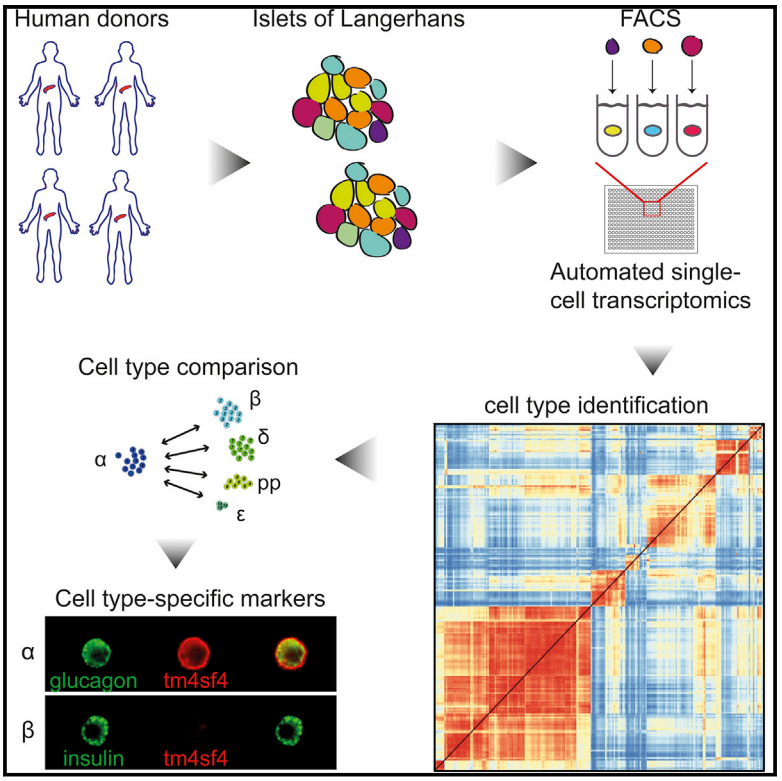
Report

Cell Systems

A Single-Cell Transcriptome Atlas of the Human Pancreas

Graphical Abstract



Authors

Mauro J. Muraro, Gitanjali Dharmadhikari, Dominic Grün, ..., Francoise Carlotti, Eelco J.P. de Koning, Alexander van Oudenaarden

Correspondence

e.koning@hubrecht.eu (E.J.P.d.K.), a.vanoudenaarden@hubrecht.eu (A.v.O.)

In Brief

Single-cell mRNA sequencing was used to describe the transcriptome of adult human pancreatic cell types. This resource was mined to find subpopulations of existing cell types and markers that can be used to purify live alpha and beta cells.

Highlights

- Single-cell sequencing of human pancreas allows in silico purification of cell types
- We provide cell-type-specific genes, transcription factors, and cell-surface markers
- StemID finds outlier populations of acinar and beta cells
- CD24 and TM4SF4 function as two markers to enrich for alpha and beta cells

Data Resources

GSE85241

Muraro et al., 2016, Cell Systems 3, 385–394
October 26, 2016
http://dx.doi.org/10.1016/j.cels.2016.09.002

Cell Systems
Report

A Single-Cell Transcriptome Atlas of the Human Pancreas

Mauro J. Muraro,[1,5] Gitanjali Dharmadhikari,[1,5] Dominic Grün,[1,2] Nathalie Groen,[4] Tim Dielen,[1] Erik Jansen,[1] Leon van Gurp,[1] Marten A. Engelse,[3] Francoise Carlotti,[4] Eelco J.P. de Koning,[1,3,*] and Alexander van Oudenaarden[1,6,*]
[1]Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences) and University Medical Center Utrecht, Cancer Genomics Netherlands, 3584 CT Utrecht, the Netherlands
[2]Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany
[3]Section of Nephrology and Section of Endocrinology, Department of Medicine, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[4]Department of Internal Medicine, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[5]Co-first author
[6]Lead Contact
*Correspondence: e.koning@hubrecht.eu (E.J.P.d.K.), a.vanoudenaarden@hubrecht.eu (A.v.O.)
http://dx.doi.org/10.1016/j.cels.2016.09.002

SUMMARY

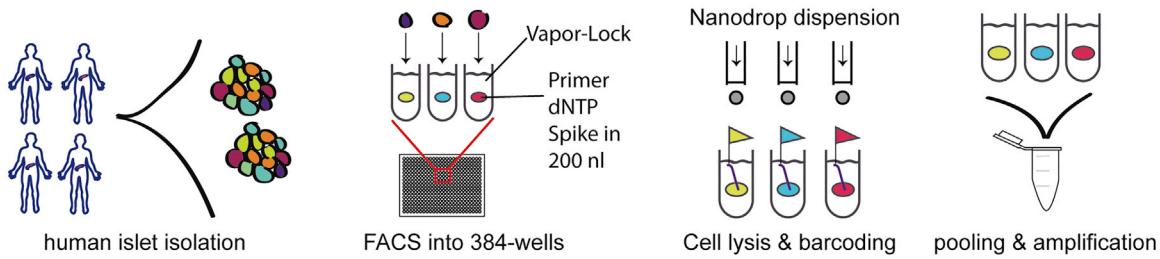
To understand organ function, it is important to have an inventory of its cell types and of their corresponding marker genes. This is a particularly challenging task for human tissues like the pancreas, because reliable markers are limited. Hence, transcriptome-wide studies are typically done on pooled islets of Langerhans, obscuring contributions from rare cell types and of potential subpopulations. To overcome this challenge, we developed an automated platform that uses FACS, robotics, and the CEL-Seq2 protocol to obtain the transcriptomes of thousands of single pancreatic cells from deceased organ donors, allowing in silico purification of all main pancreatic cell types. We identify cell type-specific transcription factors and a subpopulation of REG3A-positive acinar cells. We also show that CD24 and TM4SF4 expression can be used to sort live alpha and beta cells with high purity. This resource will be useful for developing a deeper understanding of pancreatic biology and pathophysiology of diabetes mellitus.

INTRODUCTION

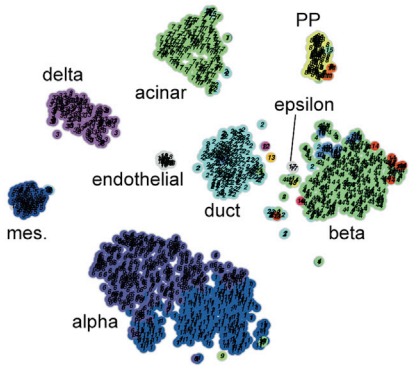
Most organs consist of a variety of cell types with interdependent functions. To understand organ function and disease, genome-wide information on each cell type is crucial. Studies on pooled material detect global gene expression patterns but represent an average dominated by the most abundant cell types. With the advent of single-cell transcriptomics, it is possible to determine the transcriptome of individual cells, allowing the identification of cell types in an unbiased manner (Grün and van Oudenaarden, 2015; Kolodziejczyk et al., 2015; Trapnell, 2015; Wang and Navin, 2015). Initial single-cell transcriptomic studies were performed on cultured cells (Deng et al., 2014; Hashimshony et al., 2012; Islam et al., 2011; Klein et al., 2015; Shalek et al., 2013; Tang et al., 2010). Subsequent studies described cell types in the mouse lung (Treutlein et al., 2014), spleen (Jaitin et al., 2014), brain (Zeisel et al., 2015), retina (Macosko et al., 2015), small intestine (Grün et al., 2015), and pancreas (Xin et al., 2016). Studies on human tissue have so far been limited to fetal neurons (Johnson et al., 2015), glioblastomas (Patel et al., 2014), and two sets of human pancreas cells (Li et al., 2016; Wang et al., 2016). We have used single-cell sequencing of the human pancreas to reveal subpopulations of cells that show potential as progenitors (Grün et al., 2016). These studies described manually processed samples and/or low numbers of cells, which limited the number of detected genes. Here, we developed a more efficient, high-throughput method to sequence primary human cells of all pancreatic cell types.

The pancreas functions as an exocrine and endocrine gland. The exocrine compartment consists of acinar cells producing digestive enzymes and ductal cells forming channels that drain into the duodenum. The endocrine compartment consists of alpha, beta, delta, pancreatic polypeptide (PP), and epsilon cells that are found in the islets of Langerhans. Insulin-producing beta cells and glucagon-producing alpha cells play major roles in glucose homeostasis, and islet dysfunction is the hallmark of diabetes mellitus, a chronic metabolic disorder affecting approximately 9% of the adult population worldwide (WHO, 2014). Functional analysis and genetic profiling are typically performed on whole islets, masking the contribution of individual cell types to pancreas biology and disease (Bugliani et al., 2013; Cnop et al., 2014; Eizirik et al., 2012). To study heterogeneity and classify subpopulations within known cell types, single-cell resolution is essential. We developed a high-throughput approach for single-cell sequencing based on the single-cell RNA-seq by multiplexed linear amplification2 (CEL-seq2) protocol (Hashimshony et al., 2016) to create a single-cell transcriptome atlas of the human pancreas. Our method implements fluorescence-activated cell sorting (FACS), which allows the user to work with low amounts of starting material. This dataset provides an unbiased view of cell types in the human pancreas at single-cell resolution, enabling comparison of gene expression patterns among cell types and detection of subpopulations within them. This resource can be mined for genes involved in pancreatic function to define novel therapeutic targets for diseases such as diabetes mellitus.

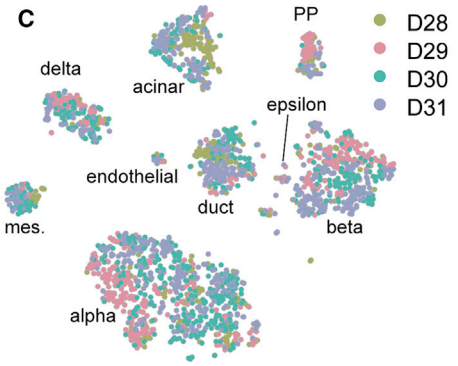
A



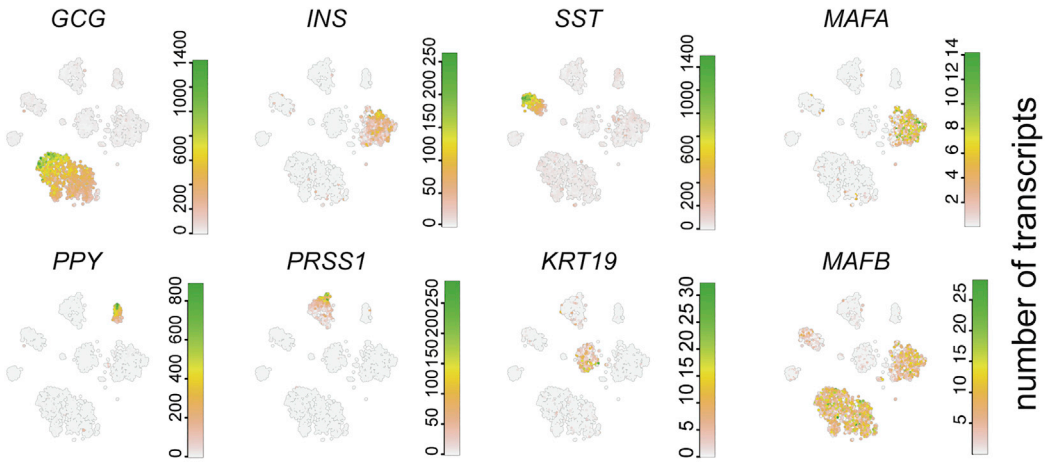
B



C



D



E

alpha cells		beta cells		delta cells		PP cells	
genes	TF	genes	TF	genes	TF	genes	TF
GCG	*IRX2*	*INS*	*MAFA*	*SST*	*HHEX*	*PPY*	*ETV1*
LOXL4	*FEV*	*IAPP*	*PDX1*	*PRG4*	*ERBB4*	*SERTM1*	*MEIS2*
PLCE1	*ARX*	*MAFA*	*SMAD9*	*LEPR*	*POU3F1*	*CARTPT*	*ID2*
IRX2	*PTGER3*	*NPTX2*	*CDKN1C*	*RBP4*	*ISL1*	*SLITRK6*	*EGR3*
GC	*HMGB3*	*DLK1*	*TFCP2L1*	*BCHE*	*PSIP1*	*ETV1*	*LMO3*
KLHL41	*RFX6*	*ADCYAP1*	*SIX3*	*HHEX*	*BHLHE41*	*THSD7A*	*MEIS1*
CRYBA2	*MAFB*	*PFKFB2*	*SIX2*	*FRZB*	*PDLIM4*	*AQP3*	*ID4*
TTR	*SMARCA1*	*PDX1*	*MNX1*	*PCSK1*	*EHF*	*ENTPD2*	*ARX*
TM4SF4	*PGR*	*TGFBR3*	*BMP5*	*RGS2*	*LCORL*	*PTGFR*	*PAX6*
RGS4	*LDB2*	*SYT13*	*PIR*	*GABRG2*	*ETV1*	*CHN2*	*ZNF503*

