Yue J. Wang,[1] Jonathan Schug,[1] Kyoung-Jae Won,[1] Chengyang Liu,[2] Ali Naji,[2] Dana Avrahami,[3] Maria L. Golson,[1] and Klaus H. Kaestner[1]

Single-Cell Transcriptomics of the Human Endocrine Pancreas

Diabetes 2016;65:3028–3038 | DOI: 10.2337/db16-0405

Human pancreatic islets consist of multiple endocrine cell types. To facilitate the detection of rare cellular states and uncover population heterogeneity, we performed single-cell RNA sequencing (RNA-seq) on islets from multiple deceased organ donors, including children, healthy adults, and individuals with type 1 or type 2 diabetes. We developed a robust computational biology framework for cell type annotation. Using this framework, we show that α- and β-cells from children exhibit less well-defined gene signatures than those in adults. Remarkably, α- and β-cells from donors with type 2 diabetes have expression profiles with features seen in children, indicating a partial dedifferentiation process. We also examined a naturally proliferating α-cell from a healthy adult, for which pathway analysis indicated activation of the cell cycle and repression of checkpoint control pathways. Importantly, this replicating α-cell exhibited activated Sonic hedgehog signaling, a pathway not previously known to contribute to human α-cell proliferation. Our study highlights the power of single-cell RNA-seq and provides a stepping stone for future explorations of cellular heterogeneity in pancreatic endocrine cells.

The endocrine pancreas plays a critical role in the control of blood glucose homeostasis. The endocrine cells are organized into the islets of Langerhans, approximately spherical groups of 500–1,000 cells, which together constitute only 1–2% of total pancreas mass. Pancreatic endocrine cells are characterized by their most abundant hormone, namely insulin (INS; β-cells), glucagon (GCG; α-cells), somatostatin (SST; δ-cells), pancreatic polypeptide (PPY; PP cells), and ghrelin (GHRL; ε-cells). The proportion of the assorted endocrine cell types and their arrangement within the islets varies widely among different mammalian species (1). For instance, whereas rodent islets are comprised of up to 90% insulin-producing β-cells in a distinct islet core, human islets display intermingled endocrine cells, with only ∼54% β-cells (2). Recently, diabetes researchers have renewed their focus on endocrine cellular heterogeneity (3). It is well accepted that not all β-cells are identical, especially in conditions of metabolic stress, such as obesity or type 2 diabetes (3–6). Moreover, it has been reported that in certain conditions of type 2 diabetes, a subset of pancreatic cells malfunction by reduction of glucose-stimulated insulin secretion or through dedifferentiation (7–9).

Individual cellular changes are diluted and therefore missed when analyzed at the level of the whole islet, or even when using sorted cell populations in bulk. Moreover, single-cell measurements can uncover unanticipated subpopulations, rare cellular states, or novel transcriptional mechanisms (10,11). Thus, methods to probe expression changes at the single-cell level are highly desirable (12–14). RNA sequencing (RNA-seq) can now be performed at the single-cell scale and, when applied in this manner, is an effective methodology for the analysis of gene expression variation among a population of apparently near-identical cells.

Here, we use single-cell RNA-seq to determine the transcriptomes of human pancreatic endocrine cells in four distinct developmental and physiological states: early childhood, normal adulthood, type 1 diabetes, and type 2 diabetes (see Fig. 1*A* for workflow and Table 1 for donor information). We discover that the transcriptional states of α- and

[1]Department of Genetics and Institute for Diabetes, Obesity, and Metabolism, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA
[2]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
[3]Endocrinology and Metabolism Service, Hadassah-Hebrew University Medical Center, Jerusalem, Israel

Corresponding authors: Klaus H. Kaestner, kaestner@mail.med.upenn.edu, Dana Avrahami, dana.tzfati@mail.huji.ac.il, and Maria L. Golson, mgolson@mail.med.upenn.edu.

Received 28 March 2016 and accepted 25 June 2016.

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db16-0405/-/DC1.

