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

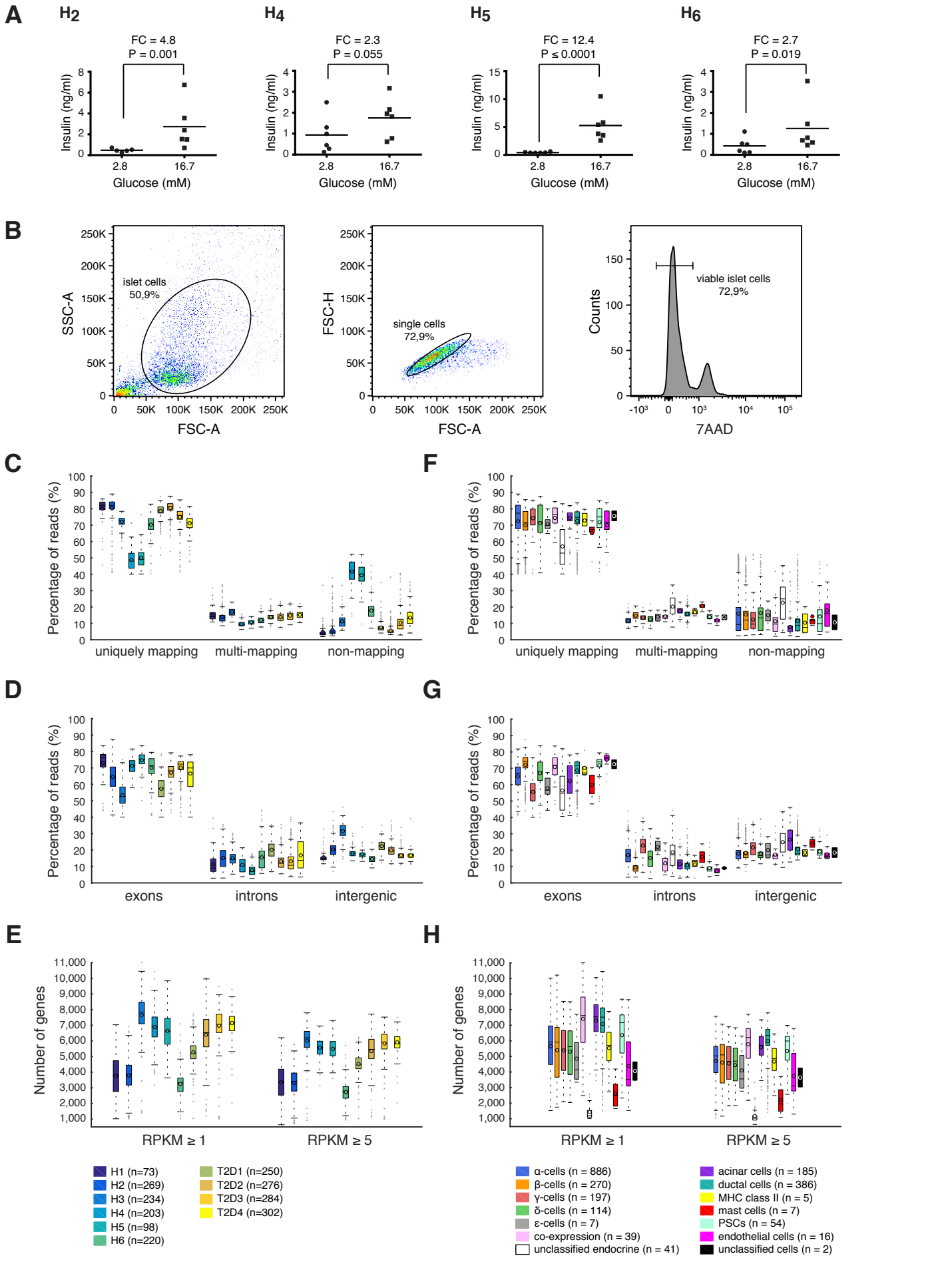

A
H2
FC = 4.8
P = 0.001
H4
FC = 2.3
P = 0.055
H5
FC = 12.4
P ≤ 0.0001
H6
FC = 2.7
P = 0.019
Insulin (ng/ml)
Glucose (mM)
2.8
16.7
B
islet cells
50,9%
SSC-A
FSC-A
single cells
72,9%
FSC-H
viable islet cells
72,9%
Counts
7AAD
C
F
Percentage of reads (%)
uniquely mapping
multi-mapping
non-mapping
D
G
exons
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

