtained embedding was dominated by donor differences (Figure S6). We applied a parametric empirical Bayesian framework to remove donor differences and repeated the t-SNE dimensionality reduction using the adjusted expression values.

Analyses of Gene Expression against BMI

Spearman's ρ coefficients were calculated separately for each gene and cell type based on the $\log_2$-transformed expression values. The results were adjusted using Benjamini-Hochberg to control the false discovery rate ($\alpha = 0.01$). The five healthy male donors were used, since there is a positive correlation between BMI and sex in our donors ($\rho = 0.49$). The lists containing genes with an absolute correlation coefficient with BMI greater than 0.5 are included in Table S5. We also generated gene correlations using all cells (not separated per cell type) from the same five male donors to simulate the correlations one would obtain in "whole-islet" analyses.

Differential Expression Analysis between Cells from Healthy and T2D Donors

We identified statistically significant gene expression differences between healthy and T2D individuals for the major cell types: α, β, γ, δ, acinar, and ductal cells. Non-parametric one-way ANOVA (Kruskal-Wallis test) was used after dividing samples into four groups based on disease status (healthy or T2D) and sex. Bonferroni adjustment was used for the multiple comparison correction and the Benjamini-Hochberg method to control the false discovery rate at significance level $\alpha = 0.01$. The genes that showed significant differences in respect to sex were excluded from the results in both tests in order to identify the differentially expressed (DE) genes related to T2D.

GSEA in Cell Types

GSEA was used to examine whether the genes identified as DE between the healthy and T2D cells for each cell type are members of categories with specific functions. We used the pre-ranked version, providing the difference in median expression values between the healthy and T2D cells as the gene ranking metric. The significant categories (FDR $\leq$ 1%) that are enriched in each cell type for the two conditions are reported in Table S7.

ACCESSION NUMBERS

Raw data (Fastq files) for single-cell and whole-islet RNA-seq have been submitted to ArrayExpress (EBI) with accession numbers ArrayExpress: E-MTAB-5061 and E-MTAB-5060, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.cmet.2016.08.020.

AUTHOR CONTRIBUTIONS

P.E., A.-C.A., E.-M.A., M.C., and A.S. handled islets and cell sorting, and performed functional in vitro and ex vivo assessments. M.K.B., D.M.S., C.Ä. and

R.S. initiated and supervised the work. Å.S. and S.P. prepared sequencing libraries. X.S. and Å.S. performed RNA FISH. M.K., X.S., Å.S., and A.S. analyzed the imaging data. A.P. performed computational experiments. Å.S., A.P., and R.S. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Olov Andersson for helpful comments on our study, the Eukaryotic Single-Cell Genomics facility at SciLifeLab Stockholm for generating single-cell sequencing libraries, and the Live Cell Imaging facility/Nikon Center of Excellence at Karolinska Institutet, Huddinge. The c-peptide hybridoma monoclonal antibody, developed by Madsen, O.D., was obtained from the Developmental Studies Hybridoma Bank, University of Iowa. This work was supported by grants from the Swedish Research Council (R.S.), European Research Council (648842; R.S.) and the Swedish Foundation for Strategic Research (M.K. and R.S.), Swedish Cancer Society (M.K.), and Center for Innovative Medicine (M.K.). P.E., E.-M.A., A.-C.A., A.S., M.C., M.K.B., and C.Ä. are employees and shareholders of AstraZeneca AB. D.M.S. is an employee and shareholder of AstraZeneca Plc.

Received: March 6, 2016
Revised: June 13, 2016
Accepted: August 26, 2016
Published: September 22, 2016

REFERENCES

Arystarkhova, E., Liu, Y.B., Salazar, C., Stanojevic, V., Clifford, R.J., Kaplan, J.H., Kidder, G.M., and Sweadner, K.J. (2013). Hyperplasia of pancreatic beta cells and improved glucose tolerance in mice deficient in the FXYD2 subunit of Na,K-ATPase. J. Biol. Chem. *288*, 7077–7085.

Benner, C., van der Meulen, T., Cacéres, E., Tigyi, K., Donaldson, C.J., and Huising, M.O. (2014). The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics *15*, 620.

Blodgett, D.M., Nowosielska, A., Afik, S., Pechhold, S., Cura, A.J., Kennedy, N.J., Kim, S., Kucukural, A., Davis, R.J., Kent, S.C., et al. (2015). Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. Diabetes *64*, 3172–3181.

Braks, J.A., and Martens, G.J. (1994). 7B2 is a neuroendocrine chaperone that transiently interacts with prohormone convertase PC2 in the secretory pathway. Cell *78*, 263–273.

Bramswig, N.C., Everett, L.J., Schug, J., Dorrell, C., Liu, C., Luo, Y., Streeter, P.R., Naji, A., Grompe, M., and Kaestner, K.H. (2013). Epigenomic plasticity enables human pancreatic α to β cell reprogramming. J. Clin. Invest. *123*, 1275–1284.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095.

Cabrera, O., Berman, D.M., Kenyon, N.S., Ricordi, C., Berggren, P.-O., and Caicedo, A. (2006). The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc. Natl. Acad. Sci. USA *103*, 2334–2339.

Chu, Z.-L., Carroll, C., Alfonso, J., Gutierrez, V., He, H., Lucman, A., Pedraza, M., Mondala, H., Gao, H., Bagnol, D., et al. (2008). A role for intestinal endocrine cell-expressed g protein-coupled receptor 119 in glycemic control by enhancing glucagon-like peptide-1 and glucose-dependent insulinotropic peptide release. Endocrinology *149*, 2038–2047.

Collombat, P., Hecksher-Sørensen, J., Krull, J., Berger, J., Riedel, D., Herrera, P.L., Serup, P., and Mansouri, A. (2007). Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. J. Clin. Invest. *117*, 961–970.

Date, Y., Nakazato, M., Hashiguchi, S., Dezaki, K., Mondal, M.S., Hosoda, H., Kojima, M., Kangawa, K., Arima, T., Matsuo, H., et al. (2002). Ghrelin is present in pancreatic alpha-cells of humans and rats and stimulates insulin secretion. Diabetes *51*, 124–129.

Dorrell, C., Schug, J., Lin, C.F., Canaday, P.S., Fox, A.J., Smirnova, O., Bonnah, R., Streeter, P.R., Stoeckert, C.J., Jr., Kaestner, K.H., and Grompe, M. (2011). Transcriptomes of the major human pancreatic cell types. Diabetologia *54*, 2832–2844.

Eto, K., Tsubamoto, Y., Terauchi, Y., Sugiyama, T., Kishimoto, T., Takahashi, N., Yamauchi, N., Kubota, N., Murayama, S., Aizawa, T., et al. (1999). Role of NADH shuttle system in glucose-induced activation of mitochondrial metabolism and insulin secretion. Science *283*, 981–985.

Fasshauer, M., and Blüher, M. (2015). Adipokines in health and disease. Trends Pharmacol. Sci. *36*, 461–470.

Gao, J., Tian, L., Weng, G., O'Brien, T.D., Luo, J., and Guo, Z. (2011). Stimulating β-cell replication and improving islet graft function by AR231453, A GPR119 agonist. Transplant. Proc. *43*, 3217–3220.

Gilon, P., Chae, H.-Y., Rutter, G.A., and Ravier, M.A. (2014). Calcium signaling in pancreatic β-cells in health and in Type 2 diabetes. Cell Calcium *56*, 340–361.

Granata, R., Settanni, F., Biancone, L., Trovato, L., Nano, R., Bertuzzi, F., Destefanis, S., Annunziata, M., Martinetti, M., Catapano, F., et al. (2007). Acylated and unacylated ghrelin promote proliferation and inhibit apoptosis of pancreatic β-cells and human islets: involvement of 3′,5′-cyclic adenosine monophosphate/protein kinase A, extracellular signal-regulated kinase 1/2, and phosphatidyl inositol 3-Kinase/Akt signaling. Endocrinology *148*, 512–529.

Jiang, Z., Song, J., Qi, F., Xiao, A., An, X., Liu, N.-A., Zhu, Z., Zhang, B., and Lin, S. (2008). Exdpf is a key regulator of exocrine pancreas development controlled by retinoic acid and ptf1a in zebrafish. PLoS Biol. *6*, e293.

Kahn, S.E., Hull, R.L., and Utzschneider, K.M. (2006). Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature *444*, 840–846.

Kambayashi, T., and Laufer, T.M. (2014). Atypical MHC class II-expressing antigen-presenting cells: can anything replace a dendritic cell? Nat. Rev. Immunol. *14*, 719–730.

Kieffer, T.J., Heller, R.S., and Habener, J.F. (1996). Leptin receptors expressed on pancreatic beta-cells. Biochem. Biophys. Res. Commun. *224*, 522–527.

Kieffer, T.J., Heller, R.S., Leech, C.A., Holz, G.G., and Habener, J.F. (1997). Leptin suppression of insulin secretion by the activation of ATP-sensitive K+ channels in pancreatic β-cells. Diabetes *46*, 1087–1093.

Klöting, N., Graham, T.E., Berndt, J., Kralisch, S., Kovacs, P., Wason, C.J., Fasshauer, M., Schön, M.R., Stumvoll, M., Blüher, M., and Kahn, B.B. (2007). Serum retinol-binding protein is more highly expressed in visceral than in subcutaneous adipose tissue and is a marker of intra-abdominal fat mass. Cell Metab. *6*, 79–87.

Ku, G.M., Kim, H., Vaughn, I.W., Hangauer, M.J., Myung Oh, C., German, M.S., and McManus, M.T. (2012). Research resource: RNA-seq reveals unique features of the pancreatic β-cell transcriptome. Mol. Endocrinol. *26*, 1783–1792.

Li, J., Klughammer, J., Farlik, M., Penz, T., Spittler, A., Barbieux, C., Berishvili, E., Bock, C., and Kubicek, S. (2016). Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. EMBO Rep. *17*, 178–187.

Ling, F., Kang, B., and Sun, X.-H. (2014). Id proteins: small molecules, mighty regulators. Curr. Top. Dev. Biol. *110*, 189–216.

Liu, Y., and Du, L. (2015). Role of pancreatic stellate cells and periostin in pancreatic cancer progression. Tumour Biol. *36*, 3171–3177.

Montanya, E. (2012). A comparison of currently available GLP-1 receptor agonists for the treatment of type 2 diabetes. Expert Opin. Pharmacother. *13*, 1451–1467.

