Differentiated human stem cells resemble fetal, not adult, β cells

Siniša Hrvatin[a], Charles W. O'Donnell[a,b], Francis Deng[a], Jeffrey R. Millman[a], Felicia Walton Pagliuca[a], Philip DiIorio[c], Alireza Rezania[d], David K. Gifford[b], and Douglas A. Melton[a,e,1]

[a]Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute and [e]Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138; [b]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA 01605; and [d]BetaLogics Venture, Janssen Research and Development, LLC, Raritan, NJ 08869

Contributed by Douglas A. Melton, January 14, 2014 (sent for review September 14, 2013)

Human pluripotent stem cells (hPSCs) have the potential to generate any human cell type, and one widely recognized goal is to make pancreatic β cells. To this end, comparisons between differentiated cell types produced in vitro and their in vivo counterparts are essential to validate hPSC-derived cells. Genome-wide transcriptional analysis of sorted insulin-expressing (INS^+) cells derived from three independent hPSC lines, human fetal pancreata, and adult human islets points to two major conclusions: (*i*) Different hPSC lines produce highly similar INS^+ cells and (*ii*) hPSC-derived INS^+ (hPSC-INS^+) cells more closely resemble human fetal β cells than adult β cells. This study provides a direct comparison of transcriptional programs between pure hPSC-INS^+ cells and true β cells and provides a catalog of genes whose manipulation may convert hPSC-INS^+ cells into functional β cells.

transcriptional profiling | differentiation | beta cells | MARIS

Human pluripotent stem cells (hPSCs), including embryonic stem cells (hESCs) and induced pluripotent stem cells (hiPSCs), are characterized by their capacity for unlimited self-renewal and the ability to differentiate into any human cell type (1–4). Stepwise differentiation protocols, designed to mimic sequential developmental signals, attempt to generate specific cell types from hPSC lines for use in transplantation therapy and disease modeling (5–7).

Significant variation in differentiation efficiencies has been observed between different hPSC lines, with some lines more readily differentiating into a particular cell type than others (8–11). The reasons for this variation have not been completely explained, but studies point to variation in genetic, epigenetic (12), and cell cycle patterns (13). Owing to these differences in differentiation propensity, directed differentiation protocols often require laborious optimization for specific hPSC lines (14). The use of different protocols and different cell lines calls into question the degree to which the final differentiation products resemble each other. This has been difficult to address because, even with an optimized protocol and cell line, only a fraction of the hPSCs achieve the desired cell fate. Thus, direct comparison of cells, which are present in a mixed population, is not generally possible, except in those rare instances where appropriate reporter lines have been constructed to facilitate cell sorting.

In addition to questions of variability, the extent to which any differentiated cell produced in vitro resembles its counterpart produced during normal human development remains unknown. Directed differentiation protocols for human cells are often generated using mouse embryonic development as a guide. Although there are many similarities, significant differences in transcriptional regulation exist between these two species (15, 16). Although it is obvious that documenting transcriptional changes that accompany human development would greatly benefit directed differentiation of human stem cells, the lack of suitable cell surface markers makes it very difficult to isolate and purify most human fetal and adult cell types for analysis.

Pancreatic β cells are a cell type responsible for the regulation of serum glucose levels through secretion of the hormone insulin. Despite productive studies on new cell-surface markers for β cells, insulin expression remains the best specific β-cell marker. Although adult β cells express only the hormone insulin, it is important to note that some fetal β cells are polyhormonal and express other hormones, in addition to insulin, including glucagon and somatostatin (17–19). Unlike glucose-responsive adult β cells, β cells present during early fetal development do not respond to glucose with increased insulin secretion (20, 21).