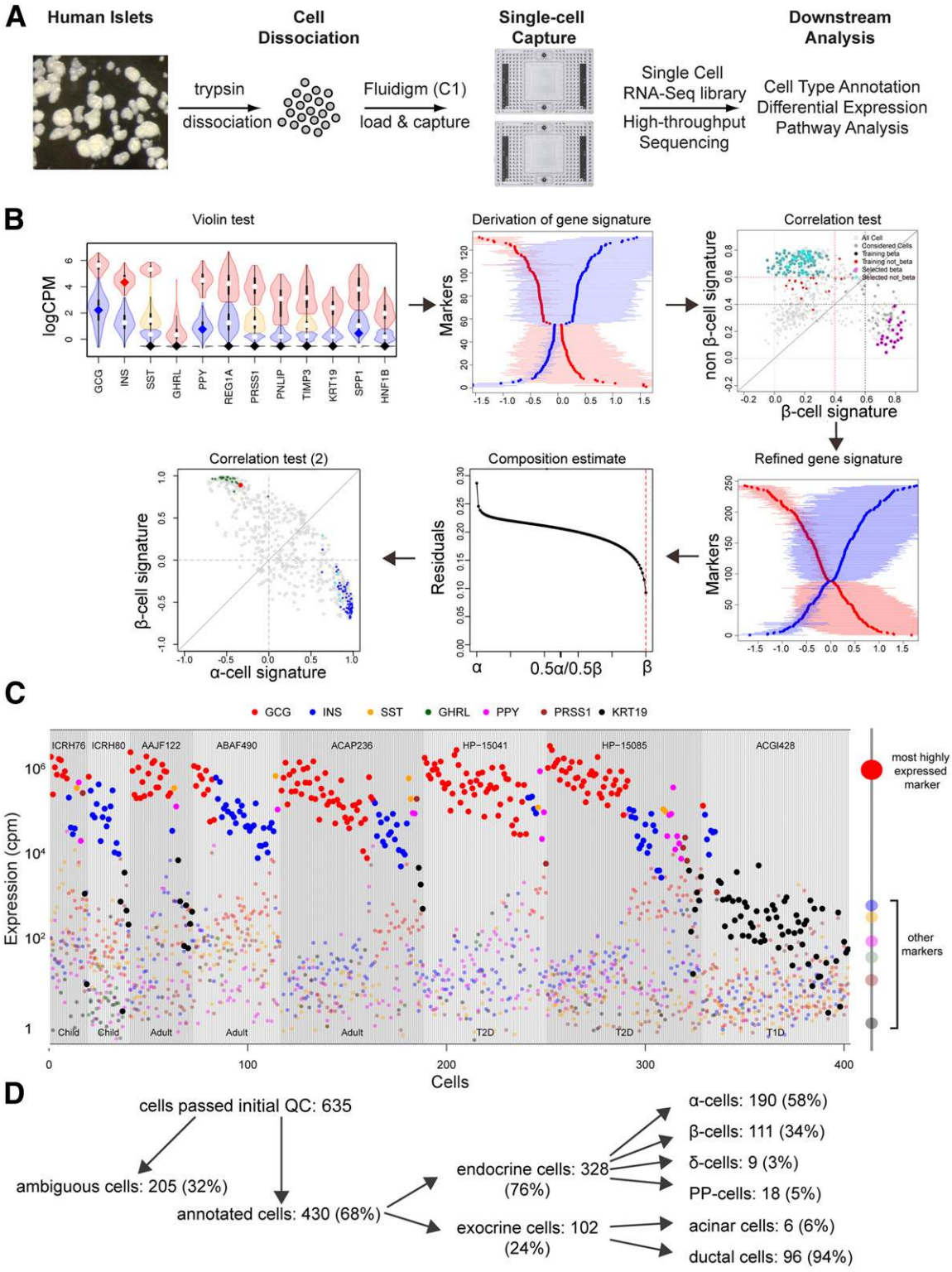

Figure 1—A microfluidics system combined with a computational pipeline results in efficient capture of single cells from human pancreatic islets and accurate annotation of their specific cell types. *A:* Schematic of experimental workflow. *B*: Cell type annotation pipeline. Results for one β-cell are displayed as an example. For details, see RESEARCH DESIGN AND METHODS. *C:* Abacus plot displaying the read counts per million (cpm) of key markers, including *GCG* (red), *INS* (blue), *SST* (orange), *GHRL* (green), *PPY* (pink), *PRSS1* (brown), and *KRT19* (black) in all annotated cells. Each vertical line in the abacus represents one cell. The most highly expressed marker in each cell is denoted by increased color opacity and size in relation to other markers. Within each sample, cells are ordered by GCG and thereafter INS levels. An example of expression of these markers in one cell is shown at the right of the abacus plot. In this example, the most highly expressed marker is *GCG* (red). All the other markers have a lower number of reads and are represented by corresponding colored dots. *D:* Overview of cell annotation frequencies. QC, quality control; T1D, type 1 diabetes; T2D, type 2 diabetes.

Table 1—Donor information

Donor ID	Age	Sex	Ethnicity	BMI (kg/m^2)	Cultured (days)	State	Cells captured						
							α	β	δ	PP	Duct	Acinar	Dropped
AAJF122	52 years	Male	Asian	29.1	6	Control	21	1	1	1	8	0	24
ABAF490	39 years	Female	White	45.2	4	Control	11	30	1	1	1	0	30
ACAP236	21 years	Male	White	39	2	Control	47	17	2	2	3	1	20
ACGI428	23 years	Male	NA	25	10	Type 1 diabetes	1	6	0	0	66	1	17
HP-15041	57 years	Male	African American	23.98	4	Type 2 diabetes	52	5	1	3	0	1	13
HP-15085	37 years	Female	White	39.3	4	Type 2 diabetes	38	16	2	7	7	2	17
HP-15085: cultured	37 years	Female	White	39.3	12	Type 2 diabetes	10	17	1	2	4	0	14
ICRH76	2 years	Male	White	13.6	2	Child	9	4	1	2	2	1	21
ICRH80	19 months	Female	White	18	3	Child	1	15	0	0	5	0	51

NA, not available.

β-cells are not fixed in early childhood but instead become more precisely defined as humans age. Furthermore, we find that in the diabetic state, α- and β-cells display a more immature gene signature, indicating a dedifferentiation process. Using this powerful technology, we find a high degree of gene expression variability within a given endocrine cell type and uncover Sonic hedgehog signaling as a mitogenic pathway potentially activated in replicating human α-cells.

RESEARCH DESIGN AND METHODS

Human Islet Sample Acquisition and Preparation

Human islets were acquired through the Diabetes Research Center of the University of Pennsylvania (National Institutes of Health DK-19525) and the Integrated Islet Distribution Program (IIDP; http://iidp.coh.org/). Prior to cell capture, islets were cultured in Prodo islet media (PIMS Standard) with 5% human albumin serum and a glucose concentration of 5.8 mmol/L. Islets were dissociated into single cells as described previously (15). Two integrated fluidic circuit chips, 5–10 and 10–17 μm in size (1006040 and 1006041; Fluidigm), were used for cell capture of each islet sample. The SMART-seq method was used for first-strand cDNA synthesis and PCR amplification (634833; Clontech). For two control adult donors (AAJF122 and ABAF490) (refer to Table 1 for details), ArrayControl spike-in (AM1780; Thermo Fisher Scientific) was applied during cell capture. Resulting cDNAs were pooled into a 96-well plate, and the Nextera XT DNA library preparation kit (FC-131-1096; Illumina) was used for RNA-seq library preparation according to the Fluidigm protocol (100-7168). Bulk β-cell RNA-seq data were obtained from a previous publication (16), with the addition of two samples from children.

Sequencing

All libraries were sequenced on the Illumina HiSeq 2500 with 100-bp single-end reads. Median read depth was 2.2 million per cell. Read alignment and gene expression quantification was performed using RNA-seq unified mapper (RUM) (17). Cells with <500,000 uniquely aligned reads were excluded from downstream analysis. A total of 82% of sequenced cells passed initial technical quality control. All sequencing data are available in the Gene Expression Omnibus (GEO) repository (accession number GSE83139; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE83139).

Cell Type Calling

In light of a recent finding that the 10–17-μm C1 chip can capture a significant number of cell doublets (Fluidigm white paper), we developed a five-step pipeline to distinguish single cells from doublets and to annotate cell types. First, we crudely classified cells into different cell types based on marker genes that have well-established cell type–restricted expression. These markers include the major hormone genes (*GCG*, *INS*, *SST*, *GHRL*, and *PPY*), genes that encode acinar cell–specific digestive enzymes (*PRSS1* and *PNLIP*), and genes associated with ductal cells (i.e., *KRT19*, *SPP1*, and *HNF1B*). The expression levels of each gene were fitted by normal mixture modeling of high, low, and (or) intermediate components using the mclust R package (cran.r-project.org/web/packages/mclust) and rendered in a "violin plot." Cells were assigned to a cell type only when they expressed specific markers in the mixture model's high expression level component and had no conflict with other markers (Fig. 1*B*, violin test). Second, we derived gene expression signatures for different cell types by comparing the transcriptome of each cell type to the rest of the cells in the same endocrine or exocrine category (Fig. 1*B*, derivation of gene signature). Third, to further refine cell type classification, previously derived gene signatures were used to calculate the correlation of each cell to its corresponding cell type–specific signature. As expected, pure cells clustered together and showed strong correlation with only one of the signatures (Fig. 1*B*, correlation test). Fourth, all cells that passed these two tests were collected and their gene expression patterns rederived (Fig. 1*B*, refined gene

signature). Fifth, we hypothesized that any cell doublets could be explained by a linear combination of two "pure" profiles. Accordingly, the new signatures from pure cell populations were used to construct in silico mixed profiles, including all the possible combinations between different cell types. To eliminate likely doublets from further analysis, all cells were examined based on the likelihood that their profiles could be explained by an artificial in silico mixture (Fig. 1*B*, composition estimate). They were also inspected based on their correlation to different pure signatures (2) (Fig. 1*B*, correlation test). The approach outlined above is similar in concept to that used in a single-cell RNA-seq study of human islets published recently (18). However, our analysis addresses the potential issue of doublets with any combination of cell types whether they are ductal, acinar, delta, etc., and not just between α- and β-cells.