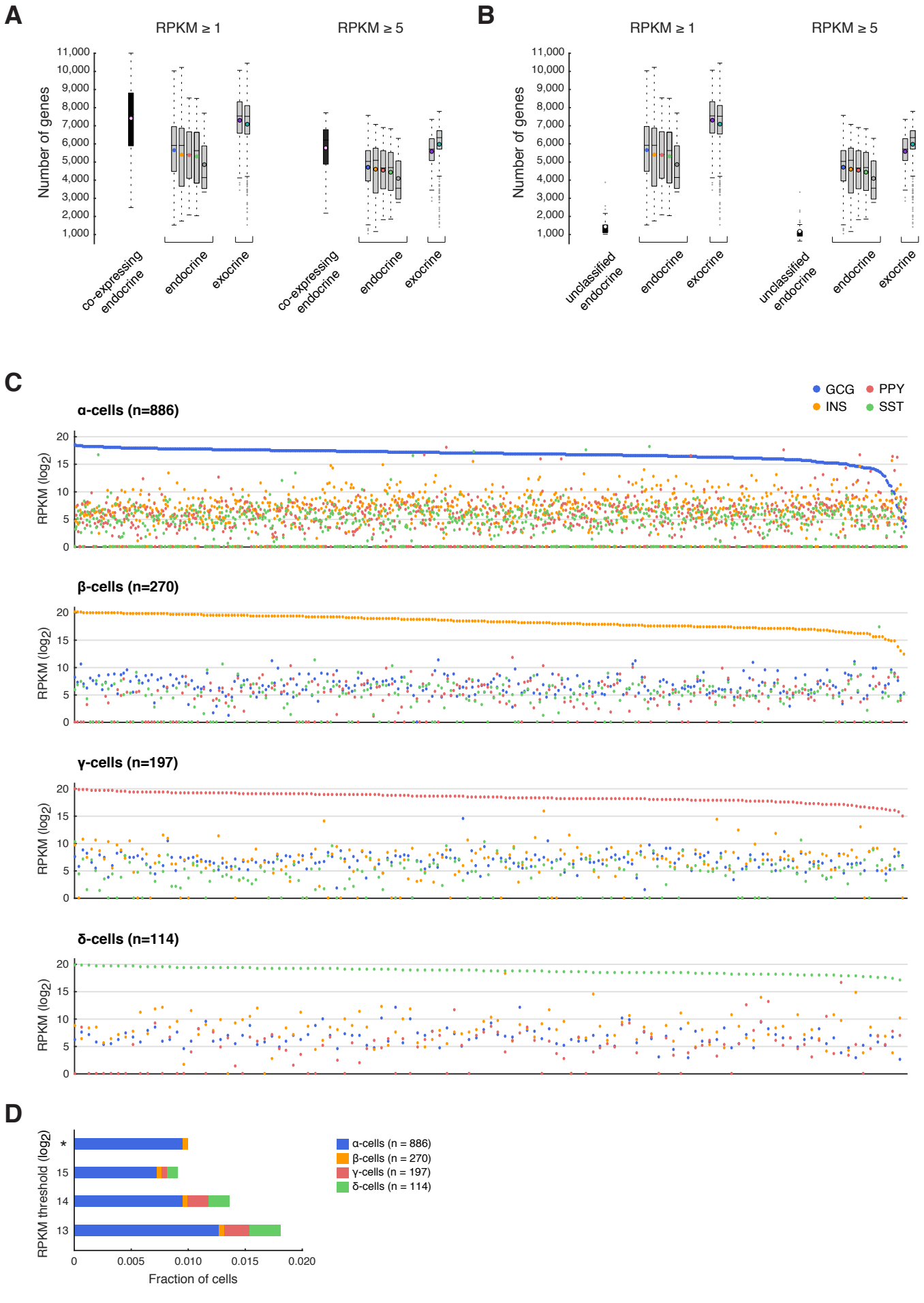

A
RPKM ≥ 1
RPKM ≥ 5
Number of genes
co-expressing endocrine
endocrine
exocrine
B
RPKM ≥ 1
RPKM ≥ 5
Number of genes
unclassified endocrine
endocrine
exocrine
C
GCG
PPY
INS
SST
α-cells (n=886)
β-cells (n=270)
γ-cells (n=197)
δ-cells (n=114)
RPKM (log2)
D
RPKM threshold (log2)
Fraction of cells
α-cells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

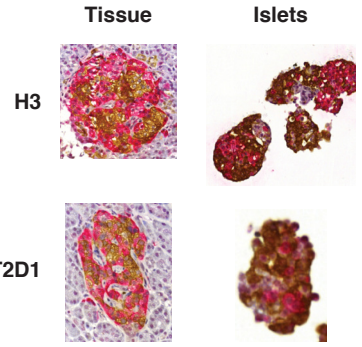
Figure S2. Filtering of single-cell RNA-sequencing data. Related to Figure 1.
(**A**) Boxplots showing the number of genes expressed across the cells from "co-expressing" endocrine, endocrine and exocrine cell types using two different thresholds (RPKM≥1 and RPKM≥5). The "co-expressing" cells have a 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