Moran, B.M., Abdel-Wahab, Y.H.A., Flatt, P.R., and McKillop, A.M. (2014). Evaluation of the insulin-releasing and glucose-lowering effects of GPR120 activation in pancreatic β-cells. Diabetes Obes. Metab. *16*, 1128–1139.

Nadkarni, P., Chepurny, O.G., and Holz, G.G. (2014). Regulation of glucose homeostasis by GLP-1. Prog. Mol. Biol. Transl. Sci. *121*, 23–65.

Nead, K.T., Li, A., Wehner, M.R., Neupane, B., Gustafsson, S., Butterworth, A., Engert, J.C., Davis, A.D., Hegele, R.A., Miller, R., et al.; BioBank Japan, AGEN-BMI, GIANT Consortium (2015). Contribution of common non-synonymous

variants in PCSK1 to body mass index variation and risk of obesity: a systematic review and meta-analysis with evidence from up to 331,175 individuals. Hum. Mol. Genet. *24*, 3582–3594.

Nica, A.C., Ongen, H., Irminger, J.-C., Bosco, D., Berney, T., Antonarakis, S.E., Halban, P.A., and Dermitzakis, E.T. (2013). Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res. *23*, 1554–1562.

Odori, S., Hosoda, K., Tomita, T., Fujikura, J., Kusakabe, T., Kawaguchi, Y., Doi, R., Takaori, K., Ebihara, K., Sakai, Y., et al. (2013). GPR119 expression in normal human tissues and islet cell tumors: evidence for its islet-gastrointestinal distribution, expression in pancreatic beta and alpha cells, and involvement in islet function. Metabolism *62*, 70–78.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat. Biotechnol. *32*, 1053–1058.

Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. *5*, e1000598.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Ruiz de Azua, I., Scarselli, M., Rosemond, E., Gautam, D., Jou, W., Gavrilova, O., Ebert, P.J., Levitt, P., and Wess, J. (2010). RGS4 is a negative regulator of insulin release from pancreatic beta-cells in vitro and in vivo. Proc. Natl. Acad. Sci. USA *107*, 7999–8004.

Sandberg, R. (2014). Entering the era of single-cell transcriptomics in biology and medicine. Nat. Methods *11*, 22–24.

Séron, K., Couturier, C., Belouzard, S., Bacart, J., Monté, D., Corset, L., Bocquet, O., Dam, J., Vauthier, V., Lecœur, C., et al. (2011). Endospanins regulate a postinternalization step of the leptin receptor endocytic pathway. J. Biol. Chem. *286*, 17968–17981.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Taneera, J., Lang, S., Sharma, A., Fadista, J., Zhou, Y., Ahlqvist, E., Jonsson, A., Lyssenko, V., Vikman, P., Hansson, O., et al. (2012). A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. Cell Metab. *16*, 122–134.

Touvier, T., Conte-Auriol, F., Briand, O., Cudejko, C., Paumelle, R., Caron, S., Baugé, E., Rouillé, Y., Salles, J.-P., Staels, B., and Bailleul, B. (2009). LEPROT and LEPROTL1 cooperatively decrease hepatic growth hormone action in mice. J. Clin. Invest. *119*, 3830–3838.

Tudurí, E., Marroquí, L., Soriano, S., Ropero, A.B., Batista, T.M., Piquer, S., López-Boado, M.A., Carneiro, E.M., Gomis, R., Nadal, A., and Quesada, I. (2009). Inhibitory effects of leptin on pancreatic alpha-cell function. Diabetes *58*, 1616–1624.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning. *9*, 2579–2605.

van der Meulen, T., Donaldson, C.J., Cáceres, E., Hunter, A.E., Cowing-Zitron, C., Pound, L.D., Adams, M.W., Zembrzycki, A., Grove, K.L., and Huising, M.O. (2015). Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. Nat. Med. *21*, 769–776.

Vauthier, V., Swartz, T.D., Chen, P., Roujeau, C., Pagnon, M., Mallet, J., Sarkis, C., Jockers, R., and Dam, J. (2014). Endospanin 1 silencing in the hypothalamic arcuate nucleus contributes to sustained weight loss of high fat diet obese mice. Gene Ther. *21*, 638–644.

Wang, J., Xu, J., Finnerty, J., Furuta, M., Steiner, D.F., and Verchere, C.B. (2001). The prohormone convertase enzyme 2 (PC2) is essential for processing pro-islet amyloid polypeptide at the NH2-terminal cleavage site. Diabetes *50*, 534–539.

Wang, Y.J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., Golson, M.L., and Kaestner, K.H. (2016). Single cell transcriptomics of the human endocrine pancreas. Diabetes, Published online June 30, 2016. http://dx.doi.org/10.2337/db16-0405.

Wei, S., Feng, Y., Che, F.-Y., Pan, H., Mzhavia, N., Devi, L.A., McKinzie, A.A., Levin, N., Richards, W.G., and Fricker, L.D. (2004). Obesity and diabetes in transgenic mice expressing proSAAS. J. Endocrinol. *180*, 357–368.

Wetzel, R.K., Pascoa, J.L., and Arystarkhova, E. (2004). Stress-induced expression of the gamma subunit (FXYD2) modulates Na,K-ATPase activity and cell growth. J. Biol. Chem. *279*, 41750–41757.

Wierup, N., Yang, S., McEvilly, R.J., Mulder, H., and Sundler, F. (2004). Ghrelin is expressed in a novel endocrine cell type in developing rat islets and inhibits insulin secretion from INS-1 (832/13) cells. J. Histochem. Cytochem. *52*, 301–310.

Yang, Q., Graham, T.E., Mody, N., Preitner, F., Peroni, O.D., Zabolotny, J.M., Kotani, K., Quadro, L., and Kahn, B.B. (2005). Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. Nature *436*, 356–362.

Cell Metabolism, Volume *24*

Supplemental Information

Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes

Åsa Segerstolpe, Athanasia Palasantza, Pernilla Eliasson, Eva-Marie Andersson, Anne-Christine Andréasson, Xiaoyan Sun, Simone Picelli, Alan Sabirsh, Maryam Clausen, Magnus K. Bjursell, David M. Smith, Maria Kasper, Carina Ämmälä, and Rickard Sandberg

Figure S1

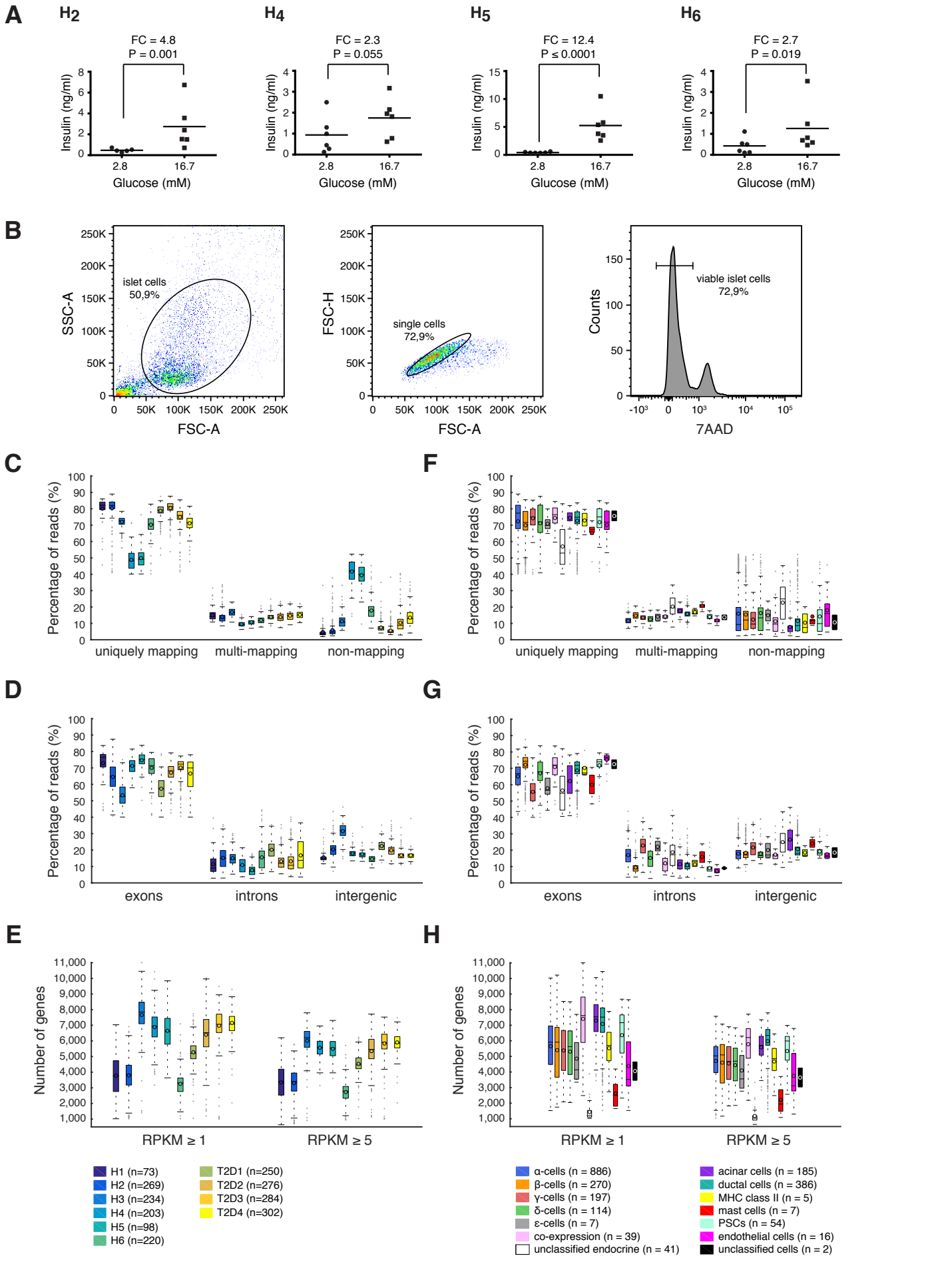

A
H2
FC = 4.8
P = 0.001
H4
FC = 2.3
P = 0.055
H5
FC = 12.4
P ≤ 0.0001
H6
FC = 2.7
P = 0.019
Insulin (ng/ml)
Glucose (mM)
2.8
16.7
B
islet cells
50,9%
SSC-A
FSC-A
single cells
72,9%
FSC-H
viable islet cells
72,9%
Counts
7AAD
C
F
Percentage of reads (%)
uniquely mapping
multi-mapping
non-mapping
D
G
exons
introns
intergenic
E
H
Number of genes
RPKM ≥ 1
RPKM ≥ 5
H1 (n=73)
H2 (n=269)
H3 (n=234)
H4 (n=203)
H5 (n=98)
H6 (n=220)
T2D1 (n=250)
T2D2 (n=276)
T2D3 (n=284)
T2D4 (n=302)
α-cells (n = 886)
β-cells (n = 270)
γ-cells (n = 197)
δ-cells (n = 114)
ε-cells (n = 7)
co-expression (n = 39)
unclassified endocrine (n = 41)
acinar cells (n = 185)
ductal cells (n = 386)
MHC class II (n = 5)
mast cells (n = 7)
PSCs (n = 54)
endothelial cells (n = 16)
unclassified cells (n = 2)