Cell line variability and the lack of a clear understanding of human early (fetal) development have hampered efforts to generate pancreatic β cells from hPSCs. Current directed differentiation protocols generate insulin-expressing (INS^+) cells (hPSC-INS^+) that lack expression of several key β-cell genes and fail to properly secrete insulin in response to glucose (14, 22–32). Additionally, careful analysis of select chromatin modifications that accompany β-cell development has shown that Polycomb-mediated differences in chromatin remodeling are deficient in in vitro differentiated cells (33). These observations led to the speculation that hPSC-INS^+ cells are more similar to fetal β cells than to adult β cells (27). It is, however, equally possible that hPSC-INS^+ cells represent a different in vitro-derived cell type unlike fetal or adult β cells. A thorough analysis as to whether hPSC-directed differentiation generates cell types found during normal human development has been long awaited.

Significance

Human pluripotent stem cells (hPSCs) can be produced from any person and have the potential to differentiate into any cell type in the body. This study focuses on the generation of insulin-expressing cells from hPSCs and compares their gene expression, as assayed by transcriptional gene products, to that of insulin-expressing β cells from human fetal and adult samples. We employ a new method to isolate and profile insulin-expressing cells and conclude that several different hPSC lines generate very similar insulin-expressing cells, cells whose transcripts resemble fetal rather than adult β cells. This study advances the possibility of directing the differentiation of stem cells into functional β cells by comparing and cataloging differences between hPSC-derived insulin-expressing cells and human β cells.

Author contributions: S.H., C.W.O., F.D., D.K.G., and D.A.M. designed research; S.H., C.W.O., F.D., J.R.M., and F.W.P. performed research; J.R.M., F.W.P., P.D., and A.R. contributed new reagents/analytic tools; S.H., C.W.O., F.D., D.K.G., and D.A.M. analyzed data; and S.H., C.W.O., and D.A.M. wrote the paper.

Conflict of interest statement: A.R. is an employee of Janssen Research and Development, LLC.

[1]To whom correspondence should be addressed. E-mail: dmelton@harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1400709111/-/DCSupplemental.

Previous attempts to analyze genome-wide transcription in enriched populations of adult β cells relied on FACS enrichment of β cells using either Newport Green dye (34) or a series of cell-surface markers (35). The extent to which the small proportion of non-β cells present in the sorted population affects transcriptional analysis, and the applicability of these sorting methods to the isolation of human fetal β cells, is unknown. Sorting of hPSC-INS$^+$ cells also has been a significant challenge. Although one hPSC insulin-GFP knock-in reporter line has been recently generated (28, 29), isolating INS$^+$ cells from multiple genetically unmodified hPSC lines is necessary to evaluate the gene expression signature of hPSC-INS$^+$ cells. Finally, to our knowledge no one has yet purified and transcriptionally profiled human fetal β cells.

Here we make use of our newly developed Method for Analyzing RNA following Intracellular Sorting (MARIS) (36) to analyze the global gene expression profile of three types of sorted INS$^+$ cells: those differentiated from hPSC lines and human fetal and human adult pancreata. We also analyze the degree of similarity between INS$^+$ cells derived from different hESC and hiPSC lines. We then document transcriptional changes that occur during human β-cell development. Finally, we compare hPSC-INS$^+$ cells to human adult and fetal β cells and identify differentially expressed genes between adult β cells and hPSC-INS$^+$ cells.

Results

Global Transcriptional Profile of INS$^+$ Cells from Several hESC and iPSC Lines. We first sought to determine the degree to which hPSC-derived INS$^+$ cells, produced from different cell lines, resemble one another at the transcriptome level. We treated hESC lines H1 and HUES 8 and hiPSC line iPS-17b with a multiple-step differentiation protocol to stage 6, at which point a small percentage of cells expressed insulin and other pancreatic hormones (Fig. S1). Given the small percentage of desired cell type that is typically produced, very little information can be gained using traditional transcriptional analyses of the whole population. Instead, stage-6 cells were fixed, stained for insulin, glucagon, and somatostatin, and sorted for RNA isolation and analysis using MARIS (Fig. 1*A*). INS$^+$ cells comprised only about 1% of all stage-6 cells. A large proportion (40–70%) of INS$^+$ cells coexpressed glucagon and somatostatin, consistent with previous reports (22, 27). Notably, the RNA isolated from hPSC-INS$^+$ cells by MARIS achieved RNA integrity numbers of 8.1 ± 0.9. Quantitative RT-PCR for insulin, glucagon, and somatostatin indicated significant enrichment of all three endocrine hormones in the sorted populations, confirming successful purification of INS$^+$ cells (Fig. 1*B*).