(legend on next page)

RESULTS

SORT-Seq Allows Deep Sequencing of Human Pancreas Cells

To assay the transcriptomes of the various human pancreatic cell types, we obtained human pancreas material from four deceased organ donors (Figure 1A). Isolation of the islets of Langerhans yielded 55%–95% islet purity (Table S1). The non-islet cells in these preparations mainly consisted of exocrine cells. After a culture period of 3–5 days, the islets were dispersed for FACS, followed by single-cell sequencing. Previously, we sorted cells from five donors, which were processed manually by single-cell RNA-seq by multiplexed linear amplification (CEL-Seq) as described (Grün et al., 2016). These yielded an average of 4,262 unique transcripts and a median of 1,958 detected genes per cell (Figures S1A and S1B) when sequencing at approximately 90.000 reads per cell. While useful to determine interesting progenitor cells and describe general differences between cell types, this dataset lacked the depth to fully describe the transcriptome of each pancreatic cell type. For example, comparing expression across endocrine cell types resulted in low numbers of differentially expressed genes (Figure S1C).

To more efficiently capture single-cell transcriptomes, we used FACS and robotics liquid handling to perform automated single-cell sequencing based on the CEL-Seq2 protocol (Hashimshony et al., 2016). We refer to this platform as SORT-seq (SOrting and Robot-assisted Transcriptome SEQuencing) (Figure 1A). Briefly, live single cells (based on DAPI and scatter properties) are sorted into 384-well plates with 5 μL of Vapor-Lock oil containing a droplet of 100 nL of CEL-seq primers, spike-ins, and dinucleotide triphosphates (dNTPs). For cDNA construction, cells are first lysed by heat, after which a robotic liquid handler dispenses reverse transcription (RT) and second-strand mix. Cells are then pooled, and the aqueous phase is extracted from the oil. The CEL-Seq2 protocol can be followed from this point onward. Compared to the manual method, the percentage of reads that could be mapped to the reference transcriptome increased from 15% to 35%. In addition, the number of unique transcripts per cell increased (median of 14,604 compared to 4,262) (Figure S1D), as did the number of genes detected per cell (median of 4,497 compared to 1,958) while sequencing an average of 41.000 reads per cell, half the depth used for the manual method (Figure S1E). This resulted in more complex single-cell libraries with more differentially expressed genes between cell types (Figure S1F).

To investigate whether we could detect the expected pancreatic cell types, we used StemID, an approach we developed for inferring the existence of stem cell populations from single-cell transcriptomic data (Grün et al., 2016). StemID calculates all pairwise cell-to-cell distances (1 − Pearson correlation) and uses this to cluster similar cells into clusters that correspond to the cell types present in the tissue (Figure S1G). This resulted in well-separated cell clusters with low intra-cluster and high inter-cluster cell-to-cell distances, as visualized in t-distributed stochastic neighbor embedding (t-SNE) maps (Figure 1B) (van der Maaten and Hinton, 2008). These maps were also used to highlight expression of specific genes across all cells (Figure 1D). To test whether the donor source influenced cluster formation, we plot donor contribution to the clusters in Figure 1C, showing that none of the clusters consist of cells from only one donor. When we compared all cells from each cell type of one donor to that of all others, we did not find major differences between donors. Most differentially expressed genes differed by less than 2-fold. As expected, *XIST* was upregulated in all cell types of D30 (Table S2), the only female donor of the set. The donor-independent clustering shows StemID groups cells based on cell type, rather than donor.

We found the clusters to highly express markers for all pancreatic cell types (Figure 1D). We found cluster-specific expression of *GCG* (alpha cells), *INS* (beta), *SST* (delta), *PPY* (PP), *PRSS1* (acinar), *KRT19* (duct), and *COL1A1* (mesenchyme) (Figures 1D and S1H). Because the algorithm did not distinguish clusters with either epsilon or endothelial cells, we looked for expression of the markers *GHRL* or *ESAM*. We found two clusters of cells exclusively expressing these markers and manually annotated them as epsilon and endothelial cells (Figure 1B).

We also detected the expression of *MAFA* and *MAFB*, transcriptional regulators important for determining the identity of endocrine cell types (Nishimura et al., 2006) (Figure 1D). *MAFA* expression is restricted to beta cells, while *MAFB* expression is found in both alpha and beta cells, as previously reported in mice (Dai et al., 2012).

We next set out to generate a resource with which to compare pancreatic cell types and mine their transcriptomes for interesting genes. To this end, we compared all alpha (clusters expressing high *GCG*), beta, epsilon, delta, PP, duct, acinar, mesenchymal, and endothelial cells based on their distinct clustering from other cell types. Each group of cells was compared to all other cell groups, yielding a list of differentially expressed genes. The top ten of each list can be found in Figures 1E and S1I, and the full list is in Table S3. We then selected only those genes that have been reported to function as transcription factors using the TFcheckpoint database (Table S4) (Chawla et al., 2013). Several genes and transcription factors found here have never been reported as markers for specific cell types of the human pancreas (Figure 1E).

Figure 1. SORT-Seq Allows for Deep Sequencing of Human Pancreas Cells

(A) Experimental workflow for SORT-seq. Islets were isolated from human donors. Cells were dispersed and sorted into 384-well plates with mineral oil, containing 100 nL of CEL-seq2 primers, dNTPs, and spike-ins. The RT mix was then distributed by Nanodrop II. After second-strand synthesis, material was pooled and amplified before RNA library preparation.

(B) Visualization of k-medoid clustering and cell-to-cell distances using t-SNEs. Each dot represents a single cell. Colors and numbers indicate clusters, and cell-type names are indicated with their corresponding cluster or clusters.

(C) t-SNE map highlighting donor source. Each color represents one donor.

(D) t-SNE maps highlighting the expression of marker genes for each of the six main pancreatic cell types. Transcript counts are given in a linear scale.

(E) Tables denoting the top ten differentially expressed genes and transcription factors (TFs) when comparing one cell type to all other cells in the dataset ($p < 10^{-6}$). Genes whose cell-type specificity was previously unknown in the human pancreas are marked in red.

Apart from the classically known alpha cell transcription factors *IRX2*, *ARX* (Dorrell et al., 2011b), and *PGR* (Doglioni et al., 1990), our analysis reveals transcription factors *FEV*, *PTGER3* (Kimple et al., 2013), *SMARCA1* (Rankin and Kushner, 2010), *HMGB3*, and *RFX6* (Piccand et al., 2014) that, to our knowledge, have not been reported to be enriched in human alpha cells and have been previously implicated in beta cell function. Some of these factors have broader expression across other endocrine cell types, such as *RFX6*, but they are most highly expressed in alpha cells.

Classical beta cell markers like *INS*, *MAFA*, and *PDX1* (Kulkarni, 2004) top the beta cell list, and we detect *PFKFB2* (Arden et al., 2008), a gene thought to regulate insulin secretion, and the transcription factor *SIX2*. To our knowledge, neither *PFKFB2* nor *SIX2* have been reported previously in human beta cells. *SIX2* is known to interact with the transcription factor *TCF7L2* (Xu et al., 2014), a well-known SNP for type 2 diabetes (Grant et al., 2006). This makes it interesting for further investigations in the context of beta cell function.

Apart from the classical *SST* and *HHEX* expression in delta cells (Zhang et al., 2014), genes like *LEPR* and *GHSR* imply a possible role of leptin and ghrelin on delta cell function. PP cells have substantial expression of genes related to neuronal cells, which hints at the developmental proximity of PP and neuronal cells. This has been previously described by others in the context of beta cells (Arntfield and van der Kooy, 2011; Le Roith et al., 1982)

In summary, these gene lists confirm markers and reveal gene expression patterns in the endocrine cell types that can be further investigated for their roles in cellular identity and function.

Cluster-Restricted Gene Expression Patterns and Identification of Cell-Type-Specific Genes

We next analyzed each cluster in detail to see whether the remaining differentially expressed genes corroborated the initial identification of the six major pancreatic cell types. To investigate to what extent gene expression patterns are shared among cell types, we focused on the expression of both the top differentially expressed genes and the classical marker genes (Figure 2A). In particular, the expression of hormones was restricted to individual clusters, taking up one-fifth of the transcriptome, while being near zero in other clusters. For most clusters, the top differentially expressed genes were documented markers (Table S3). For example, *INS* and *IAPP* were co-expressed in beta cells, *LOXL4* was expressed with *GCG* (alpha cells), and *PNLIP* was expressed with *PRSS1*. *PRG4* was most highly expressed in delta cells after *SST*. Ductal markers *SPP1* and *KRT19* were relatively lowly expressed but limited to the ductal cluster. Further inspection of the top differentially expressed genes per cluster yielded new cell-type-specific genes, such as *ALDH1A1*, which was enriched in alpha cells and co-expressed with *GCG* (Figures S2C and S2D).

Going further down the list of differentially expressed genes continued to show cell type-restricted patterns (Figure S2A). To test whether we could use StemID clustering to compare different types of cells, we determined differentially expressed genes between all endocrine and exocrine cells. This yielded 2,858 genes that were differentially expressed ($p < 10^{-6}$ after Benjamini-Hochberg correction). Clear separation of endocrine and exocrine was visible by plotting the top 100 differentially expressed genes (Figure S2B). This list consisted of many genes related to endocrine function, proving single-cell sequencing yields useful data on specific pancreatic cell types. This allowed us to continue exploring differences between more closely related cell types such as alpha and beta cells, which yielded a list of 1,376 differentially expressed genes ($p < 10^{-6}$) (Figure 2B).

Plotting these differences in expression patterns showed clear cell-type-specific patterns (Figure 2C). Not surprisingly, canonical marker genes for alpha and beta cells (*GCG*, *MAFA*, *IAPP*, *CHGB*, *INS*, *INS-IGF2*, *SCG2*, *PCSK1*, and *PCSK2*) were in the list, as were genes found in studies that analyzed enriched populations of alpha or beta cells, such as *TTR*, which is specific in mouse alpha cells (Dorrell et al., 2011a); *NPTX2* in beta cells (Figure S2A) (Nica et al., 2013); and group-specific component (*GC*) in human alpha cells (Ackermann et al., 2016). We also identified several previously unreported cell-type-specific genes for both alpha cells (*CRYBA2*, *TM4SF4*, and *ALDH1A1*) and beta cells (*ID1*, *RBP4*, *SQSTM1*, *MT1X*, *FTL*, and *FTH1*) (Ackermann et al., 2016; Benner et al., 2014). Many of these beta cell-specific genes have been linked to Type 2 diabetes (T2D) or to oxidative and/or endoplasmic reticulum (ER) stress responses (Åkerfeldt and Laybutt, 2011; Chen et al., 2001; Orino et al., 2001; Yang et al., 2005). To validate our results, we visualized protein levels of *FTL* and *ALDH1A1* in tissue sections of human pancreas. FTL expression was visible in insulin-producing cells and absent from *GCG*-positive alpha cells in the islets of Langerhans (Figure 2D). *ALDHI1A1* expression appeared to be quite similar in acinar cells and alpha cells, whereas in general, higher mRNA expression was observed in alpha cells (Figures S2C and S2D). Within the islets of Langerhans, we detected ALDH1A1 expression only in glucagon-positive alpha cells, not in other cells in the islets.