Figure S1. **Functionality of pancreatic islets and sequence library statistics. Related to Experimental Procedures.**

(**A**) Dot plots showing glucose-stimulated insulin secretion (GSIS) of human islets after 6 to 8 days in culture to assess functionality. (**B**) FACS analysis and sorting of dissociated human islets. Cells were distinguished from debris in a forward scatter (FSC)-side scatter (SSC) plot followed by discrimination of aggregates using FSC-H and FSC-A. Exclusion of nonviable cells was performed after staining cells with 7-aminoactinomycin D (7AAD). **(C)** Boxplots showing the percentage of sequenced reads that aligned uniquely to human genome, multi-mapping or non-mapping for each donor. **(D)** Boxplots showing the percentage of uniquely aligned reads that overlap annotated RefSeq exons, introns or between gene annotations for each donor. **(E)** Boxplots with the number of genes expressed across the cells from each donor using two different expression thresholds (RPKM≥1 and RPKM≥5). **(F-H)** Boxplots illustrating the mapping statistics as in (C-E) respectively, but cells are grouped according to the identified cell types.

Figure S2

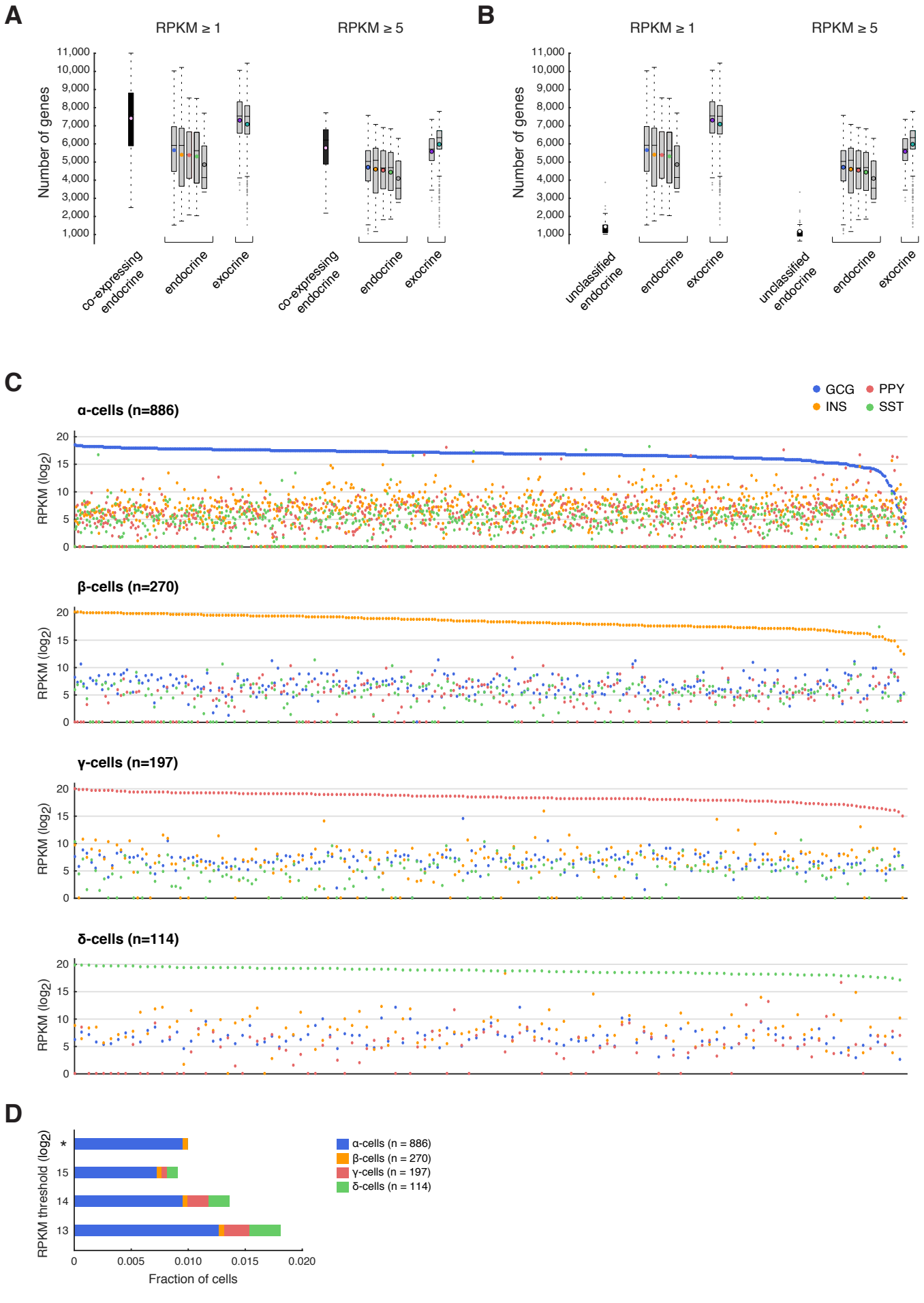

A
RPKM ≥ 1
RPKM ≥ 5
Number of genes
co-expressing endocrine
endocrine
exocrine
B
RPKM ≥ 1
RPKM ≥ 5
Number of genes
unclassified endocrine
endocrine
exocrine
C
GCG
PPY
INS
SST
α-cells (n=886)
β-cells (n=270)
γ-cells (n=197)
δ-cells (n=114)
RPKM (log2)
D
RPKM threshold (log2)
Fraction of cells
α-cells (n = 886)
β-cells (n = 270)
γ-cells (n = 197)
δ-cells (n = 114)

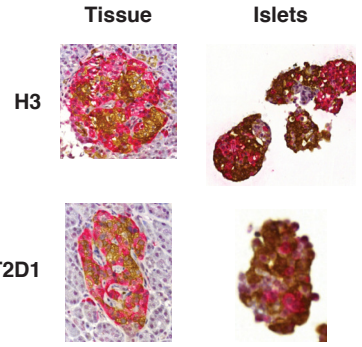
Figure S2. Filtering of single-cell RNA-sequencing data. Related to Figure 1.
(**A**) Boxplots showing the number of genes expressed across the cells from "co-expressing" endocrine, endocrine and exocrine cell types using two different thresholds (RPKM≥1 and RPKM≥5). The "co-expressing" cells have a larger number of genes expressed compared to the other endocrine cells, in line with the potential sorting of cell doublets. (**B**) Boxplots showing the number of genes expressed across the cells from unclassified endocrine, endocrine and exocrine cell types using two different thresholds (RPKM≥1 and RPKM≥5). The unclassified endocrine cells express very few genes, which could have resulted from failures to accurately amplify the cellular RNA or that only parts of a cell was deposited into these wells of the plate during FACS distribution. Both the "co-expressing" and unclassified endocrine cells were excluded from the analyses of the paper. **(C)** Scatter plots showing the expression of the four endocrine hormones: *GCG* (blue), *INS* (orange), *PPY* (red) and *SST* (green), across α, β, γ and δ-cells. In each graph, cells (x-axis) are sorted based on decreasing expression of the cell-type specific hormone. **(D)** Bars showing the fraction of cells in which the expression of at least one hormone of other cell types is above the corresponding threshold (log_2RPKM). Bars denoted with asterisk (*) show the number of cells in which the expression of at least one hormone of other cell types is higher than the cell-type specific hormone expression. Calculations were performed considering only cells from the α, β, γ and δ cell types. Colors indicate the fraction of cells in each cell type. Fractions were computed using the number of all sequenced cells (n=2,209).

Figure S3

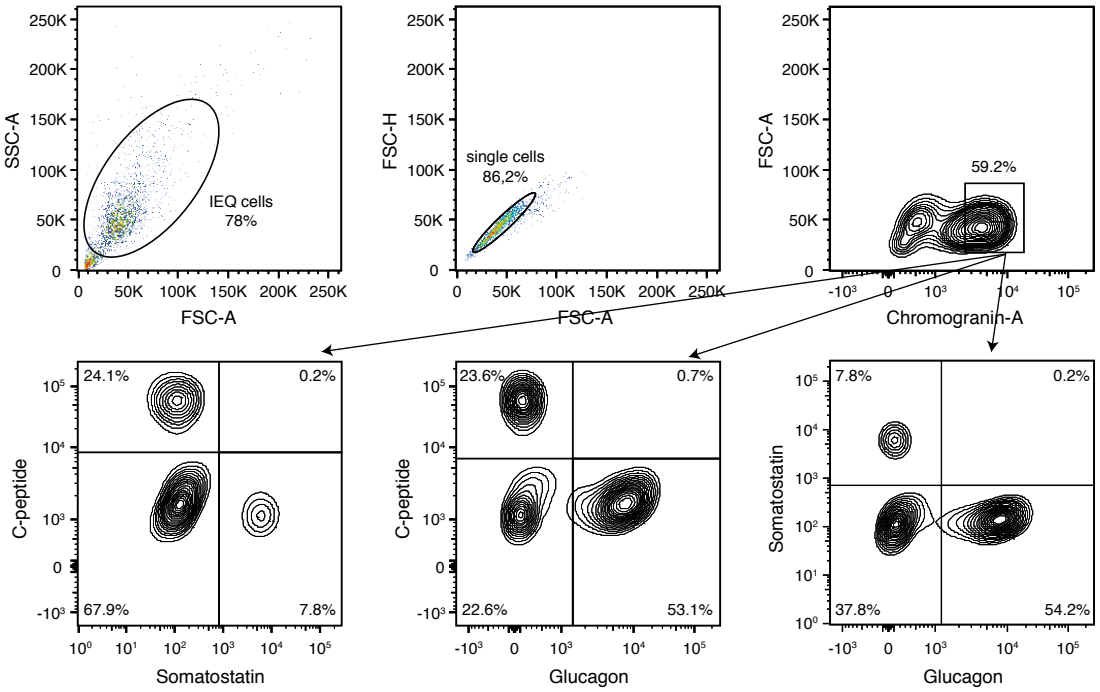
A

	H_1	H_2	H_3	H_4	H_5	H_6	$T2D_1$	$T2D_2$	$T2D_3$	$T2D_4$	total (cell type)
α-cells	28	117	26	136	44	92	141	119	87	96	886
β-cells	12	48	32	34	10	35	10	14	11	64	270
γ-cells	7	19	15	2	1	31	70	8	10	34	197
δ-cells	7	21	2	7	10	12	9	6	5	35	114
ε-cells	0	1	1	0	0	3	1	0	0	1	7
co-expression	3	3	5	6	3	6	1	5	1	6	39
unclassified endocrine cells	5	15	4	0	0	5	3	3	6	0	41
acinar cells	4	20	80	3	2	3	8	28	24	13	185
ductal cells	4	19	67	8	23	14	3	76	125	47	386
antigen presenting-MHC class II	1	0	0	0	0	0	0	2	1	1	5
mast cells	0	4	0	0	0	0	0	2	0	1	7
PSCs	1	1	2	6	3	10	2	12	13	4	54
endothelial cells	1	1	0	1	2	8	1	1	1	0	16
unclassified exocrine cells	0	0	0	0	0	1	1	0	0	0	2
total (donor)	73	269	234	203	98	220	250	276	284	302	2209