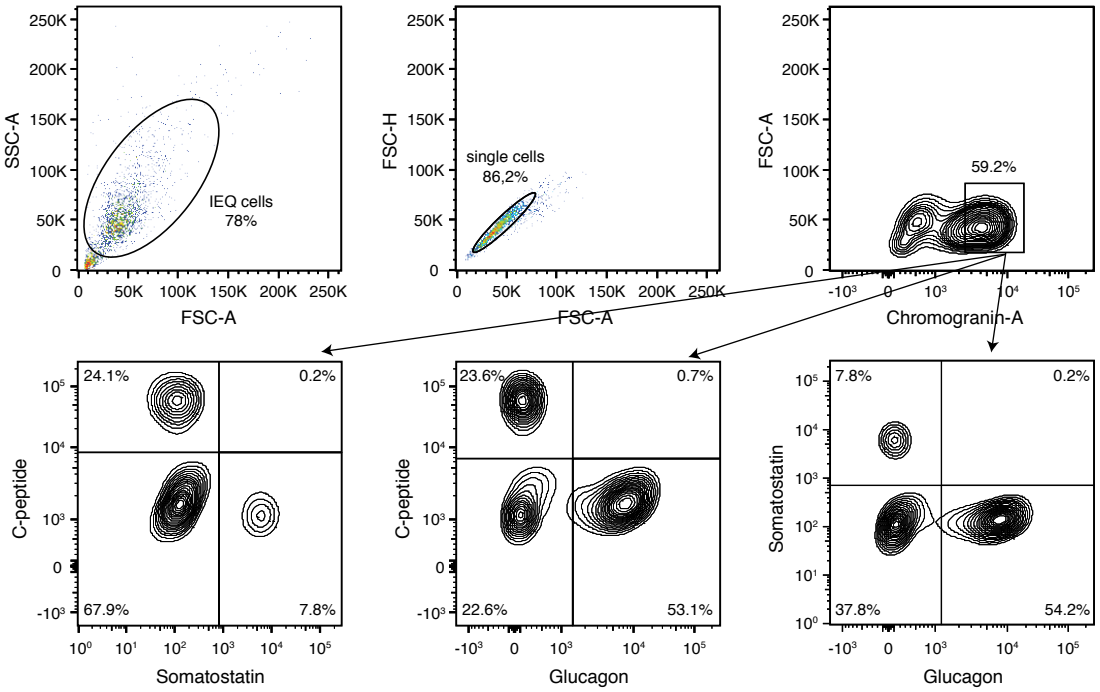
A

	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