The RNA isolated from hPSC-INS$^+$ cell populations was also analyzed using Illumina microarrays. RNA isolated from unfixed, undifferentiated pluripotent cells from each line and from unsorted stage-6 populations were included as controls. Hierarchical clustering across all genes identified three distinct, statistically confirmed groups of samples (Fig. 1*C*). All INS$^+$ cells clustered together, suggesting that there were fewer differences between INS$^+$ cells derived from different cell lines than differences between unsorted stage-6 cells and sorted INS$^+$ cells within each cell line. Moreover, expression profiles between INS$^+$ cells derived from different hPSC lines were as highly correlated ($r^2 = 0.93 \pm 0.02$) as those between hPSC lines at the pluripotent stage ($r^2 = 0.94 \pm 0.02$) (Fig. 1*D*).

To further analyze hPSC-INS$^+$ cells we focused on 152 endocrine lineage genes known for their role in pancreatic development, endocrine hormone secretion, and glucose metabolism (27, 28, 37) (Table S1). Hierarchical clustering and correlation values based on these endocrine lineage genes confirmed a high degree of similarity between hESC-derived and hiPSC-derived insulin$^+$ cells (Fig. S2 *A* and *B*). Together, these data suggest that INS$^+$ cells derived from different hPSC lines are very similar.

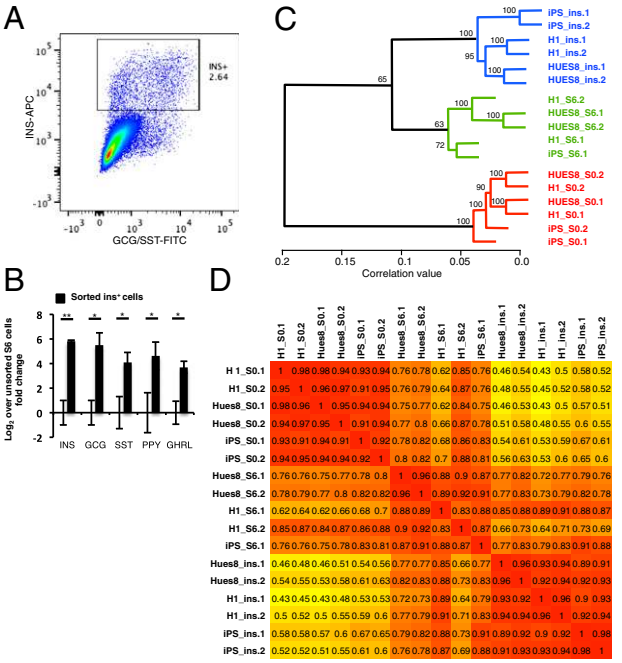


Fig. 1. RNA profiling of sorted hPSC-derived insulin-expressing cells. (*A*) FACS plot of stage-6 H1-derived cells sorted for insulin-APC. (*B*) Quantitative RT-PCR of unsorted and insulin-sorted stage-6 hPSC-derived cells for pancreatic hormone genes INS (insulin), GCG (glucagon), SST (somatostatin), PPY (pancreatic polypeptide), and GHRL (ghrelin) suggests significant enrichment of mRNA specific for pancreatic hormones in the insulin-APC sorted population (*$P < 0.05$, **$P < 0.01$). (*C*) Three human pluripotent stem cell lines, HUES8, H1, and iPS-17b, were differentiated to stage 6 and sorted for INS$^+$ cells. RNA was isolated from undifferentiated cells, stage-6 cells, and sorted INS$^+$ cells for all three cell lines. Global gene expression for each sample was analyzed using the Illumina microarray platform. Hierarchical clustering identified three major groups of samples. Lengths in the dendrogram represent correlation value. Approximately unbiased (AU) *P* values are displayed. INS$^+$ cells from different cell lines form a statistically significant cluster. (*D*) r^2 values based on microarray data across all genes are shown. The average r^2 value between stage-0 cells, 0.94, is similar to the average r^2 value between sorted INS$^+$ cells, 0.93. ins, insulin$^+$ MARIS-sorted stage-6 differentiated, pluripotent stem cells; S0, unsorted, undifferentiated pluripotent stem cells; S6, unsorted stage-6 differentiated pluripotent stem cells.