B



C



D

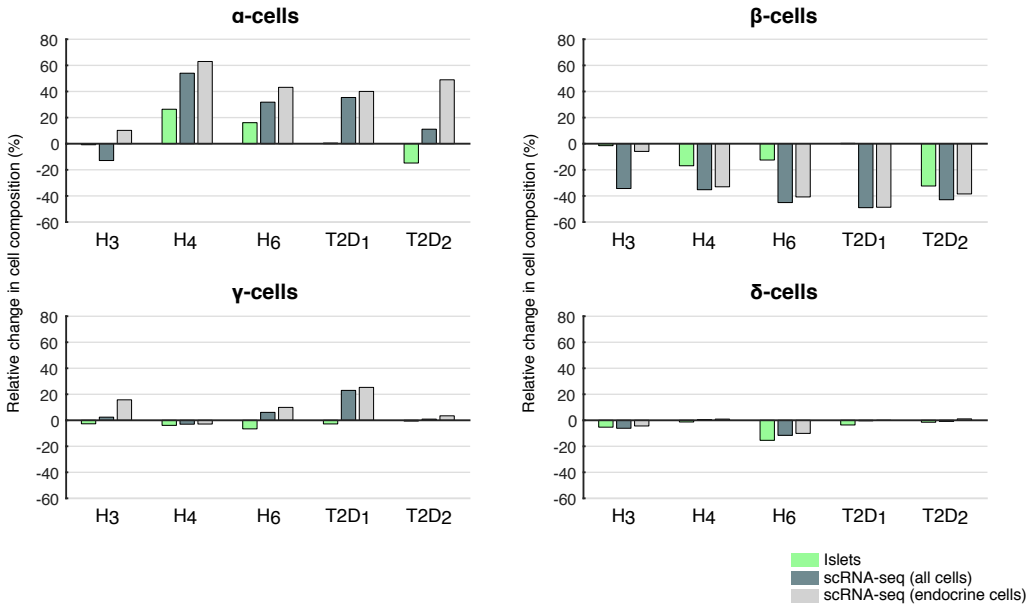
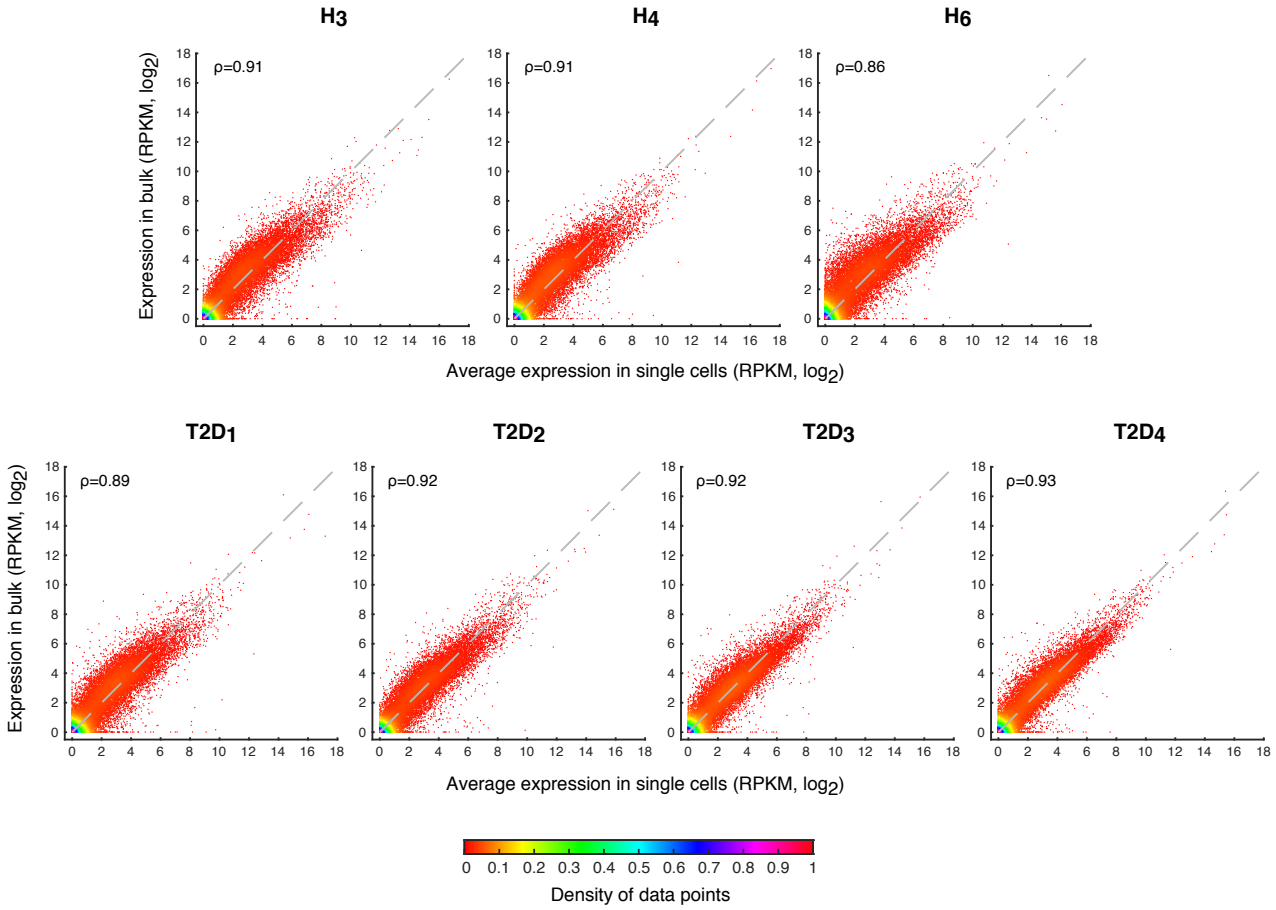


Figure S3. Cell-type composition in single-cell sequencing, tissue and dissociated islets. Related to Figure 1.

(**A**) Table with the number of cells for each cell type and donor. (**B**) (Left) Pancreatic tissue from a healthy (H3) and a T2D (T2D1) donor, respectively. (Right) Isolated islet from the two donors. Immunohistochemistry (IHC) of pancreatic tissue and isolated pancreatic endocrine islets stained for glucagon (red) and insulin (brown), labeling α and β-cells, respectively. (**C**) FACS analysis of human islet cells labeled with anti-chromogranin A, anti-c-peptide, anti-somatostatin and anti-glucagon antibodies. Percentage of each parental population is indicated in gates for each graph. (**D**) The relative change in abundance of α, β, γ and δ cell types based on islet and single-cell RNA-sequencing material from three healthy and two T2D donors in respect to the tissue. Tissue and islets were stained with IHC and quantified with BioPix software. FACS-sorted single pancreatic cells were analyzed using single-cell RNA-sequencing. Y-axis shows relative change of expression in relation to the tissue signal for the different cell types and donors. For the single-cell RNA-sequencing data, α, β, γ and δ-cells from individual donors were grouped and quantified based on their gene expression profiles in relation to all cells (all cells) or only the endocrine part (endocrine cells) of the donors.

Figure S4

A


H3
H4
H6
ρ=0.91
ρ=0.91
ρ=0.86
Expression in bulk (RPKM, log2)
Average expression in single cells (RPKM, log2)
T2D1
T2D2
T2D3
T2D4
ρ=0.89
ρ=0.92
ρ=0.92
ρ=0.93
Expression in bulk (RPKM, log2)
Average expression in single cells (RPKM, log2)
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Density of data points