To assess background contamination levels, we sequenced 25 wells that lacked any cells by microscopic inspection after cell capture. The ratios of all transcripts versus RNA spike-ins for all sequenced wells, together with the ratio of spike-ins versus transcripts of interest (e.g., GCG) of empty wells, were modeled by mclust (cran.r-project.org/web/packages/mclust) (Supplementary Fig. 5*A* and *B*). Based on this estimate, the predicted contamination level for transcripts of interest (e.g., GCG) was calculated (Supplementary Fig. 5*C*) using the following equation: predicted proportion of reads from GCG contamination = [(reads from all transcripts/reads from RNA spike-ins)$_{\text{non-empty wells}}$ × (reads from RNA spike-ins/reads from GCG)$_{\text{empty wells}}$]$^{-1}$.

Downstream Analysis

EdgeR and DESeq2 were used with default settings for differential expression analysis (19,20). Pathway analyses were performed with Gene Set Enrichment Analysis (GSEA; http://software.broadinstitute.org/gsea/index.jsp), Ingenuity Pathway Analysis (IPA; http://www.ingenuity.com/), and the Database for Annotation, Visualization, and Integrated Discovery (DAVID; https://david.ncifcrf.gov/) (21,22). Heatmaps were generated by the "aheatmap" function from the NMR R package (http://renozao.github.io/NMF), or by the "heatmap2" function from gplots R package (https://cran.r-project.org/web/packages/gplots/), with Pearson correlation as the distance function.

RESULTS

Single-Cell RNA-seq Analysis of Human Islets

We performed single-cell RNA-seq of islets from eight deceased organ donors, of whom there were three adults without diabetes, one subject with type 1 diabetes, two subjects with type 2 diabetes, and two children (Table 1). We dissociated human islets into single-cell suspensions and loaded the cell mixture onto Fluidigm C1 chips without preselection for particular cell types (Fig. 1*A*). The mean diameter of human pancreatic endocrine cells is 9 μm, with a range from 6 to 13 μm (data not shown). In an effort to reduce size-related sampling bias, for each donor sample, we captured cells on both the 5–10- and 10–17-μm chips. Amplified cDNAs were then indexed, pooled, and further processed to construct RNA-seq libraries (Fig. 1*A*). A total of 635 cells were sequenced and passed our initial sequencing quality control (Fig. 1*D* and Table 1).

Islets are often received after different culturing times. Since in vitro culture conditions may alter islet cell function (23–25), we assessed whether culture duration influences gene expression by comparing islet transcriptomes from a single donor with type 2 diabetes at two different time points: after 4 days in culture and after a total of 12 days in culture. Remarkably, RNA expression levels between cultured and fresh cells displayed good correlations (Spearman correlation coefficient 0.863 for α-cells and 0.857 for β-cells) (Supplementary Fig. 1). Whereas other groups have reported decreased expression of differentiation and maturity markers such as *ARX*, *MAFA*, and *NEUROD1* with increased culture time (25), we observed either no change or an increase in these markers either in α- or β-cells (Supplementary Figs. 2 and 3). It is possible that the current culture conditions are improved over those used in previous studies. Furthermore, our expression profiles are based on pure α- and β-cells, whereas the previous study used laser capture microdissected bulk tissues (25). We thus conclude that differences in culture time between different donor samples have only a minimal impact on our downstream analysis.

Cell Type Annotation

We developed a computational pipeline for cell type annotation and for the removal of potential cell doublets (see RESEARCH DESIGN AND METHODS) (Fig. 1*B*). In brief, we first separated our sequenced cells into different cell types based on key marker gene expressions. Subsequently, we derived cell type–specific gene signatures for pure α-, β-, δ-, PP, ductal, and acinar cells from our samples and calculated an in silico mix of profiles for each potential combination of cell types (Fig. 1*B*). Next, we calculated the distance of the expression profile of each cell to all of the mixture profiles, as well as the correlation of the cell to different signatures (Fig. 1*B*). We successfully assigned a specific cell type to 430 cells. For a general sample overview, we plotted the marker gene expressions in each of the successfully annotated cells in an abacus plot (Fig. 1*C*). In total, we annotated 190 α-cells, 111 β-cells, 9 δ-cells, 18 PP-cells, 96 ductal cells, and 6 acinar cells. Due to ambiguous profiles, 205 cells were initially excluded from analysis (Fig. 1*D* and Table 1). A high proportion of ductal cells were obtained from the donor with type 1 diabetes, unsurprisingly since islet purity was low for this donor. Interestingly, whereas all donor samples contained some cells with undetermined annotation, most cells with an undetermined phenotype were from juvenile donors. Notably, these cells tended to have a conflicted endocrine/exocrine nature. This result could potentially be explained either by the phenomenon that islet samples from children are highly embedded in the surrounding exocrine and stromal tissue (26,27) or

by less-defined gene expression patterns in children (28,29). In contrast, most of the undetermined cells from the donors with type 2 diabetes demonstrated a mixture of endocrine cell signatures, mainly α- versus β-cells.

Validation of Cell Type Analysis

To evaluate the cell type classification strategy, we first plotted the aggregated single β-cell gene expression values against RNA-seq data from sorted bulk β-populations for both adult and child samples (Fig. 2*A* and *B*). We observed remarkably good agreement between the two methodologies, with a Spearman correlation of >0.80. The most abundant transcripts are *INS* and *INS-IGF2* (Fig. 2*A* and *B*). As expected, very low-abundance genes are only detected by bulk cell expression analysis. These genes are often missed in single-cell RNA-seq experiments due to stochastic sampling during the cDNA amplification process. Interestingly, some of the discordant genes between bulk and single-cell RNA-seq are non–β-cell hormone markers, likely reflecting the limited purity of FACS-sorted human islet cells (Fig. 2*A* and *B*, labeled dots). This result demonstrates the advantage of single-cell RNA-seq for gene expression analyses of human islet cell populations, since we can retrospectively confirm cell type, thereby ensuring sample purity and eliminating gene expression signals from contaminating cells present in bulk cell fractions.

We subsequently performed hierarchal clustering and constructed a heatmap using cell type markers, important transcription factors, and signaling pathway indicators (Fig. 2*C*). We confirmed that well-established cell type markers were highly expressed, as expected, in their assigned cell types. For example, β-cells had high levels of *PDX1*, *IAPP*, *PCSK1*, *MAFA*, and *IGF2*, whereas α-cells had high expression of *IRX2* and *ARX*. Both cell types have comparable levels of *FOXA2*, *NKX2-2*, *NEUROD1*, and *MAFB*, although cell-to-cell heterogeneity is also apparent. We further examined expression of key marker genes in all endocrine cell types plus ductal and acinar cells and found that all cells classified within a given cell type expressed high levels of their expected marker gene and low levels of all others (Fig. 2*D–I* and Supplementary Fig. 4). For example, β-cells expressed high levels of *INS* and low levels of acinar and ductal markers, whereas duct cells showed high transcript levels for *KRT19* but only marginal signals for hormone genes (Fig. 2*D–I* and Supplementary Fig. 4). Furthermore, our data show a high degree of similarity to recently published single-cell transcriptomic data of 60 pancreatic islet cells from control adult donors (18). For example, *ARX* is expressed highly in both α- and PP cells, whereas *ETV1* is highly enriched in PP and δ-cells (Supplementary Fig. 4). Together, these observations demonstrate the high accuracy of our cell type annotation methodology.