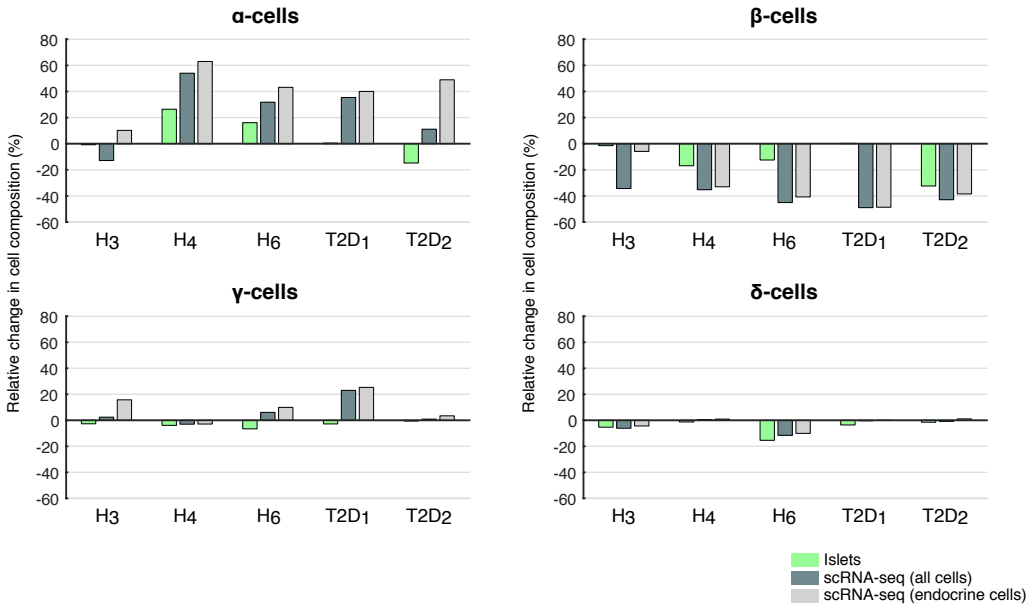
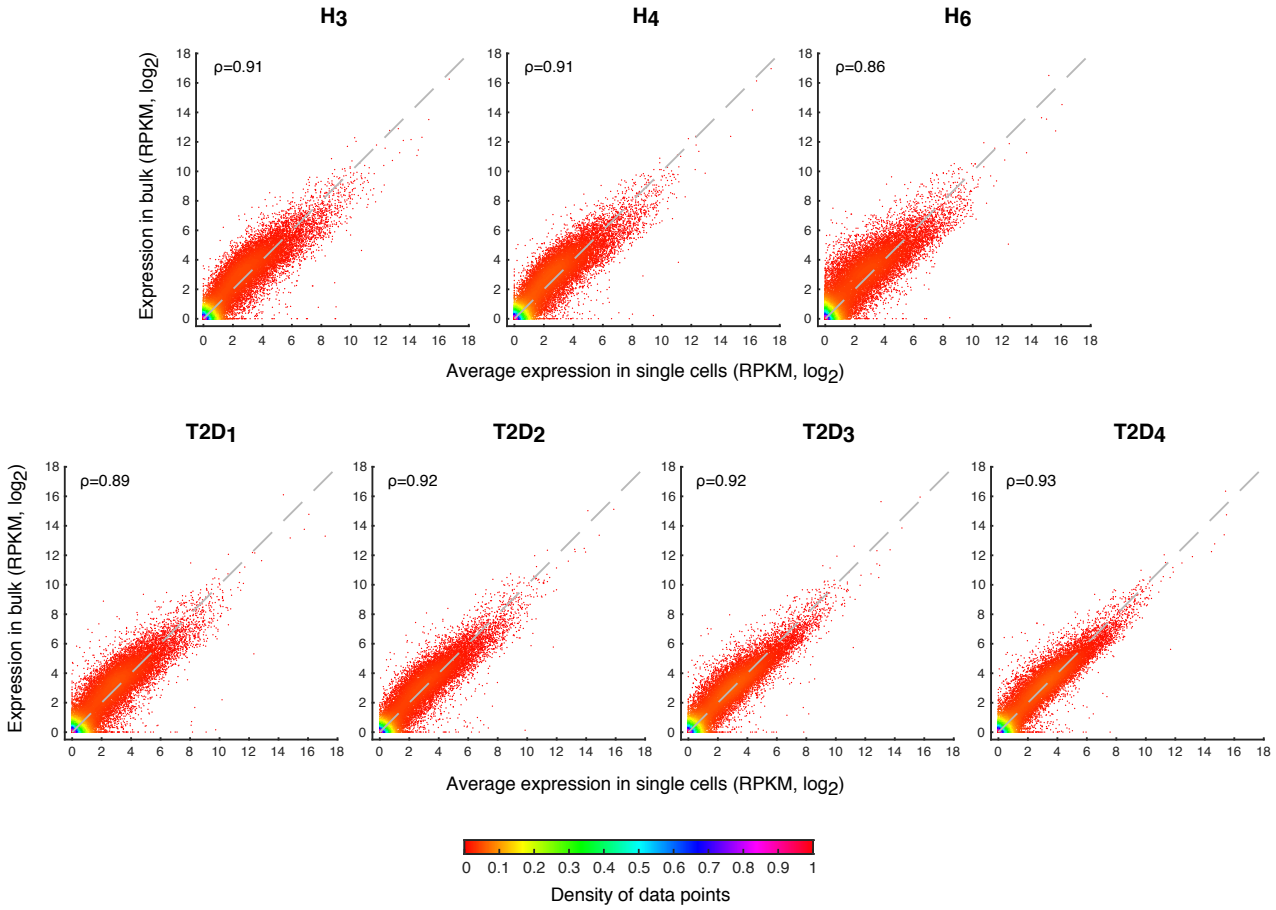