B

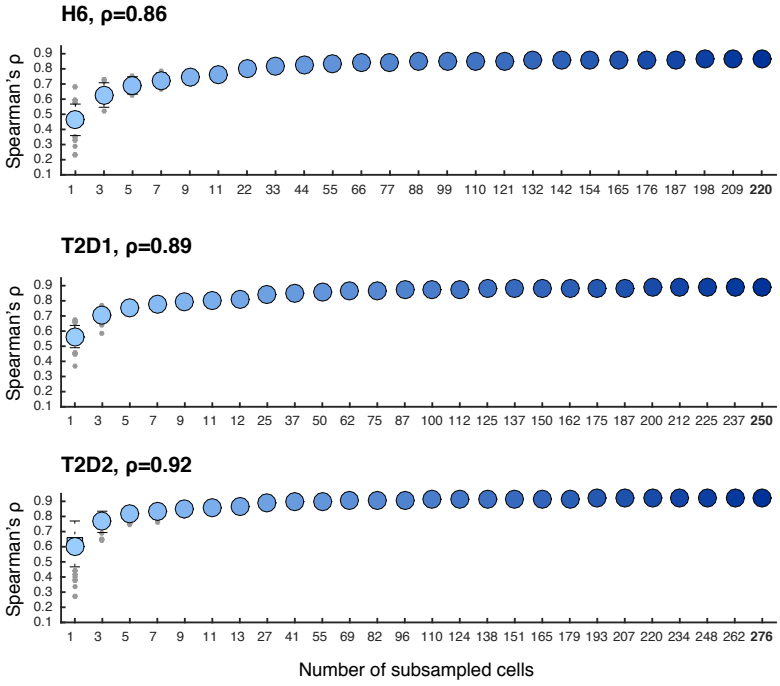
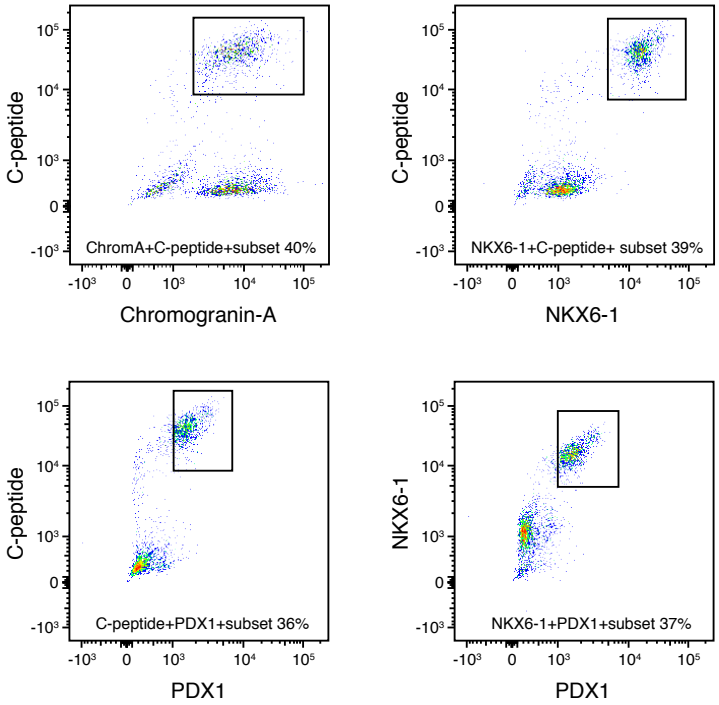

H6, ρ=0.86
T2D1, ρ=0.89
T2D2, ρ=0.92
Spearman's ρ
Number of subsampled cells

Figure S4. Comparison of single-cell and whole-islet RNA-sequencing data. Related to Experimental Procedures.

(A) Scatter plots showing the gene expression levels averaged across single cells (x-axis) against the whole-islet RNA-sequencing data (y-axis) of the same individual. Values are in $\log_2$-scale. Spearman's correlation (ρ) between the expression levels of the two methods is indicated in each graph. Colors correspond to the density of the data (red: sparse, pink: dense). **(B)** Investigating the (minimum) number of single cells needed to obtain a reliable view of the whole islet. Boxplots showing the Spearman's correlation coefficients between single-cell and bulk expression data, computed using different number of single cells. For each different number of cells tested, 200 iterations were performed in which cells were selected at random from each individual donor and compared with the corresponding bulk data. Data is shown for 3 donors (H6, T2D1 and T2D2).

Figure S5

A



B

α-cells: Top 25 genes

	rank	p-value
TTR	40	0,000
SSR4	3274	0,068
CRYBA2	14	0,000
SPINT2	704	0,000
PEMT	3020	0,052
CHGA	135	0,000
GPX3	167	0,000
TMEM176B	33	0,000
GC	174	0,000
FKBP2	1348	0,003
PCSK2	16	0,000
NAA20	91	0,000
TMEM176A	41	0,000
ERP29	299	0,000
DPM3	4083	0,148
TMED3	970	0,001
CNPY2	756	0,000
C10orf10	4741	0,235
TUBA1B	18591	1,000
SMIM24	-	-
F10	442	0,000
FXYD5	1543	0,005
CD46	258	0,000
SLC22A17	141	0,000
FXYD3	4284	0,173

γ-cells: Top 25 genes

	rank	p-value
MALAT1	15862	1,000
SCG2	1409	1,000
SCGB2A1	1414	1,000
PAM	1770	1,000
GPC5-AS1	-	-
STMN2	7552	1,000
PAX6	13384	1,000
MEIS2	10262	1,000
CMTM8	13372	1,000
TTC3	17602	1,000
ARX	8448	1,000
FGFR1	16562	1,000
AKAP9	14793	1,000
ETV1	5625	1,000
PPY2	8	1,000
NCKAP1	6384	1,000
INPP5F	10846	1,000
PXK	24	1,000
ID4	17260	1,000
SERTM1	-	-
SLITRK6	2	0,043
SEMA3E	4475	1,000
APOBEC2	6251	1,000
ABCC9	13932	1,000
REV3L	14924	1,000

δ-cells: Top 25 genes

	rank	p-value
RBP4	388	0,000
SEC11C	1273	0,000
PCP4	1927	0,000
RGS2	15254	1,000
HHEX	12902	1,000
TPPP3	1129	0,000
UCP2	1001	0,000
LEPR	486	0,000
BAIAP3	834	0,000
MS4A8	-	-
CASR	450	0,000
PSIP1	3648	0,018
BCHE	265	0,000
GABRB3	115	0,000
LY6H	17259	1,000
UNC5B	9173	1,000
EDN3	2478	0,002
OGDHL	866	0,000
NSG1	-	-
FFAR4	-	-
LINC00643	-	-
LINC01014	-	-
TMEM130	2785	0,003
PRG4	9517	1,000
TENM3	-	-

acinar cells: Top 25 genes

	rank	p-value
SPINK1	6	0,000
PRSS1	36	0,000
REG3A	2942	0,092
CTRB2	333	0,000
SERPINA3	245	0,000
RNASE1	142	0,000
IL32	60	0,000
PRSS3	489	0,000
REG1B	1	0,000
PRSS3P2	-	-
CTRB1	509	0,000
CFB	4234	0,348
GDF15	1858	0,018
MUC1	6894	1,000
C15orf48	3929	0,262
DUOXA2	3753	0,224
AKR1C3	1470	0,008
CPA2	213	0,000
OLFM4	4645	0,461
GSTA1	2646	0,063
LGALS2	6807	1,000
MGST1	44	0,000
PDZK1IP1	1194	0,003
SOD2	451	0,000
RARRES2	518	0,000

ductal cells: Top 25 genes

	rank	p-value
SPP1	15616	1,000
MMP7	1284	0,002
ANXA4	6526	1,000
ANXA2	3	0,000
DEFB1	3055	0,067
SERPING1	7153	1,000
ATP5B	10414	1,000
ANXA2P2	23	0,000
TSPAN8	3231	0,083
CLDN10	4363	0,279
IFITM2	1668	0,005
CTSH	13512	1,000
SERPINA1	13999	1,000
CD59	704	0,000
SLPI	1671	0,005
S100A16	10	0,000
CD9	1000	0,001
MIR492	-	-
SERPINA5	18157	1,000
PIGR	7624	1,000
PPAP2C	190	0,000
CFTR	9415	1,000
S100A13	3791	0,162
AMBP	13211	1,000
CLDN1	80	0,000

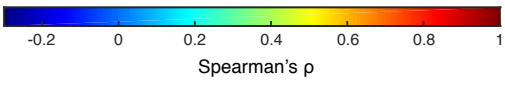


Figure S5. FACS analysis of Pdx1 and Nkx6-1 positive β-cells and assessment of cell-type enriched expression inferred from whole-islet data. Related to Experimental Procedures.
(A) FACS analysis of dissociated human islets cells labelled with anti-c-peptide, anti-chromogranin-A, anti-Nkx6-1 and anti-Pdx1 antibodies. The β-cell population was distinguished in the double-positive scatter plots for c-peptide and chromogranin-A (top left FACS graph). Nkx6-1 and Pdx1 positive scatters coincide with the c-peptide positive β-cell population (top right and bottom left FACS graphs). The same cell population is also Nkx6-1 and Pdx1 double-positive (bottom right FACS graph). **(B)** Correlation between cell type hormone expression and gene expression using the data published in Taneera et al., 2012. Spearman's correlation coefficients are shown for the top 25 genes in each cell type with the corresponding hormone or marker gene (*GCG* for α, *PPY* for γ, *SST* for δ, *REG1A* for acinar and *KRT19* for ductal top genes). The rank obtained from sorting the genes according to correlation magnitude (absolute) in descending order and the adjusted p-value for each gene are displayed on the right of the heat maps. Colors in the heat map correspond to Spearman's correlation coefficients (ρ).