GO-Term Analysis Reveals Cell-Type-Specific Gene Expression Patterns Relevant to Diabetes and Glucose Metabolism

We used EnrichR (Chen et al., 2013) to perform gene ontology (GO)-term analysis on the full list of genes differentially expressed in each cell type compared to all other pancreatic cell types. We determined the top 15 enriched GO terms for alpha, beta, delta, and PP cells (Figure S3A). In addition, we provide the lists of GO terms for each type, along with the genes that are involved in this GO term (Table S5). Parsing the file for alpha cell-related GO terms shows the inositol receptor *ITPR1* is involved in insulin secretion. *ITPR1* has previously been associated with a diabetic phenotype in mice (Figure S3C) (Ye et al., 2011). GO terms, like negative regulation of nervous system development, are highest in PP cells, indicating these cells have a more neuronal nature than other cells. The serotonin transporter *SLC6A4* is found predominantly in PP cells (Figure S3C) and has well-documented roles in neurons and behavior (Murphy and Lesch, 2008). To focus on differences between cell types in more detail, we performed GO-term analysis on gene sets obtained after comparing beta cells to alpha, delta, and PP cells separately (Figure S3B). In particular, delta cells show more hits in behavior and synaptic

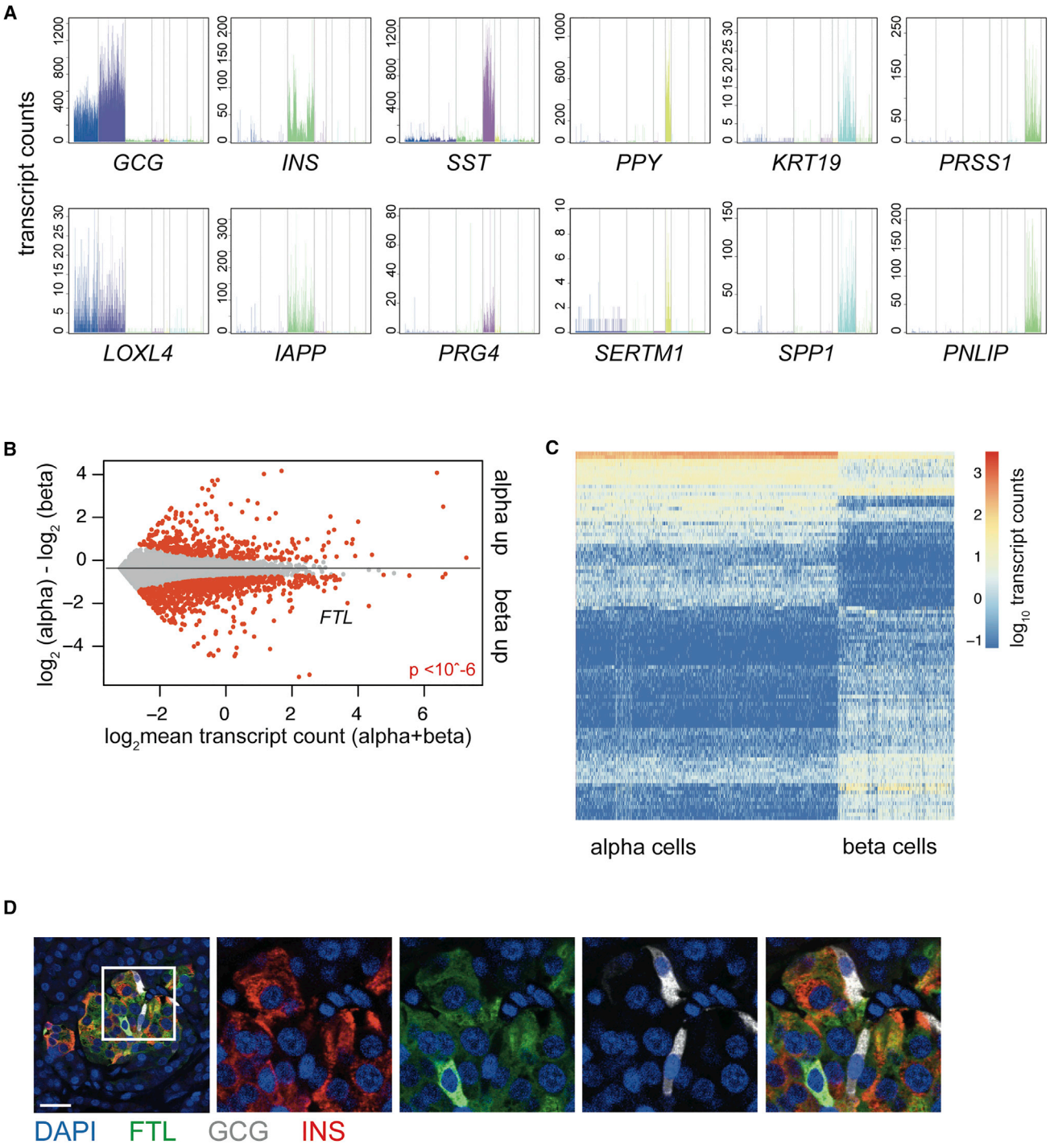


Figure 2. Cluster-Restricted Gene Expression Patterns and Identification of Cell-Type-Specific Genes

(A) Expression of well-known marker genes (top) and the most differentially expressed gene (bottom) in each of the six main pancreatic cell types. If the most differentially expressed gene is also a canonical marker gene, the next most differentially expressed gene is shown. Gene expression values are plotted on the y axis. Each bar represents a cell, and cells are grouped by cluster with a specific color in the following order: alpha, beta, delta, PP, duct, and acinar.

(B) Differential gene expression analysis between alpha and beta cells reveals 1,376 differentially expressed genes. Gray dots indicate genes; red dots indicate significant genes ($p < 10^{-6}$). The y axis indicates the $\log_2$ fold change, and the x axis shows the mean transcript count over both groups of cells.

(C) Heatmap of the top 100 differentially expressed genes between alpha and beta cells. Rows are genes, and columns are cells. Log_2 expression of transcript counts for genes is plotted. Columns are ordered based on cell type (alpha on the left and beta on the right). Genes are grouped based on hierarchical clustering.

(D) Immunohistochemistry for ferritin light subunit (FTL, green) glucagon (GCG, gray), and insulin (INS, red) with counterstaining for DAPI (blue) on human pancreatic tissue sections. Scale bar represents 25 μM.

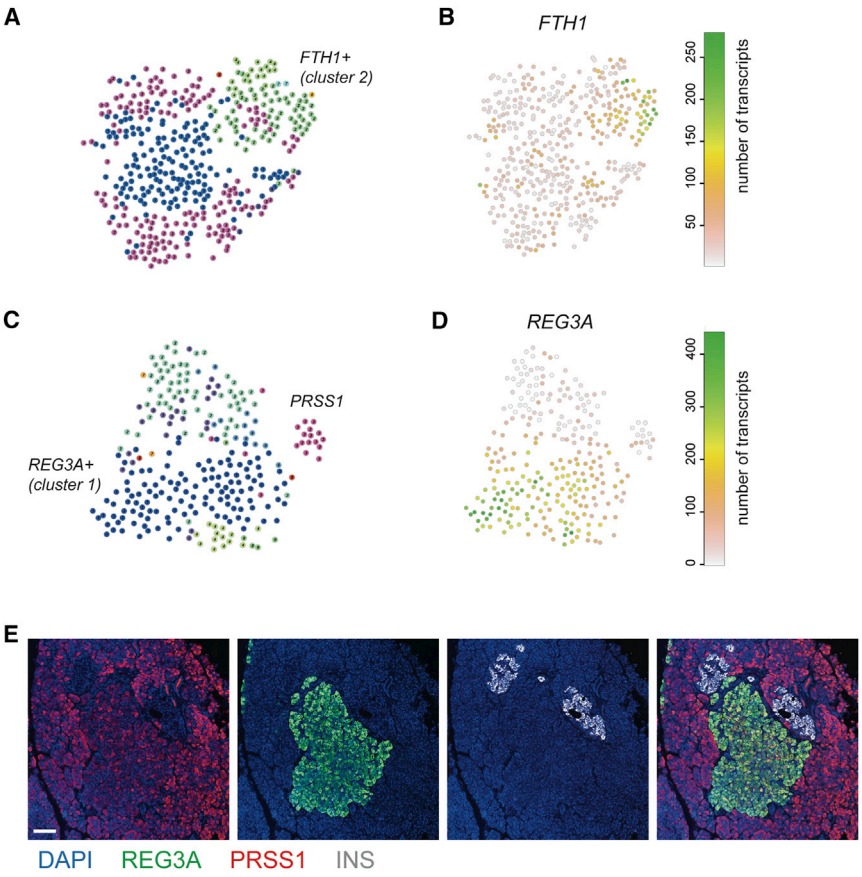


Figure 3. Outlier Identification Reveals Heterogeneity within Acinar and Beta Cells

(A) t-SNE map of RaceID clusters after clustering of only beta cells.

(B) t-SNE map highlighting the expression of *FTH1*.

(C) t-SNE map of RaceID clusters after clustering of only acinar cells.

(D) t-SNE map highlighting the expression of *REG3A*.

(E) Immunohistochemistry for REG3A (green), trypsin (red), and insulin (gray), with counterstaining for DAPI (blue). Scale bar represents 75 μM.

transmission. The ghrelin receptor *GHSR* is involved in several of these processes. This receptor is only present in delta cells (Figure S3C), indicating a role for ghrelin in delta cell function, which has indeed been demonstrated in mice (DiGruccio et al., 2016). These results are an example of how genes obtained in this resource can be used for GO-term analysis. By zooming in on specific genes from interesting terms, we can generate hypotheses regarding cell-type-specific processes in the human pancreas.

Outlier Identification Reveals Heterogeneity within Acinar and Beta Cells

We set out to analyze cellular heterogeneity by detecting outliers within specific populations of cells using the RaceID algorithm (Grün et al., 2015). The most striking results were found in beta and acinar cells, in which we found subpopulations of cells with distinct gene expression patterns. In beta cells, the most significant genes dictating this heterogeneity were *SRXN1*, *SQSTM1*, and three ferritin subunits: *FTH1P3*, *FTH1*, and *FTL* (Figures 3A and S4A). All these genes were highly expressed in cluster 2 (Figure S4A) and are implicated in response to ER and oxidative stress (Orino et al., 2001; Zhou et al., 2015; Rantanen et al., 2013). The main acinar cluster split into four clusters, of which cluster 2 showed the high levels of *REG3A* expression (Figures 3C, 3D, and S4B). Meanwhile, the acinar marker *PRSS1* was expressed in all clusters but was highest in a group of cells in clusters 3 and 4 (Figure S4C).

To confirm the existence of subpopulations of *REG3A*-positive acinar cells, we stained sections of human pancreas for REG3A and PRSS1. Scattered individual REG3A/PRSS1 double-positive cells were observed (Figure S4D) in acinar tissue. Interestingly, we also detected large clusters of brightly REG3A/PRSS1-positive acinar cells close to the islets of Langerhans (Figure 3E).

To characterize subpopulations obtained in silico in more depth, we averaged the expression profiles of all single cells belonging to the different subpopulations. By averaging and pooling the transcriptomes from these groups of cells, we achieve transcriptome coverage that is similar to that of bulk sequencing experiments (Table S6).