Assessment of Background Contamination Levels

Some of the assigned cells demonstrate mRNA expression of markers not normally associated with that cell type. For example, low levels of *GCG* expression were also noted in non–α-cells (Fig. 1*C*, faint, small red dots). Since we rigorously excluded cells that could represent doublets, we next examined whether the secondary markers could be a result of RNA contamination originating from lysed cells prior to capture. To this end, we sequenced cDNA from 25 empty wells and modeled the relationship between the number of all mapped transcripts, reads from spike-in RNA, and reads from *GCG* transcripts to predict the level of *GCG* mRNA expression arising from contamination (see RESEARCH DESIGN AND METHODS and Supplementary Fig. 5). Based on this analysis, we concluded that the low levels of *GCG* found across all cell types most likely resulted from contamination.

Gene Expression Profiles of α- and β-Cells From Donors With Type 2 Diabetes Show Features of Juvenile Gene Activation

Next, we derived α-cell and β-cell gene signatures from adult donors without diabetes. By design, when these genes (indicated as black bars in Fig. 3*A* and *B*) are projected back on the rank-ordered differentially expressed genes between these two cell types, α-cell signature genes cluster on the α-cell side (Fig. 3*A*), and β-cell signatures genes are expressed only in β-cells (Fig. 3*B*). When we then analyzed the expression pattern of these adult signature genes in α- and β-cells from young children, many of them were no longer expressed in the expected cell type–specific fashion (Fig. 3*C* and *D*). Multiple α-cell signature genes were in fact preferentially expressed in juvenile β-cells (Fig. 3*C*). This inappropriate gene activation profile is even more pronounced for α-cells of young children; a very large fraction of adult β-cell genes are expressed at high levels in α-cells (Fig. 3*D*). Consequently, the gene set enrichment scores for the adult endocrine cell gene signatures are much lower in the young α- and β-cells, indicative of an only partially complete differentiation program in juvenile islets. The lists of genes that are misexpressed in children are presented in Supplementary Table 1. For lists of genes used to define α- and β-cell signatures see Supplementary Table 2.

Subsequently, we analyzed the enrichment of the nondiabetic adult islet cell gene signatures in α- and β-cells from individuals with type 2 diabetes. Gene sets from donors with type 2 diabetes had intermediate enrichment scores between those from adult control subjects and children (Fig. 3*E* and *F*). Similarly, several of the α- or β-cell signature genes were expressed at high levels in the inappropriate endocrine cell type (Fig. 3*E* and *F*). This feature is reminiscent of the pattern seen when comparing gene sets between child donors and adult control donors (Fig. 3*C* and *D*). This resemblance of endocrine cells between type 2 diabetic samples and child samples was confirmed when we derived α-and β-cell gene signatures from our juvenile organ donors and compared them to the gene expression patterns present in normal and type 2 diabetic adult endocrine cells (Fig. 4). Again, gene expression profiles of both α- and β-cells from donors with type 2 diabetes exhibited features of gene activation seen in children (Fig. 4*A*

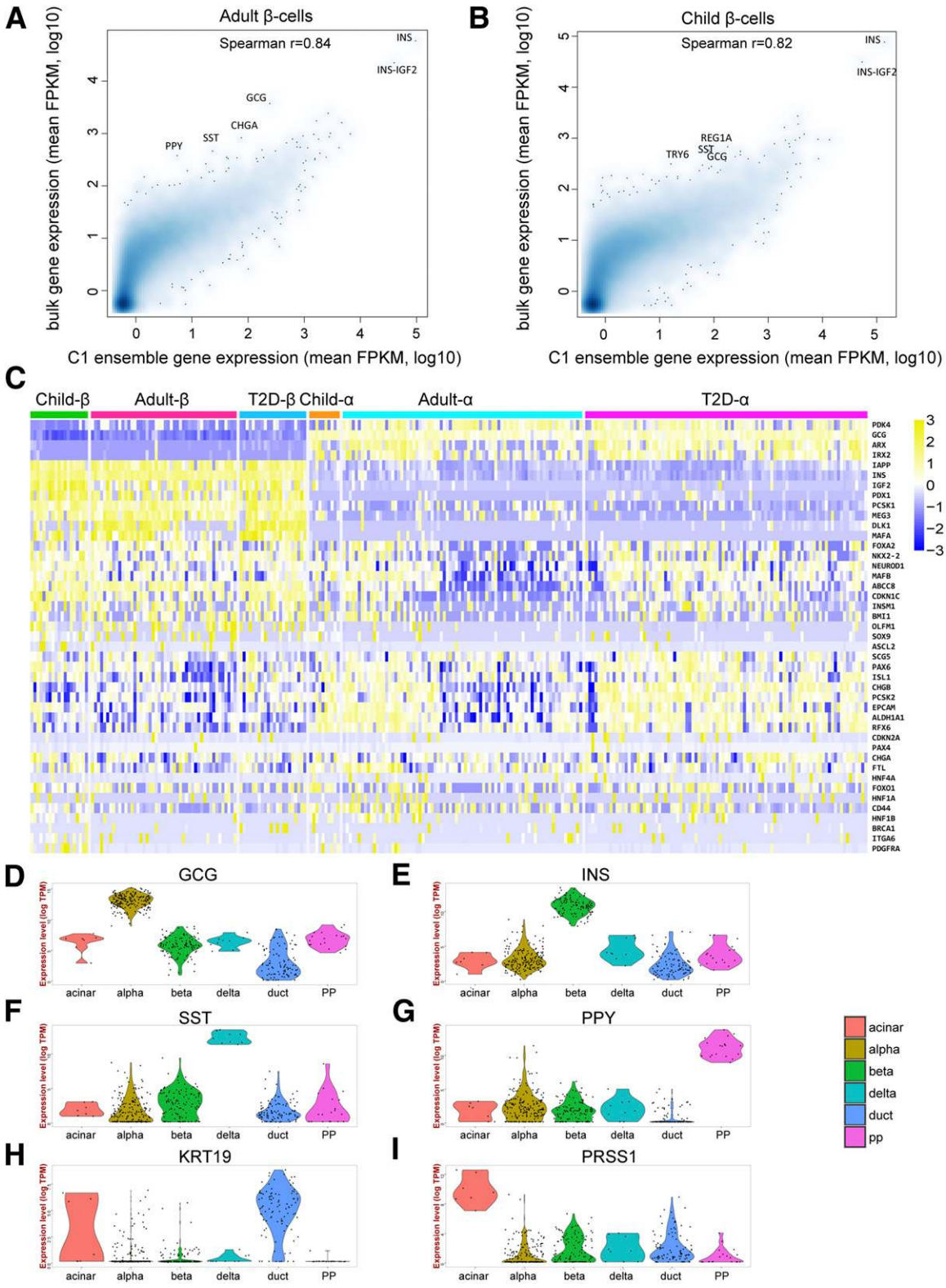

Figure 2—Single-cell RNA-seq of human islets identifies pancreatic cell types and gene signatures. Single-cell expression data from β-cells were aggregated and compared with expression profiles from sorted bulk β-cells in adult (*A*) and child (*B*) samples. Outlier genes indicative of contaminating cells in the bulk samples are labeled. FPKM, fragments per kilobase of transcript per million. *C:* Heatmap showing hierarchical clusters of α- and β-cells from different donor types with selected genes. *D–I*: Violin plots confirm that annotated cell types have mutually exclusive marker gene expressions. Each dot inside the violin represents one cell. T2D, type 2 diabetes.