Figure S6

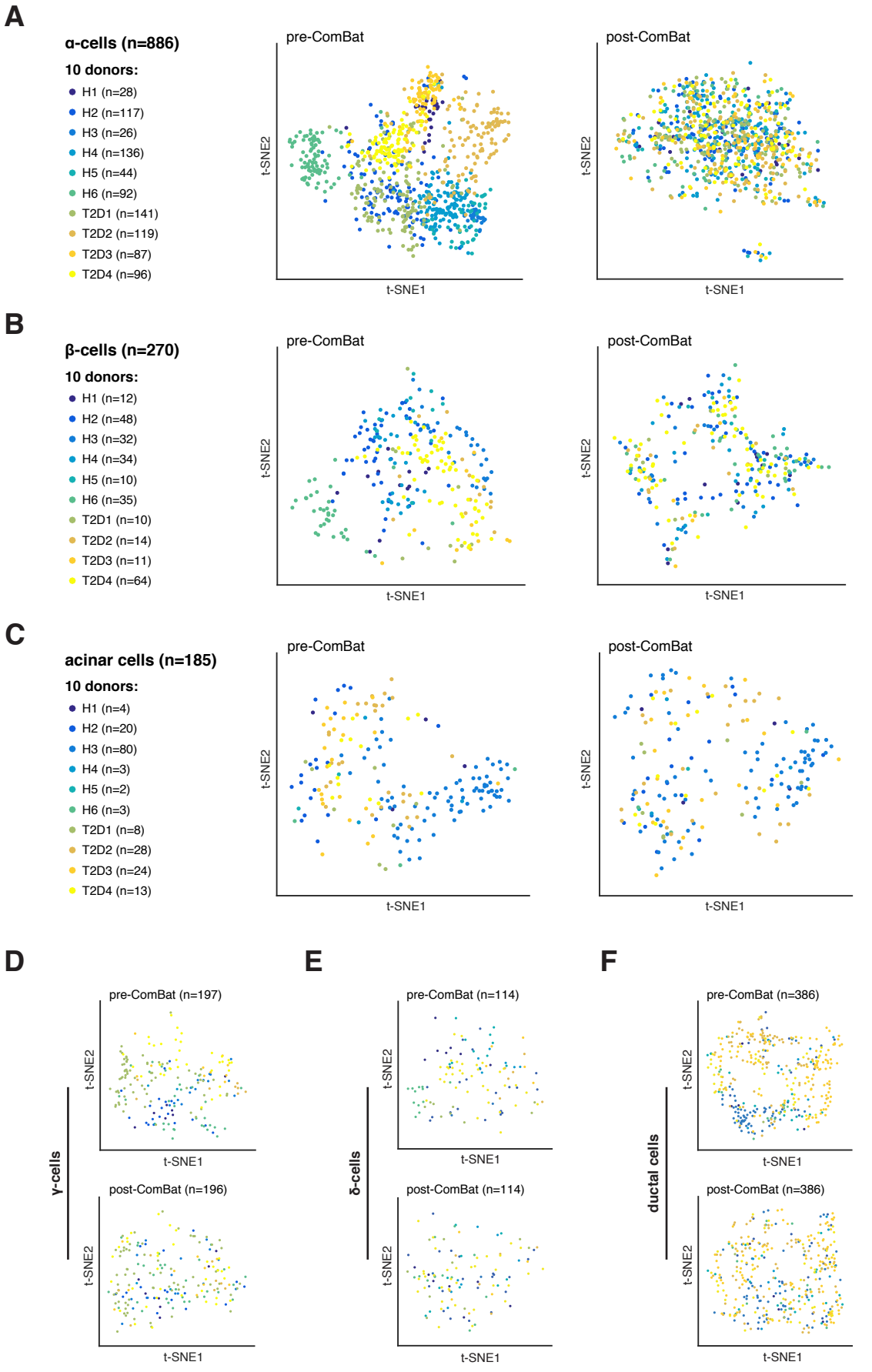

A
α-cells (n=886)
10 donors:
H1 (n=28)
H2 (n=117)
H3 (n=26)
H4 (n=136)
H5 (n=44)
H6 (n=92)
T2D1 (n=141)
T2D2 (n=119)
T2D3 (n=87)
T2D4 (n=96)
pre-ComBat
post-ComBat
t-SNE1
t-SNE2
B
β-cells (n=270)
10 donors:
H1 (n=12)
H2 (n=48)
H3 (n=32)
H4 (n=34)
H5 (n=10)
H6 (n=35)
T2D1 (n=10)
T2D2 (n=14)
T2D3 (n=11)
T2D4 (n=64)
C
acinar cells (n=185)
10 donors:
H1 (n=4)
H2 (n=20)
H3 (n=80)
H4 (n=3)
H5 (n=2)
H6 (n=3)
T2D1 (n=8)
T2D2 (n=28)
T2D3 (n=24)
T2D4 (n=13)
D
γ-cells
pre-ComBat (n=197)
post-ComBat (n=196)
E
δ-cells
pre-ComBat (n=114)
post-ComBat (n=114)
F
ductal cells
pre-ComBat (n=386)
post-ComBat (n=386)

Figure S6. Sub-clustering of cells before and after donor normalization. Related to Figure 4.
Projection of (**A**) α-cells, (**B**) β-cells, (**C**) acinar cells, (**D**) γ-cells (**E**) δ-cells and (**F**) ductal cells onto two dimensions using *t*-SNE. For each cell type, the two embeddings shown on the left and right were obtained using the expression values (log_2RPKM) of the most variable genes, before and after ComBat adjustment respectively. The colors correspond to the individual donors in order to illustrate the removal of donor effect on the resulting clusters following batch correction.

Figure S7

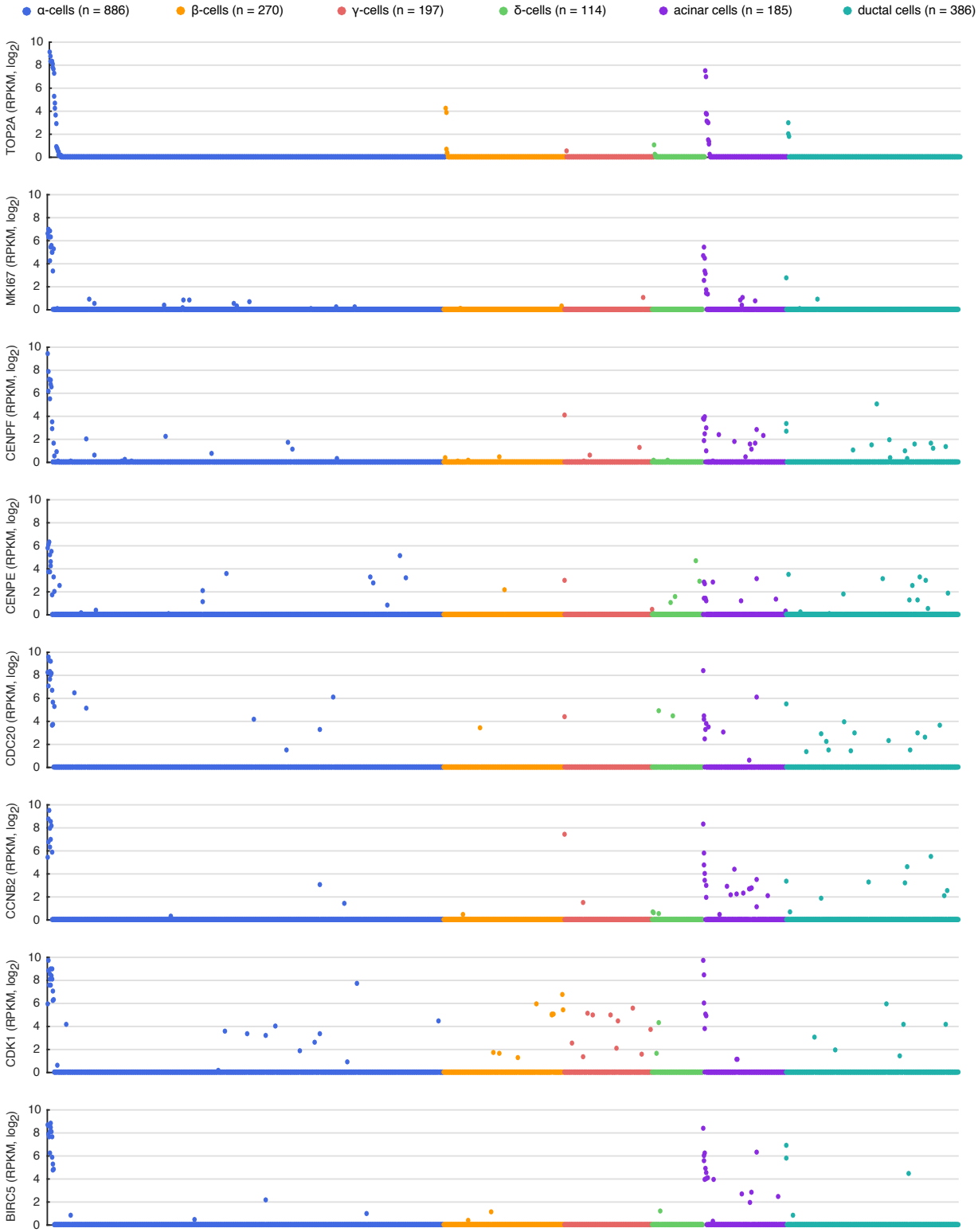

α-cells (n = 886)
β-cells (n = 270)
γ-cells (n = 197)
δ-cells (n = 114)
acinar cells (n = 185)
ductal cells (n = 386)
TOP2A (RPKM, log2)
MKI67 (RPKM, log2)
CENPF (RPKM, log2)
CENPE (RPKM, log2)
CDC20 (RPKM, log2)
CCNB2 (RPKM, log2)
CDK1 (RPKM, log2)
BIRC5 (RPKM, log2)
0
2
4
6
8
10

Figure S7. Expression of proliferation-associated genes. Related to Figure 4.
Scatter plots showing the expression of proliferation-associated genes across the six cell types: α, β, γ, δ, acinar and ductal cells. In all graphs, cells (x-axis) are sorted based on decreasing expression of *TOP2A* in each cell type. Colors correspond to cell types.

SUPPLEMENTAL TABLES

Table S1. Cell-type identification and gene expression. Related to Figure 1.
This table provides details on the identification of genes with most biological variation, genes with highest expression in cell types, donors and whole-islet RNA-sequencing of donors. Additionally, it contains the cell-type composition per donor and general statistics of each processed cell.

(supplied as Excel file: Supplemental Table 1.xlsx)

Table S2. Differential expression analysis of cell types. Related to Figure 2.
This supplemental table lists the genes identified as significantly cell-type enriched in each of the cell types (as separate sheets in the Excel file).

(supplied as Excel file: Supplemental Table 2.xlsx)

Table S3. Detailed analyses of single-molecule RNA-FISH images. Related to Figure 2.
Detailed information of the quantification of expression based on the RNA in situ hybridization experiments on *FAP*/*GCG* in α-cells and *LEPR*/*SST* in δ-cells. It summarizes the quantifications of multiple human islets in sections obtained from several donors. The images and details of the counting of each cell in each image are provided in the additional sheets in the Excel file (names by donor and RNA in situ targets).

(supplied as Excel file: Supplemental Table 3.xlsx)

Table S4. Differential expression analysis of the subpopulations within cell types. Related to Figure 4.
Lists of the genes that were identified as significantly differentially expressed between subpopulations within cell types. Results for each cell type are provided in separate sheets in the Excel file (α, β and acinar cells; the cell types for which we could identify robust subpopulations).

(supplied as Excel file: Supplemental Table 4.xlsx)

Table S5. Correlation of gene expression and BMI in the cell types. Related to Figure 5.
Details on the correlations of gene expression with BMI either for cells of each cell type or using all cells per donor.

(supplied as Excel file: Supplemental Table 5.xlsx)

Table S6. Differential expression analysis between healthy and T2D cells in each cell type. Related to Figure 6.
Lists of the genes identified as differentially expressed between healthy individuals and type 2 diabetes.

(supplied as Excel file: Supplemental Table 6.xlsx)

Table S7. Gene set enrichment analysis (GSEA). Related to Figure 6.
Detailed results from the Gene Set Enrichment Analysis (GSEA) performed on each cell type that are listed in separate sheets in the Excel file.