Figure S3. Cell-type composition in single-cell sequencing, tissue and dissociated islets. Related to Figure 1.

(**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

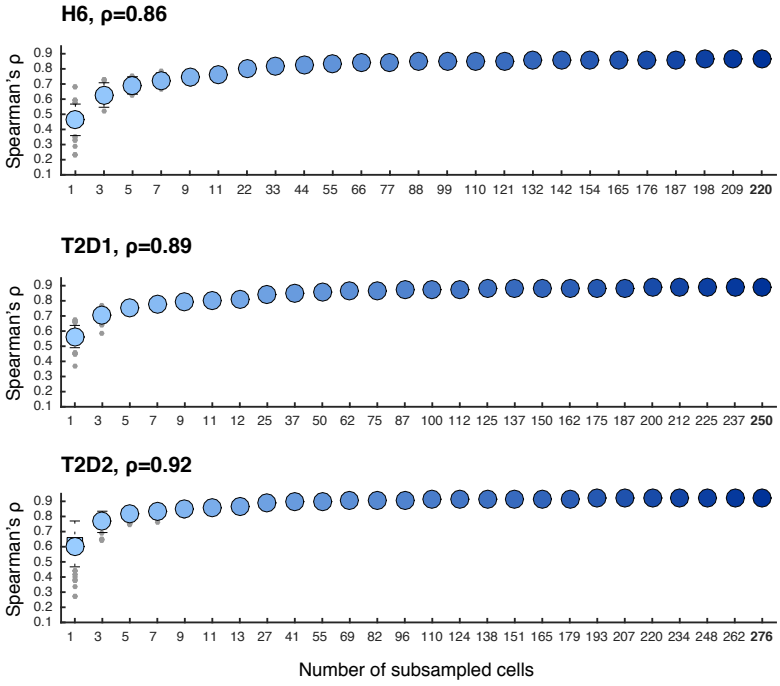
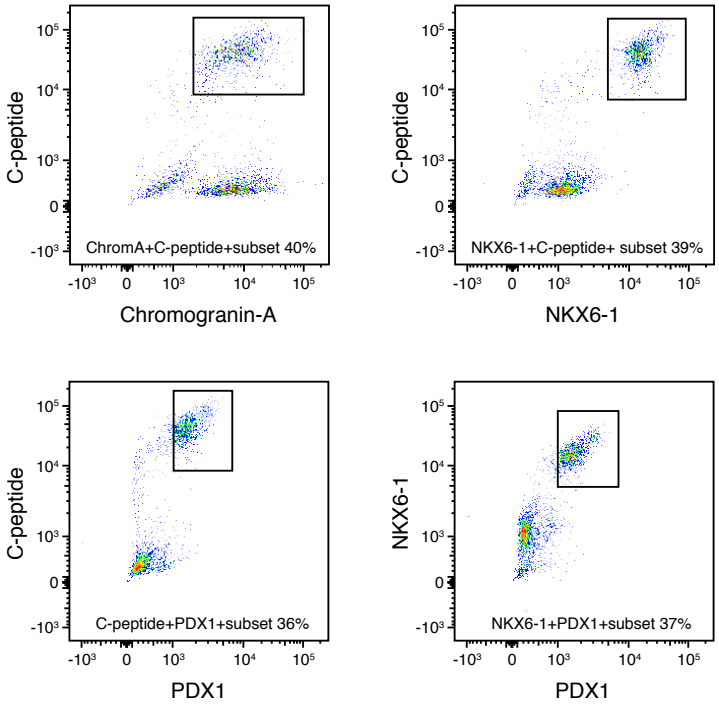

H6, ρ=0.86
T2D1, ρ=0.89
T2D2, ρ=0.92
Spearman's ρ
Number of subsampled cells

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

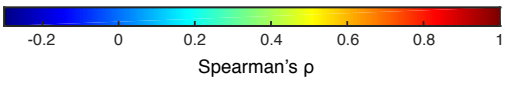


Figure S5. FACS analysis of Pdx1 and Nkx6-1 positive β-cells and assessment of cell-type enriched expression inferred from whole-islet data. Related to Experimental Procedures.
(A) FACS analysis of dissociated human islets cells labelled with anti-c-peptide, anti-chromogranin-A, anti-Nkx6-1 and anti-Pdx1 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