In summary, we detected subpopulations of beta cells expressing higher levels of *FTH1* and validated acinar subpopulations expressing high levels of REG3A. This subtype of acinar cell merits more investigation, because the role of *REG3A* in pancreas biology is unclear.

Enrichment of Alpha and Beta Cells Based on Cell-Surface Markers

We next mined our transcriptome resource for novel cell-surface markers to enrich specific pancreatic cell types using live-cell sorting. As a proof of principle, we set out to deplete the exocrine fraction from islet isolations of low purity. We found cell-surface markers *CD24* and *CD44* were restricted to acinar and ductal clusters (Figure 4A). Next, we prepared six FACS libraries, two with only live cells and four with negative selection for *CD24* and *CD44* (Figure 4B). This yielded compact clusters of cells that correspond to the main pancreatic cell types (Figure S5B). Nearly all endocrine cells were derived from the negatively selected libraries (Figure S5A), demonstrating the efficiency of the predicted cell-surface markers. Alpha cells seemed to be preferentially enriched with this strategy (Figure S5A).

To test whether we could enrich for one pancreatic cell type, we explored alpha cell-surface markers, finding *TM4SF4*, a tetraspanin family member that has been linked to pancreatic development (Anderson et al., 2011) and to be specifically

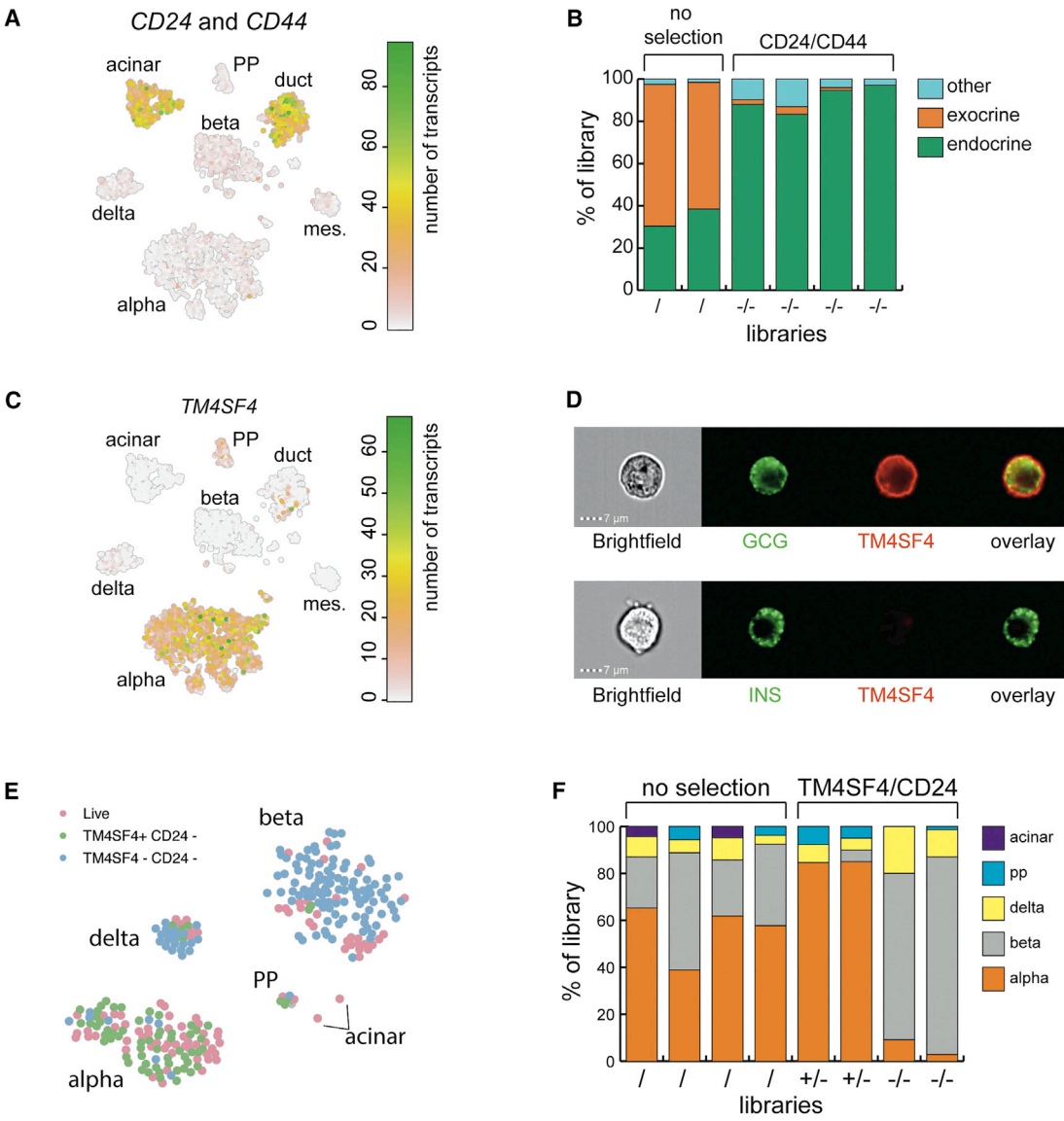


Figure 4. Enrichment of Alpha and Beta Cells Based on Cell-Surface Markers

(A) t-SNE map highlighting the combined expression of *CD24* and *CD44*.

(B) Results of FACS enrichment based on selection against CD24/CD44. Two libraries were selected for live staining (/), and four were selected against CD24 and CD44 expression (−/−). The y axis indicates the portion of the library consisting of a particular cell type. Colors indicate cell types.

(C) t-SNE map highlighting the expression of *TM4SF4*.

(D) Imagestream analysis of dispersed, fixed single-cells from human pancreas. The left panels show a bright field image of the cell and then immunostaining against glucagon (green) and Tm4sf4 (red). The lower panel shows insulin in green.

(E) t-SNE map highlighting libraries from a TM4SF4/CD24 sort. Cells that were not stained are in pink. Cells sorted for TM4SF4+/CD24− are in green and for TM4SF4−/CD24− are in blue.

(F) Results of FACS enrichment based on selection for TM4SF4/CD24. Four libraries were selected for live staining (/). Two libraries were TM4SF4+/CD24− (+/−), and two were TM4SF4−/CD24− (−/−). The y axis indicates the portion of the library consisting of a particular cell type. Colors indicate cell types.

expressed in alpha cells, with sparse expression in PP cells (Figure 4C). To verify the membrane-localized expression of TM4SF4 in alpha cells, we performed imaging flow cytometry analysis on fixed cells that were co-stained with either glucagon or insulin and TM4SF4 antibodies. We found TM4SF4 to be localized at the membrane of alpha cells but not that of beta cells (Figure 4D). To test whether this antibody can be used to enrich for alpha cells, we processed eight libraries from an endocrine-rich islet extraction (Table S1): four libraries were composed of live cells, two were CD24−/TM4SF4+, and two were CD24−/TM4SF4−. We found the main endocrine pancreatic cell types after clustering (Figure S5C). Libraries sorted for TM4SF4 consisted of >85% alpha cells. When selecting against TM4SF4 and CD24, alpha cells were depleted from the resulting population and enrichment for beta cells became possible with similar purity (Figure 4F).

In conclusion, this shows that our resource can be used to mine for genes with a specific subcellular location in a pancreatic cell type of choice. Table S7 provides a list of cell-type-enriched cell-surface markers in each of the main pancreatic cell types.

DISCUSSION

Scarcity of material, lack of reliable cell-surface markers, and analysis of pooled populations of cells often hamper analysis of human organ cell-type composition. Most importantly, methods relying on pooled cells average gene expression profiles over thousands of cells, masking any heterogeneity to be found within one cell type and potentially missing interesting intermediate cell types. To overcome these challenges, we have sequenced single cells from donor pancreata from four donors using SORT-seq, a FACS-compatible, automated version of the CEL-Seq2 protocol. We readily detected several clusters corresponding to the canonical pancreatic cell types, allowing us to purify cell types in silico for further analysis.

Due to consideration for transplantation, the islets obtained for this study were cultured for 3–5 days before dispersion to single cells and FACS. Culture conditions might affect the varied pancreatic cell types differently (progenitor cells are more likely to be affected than terminally differentiated cell types). However, shorter culture times for human islets are difficult to achieve, and we could not detect any major biases among donors, irrespective of their culture times.

Because the efficiency of single-cell sequencing (especially when using manual TRIzol-based methods) is on the order of 10% (Grün et al., 2014), lowly expressed genes are detected only sporadically. However, sequencing many cells enabled us to detect transcription factors, rare cell types, and heterogeneity within canonical pancreatic cell types such as acinar and beta cells. To further test the predictive power of this resource, we describe a panel of cell-surface markers specifically expressed in exocrine or alpha cells. Using these markers, we were able to enrich for endocrine, alpha, and beta cells.

In conclusion, we present this dataset as a resource that can be used to study pancreas composition and function. We envision broad applicability of this single-cell transcriptome atlas of the human pancreas to improve our understanding of pancreas biology and diabetes research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Human Islet Isolation, Dispersion, and Sorting
  - Single-Cell mRNA Sequencing of Single Cells
  - Immunofluorescence and Imaging Flow Cytometry
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data Analysis
- DATA AND SOFTWARE AVAILABILITY
  - Software
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, seven tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2016.09.002.

AUTHOR CONTRIBUTIONS

M.J.M., G.D., E.J.P.d.K., and A.v.O. conceived the project. M.J.M. and G.D. carried out the experiments. M.A.E. supervised the human islet isolation procedure. D.G. helped with StemID. N.G., T.D., E.J., L.v.G., and F.C. aided with the experiments. M.J.M., G.D., and A.v.O analyzed the data. M.J.M., G.D., E.J.P.d.K., and A.v.O wrote the manuscript.

ACKNOWLEDGMENTS

We thank Tamar Hashimshony and Itai Yanai for sharing CEL-Seq2. We thank USF for sequencing, Anko de Graaf for help with microscopy, and Reinier van der Linden for FACS. Many thanks to Nicola Crosetto for ideas on automation of CEL-Seq. This work was supported by an ERC Advanced grant (ERC-AdG 294325-GeneNoiseControl), a NWO VICI award, and grants from Stichting DON and the Dutch Diabetes Research Foundation. N.G. is supported by the JDRF.

Received: December 23, 2015
Revised: July 4, 2016
Accepted: September 7, 2016
Published: September 29, 2016

REFERENCES

Ackermann, A.M., Wang, Z., Schug, J., Naji, A., and Kaestner, K.H. (2016). Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol. Metab. *5*, 233–244.

Åkerfeldt, M.C., and Laybutt, D.R. (2011). Inhibition of Id1 augments insulin secretion and protects against high-fat diet-induced glucose intolerance. Diabetes *60*, 2506–2514.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. (2010). Genome Biol. *11*, R106.

Anderson, K.R., Singer, R.A., Balderes, D.A., Hernandez-Lagunas, L., Johnson, C.W., Artinger, K.B., and Sussel, L. (2011). The L6 domain tetraspanin Tm4sf4 regulates endocrine pancreas differentiation and directed cell migration. Development *138*, 3213–3224.

Arden, C., Hampson, L.J., Huang, G.C., Shaw, J.A.M., Aldibbiat, A., Holliman, G., Manas, D., Khan, S., Lange, A.J., and Agius, L. (2008). A role for PFK-2/FBPase-2, as distinct from fructose 2,6-bisphosphate, in regulation of insulin secretion in pancreatic beta-cells. Biochem. J. *411*, 41–51.

Arntfield, M.E., and van der Kooy, D. (2011). β-Cell evolution: how the pancreas borrowed from the brain: The shared toolbox of genes expressed by neural and pancreatic endocrine cells may reflect their evolutionary relationship. BioEssays *33*, 582–587.

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.

Bugliani, M., Liechti, R., Cheon, H., Suleiman, M., Marselli, L., Kirkpatrick, C., Filipponi, F., Boggi, U., Xenarios, I., Syed, F., et al. (2013). Microarray analysis of isolated human islet transcriptome in type 2 diabetes and the role of the ubiquitin-proteasome system in pancreatic beta cell dysfunction. Mol. Cell. Endocrinol. *367*, 1–10.