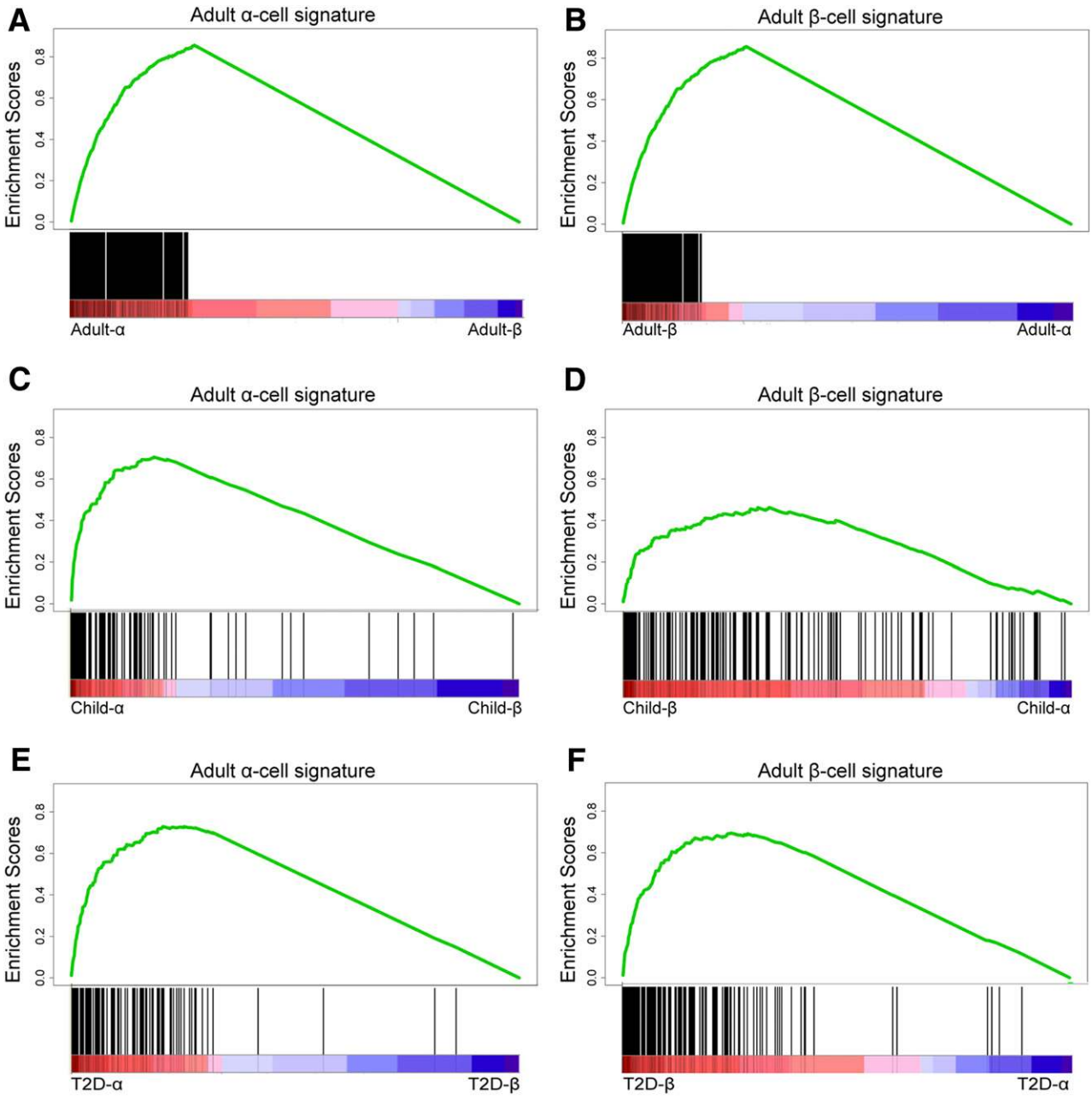

Figure 3—Child α- and β-cells exhibit a different transcriptome profile than adult α- and β-cells. Adult gene signatures for gene set enrichment analysis were derived from control adult endocrine cells with a ≥10-fold difference between α- and β-cell gene expression levels and a false discovery rate <5%. Differentially expressed genes from control adult α- and β-cells ordered by fold changes with genes more highly expressed in α-cells on the left of the *x*-axis in *A* and genes more highly expressed in β-cells on the left of the *x*-axis in *B*. *C–F*: Similar display as in *A* and *B*, except that the same nondiabetic adult gene signature was used to analyze α- and β-cells from child donors (*C* and *D*) and donors with type 2 diabetes (*E* and *F*). T2D, type 2 diabetes.

and *B*). In β-cells, these include cell cycle and regulatory genes such as *CDKN2B*, *BARD1*, and *JUNB* (Supplementary Table 1); at least one gene involved in insulin secretion, *PRKD1*, is also upregulated in type 2 diabetic β-cells (30). This finding suggests that endocrine cells in patients with type 2 diabetes are not able to maintain a fully differentiated gene expression profile, which is in line with recent publications reporting partial dedifferentiation of β-cells in diabetic states (8,31). The complete list of misexpressed genes can be found in Supplementary Table 1.

Human Proliferating α-Cells Activate the Sonic Hedgehog Pathway

One of the α-cells analyzed from an adult donor without diabetes was clearly proliferating, as indicated by a very high level of Ki67 mRNA expression. We compared the expression profile of this proliferating α-cell with quiescent