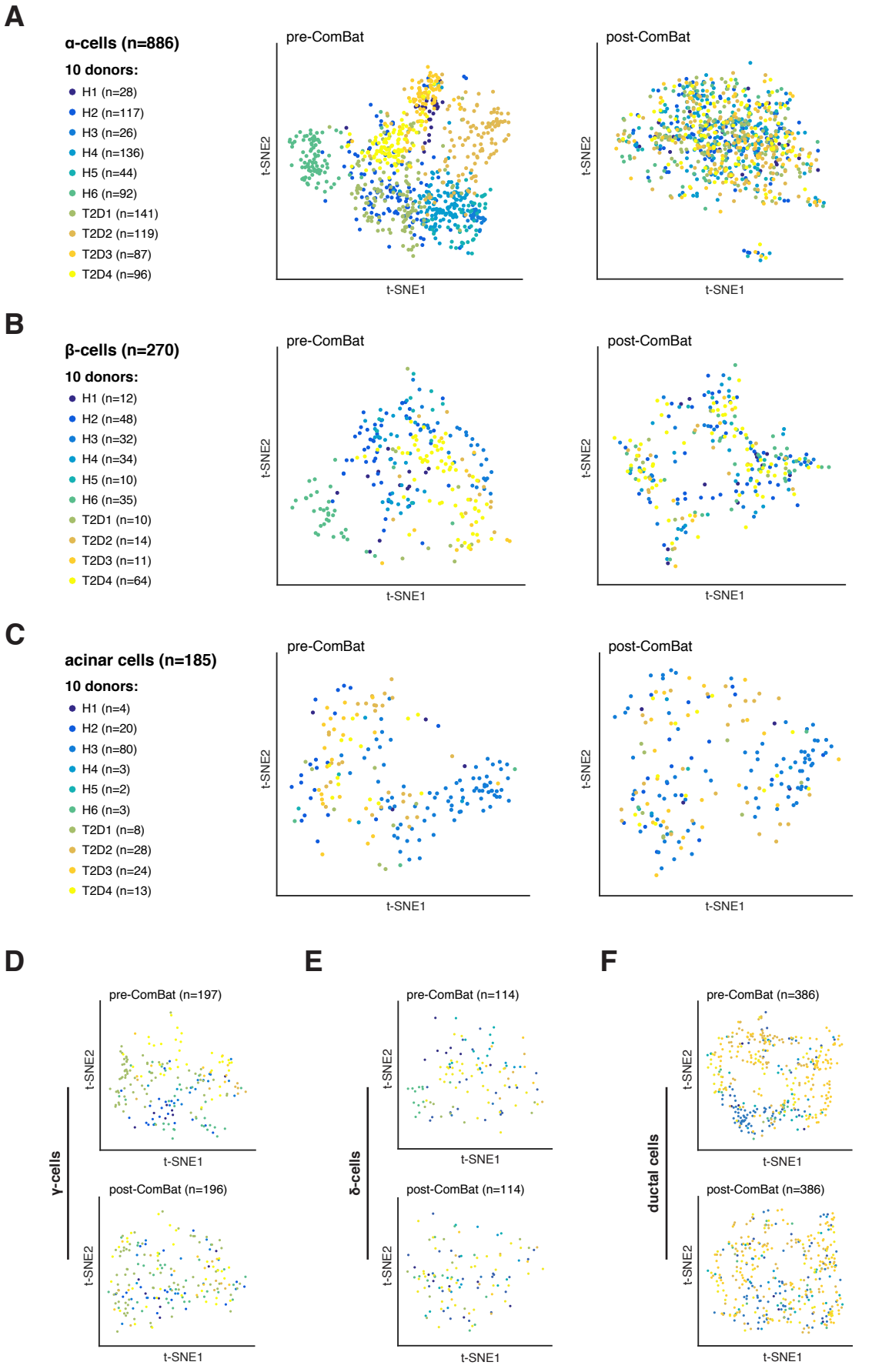

A
α-cells (n=886)
10 donors:
H1 (n=28)
H2 (n=117)
H3 (n=26)
H4 (n=136)
H5 (n=44)
H6 (n=92)
T2D1 (n=141)
T2D2 (n=119)
T2D3 (n=87)
T2D4 (n=96)
pre-ComBat
post-ComBat
t-SNE1
t-SNE2
B
β-cells (n=270)
10 donors:
H1 (n=12)
H2 (n=48)
H3 (n=32)
H4 (n=34)
H5 (n=10)
H6 (n=35)
T2D1 (n=10)
T2D2 (n=14)
T2D3 (n=11)
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