Chawla, K., Tripathi, S., Thommesen, L., Lægreid, A., and Kuiper, M. (2013). TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. Bioinformatics *29*, 2519–2520.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, H., Carlson, E.C., Pellet, L., Moritz, J.T., and Epstein, P.N. (2001). Overexpression of metallothionein in pancreatic beta-cells reduces streptozotocin-induced DNA damage and diabetes. Diabetes *50*, 2040–2046.

Cnop, M., Abdulkarim, B., Bottu, G., Cunha, D.A., Igoillo-Esteve, M., Masini, M., Turatsinze, J.V., Griebel, T., Villate, O., Santin, I., et al. (2014). RNA sequencing identifies dysregulation of the human pancreatic islet transcriptome by the saturated fatty acid palmitate. Diabetes *63*, 1978–1993.

Dai, C., Brissova, M., Hang, Y., Thompson, C., Poffenberger, G., Shostak, A., Chen, Z., Stein, R., and Powers, A.C. (2012). Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets. Diabetologia *55*, 707–718.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

DiGruccio, M.R., Mawla, A.M., Donaldson, C.J., Noguchi, G.M., Vaughan, J., Cowing-Zitron, C., van der Meulen, T., and Huising, M.O. (2016). Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. Mol. Metab. *5*, 449–458.

Doglioni, C., Gambacorta, M., Zamboni, G., Coggi, G., and Viale, G. (1990). Immunocytochemical localization of progesterone receptors in endocrine cells of the human pancreas. Am. J. Pathol. *137*, 999–1005.

Dorrell, C., Grompe, M.T., Pan, F.C., Zhong, Y., Canaday, P.S., Shultz, L.D., Greiner, D.L., Wright, C.V., Streeter, P.R., and Grompe, M. (2011a). Isolation of mouse pancreatic alpha, beta, duct and acinar populations with cell surface markers. Mol. Cell. Endocrinol. *339*, 144–150.

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. (2011b). Transcriptomes of the major human pancreatic cell types. Diabetologia *54*, 2832–2844.

Eizirik, D.L., Sammeth, M., Bouckenooghe, T., Bottu, G., Sisino, G., Igoillo-Esteve, M., Ortis, F., Santin, I., Colli, M.L., Barthson, J., et al. (2012). The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet. *8*, e1002552.

Grant, S.F.A., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., et al. (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat. Genet. *38*, 320–323.

Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. Cell *163*, 799–810.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Grün, D., Muraro, M.J., Boisset, J.-C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., van Es, J., Jansen, E., Clevers, H., et al. (2016). De novo prediction of stem cell identity using single-cell transcriptome data. Cell Stem Cell *19*, 266–277.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-seq: single-cell RNA-seq by multiplexed linear amplification. Cell Rep. *2*, 666–673.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-seq2: sensitive highly-multiplexed single-cell RNA-seq. Genome Biol. *17*, 77.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779.

Johnson, M.B., Wang, P.P., Atabay, K.D., Murphy, E.A., Doan, R.N., Hecht, J.L., and Walsh, C.A. (2015). Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. Nat. Neurosci. *18*, 637–646.

Kimple, M.E., Keller, M.P., Rabaglia, M.R., Pasker, R.L., Neuman, J.C., Truchan, N.A., Brar, H.K., and Attie, A.D. (2013). Prostaglandin E2 receptor, EP3, is induced in diabetic islets and negatively regulates glucose- and hormone-stimulated insulin secretion. Diabetes *62*, 1904–1912.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol. Cell *58*, 610–620.

Kulkarni, R.N. (2004). The islet beta-cell. Int. J. Biochem. Cell Biol. *36*, 365–371.

Le Roith, D., Shiloach, J., and Roth, J. (1982). Is there an earlier phylogenetic precursor that is common to both the nervous and endocrine systems? Peptides *3*, 211–215.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

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.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Murphy, D.L., and Lesch, K.-P. (2008). Targeting the murine serotonin transporter: insights into human neurobiology. Nat. Rev. Neurosci. *9*, 85–96.

Nica, A.C., Ongen, H., Irminger, J., 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.

Nishimura, W., Kondo, T., Salameh, T., El Khattabi, I., Dodge, R., Bonner-Weir, S., and Sharma, A. (2006). A switch from MafB to MafA expression accompanies differentiation to pancreatic beta-cells. Dev Biol. *293*, 526–539.

Orino, K., Lehman, L., Tsuji, Y., Ayaki, H., Torti, S.V., and Torti, F.M. (2001). Ferritin and the response to oxidative stress. Biochem. J. *357*, 241–247.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Piccand, J., Strasser, P., Hodson, D.J., Meunier, A., Ye, T., Keime, C., Birling, M.-C., Rutter, G.A., and Gradwohl, G. (2014). Rfx6 maintains the functional identity of adult pancreatic β cells. Cell Rep. *9*, 2219–2232.

Rankin, M.M., and Kushner, J.A. (2010). Aging induces a distinct gene expression program in mouse islets. Islets *2*, 345–352.

Rantanen, K., Pursiheimo, J.-P., Högel, H., Miikkulainen, P., Sundström, J., and Jaakkola, P.M. (2013). p62/SQSTM1 regulates cellular oxygen sensing by attenuating PHD3 activity through aggregate sequestration and enhanced degradation. J. Cell Sci. *126*, 1144–1154.

Ricordi, C., Lacy, P.E., Finke, E.H., Olack, B.J., and Scharp, D.W. (1988). Automated method for isolation of human pancreatic islets. Diabetes *37*, 413–420.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-seq analysis. Cell Stem Cell *6*, 468–478.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Wang, Y., and Navin, N.E. (2015). Advances and applications of single-cell sequencing technologies. Mol. Cell *58*, 598–609.

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, db160405, http://dx.doi.org/10.2337/db16-0405, Epub ahead of print.

WHO (2014). Global status report on noncommunicable diseases 2014. http://www.wsmi.org/wp-content/uploads/2015/01/Global-Status-Report-NCDs-2014.pdf.

Xin, Y., Kim, J., Ni, M., Wei, Y., Okamoto, H., Lee, J., Adler, C., Cavino, K., Murphy, A.J., Yancopoulos, G.D., et al. (2016). Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. Proc. Natl. Acad. Sci. USA *113*, 3293–3298.

Xu, J., Liu, H., Park, J.-S., Lan, Y., and Jiang, R. (2014). Osr1 acts downstream of and interacts synergistically with Six2 to maintain nephron progenitor cells during kidney organogenesis. Development *141*, 1442–1452.

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.

Ye, R., Ni, M., Wang, M., Luo, S., Zhu, G., Chow, R.H., and Lee, A.S. (2011). Inositol 1,4,5-trisphosphate receptor 1 mutation perturbs glucose homeostasis and enhances susceptibility to diet-induced diabetes. J. Endocrinol. *210*, 209–217.

Zhang, J., McKenna, L.B., Bogue, C.W., and Kaestner, K.H. (2014). The diabetes gene Hhex maintains δ-cell differentiation and islet function. Genes Dev. *28*, 829–834.

Zhou, Y., Duan, S., Zhou, Y., Yu, S., Wu, J., Wu, X., Zhao, J., and Zhao, Y. (2015). Sulfiredoxin-1 attenuates oxidative stress via Nrf2/ARE pathway and 2-Cys Prdxs after oxygen-glucose deprivation in astrocytes. J. Mol. Neurosci. *55*, 941–950.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-Ftl	Abcam	ab69090; RRID: AB_1523609
Mouse anti-Glucagon	Abcam	ab10988; RRID: AB_297642
Guinea pig anti-Insulin	Abcam	ab7842; RRID: AB_306130
Mouse anti-trypsin-1	Santacruz	sc-137077; RRID: AB_2300318
Rabbit anti-Reg3a	Abcam	ab134309
Rabbit anti-Aldh1a1	Abcam	ab23375; RRID: AB_2224009
TM4SF4-APC	BD	FAB7998A
FITC-mouse anti Human CD24 Clone ML5	BD	560992; RRID: AB_10562033
PE-mouse anti CD44(156-3C11)	Cell Signaling	8724S; RRID: AB_10829611
Chemicals, Peptides, and Recombinant Proteins		
CMRL 1066 medium	Mediatech	99663-CV
Accutase	StemCell Technologies, Inc.	07920
Vapor-lock	QIAGEN	981611
Critical Commercial Assays		
BD Cytofix/Cytoperm Fixation/Permeabilization Kit	BD	554717
Thermo Scientific reagents for CEL-Seq2	Hashimshony et al., 2016	N/A
Deposited Data		
Single-cell mRNA sequencing of cells from the pancreas of 4 human donors	this paper	GEO: GSE85241
Sequence-Based Reagents		
Reagents for library prep from CEL-Seq2	Hashimshony et al., 2016	N/A
Software and Algorithms		
IDEAS software	EMD Millipore	N/A
StemID algorithm	Grün et al., 2016	N/A
Rstudio software		Version 0.99.491
Bwa	Li and Durbin, 2009	N/A
Other		
Nanodrop II liquid handling platform (or any other liquid handling platform that can dispense sub-microliter quantities quickly into 384-well plates	Innovadyne	N/A
SP8 confocal microscope	Leica	N/A
Amnis Imagestream^X Mark II Imaging Flow cytometer	EMD Millipore	N/A
384-well plates for sorting and SORT-Seq protocol	Biorad	HSP3801
FACSJazz (or any other FACS-machine that can sort into 384-well plates)	BD	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests may be directed to primary contact Alexander van Oudenaarden, Hubrecht Institute (a.vanoudenaarden@hubrecht.eu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human cadaveric donor pancreata were procured through a multiorgan donor program. Pancreatic tissue was only used if the pancreas could not be used for clinical pancreas or islet transplantation, only if research consent was given and according to national laws. In total, 4 human donor pancreata were procured (3 male, 1 female). See Table S1 for details on donor age, sex and BMI.

METHOD DETAILS

Human Islet Isolation, Dispersion, and Sorting

Human islet isolations from pancreatic tissue were performed in the islet isolation facility of the Leiden University Medical Center according to a modified protocol originally described by Ricordi et al., (1988). Islets were cultured in CMRL 1066 medium (5.5 mM glucose) (Mediatech) supplemented with 10% human serum, 20 μg/ml ciprofloxacin, 50 μg/ml gentamycin, 2 mM L-glutamin, 0.25 μg/ml fungizone, 10 mM HEPES and 1.2 mg/ml nicotinamide for 3-6 days. Islets were maintained in culture at 37°C in a 5% CO_2 humidified atmosphere. Medium was refreshed the day after isolation and every 2-3 days thereafter until cell sorting. The islets were cultured for 3.-5 days after islet isolation. Culture time depended on the decision time needed for considering islets for transplantation and FACS.

For cell sorting cultured Islets were briefly washed in cold PBS. The islet pellet was then suspended in 1 ml of Accutase (Stemcell technologies) per 5000 islet equivalents and incubated at 37 degrees with gentle intermittent shaking for 8-10 min until the islets were dispersed into single cells. The digestion process was stopped using an excess volume of cold RPMI medium containing 10% FCS. The dispersed tissue was washed briefly with cold PBS followed by filtering through a sieve to get rid of any debris and undigested material. DAPI was added to access the viability of the cells. The tissue was stored on ice until sorting using a FACS Aria II or FACSJazz (BD biosciences). Live single cells (based on DAPI exclusion and forward/side scatter properties) were sorted into 384-well hard shell plates (Biorad) with 5 μl of vapor-lock (QIAGEN) containing 100-200 nl of RT primers, dNTPs and synthetic mRNA Spike-Ins and immediately spun down and frozen to −80°C. For cells sorted on cell surface markers; filtered, dispersed cells were incubated with FITC-CD24 (BD, 560992), PE-CD44 (Cell signaling, 8724S) and/or APC-TM4SF4 (BD, FAB7998A) antibodies for 30 min post dispersion on ice, followed by brief washing and sorting as above.