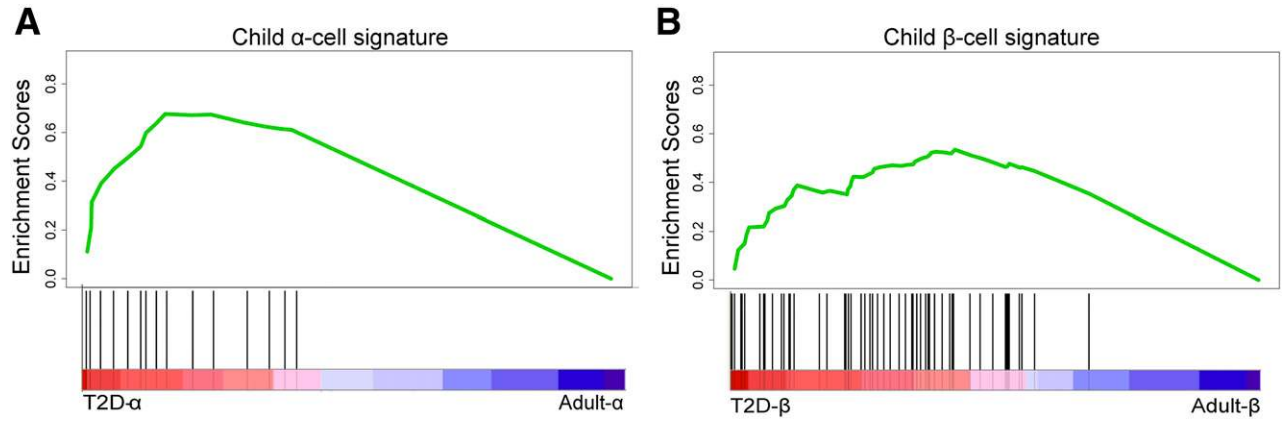

Figure 4—Endocrine cells from donors with type 2 diabetes and child donors have similarities in gene expression. Pediatric gene signatures for gene set enrichment analysis were derived from juvenile endocrine cells with a ≥10-fold difference between child and control adult gene expression levels and a false discovery rate <5%. *A*: α-Cells from donors with type 2 diabetes have an expression profile with features of child α-cells. *B*: β-Cells from donors with type 2 diabetes have an expression profile with similarity to the child β-cells. T2D, type 2 diabetes.

α-cells from the same donor and performed pathway analysis. We confirmed that this cell had, in fact, activated cell cycle pathways and inhibited cell cycle checkpoint control (Fig. 5*A*). Remarkably, the Sonic hedgehog signaling pathway was activated in the replicating α-cell (Fig. 5*A*). *DYRK1A* and *GSK3B* were both repressed in this proliferating cell, consistent with recent reports that inhibition of both molecules induces proliferation of human endocrine cells (32,33) (Fig. 5*B*). Interestingly, both of these proteins regulate GLI transcription factors, major targets of Sonic hedgehog pathway signaling (Fig. 5*B*). The ability to analyze differential expression between a proliferating α-cell and quiescent α-cells highlights the power of single-cell RNA-seq to explore rare cell populations.

Deeper Analysis of Cells With Conflicted Expression Profiles

Our original exclusion of cells with a mixed signature was designed to eliminate any cells captured as doublets. This rigorous exclusion could have the detrimental effect of omitting interesting cells, such as transdifferentiating cells and even possibly rare progenitor cells. For example, recent publications have demonstrated the plasticity of pancreatic endocrine cells in transition from one cell fate to another under experimental or pathophysiological conditions (8,34–36). This plasticity corresponds to the highly similar chromatin and epigenetic state of human α- and β-cells (16,36). We therefore re-evaluated cells that were excluded from our analysis based on a conflicted α/β-expression profile by comparing their gene expression pattern to that of annotated α- and β-cells, using hierarchical clustering of both genes and cells (Fig. 6*A*). As seen in Fig. 6*A*, many conflicted cells with a mixed α-/β-signature clustered together on the heatmap in between α- and β-cells and displayed both α- and β-cell signatures at roughly equal strength (cells in a black box); they thus represent true doublets. However, several of the conflicted cells did not cluster with these doublets but rather with either α- or β-cells (see red arrows on top of the heatmap) and thus likely represent rare endocrine cells that were excluded on the very stringent criteria we described above.

We also examined α-, β-, and conflicted α-/β-cells using principle component analysis based on the same differentially expressed genes (Supplementary Fig. 6*A*). There are distinct three clusters of cells, corresponding to pure α- and β-cells and the group of cells showing mixed α-/β-signature, based on the first two principle components, PC1 and PC2. Strikingly, both PC1 and PC2 show high correlation with the number of transcripts sequenced (Supplementary Fig. 6*A*). The observation that many of the suspected admixed cells have relatively high numbers of transcripts further corroborates the notion that they are likely to be doublets (Supplementary Fig. 6*A*). However, we again observed cells with conflicted gene expression signatures that do not cluster with the three major groups of cells, representing potential rare cell types (Supplementary Fig. 6*A*).

Similarly, we performed this analysis for conflicted β-/duct cells (Fig. 6*B*). This analysis demonstrates that in addition to the conflicted cells that are clearly the results of two cells in one well, there are other rare cells that do not display the standard α-, β-, or duct cell signature that are not doublets. Future research with larger sample sizes will likely reveal interesting properties about rare cell types.

DISCUSSION

Pancreatic transcriptome studies of bulk-sorted endocrine cell fractions have revealed important insights into cell type–specific gene expression signatures (36,37). In our study, we obtained RNA-seq data from well-annotated single cells from multiple organ donors covering children, healthy adults, and individuals with type 1 and type 2 diabetes. We developed a robust pipeline to ensure that