(supplied as Excel file: Supplemental Table 7.xlsx)

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Human tissue collection

Human primary islets from 6 healthy and 4 type 2 diabetic diseased donors were purchased from Prodo Laboratories Inc (Irvine, CA, USA) providing islet isolated from donor pancreases obtained with research consent from Organ Procurement Organizations (OPOs) (Kühtreiber et al., 2010) and kept in Prodo Islet Media Standard (PIM(S)) complete. The use and storage of human islets and tissue samples was performed in compliance with the Declaration of Helsinki, ICH/Good Clinical Practice and was approved by the independent Regional Ethics Committee (Gothenburg, Sweden).

Dissociation of islet cells

Human islet samples (85–95% pure) were cultured for 4 days in a 37°C incubator with 5% CO2 in PIM complete media to recover after arrival. Media was changed every second day. For preparation of a single-cell suspension, 300 islets were identified using a microscope and handpicked to a tube containing fresh culture media. Islets were allowed to sediment on ice for 5 minutes or gently centrifuged if necessary. Media was removed carefully and 1 mL pre-warmed TrypLE Express (Life Technologies, Darmstadt, Germany) was added. The islets were incubated for 5 minutes in a water bath at 37°C and triturated by resuspending 4-5 times during incubation. Remaining cell aggregates were removed with a 40 µm filter. Dissociated islet cells were washed once with culture media before FACS cell sorting.

FACS analysis of endocrine cells

Dissociated human islets cells were fixed with Fixation buffer I (BD Biosciences) at 37°C for 10 minutes. Cells were washed in DPBS without calcium and magnesium and permeabilized with Perm/Wash buffer (BD Biosciences). Cells were incubated over night at 4°C with primary antibodies, rabbit anti-chromogranin A (Abcam, Cambridge, UK), rat anti-c-peptide (Developmental Studies Hybridoma bank, The University of Iowa), rabbit anti-insulin (Cell Signaling technology, Beverly, MA), goat anti-somatostatin (Santa Cruz) and mouse anti-glucagon (Abcam). For labeling of GLP1R, we have used BODYPI FL dye tagged GLP1R antagonist (Exendin 9-39) conjugate produced in a similar way as previously described (Montrose-Rafizadeh et al., 1997). After washing twice in BD Perm/wash buffer, cells were stained with secondary antibodies, donkey anti-rabbit alexa fluor 350, donkey anti-rat alexa fluor 488, donkey anti-goat alexa fluor 680 and donkey anti-mouse alexa fluor 594 (Invitrogen, Carlsbad, CA). Cells were washed twice in BD Perm/wash buffer and resuspended in DPBS without calcium and magnesium containing 2% FBS and 5mM EDTA. Analysis was performed on a BD LSR Fortessa.

FACS single-cell sorting

Dissociated primary islets cells were resuspended in 1mL DPBS without calcium and magnesium containing 2% FBS and 5mM EDTA (cell density should not exceed $1.5x10^6$/mL). Dead cells were stained with 7-AAD solution (BD Biosciences, San Jose, CA, USA) 5 minutes prior to fluorescence-activated cell sorting (FACS) analysis and cell sorting using a BD FACS AriaII (BD Biosciences). FACS sorting was done using a ceramic nozzle with a size of 100 µm. Cell doublets were excluded by gating singlets on FSC-H and FSC-A. By using a single-cell discrimination mask viable individual islet cells were sorted and collected in FrameStar 384-well plates (4titude, Surrey, UK) containing 2,3 µL lysis buffer (0.4% TritonX100 (Sigma-Aldrich), 1 U RNase inhibitor (Clontech), 2,5 µM Smart dTVN30 oligos (Picelli et al., 2014), 4mM dNTP (Thermo scientific, Waltham, MA USA), 0,1 µl (H1 donor) or 0,025 µl (all other 9 donors) of 1:40 000 dilution ERCC RNA spike-in mix (Ambion, Life Technologies)). Sample plates were kept at -80°C for future preparation into Smart-seq2 cDNA libraries.

Image-based validation of single-cell sorting

Cell-permeable Hoechst 33342 diluted 1:2000 (ThermoFisher, Invitrogen) and 5 µM of the vital dye calcein-AM (Invitrogen) were incubated with the cells for 30 minutes prior to FACS sorting (for details see *FACS single-cell sorting*, *Experimental Procedures*). Single cells, or in control wells hundreds of cells, were sorted into 384-well plates (Greiner Bio-One) and imaged using an ImageXpress automated fluorescence microscope fitted with a 4x S-fluor objective (3,2 pixels/µm). Well images were masked and segmented for cell counting using MetaXpress image analysis software (Molecular Devices, 5.3.0.1).

Glucose-stimulated insulin secretion (GSIS)

Glucose-stimulated insulin secretion (GSIS) of human islets from different donors was performed after 6-8 days in culture. Briefly, islets were washed in 2.8 mM Krebs Ringer phosphate hepes (KRH) buffer.

Five islets/well were transferred into a 96 well plate containing KRH buffer with 2.8 or 16.7 mM glucose. When testing the effect of a GLP1 receptor agonist, 10nM of exenatide (Ex4) (Bachem, Bubendorf, Switzerland) was added to KRH buffer containing 16.7mM glucose. Secreted insulin was measured in six replicates of each condition after 1 hour of incubation. Human Insulin ELISA (Mercodia AB, Uppsala, Sweden, article number 10-1113-01) was run according to manufacturer's protocol.

Immunohistochemistry

Histological staining was performed on 4 μm thick paraffin-embedded pancreatic tissue sections or isolated pancreatic islets using an IntelliPath FLX automated immunostainer (Biocare) as described previously (Walsh et al., 2014). Primary antibodies, rabbit anti-secretogranin III (Atlas antibodies, Stockholm Sweden) diluted 1:350, mouse anti-glucagon (Sigma-Aldrich, Stockholm, Sweden) diluted 1:8000, guinea pig anti-insulin (Dako, Glostrup, Denmark) diluted 1:8000 and rabbit anti-GLP1R (Abcam), diluted 1:1000 was incubated for 1 hour following 1 or 2 steps of Polymer kit (Biocare Medical, Concord, CA, USA). Following anti-insulin labeling a secondary biotinylated donkey anti-Guinea Pig antibody (Jackson Laboratory, Bar Harbor, Maine, USA) was used. This was followed by 4+streptavidin horse radish peroxidase labeling for 10 minutes using 3–3'-diaminobenzidine as the chromogen (Biocare Medical). Stained sections were scanned and digitized at a magnification of ×20 with the use of the Carl Zeiss MIRAX slide scanner (Zeiss, Goettingen, Germany). For calculation of the region of interest and the area of staining the whole section from each donor was analyzed using BioPix Image software (BioPix, Gothenburg, Sweden).

Preparation and sequencing of single-cell RNA-sequencing libraries

Single-cell RNA-sequencing libraries were produced in half the reaction volumes compared to the Smart-seq2 protocol previously described (Picelli et al., 2014). The protocol was executed with either a liquid handling robot (Biomek FXP, Beckman Coulter) or by manual preparation. Adjustments to the original Smart-seq2 protocol were the following: cDNA was synthesized with Superscript II (Invitrogen) and 2μM TSO strand switch oligo and further amplified with ISPCR primers at a concentration of 0.08μM with KAPA High Fidelity Hot Start polymerase (Kapa Biosystems). The cDNA was purified using Sera-Mag magnetic SpeedBeads, carboxylate-modified (GE Healthcare Biosciences) in the presence of 19.5% PEG8000 at a 0.8:1 ratio beads:cDNA. The quality of the cDNA was assessed for random samples with an Agilent 2100 Bioanalyser and High Sensitivity DNA Chip (Agilent Technologies Inc.). Tn5 transposase directed tagmentation of 0.5-1 ng cDNA was performed with recombinant Tn5 in 10% PEG8000, 10mM TAPS-NaOH (pH 8.3), 5mM $MgCl_2$ (Picelli et al., 2014b). The Tn5 enzyme was removed from the DNA with 0.04% SDS. Sequencing libraries were generated with Nextera XT Index kit v2 (Illumina Inc.) and KAPA High Fidelity amplification (Kapa Biosystems). Sequencing libraries were multiplexed with 192 cells in each and purified with Sera-Mag magnetic SpeedBeads, carboxylate-modified (GE Healthcare Biosciences) in the presence of 24% PEG8000 at 1:1 ratio beads:DNA. The quantity and quality of the sequencing libraries were analysed with an Agilent 2100 Bioanalyser and Qubit 2.0 Fluorometer (Invitrogen). Sequencing was carried out with an Illumina HiSeq 2000 generating 43 bp single-end reads.

RNA-sequencing of whole islets

RNA was isolated from seven whole islets (healthy donors H3, H4 and H6 and all T2D donors) using Qiagen RNeasy microkit with on-column DNAse digestion (Qiagen, Hombrechtikon, Swizerland). The RNA samples were processed with Illumina TruSeq Stranded mRNA Library prep kit following the manufacturer's recommendations. Libraries were quantified with Qubit HS (ThermoFisher, MA, USA) and Fragment Analyzer (Advanced Analytical Technologies, Iowa, USA) adjusted to the appropriate concentration for sequencing. Indexed libraries were pooled and sequenced at a final concentration of 1.6 pM on an Illumina NextSeq 500 high-output run using paired-end chemistry with 75 bp read lengths.

Processing, quality control and filtering of RNA-sequencing data

Sequence reads were aligned towards the human genome (hg19 assembly) using STAR (v2.3.0e) and uniquely aligned reads within RefSeq gene annotations were used to quantify gene expression as reads per kilobase transcript and million mapped reads (RPKMs) using rpkmforgenes (Ramsköld et al., 2009). The requirements for retaining cells in the analysis were: ≥ 50,000 sequenced reads, ≥ 40% of reads aligning uniquely to the genome and ≥ 40% of them aligning within annotated RefSeq exons and detection of at least 1,000 genes at the expression threshold of RPKM ≥ 1. Samples that failed to meet these criteria were considered of low quality and therefore excluded from the downstream analysis.

Correlation between single-cell and whole-islet expression data

Spearman's correlation (ρ) between single-cell and bulk expression data was computed for H3, H4, H6 and all T2D donors. Gene expression was averaged across single cells of the same individual and correlated against the corresponding whole-islet RNA-sequencing data. The correlations were computed using log_2-transformed expression data. Only the genes with expression greater than zero either in bulk or in at least one single cell (Mean>0) were included in the calculations and in the graphs of Figure S4. For the down-sampling procedure, Spearman's correlation coefficients were computed as described above, but using a subset of single cells for each individual. For each different number of cells tested, 200 iterations were performed in which cells were selected at random from each individual donor and correlated against the corresponding bulk data.