Single-Cell mRNA Sequencing of Single Cells

For SORT-seq, cells were lysed by 5 min at 65°C, after which RT and second strand mixes were dispersed with the Nanodrop II liquid handling platform (GC biotech). Aqueous phase was separated from the oil phase after pooling all cells in one library, followed by IVT transcription. The CEL-Seq2 protocol was used for library prep. Primers consisted of a 24 bp polyT stretch, a 4bp random molecular barcode (UMI), a cell-specific 8bp barcode, the 5′ Illumina TruSeq small RNA kit adaptor and a T7 promoter. mRNA of each cell was then reverse transcribed, converted to double-stranded cDNA, pooled and in vitro transcribed for linear following the CEL-Seq 2 protocol (Hashimshony et al., 2016). Illumina sequencing libraries were then prepared with the TruSeq small RNA primers (Illumina) and sequenced paired-end at 75 bp read length the Illumina NextSeq.

Immunofluorescence and Imaging Flow Cytometry

Pancreatic tissue samples were fixed overnight in 4% formaldehyde (Klinipath), stored in 70% ethanol, and subsequently embedded in paraffin. Sections were deparaffinized in xylene and rehydrated in a series of ethanol, followed by heat assisted antigen retrieval in citric buffer (pH 6.0). Sections were blocked by incubating with 2% normal donkey serum and 1% lamb serum in PBS. Primary antibodies included rabbit anti-FTL (ab69090), mouse anti-Glucagon (ab10988) and guinea pig anti-Insulin (ab7842), mouse anti-trypsin-1 (sc-137077), rabbit anti-REG3a (ab134309) and rabbit anti-ALDH1A1 (ab23375). Sections were incubated in with primary antibody in PBS/1% lamb serum at 4°C overnight. Alexa Fluor 488-, 568- and 647- conjugated secondary antibodies against rabbit, mouse, and guinea pig IgG as appropiate (Life Technologies A11008, A10037 and A21450) were diluted 1:200 and incubated at room temperature for 1 hr. Nuclear counterstaining was done with DAPI and by additionally embedding with DAPI vectashield (Vector Laboratories #H-1500). Imaging was done on a Leica SP8 confocal microscope using hybrid detectors.

TM4SF4 staining on alpha versus the beta cells was performed on fixed, stained single cells from dispersed human islets. Dispersed Islet cells were fixed with 4%PFA and washed using 2% FCS/PBS, followed by permeabilization using Perm/Wash buffer from BD Cytofix/Cytoperm Fixation/Permeabilization Kit (Cat. 554717) 15 min at room temperature. The samples were incubated with antibodies diluted in Perm/Wash buffer rabbit anti glucagon (1:200) or guinea pig anti insulin (1:200) or anti TM4SF4-APC (1:50) for 30 min at room temperature. Alexa Fluor 488- conjugated secondary antibodies (directly or in biotin-streptavidin system) against rabbit, and guinea pig as appropriate (Life Technologies A11008) were diluted 1:200 and incubated at room temperature for 30 min. These samples were imaged using Amnis Imagestream^X Mark II Imaging Flow cytometer (EMD Millipore, WA USA) with 488 nm and 642 nm lasers respectively. Analysis was done using the IDEAS software.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data Analysis

Paired-end reads from illumina sequencing were aligned to the human transcriptome with BWA (Li and Durbin, 2009). Read 1 was used for assigning reads to correct cells and libraries, while read 2 was mapped to gene models. Reads that mapped equally well to multiple locations were discarded. Read counts were first corrected for UMI barcode by removing duplicate reads that had identical combinations of library, cellular, and molecular barcodes and were mapped to the same gene. Transcript counts were then adjusted to the expected number of molecules based on counts, 256 possible UMI's and poissonian counting statistics.

Samples were normalized by downsampling to a minimum number of 6000 transcripts. StemID was used to cluster cells and to perform outlier analysis. Differentially expressed genes between two subgroups of cells were identified similar to a previously

published method (Anders and Huber, 2010). First, a negative binomial distribution was calculated reflecting the gene expression variability within each subgroup based on the background model for the expected transcript count variability computed by StemID (Grün et al., 2016). Using these distributions a p value for the observed difference in transcript counts between the two subgroups is computed as described in Anders and Huber (2010). These p values were corrected for multiple testing by the Benjamini-Hochberg method.

DATA AND SOFTWARE AVAILABILITY

Software

The StemID algorithm and custom scripts (Data S1) were run with RStudio, version 0.99.491.

Data Resources

The accession number for the single-cell sequencing data described in this study is GEO: GSE85241.

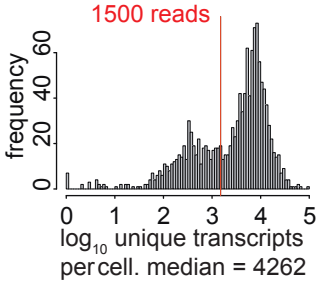
Cell Systems, Volume *3*

Supplemental Information

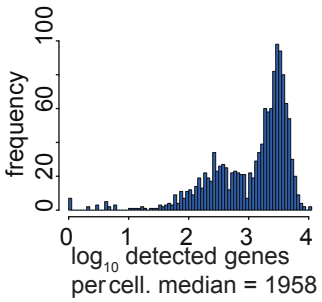
A Single-Cell Transcriptome Atlas of the Human Pancreas

Mauro J. Muraro, Gitanjali Dharmadhikari, Dominic Grün, Nathalie Groen, Tim Dielen, Erik Jansen, Leon van Gurp, Marten A. Engelse, Francoise Carlotti, Eelco J.P. de Koning, and Alexander van Oudenaarden

A



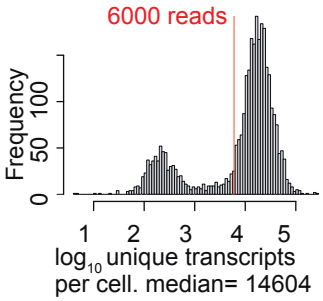
B



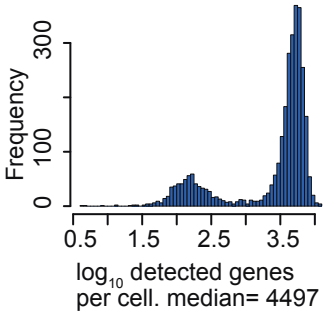
C

	alpha	beta	delta	PP
alpha		71	26	8
beta	6		30	12
delta	3	7		5
PP	2	1	1	

D



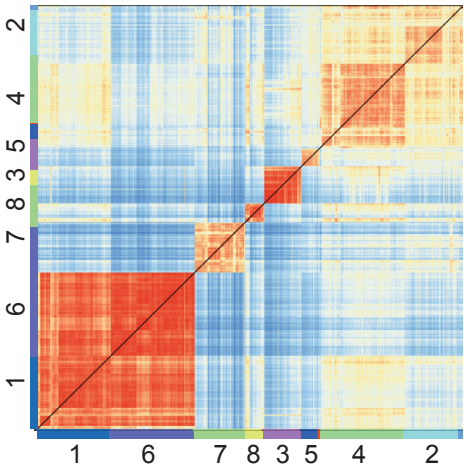
E



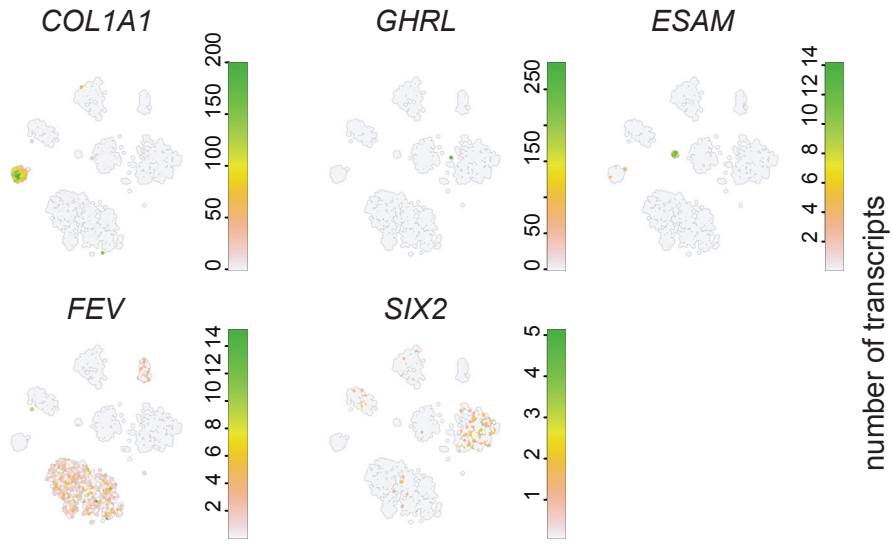
F

	alpha	beta	delta	PP
alpha		1376	533	259
beta	233		428	373
delta	71	74		160
PP	47	66	22	

G



H



I

epsilon cells		duct cells		acinar cells		mesenchymal cells		endothelial cells	
genes	TF	genes	TF	genes	TF	genes	TF	genes	TF
GHRL	*VTN*	*SPP1*	*ONECUT2*	*PNLIP*	*GATA4*	*COL1A1*	*WNT5A*	*FLT1*	*SOX18*
ANXA13	*EBF1*	*CFTR*	*LITAF*	*REG1B*	*MECOM*	*COL1A2*	*SNAI2*	*KDR*	*RGCC*
PHGR1	*BMP7*	*AQP1*	*SOX4*	*PRSS1*	*NR5A2*	*COL3A1*	*NOTCH3*	*CD93*	*SMAD6*
ACSL1	*CDKN2A*	*ALDH1A3*	*DAB2*	*ALB*	*ZFP36L1*	*COL6A3*	*FBN1*	*ESAM*	*ERG*
FRZB	*PROX1*	*KRT19*	*CREB5*	*PRSS3P2*	*CSDA*	*FN1*	*HEYL*	*SOX18*	*PRDM1*
SPTSSB	*ARX*	*CRP*	*HLA-DQB1*	*CPA2*	*CEBPD*	*SFRP2*	*PRRX1*	*PECAM1*	*TCF4*
ASGR1	*ZKSCAN1*	*DEFB1*	*WWTR1*	*CTRB2*	*CREB3L1*	*COL5A1*	*UACA*	*ESM1*	*NOTCH4*
HEPACAM2		*CEACAM6*	*PPARGC1A*	*CEL*	*XBP1*	*SPARC*	*AEBP1*	*PASK*	*SNAI1*
VTN		*MMP7*	*PKHD1*	*PLA2G1B*	*LGR4*	*COL15A1*	*TBX3*	*SLCO2A1*	*NKX2-3*
SERPINA1		*TSPAN8*	*NFIB*	*CELA3A*	*NUPR1*	*SERPINE1*	*FOXF2*	*PLVAP*	*ETS1*

Figure S1

Figure S1. SORT-Seq allows for deep sequencing of human pancreas cells, Related to Figure 1.

(A) Histogram of the total detected transcripts per cell for cells of the five donors processed by manual CEL-Seq. On the X-axis are the $\log_{10}$ detected unique transcripts per cell. Y-axis is the frequency. The minimum number of unique transcripts per cell used as cutoff for downsampling and analysis is indicated in red (1500).

(B) Histogram of genes detected per cell for cells of the first five donors processed by manual CEL-Seq. X-axis are the genes detected per cell. Y-axis is the frequency.

(C) Table indicating the differentially expressed genes (blue) and transcription factors (green) when comparing across the different endocrine cell types from data prepared by manual CEL-Seq.

(D) Histogram of the total detected transcripts per cell for cells of the four donors (SORT-Seq) used in this study. On the X-axis are the $\log_{10}$ detected unique transcripts per cell. Y-axis is the frequency. . The minimum number of unique transcripts per cell used as cutoff for downsampling and analysis is indicated in red (6000).