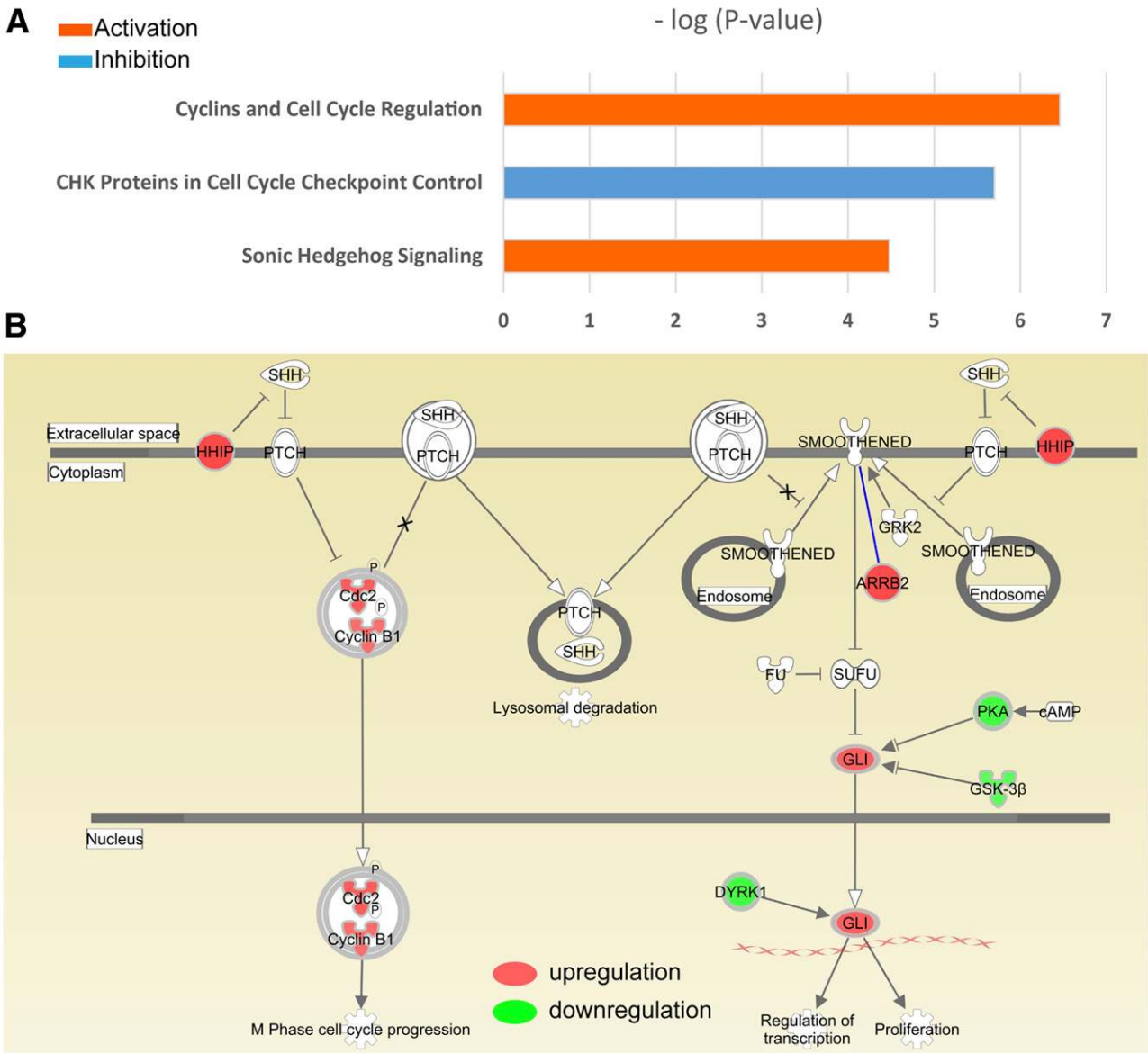

Figure 5—Sonic hedgehog signaling may contribute to endocrine cell replication. *A:* A replicating α-cell found by single-cell RNA-seq analysis showed activation of the cell cycle and repression of checkpoint control genes and, unexpectedly, activation of the Sonic hedgehog pathway. *B:* Differentially expressed genes in the Sonic hedgehog pathway are depicted with a circle with color-filled interior (red, upregulation; green, downregulation in the replication α-cell, compared with quiescent α-cells from the same donor).

each cell is correctly assigned to a specific cell type and that artifacts and cell doublets were removed. We also demonstrate several clear advantages in analyzing data from single cells versus the entirety of a population, even when cells are FACS sorted prior to RNA extraction.

A particularly striking result from our study is the discovery that the transcriptomes of both α- and β-cells from donors with type 2 diabetes exhibit features of expression profiles of their juvenile counterparts, indicating partial regression to an immature state, as exhibited by upregulation of several cell cycle genes (Figs. 3 and 4 and Supplementary Table 1). This was made possible by the derivation of α- and β-cell signature gene lists derived from 100% pure cells identified by single-cell technology, which is not possible through bulk cell sequencing analysis. The mechanisms driving this dedifferentiation will need to be explored further, although the transcription factor FOXO1 clearly plays a role in this process (8).

A second major advantage of single-cell technology is the ability to detect rare cell populations. We were able to discover a single proliferating α-cell, thus showcasing the power of single-cell RNA-seq methodology (Fig. 5). We demonstrate the presence of the expected activation of cell cycle genes, as well as downregulation of both *DYRK1A* and *GSK3B* in this replicating cell, both of which contribute to endocrine cell replication (32,33) and

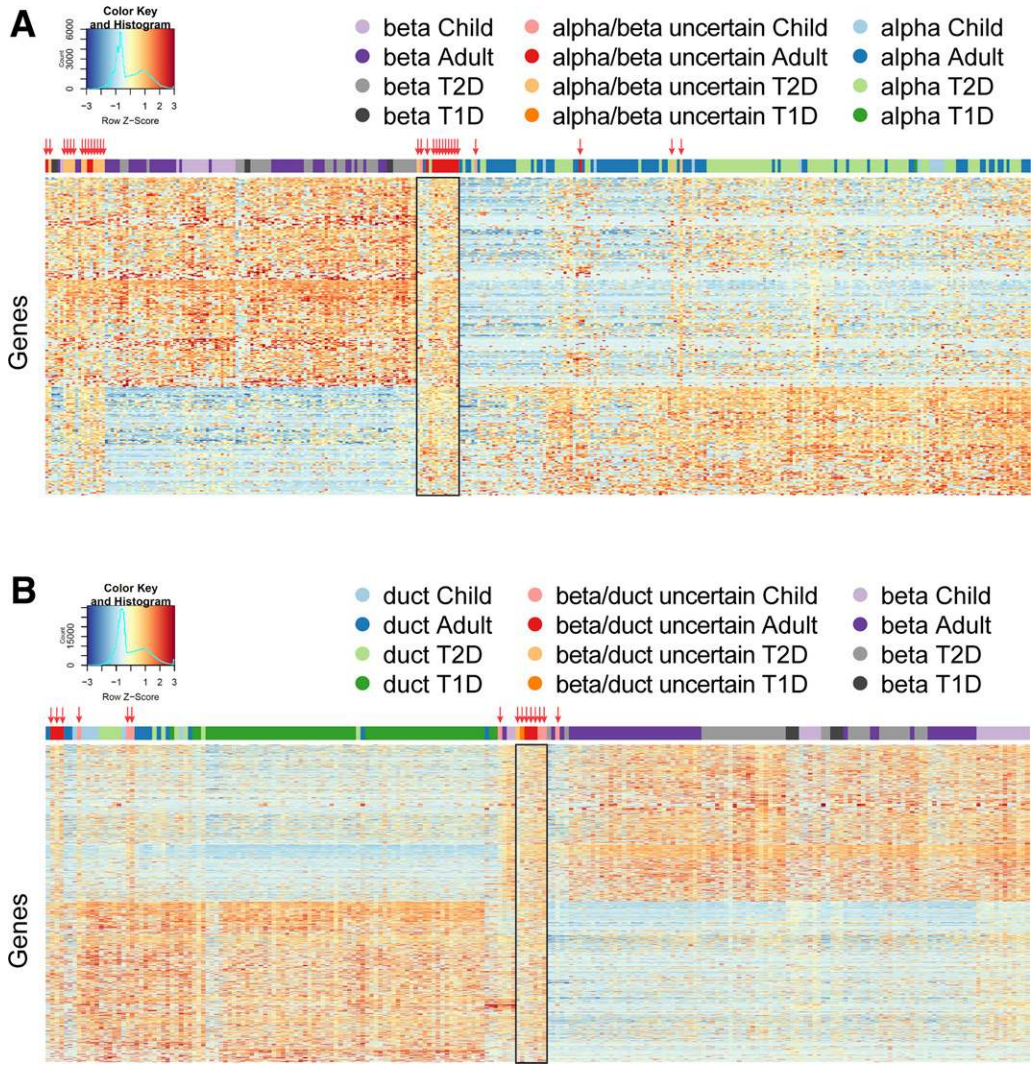

Figure 6—Heatmaps help distinguish doublets from rare variant cells. *A*: Annotated α- and β-cells, together with suspected α- and β-doublets, are clustered using genes that are >6.5-fold differentially expressed between α- and β-cells. Each column represents one cell and each row represents one gene. Cells from different sample types are color coded and indicated by the color bar on top of the heatmap. Cells with an ambiguous signature are depicted with a red arrow. Many ambiguous cells cluster together between α- and β-cells and display clear evidence of a combined signature (black box). Other ambiguous cells, however, exhibit expression profiles more closely related to either the α- or β-cell gene expression pattern and thus could represent rare transitional cells. *B*: Similar organization as in *A*, except that annotated β-cells and duct cells, together with suspected β-cell and duct cell doublets, are clustered using genes that are >3.2-fold differentially expressed between β-cells and duct cells. T1D, type 1 diabetes; T2D, type 2 diabetes.

therefore implicate a role for Sonic hedgehog signaling in this process (Fig. 5).