Identification of highly variable genes

Gene expression was ranked in descending order of variance across cells, while controlling for the relation between the expression magnitude and the variability arising from technical noise (Brennecke et al., 2013). For this purpose, we used an in-house implementation of the method in R with the additional option of specifying the number of cells considered as outliers (winsorization). The winsorization parameter was set to 1 in order to prevent genes with extreme expression values in only one cell from being high in the ranking. Ranked lists of genes for the different cell groups analyzed are included in Table S1.

Dimensionality reduction

The t-Distributed Stochastic Neighbor Embedding (*t*-SNE) method was employed to reduce the dimensions of the gene expression data and project the cells onto a two dimensional space (using the MATLAB implementation of *t*-SNE). The normalized expression values (log_2RPKM) were used as input to the algorithm, while the number of principal components employed internally for the reduction was adjusted at every different run to retain the 60% of the total variability of the input data. The perplexity parameter was set to 50 when the number of cells exceeded 800, 30 for sample sizes up to 800 and 15 in the cases where the number of cells was less than 60.

Cell-type classification

The basic approach for the cell-type classification was the projection of cells with *t*-SNE onto the two dimensional space spanned by a set of genes adequate to capture the overall variability within the dataset. The assignment of the formed clusters to cell classes was obtained based on the expression levels of hormones or other known marker genes. The cell-type classification of all cells analyzed is included in Table S1.

Differential expression analysis of the endocrine and exocrine cell types

For the differential expression analysis among the six major cell types (α, β, γ, δ, acinar and ductal cells) we performed one-way Analysis of variance (ANOVA) using the log_2-transformed expression data from the five healthy male donors. The analysis was followed by multiple comparisons testing to further identify the cell types that displayed significant differences in their expression out of all the possible pairwise combinations. The reason for using only a subset from the available samples was to rule out any variations introduced in the data due to the sex and disease associated differences. Although such factors of variation could be modeled with a multifactorial design, the existence of small and unequal sample sizes arising from the different donors in each cell class would reduce the power of the statistical analysis. The criteria to include a gene in the differential analysis were based on the average magnitude (Mean log2RPKM >= 1 in at least 1 cell type) and variation of the expression in the cell types tested. To further reduce the risk of erroneously reporting individual donor differences as cell-type differential expression, we added two more requirements that every gene should meet in order to be tested. The first was that the expression detected in each cell type should come from cells belonging to at least two different individuals of the same cell type (Mean log2RPKM >= 1 in at least 2 donor groups of the same cell type).. Based on the second, the variation in at least two different donor groups of the same cell type should not be greater than the overall variation across the cell type. The Bonferroni adjustment was applied in order to correct for the multiple comparisons performed for each of the genes tested. Also, the Benjamini-Hochberg method was used to control the false discovery rate at significance level α=0.01. Similar results were obtained using a non-parametric test, but the results from the ANOVA differential expression analysis are summarized in Table S2. The exact same procedure was repeated including the ε-cells from the five healthy male donors in order to identify cell type specific expression in this rare population. The non-parametric one-way ANOVA (Kruskal Wallis) test was additionally performed due

to the difference in size of the samples being compared. The two tests gave similar results and we report the genes identified as up-regulated in ε-cells based on both analyses in Table S2.

Single-molecule mRNA FISH

mRNAs were visualized by single molecule FISH (smFISH) using the RNAscope Fluorescent Multiplex Kit (Advanced Cell Diagnostics, Inc.) according to the manufacturer's instructions. SmFISH stainings were performed on formalin-fixed, paraffin-embedded (FFPE) sections of pancreas from 4 healthy donors. The following RNAscope probes were used: Hs-GCG (ACD556741), hs-SST (ACD310591), hs-FAP (ACD411971) and hs-LEPR (ACD406371). Images were acquired on a Nikon A1R confocal microscope. The DAPI and probe signals within islets cells were manually appointed to either of four groups; i) negative for both hormone and gene of interest (GOI) ii) hormone positive and GOI negative, iii) hormone negative and GOI positive iv) hormone positive and GOI positive (Table S3). Detailed information on cell counting and statistics per slide and donor are included in TableS3.

Differential expression analysis between the pancreatic stellate cells and endothelial cells

In order to identify the genes that are differentially expressed between the clusters consisting of pancreatic stellate and endothelial cells, the R/Bioconductor package SCDE (Kharchenko et al, 2014) was used. All genes expressed within the two cell classes were tested and the analysis was performed with the raw read counts as input. The grid of expression magnitude was set to 450 for increased sensitivity and the independent fit was selected for the error modeling. The results are summarized in Table S2.

Comparison between single-cell and whole-islet RNA-sequencing cell-type enriched expression.

In order to assess the performance of previously used strategies in identifying cell type specific gene expression, we correlated the differentially expressed genes from the single-cell analysis with the cell type hormones or marker genes using the expression data published in Taneera et al. 2012. The genes were selected based on the magnitude of mean expression in each cell type. Spearman's correlation coefficients and p-values (adjusted) were calculated for the 25 strongest genes in each cell type with the corresponding hormone or marker gene: *GCG* for α, *PPY* for γ, *SST* for δ, *REG1A* for acinar and *KRT19* for ductal top genes. Data for *INS* was not available in the study. The rank of each gene was obtained from sorting all the genes measured in the study in decreasing magnitude of correlation coefficient.

Analysis of heterogeneity in the cell types

To further explore the heterogeneity of the identified cell types, we analyzed their transcriptomes in isolation following the procedure as described above. The expression data of each cell type was ranked based on biological variability and using this as input the cells were projected onto two dimensions with the *t*-SNE. In all cases the obtained embedding was dominated by donor differences (Figure S6). In order to correct for such effects that might confound biological differences, we used ComBat (Johnson et al, 2007), more specifically a Python implementation of the ComBat function in R/Bioconductor package SVA. The method provides a robust adjustment even for small sample sizes as in the case of our dataset. We used the parametric empirical Bayesian framework, setting the adjustment variable to denote the donor individuals. The *t*-SNE dimensionality reduction was repeated with input the adjusted expression values. Following batch correction, the donor effect was completely removed from the resulting clusters in each cell type.

Differential expression analysis within the cell types

We performed differential expression analysis to identify the genes that are responsible for the resulting sub-clusters in each cell type: α, β and acinar cells. SCDE method was applied as previously described. The results obtained from the analysis in each cell type are summarized in Table S4.

Relationship of BMI and gene expression in different cell types

The relationship between body mass index (BMI) and gene expression in each cell type was measured using rank correlation statistics. Spearman's ρ coefficients were calculated separately for each gene based on the $\log_2$-transformed expression values of the samples in each cell class. Since multiple tests were performed (one for every gene), the results were adjusted with the Benjamini-Hochberg method to control the false discovery rate at significance level $\alpha=0.01$. Only the samples from the healthy male donors were used, since there is a positive correlation between BMI and sex in our donors ($\rho=0.49$). The lists containing genes with an absolute correlation coefficient with BMI greater than 0.5 are included in

Table S5. We also generated correlations per gene using all cells (not separated per cell type) from the same five male donors to simulate the correlations one would obtain in "whole-islet" analyses.

Differential expression analysis between cells from healthy and T2D donors

We compared the expression between the healthy and T2D samples in order to identify disease associated genes for the major cell types: α, β, γ, δ, acinar and ductal cells. For this purpose, we conducted non-parametric one-way Analysis of variance (Kruskal-Wallis test) using the samples from all donors in each cell type tested. The samples were divided into four groups based on the status (healthy or T2D) and sex, resulting into the four groups being compared: healthy male, healthy female, T2D male and T2D female. We also performed the analysis using a parametric multifactorial design: two-way Analysis of variance with the first factor indicating the status (healthy or T2D) and the second the sex of the samples. Bonferroni adjustment was used for the multiple comparison correction and the Benjamini-Hochberg method to control the false discovery rate at significance level $\alpha=0.01$. The genes that showed significant differences in respect to sex were excluded from the results in both tests in order to identify the differentially expressed genes related to T2D. The two tests yielded overlapping results, but the gene sets obtained from the non-parametric design were less affected by the non-normal distributions of gene expression, reflecting more robust disease related differences. Therefore, we report and use in the downstream analysis the results provided by the non-parametric test. In the case of δ-cells, both tests reported no significant results due to the existence of a group with a very small sample size (healthy female group consisting of only two cells). To overcome this limitation, we performed one-way Analysis of variance (parametric and non-parametric) using only the cells from the male donors, but no interesting genes were reported (*SNORD110* and *COL6A2*). The differentially expressed genes between the healthy and T2D cells for each cell type are listed in Table S6.

Gene set enrichment analysis in cell types

Gene set enrichment analysis (GSEA) was used to examine whether the genes identified as differentially expressed between the healthy and T2D cells for each cell type are members of categories with specific functions. We used the pre-ranked version, providing the difference in median expression values between the healthy and T2D cells as the gene ranking metric. For each cell type, the ranked list of genes was tested for overlaps with all the gene sets belonging to four major collections of the Molecular Signatures Database: curated gene sets (C2), GO gene sets (C5), oncogenic signatures (C6) and immunologic signatures (C7). The significant categories (FDR≤1%) that are enriched in each cell type for the two conditions are reported in Table S7.

SUPPLEMENTAL REFRENCES

Johnson, W.E., Li, C., and Rabinovic, A., (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostat 8, 118–127.

Kühtreiber, W.M., Ho, L.T., Kamireddy, A., Yacoub, J.A.W., and Scharp, D.W., (2010). Islet isolation from human pancreas with extended cold ischemia time. Transplant. Proc. 42, 2027–2031.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T., (2014). Bayesian approach to single-cell differential expression analysis. Nat Methods 11:740-2.

Montrose-Rafizadeh, C., Yang, H., Rodgers, B.D., Beday, A., Pritchette, L.A., and Eng, J., (1997). High potency antagonists of the pancreatic glucagon-like peptide-1 receptor. J. Biol. Chem. 272, 21201–21206.

Picelli, S., Björklund, Å.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R., (2014b). Tn5 transposase and tagmentation procedures for massively-scaled sequencing projects. Genome Res. 24, 2033-2040.

Walsh, S.K., Hector, E.E., Andréasson, A.-C., Jönsson-Rylander, A.-C., and Wainwright, C.L., (2014). GPR55 deletion in mice leads to age-related ventricular dysfunction and impaired adrenoceptor-mediated inotropic responses. PLoS ONE 9, e108999.