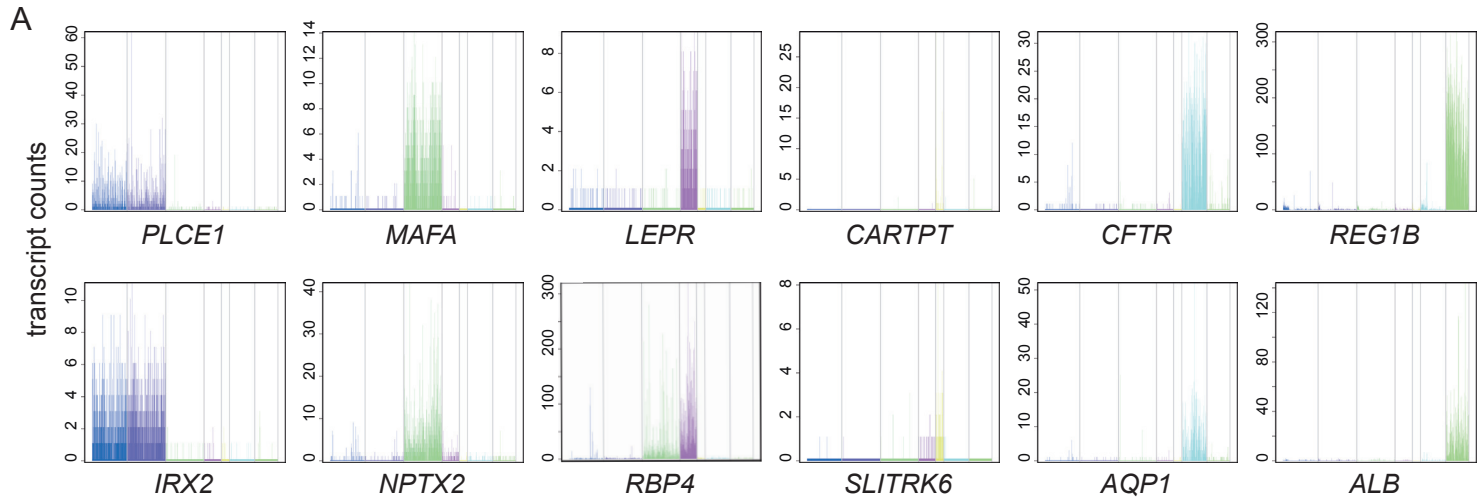
(E) Histogram of genes detected per cell for cells of the four donors (SORT-Seq) used in this study. X-axis are the genes detected per cell. Y-axis is the frequency. On average 1891 genes were detected.

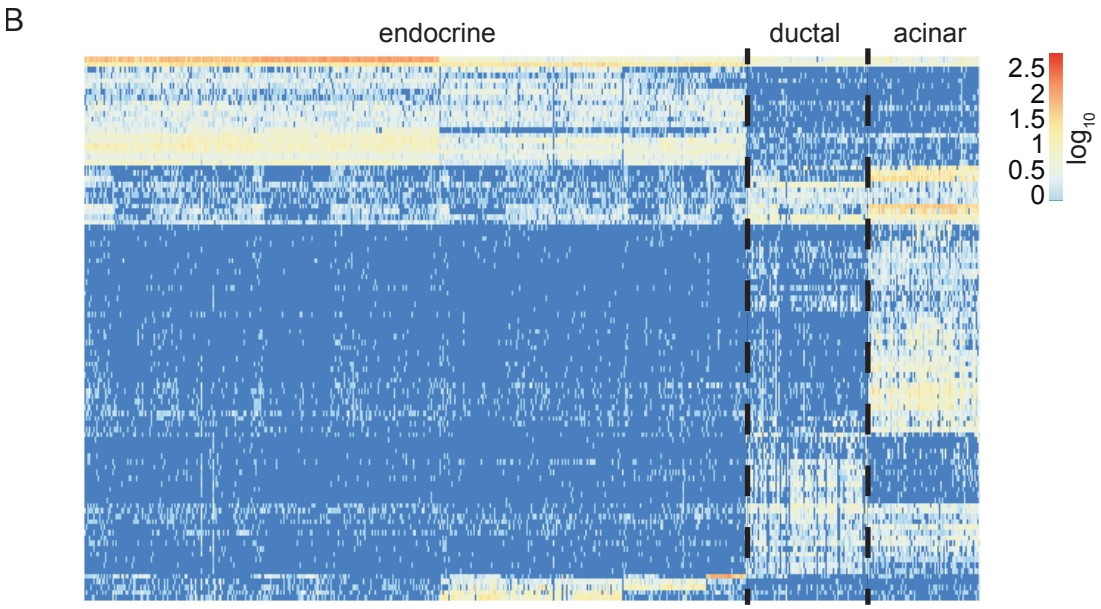
(F) Table indicating the differentially expressed genes (blue) and transcription factors (green) when comparing across the different endocrine cell types from data prepared by SORT-Seq.

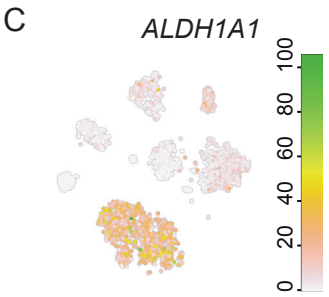
(G) Heat map showing distances between cellular transcriptomes obtained by sequencing. Clustering was performed by StemID (Grün et al, 2016). Distances are calculated as 1 – Pearson correlation and used as input for k-medoid clustering. Each line represents a cell and cells are grouped by cluster. Black lines indicate clusters, as do color bars and numbers on the axes, which match the colors and numbers in Figure 1B.

(H) t-SNE maps highlighting cell type-specific expression of pancreatic marker genes. Transcript counts are given in linear scale. Green indicates high expression.

(I) Tables denoting the top 10 differentially expressed genes and transcription factors (TF) when comparing one of the pancreatic cell types to all other cells in the dataset ($P<10^{-6}$). Continuation of Figure 1E.


A
transcript counts
PLCE1
MAFA
LEPR
CARTPT
CFTR
REG1B
IRX2
NPTX2
RBP4
SLITRK6
AQP1
ALB


B
endocrine
ductal
acinar
log10


C
ALDH1A1

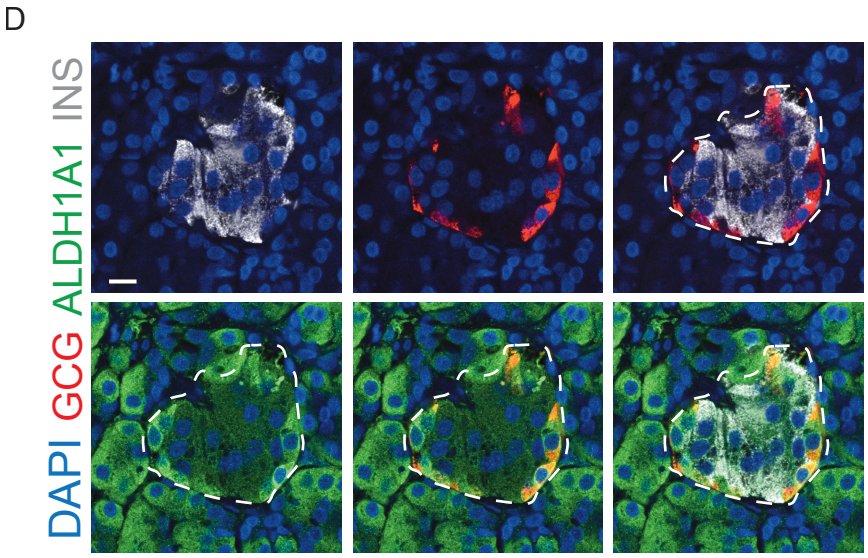

D
DAPI GCG ALDH1A1 INS

Figure S2

Figure S2. Cluster-restricted gene expression patterns and identification of new cell-type specific genes, Related to Figure 2.

(A) Expression of second (top) and third (bottom) most differentially expressed genes in each of six of the main pancreatic cell types. Down-sampled gene expression values are plotted on the Y-axis. Each bar represents a cell and cells are grouped by cluster with a specific color in the following order: alpha, beta, delta, PP, duct and acinar cells. If the most differentially expressed gene was also a canonical marker gene, the third and fourth most differentially expressed genes are shown.

(B) Heat map of the top 100 differentially expressed genes between endocrine and exocrine cell types. Rows are genes, columns are cells. Dashed lines indicate separation between acinar, ductal and endocrine cells. Log_2 expression of transcript counts for genes is plotted where red is high expression. Genes are grouped based on hierarchical clustering.

(C) t-SNE map highlighting the expression *ALDH1A1*. Transcript counts are given in linear scale. Green indicates high expression.

(D) Immunohistochemistry for ALDH1A1 (green) glucagon (red) and insulin (gray) with counterstaining for DAPI (blue) on human pancreatic tissue sections. Co-staining for INS and GCG identifies an Islet of Langerhans (marked by white dashed line). Co-staining of ALDH1A1 with GCG and INS shows overlap in the alpha cells, but not the beta cells inside the islet of Langerhans. Surrounding acinar cells express ALDH1A1 as well. Scale bar is 25 μM.