Furthermore, we performed a detailed analysis of the cells that showed a conflicted signature based on their individual expression profiles. While many clustered together, some of these cells with a somewhat "mixed" gene expression signature clustered within the bona fide α-, β-, or duct cells and thus possibly represent transitional states. Further research is required to uncover the biological significance of these cells.

In conclusion, by applying single-cell RNA-seq technology to profile pancreatic islets from multiple human organ donors, comparing both pediatric and adult organ donors with and without diabetes, we uncover several previously unknown features of the human endocrine pancreas. First, gene expression even of the major hormone genes is quite variable among human α- and β-cells. Second, α- and β-cells from type 2 diabetes exhibit transcriptome features of pediatric endocrine cells, indicating a reversal to a more immature state. In light of our recent findings that aged mouse islets have enhanced insulin secretion compared with young islets, the similarity of type 2 diabetic and young α- and β-cells may indicate that type 2 diabetes could in part result from a failure of compensatory mechanisms that normally occur with age (29). To what degree these abnormalities contribute to the etiology of diabetes will have to be established in the future; nevertheless, the data presented here illustrate

the power of a well-controlled single-cell transcriptomic analysis.

Acknowledgments. The authors thank all the islet donors and their families who make this work possible. The authors also thank Olga Smirnova, Haleigh Zillges, and Christina Theodouru (University of Pennsylvania) for excellent technical support and the University of Pennsylvania Diabetes Research Center for the use of the Functional Genomics and Islet Cell Biology Cores (National Institute of Diabetes and Digestive and Kidney Diseases [NIDDK] P30-DK19525).

Funding. This study was supported by the BIRAX Regenerative Medicine Initiative (14BX14NHBG to D.A.) and the NIDDK (UC4DK104119 to K.H.K.).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. Y.J.W., D.A., and M.L.G. performed research and wrote and edited the manuscript. J.S. and K.-J.W. provided computational analyses. C.L. and A.N. provided human islets and discussion. K.H.K. wrote and edited the manuscript. M.L.G., D.A., and K.H.K. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Steiner DJ, Kim A, Miller K, Hara M. Pancreatic islet plasticity: interspecies comparison of islet architecture and composition. Islets 2010;2:135–145
2. Brissova M, Fowler MJ, Nicholson WE, et al. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem 2005;53:1087–1097
3. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res 2005;15:1388–1392
4. Pipeleers DG. Heterogeneity in pancreatic beta-cell population. Diabetes 1992;41:777–781
5. Chiang MK, Melton DA. Single-cell transcript analysis of pancreas development. Dev Cell 2003;4:383–393
6. Benninger RK, Piston DW. Cellular communication and heterogeneity in pancreatic islet insulin secretion dynamics. Trends Endocrinol Metab 2014;25:399–406
7. Gunasekaran U, Gannon M. Type 2 diabetes and the aging pancreatic beta cell. Aging (Albany, NY) 2011;3:565–575
8. Talchai C, Xuan S, Lin HV, Sussel L, Accili D. Pancreatic β cell dedifferentiation as a mechanism of diabetic β cell failure. Cell 2012;150:1223–1234
9. Kushner JA. The role of aging upon β cell turnover. J Clin Invest 2013;123:990–995
10. Grün D, Lyubimova A, Kester L, et al. Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 2015;525:251–255
11. Shalek AK, Satija R, Adiconis X, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 2013;498:236–240
12. Toriello NM, Douglas ES, Thaitrong N, et al. Integrated microfluidic bioprocessor for single-cell gene expression analysis. Proc Natl Acad Sci USA 2008;105:20173–20178
13. Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. Trends Biotechnol 2010;28:281–290
14. Kalisky T, Blainey P, Quake SR. Genomic analysis at the single-cell level. Annu Rev Genet 2011;45:431–445
15. Dorrell C, Abraham SL, Lanxon-Cookson KM, Canaday PS, Streeter PR, Grompe M. Isolation of major pancreatic cell types and long-term culture-initiating cells using novel human surface markers. Stem Cell Res (Amst) 2008;1:183–194
16. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. Mol Metab 2016;5:233–244
17. Grant GR, Farkas MH, Pizarro AD, et al. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics 2011;27:2518–2528
18. Li J, Klughammer J, Farlik M, et al. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. EMBO Rep 2016;17:178–187
19. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–140
20. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550
21. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57
22. Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1–13
23. Daoud JT, Petropavlovskaia MS, Patapas JM, et al. Long-term in vitro human pancreatic islet culture using three-dimensional microfabricated scaffolds. Biomaterials 2011;32:1536–1542
24. Zhang Y, Jalili RB, Warnock GL, Ao Z, Marzban L, Ghahary A. Three-dimensional scaffolds reduce islet amyloid formation and enhance survival and function of cultured human islets. Am J Pathol 2012;181:1296–1305
25. Negi S, Jetha A, Aikin R, Hasilo C, Sladek R, Paraskevas S. Analysis of beta-cell gene expression reveals inflammatory signaling and evidence of de-differentiation following human islet isolation and culture. PLoS One 2012;7:e30415
26. Lakey JR, Warnock GL, Rajotte RV, et al. Variables in organ donors that affect the recovery of human islets of Langerhans. Transplantation 1996;61:1047–1053
27. Balamurugan AN, Chang Y, Bertera S, et al. Suitability of human juvenile pancreatic islets for clinical use. Diabetologia 2006;49:1845–1854
28. Rankin MM, Kushner JA. Aging induces a distinct gene expression program in mouse islets. Islets 2010;2:345–352
29. Avrahami D, Li C, Zhang J, et al. Aging-dependent demethylation of regulatory elements correlates with chromatin state and improved β cell function. Cell Metab 2015;22:619–632
30. Sumara G, Formentini I, Collins S, et al. Regulation of PKD by the MAPK p38delta in insulin secretion and glucose homeostasis. Cell 2009;136:235–248
31. Wang Z, York NW, Nichols CG, Remedi MS. Pancreatic β cell dedifferentiation in diabetes and redifferentiation following insulin therapy. Cell Metab 2014;19:872–882
32. Wang P, Alvarez-Perez JC, Felsenfeld DP, et al. A high-throughput chemical screen reveals that harmine-mediated inhibition of DYRK1A increases human pancreatic beta cell replication. Nat Med 2015;21:383–388
33. Shen W, Taylor B, Jin Q, et al. Inhibition of DYRK1A and GSK3B induces human β-cell proliferation. Nat Commun 2015;6:8372
34. Collombat P, Xu X, Ravassard P, et al. The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. Cell 2009;138:449–462
35. Thorel F, Népote V, Avril I, et al. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature 2010;464:1149–1154
36. Bramswig NC, Everett LJ, Schug J, et al. Epigenomic plasticity enables human pancreatic α to β cell reprogramming. J Clin Invest 2013;123:1275–1284
37. Nica AC, Ongen H, Irminger JC, et al. Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res 2013;23:1554–1562