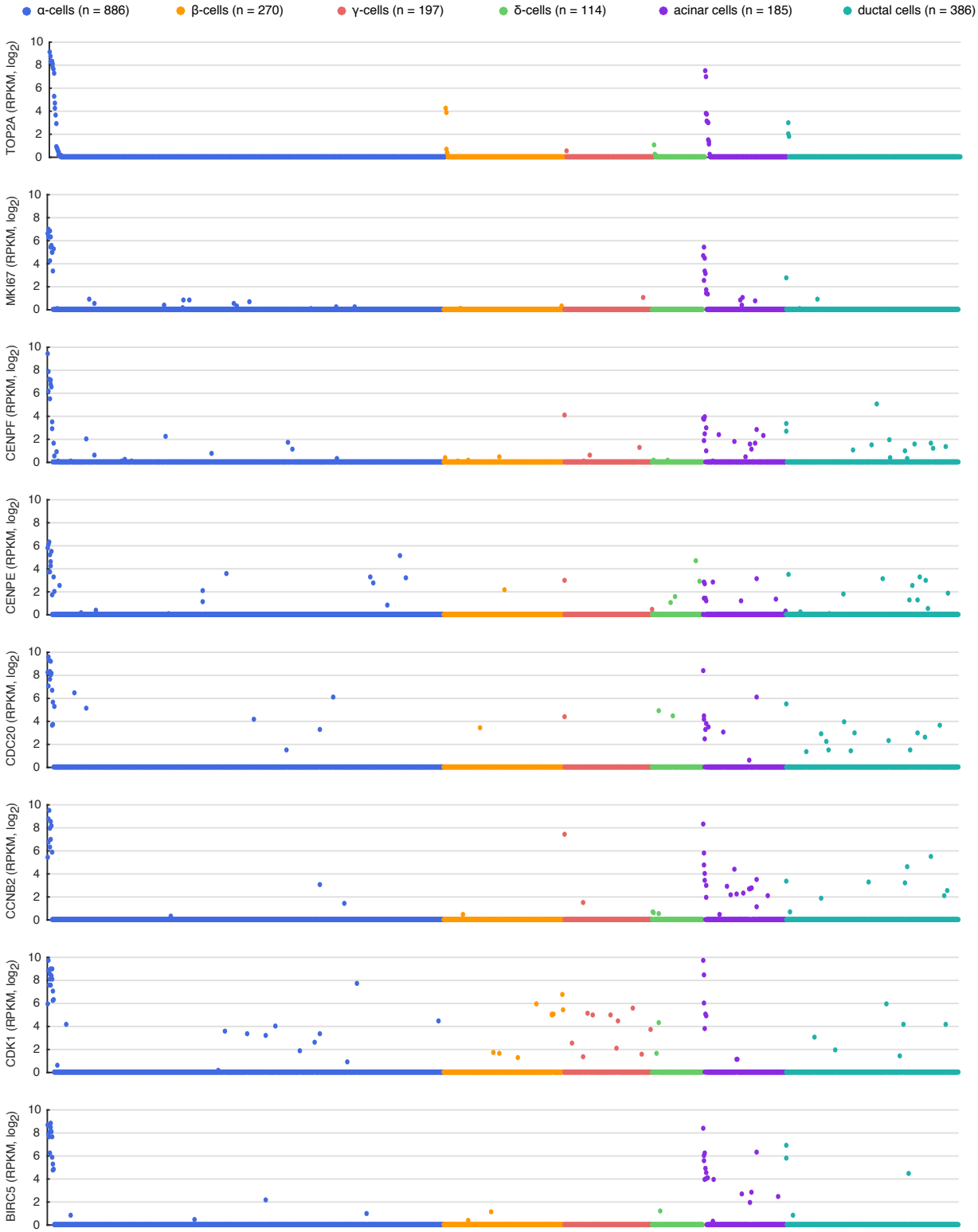

α-cells (n = 886)
β-cells (n = 270)
γ-cells (n = 197)
δ-cells (n = 114)
acinar cells (n = 185)
ductal cells (n = 386)
TOP2A (RPKM, log2)
MKI67 (RPKM, log2)
CENPF (RPKM, log2)
CENPE (RPKM, log2)
CDC20 (RPKM, log2)
CCNB2 (RPKM, log2)
CDK1 (RPKM, log2)
BIRC5 (RPKM, log2)
0
2
4
6
8
10

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.