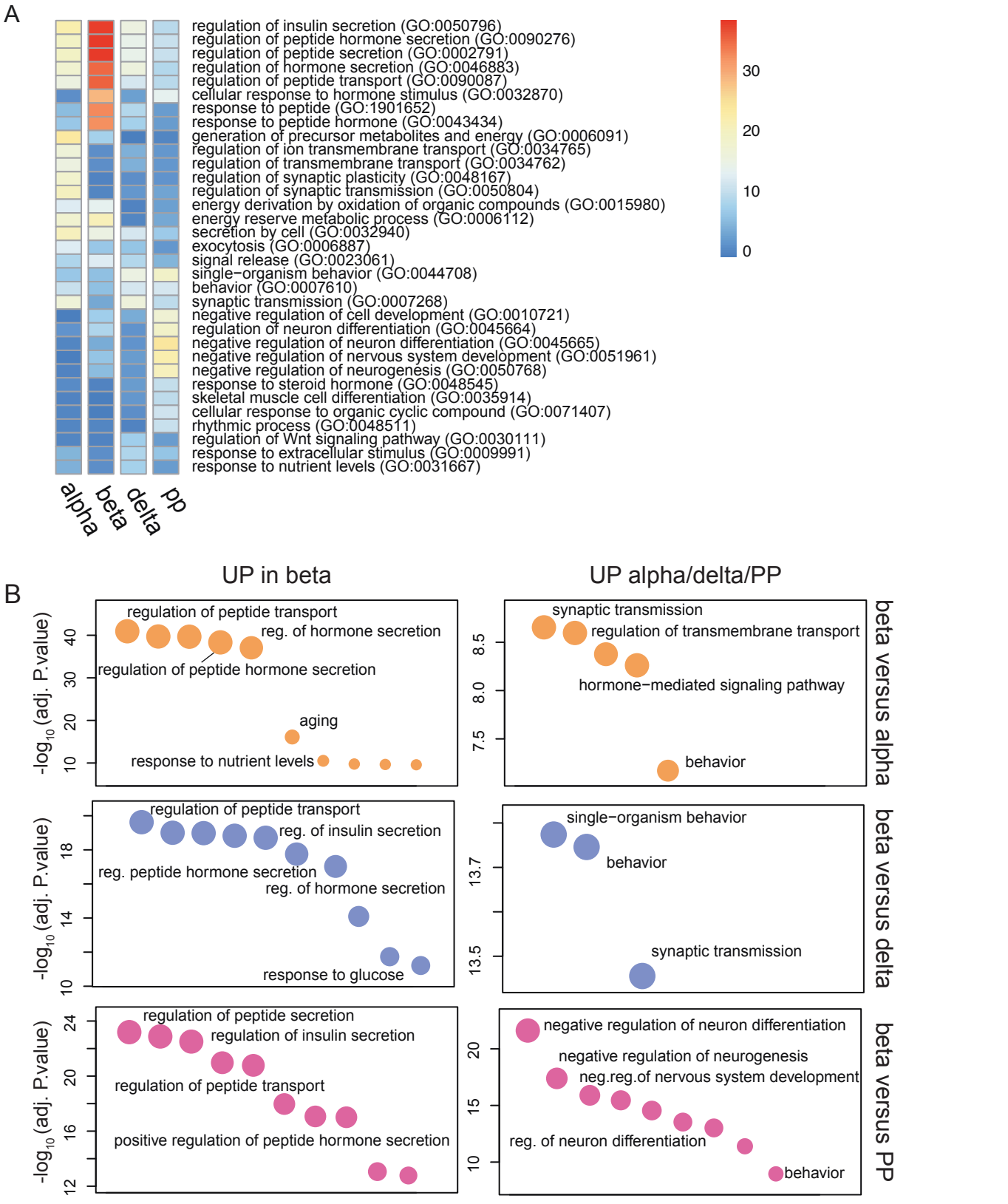

A
regulation of insulin secretion (GO:0050796)
regulation of peptide hormone secretion (GO:0090276)
regulation of peptide secretion (GO:0002791)
regulation of hormone secretion (GO:0046883)
regulation of peptide transport (GO:0090087)
cellular response to hormone stimulus (GO:0032870)
response to peptide (GO:1901652)
response to peptide hormone (GO:0043434)
generation of precursor metabolites and energy (GO:0006091)
regulation of ion transmembrane transport (GO:0034765)
regulation of transmembrane transport (GO:0034762)
regulation of synaptic plasticity (GO:0048167)
regulation of synaptic transmission (GO:0050804)
energy derivation by oxidation of organic compounds (GO:0015980)
energy reserve metabolic process (GO:0006112)
secretion by cell (GO:0032940)
exocytosis (GO:0006887)
signal release (GO:0023061)
single−organism behavior (GO:0044708)
behavior (GO:0007610)
synaptic transmission (GO:0007268)
negative regulation of cell development (GO:0010721)
regulation of neuron differentiation (GO:0045664)
negative regulation of neuron differentiation (GO:0045665)
negative regulation of nervous system development (GO:0051961)
negative regulation of neurogenesis (GO:0050768)
response to steroid hormone (GO:0048545)
skeletal muscle cell differentiation (GO:0035914)
cellular response to organic cyclic compound (GO:0071407)
rhythmic process (GO:0048511)
regulation of Wnt signaling pathway (GO:0030111)
response to extracellular stimulus (GO:0009991)
response to nutrient levels (GO:0031667)
alpha
beta
delta
pp
30
20
10
0
B
UP in beta
UP alpha/delta/PP
-log10 (adj. P.value)
regulation of peptide transport
reg. of hormone secretion
regulation of peptide hormone secretion
aging
response to nutrient levels
synaptic transmission
regulation of transmembrane transport
hormone−mediated signaling pathway
behavior
beta versus alpha
reg. of insulin secretion
reg. peptide hormone secretion
response to glucose
single−organism behavior
synaptic transmission
beta versus delta
regulation of peptide secretion
regulation of insulin secretion
positive regulation of peptide hormone secretion
negative regulation of neuron differentiation
negative regulation of neurogenesis
neg.reg.of nervous system development
reg. of neuron differentiation
beta versus PP
C
ITPR1
SLC6A4
GHSR

Figure S3

Figure S3. GO-term analysis reveals cell-type specific gene expression patterns relevant to endocrine biology and glucose metabolism.

(A) Heatmap showing the combined list of top 15 enriched GO terms for genes differentially expressed in endocrine cell types. Color indicates 1/p-value value so that red indicates a high score.

(B) Plot showing top 10 enriched GO terms for genes differentially expressed between beta cells compared to the three other endocrine cell types. The left column shows GO terms for genes with higher expression in beta cells, the right column shows GO terms of genes with higher expression in each of the other endocrine cell types. Terms are ordered on p-value on the x-axis, with the most significant on the left. Names of relevant terms are highlighted.

(C) t-SNE map genes found upon GO-term analysis with alpha, beta or delta specific expression. Green indicates high expression.

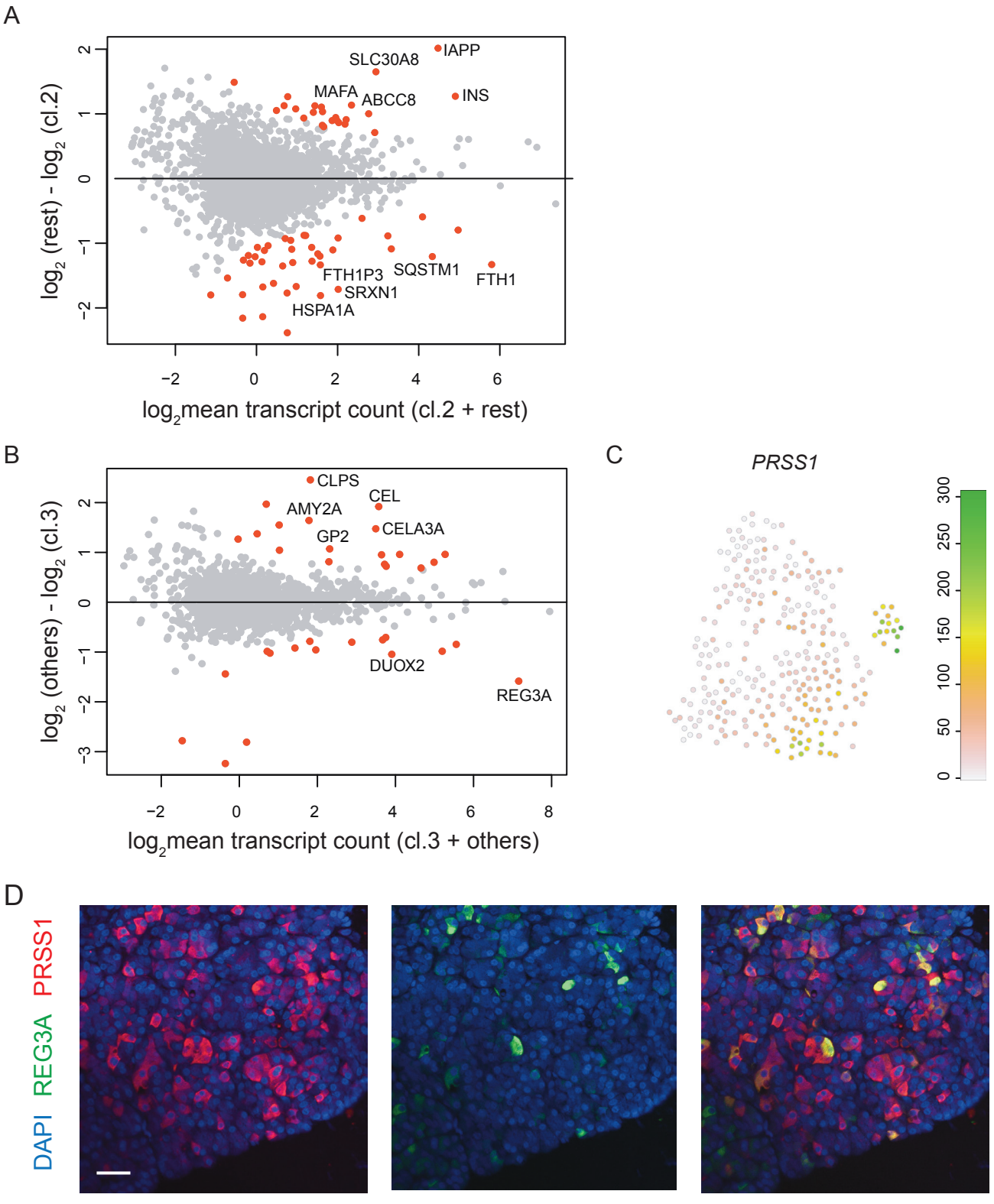

A
SLC30A8
IAPP
MAFA
ABCC8
INS
FTH1P3
SQSTM1
FTH1
SRXN1
HSPA1A
log2 (rest) - log2 (cl.2)
log2mean transcript count (cl.2 + rest)
B
CLPS
CEL
AMY2A
CELA3A
GP2
DUOX2
REG3A
log2 (others) - log2 (cl.3)
log2mean transcript count (cl.3 + others)
C
PRSS1
D
DAPI
REG3A
PRSS1

Figure S4

Figure S4. Outlier identification shows heterogeneity within acinar and beta cells, Related to Figure 3.

(A) Differential gene expression analysis of beta subclusters (*FTH1*-high cluster 2 versus the rest of the cells). Grey dots indicate genes, red dots indicate significant genes (*P*<10ˆ-6).

(B) Differential gene expression analysis between the acinar subclusters (*REG3A*-high cluster 1 versus the rest of the cells). Grey dots indicate genes, red dots indicate significant genes (*P*<10ˆ-6).

(B) t-SNE map highlighting the expression *PRSS1* across all acinar cells. Transcript counts are given in linear scale. Green indicates high expression.

(D) Immunohistochemistry showing protein expression for REG3A (green), and PRSS1 (red) with counterstaining for DAPI (blue). Scale bar is 25 μM.


A
Live
CD24 - / CD44 -
alpha
PP
endothelial
duct
delta
acinar
beta
B
GCG
INS
SST
PPY
PRSS1
KRT19
ESM1
C
GCG
INS
SST
PPY
TM4SF4

Figure S5

Figure S5. FACS Enrichment of endocrine and alpha cells based on novel cell-surface markers, Related to Figure 4

(A) t-SNE map highlighting the cells coming from the different FACS gating strategies. Each strategy is one color. Names of cell types are indicated next to their corresponding clusters. Cells sorted on only live (DAPI) marker are pink. Cells sorted against CD24 and CD44 expression are green.

(B) t-SNE maps highlighting the expression of the main pancreatic marker genes in libraries obtained by sorting for live or CD24/CD44 negative cells. Green indicates high expression.

(C) t-SNE map highlighting the expression of the main pancreatic marker genes in libraries obtained by sorting for live or CD24-/TM4SF4+/- cells. Green means high expression.

Supplemental tables and data

Donor	Sex	Age	BMI	Purity	Cell #	Type of sort
D28	Male	54	26	65	768	Live
D29	Male	23	22	95	768	Live
D30	Female	48	26	60	768	Live
D31	Male	59	25	60	768	Live
D16	Male	53	25	90	576	CD24/CD44
D25	male	30	18	55	768	CD24/TM4SF4

Table S1: Donor information (related to Figure 1)
Donor sex, age, BMI, purity of islet isolation procedure, number of processed cells and FACS gating parameters are indicated.

Table S2: **D30 Donor comparison (related to Figure 1)**
List of differentially expressed genes in each cell type between donor D30 and the other 3 donors.

Table S3: **Differential gene expression (related to figure 1)**
Differentially expressed genes between each cell type compared to all others (across all donors)

Table S4: **Differential transcription factor expression (related to figure 1)**
Differentially expressed transcription factors between each cell type compared to all others (across all donors)

Table S5: **GO-term analysis for endocrine cell types**
(Related to main text section: GO-term analysis reveals cell-type specific gene expression patterns relevant to diabetes and glucose metabolism")
GO-term analysis for alpha, beta, delta and PP cells compared to all others (across al donors)

Table S6: **Mean expression of subpopulations (related to figure 3)**
Average of gene expression across all cells of acinar and beta subpopulations.

Table S7: **Differential cell-surface marker expression (related to figure 4)**
Differentially expressed cell-surface markers factors between each cell type compared to all others (across all donors)

Data S1: (Related to Figures 1, 2, 3 and 4)
Data analysis script detailing StemID parameters and differential gene expression analysis between one cell type and all others.