Recently, Micallef et al. (29) reported the generation of an insulin-GFP knock-in hESC-reporter line. Basford et al. (28) performed microarray analysis with this cell line and described 28 genes that were differentially expressed between insulin-positive and insulin-negative cells. In our analysis, 27 of the 28 identified genes had the same direction of enrichment in each of the three hPSC lines (Fig. S2*C*) (28). These data further strengthen the conclusion that INS$^+$ cells derived from different hPSC lines display highly similar molecular signatures.

Human β-Cell Maturation. Study of human fetal development has been hampered by the absence of cell-surface markers that allow for sorting of cell types produced in vivo, as well as the scarcity of human fetal material for study. Researchers have thus relied on studies in model organisms, primarily the mouse, as the basis for understanding human development and optimizing directed differentiation. Recently, our laboratory identified differential gene expression patterns that distinguish fetal β cells from adult β cells in the mouse (38). To uncover markers for human β-cell maturation, we performed a purification and transcriptome-wide molecular characterization of human fetal and adult β cells.

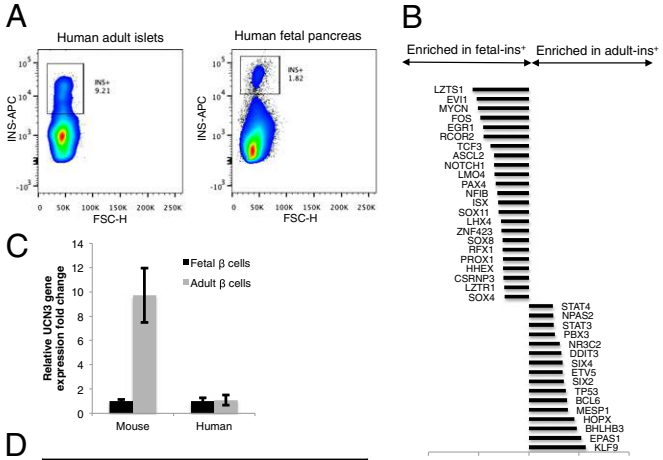
Human pancreata at 15–16 wk gestational age were used, because β cells at this stage are immature and glucose-nonresponsive

(20, 21). Adult human cadaveric islets and fetal pancreata were dispersed, stained for insulin, and FACS-sorted. RNA was isolated and analyzed by Illumina microarrays (Fig. 2*A*).

The analysis shows that human β-cell maturation, between gestational week 16 and adulthood, is characterized by gene expression changes in 643 genes, of which 39 were transcription factors ($P < 0.05$, fold change >3) (Fig. 2*B*). Urocortin 3 (UCN3), which we identified previously as a marker of mouse β-cell maturation, was expressed only 1.1x-fold higher in sorted human adult β cells over fetal β cells, indicating that UCN3 expression does not significantly change between human week-16 and adult β cells (Fig. 2*C*). The differentially expressed genes presented in this study could be used as genetic markers of human β-cell maturation. Down-regulation or up-regulation of several genes such as NFIB, prospero homeobox 1 (PROX1), HHEX, and KLF9 has been implicated in the maturation of other cellular lineages, suggesting that these may be transcription factors involved in the pan-tissue transitions from fetal to adult transcriptional programs (39–42).

Gene Ontology analysis of all differentially expressed genes ($P < 0.05$) suggests that several metabolic and secretory biological processes are significantly enriched in adult β cells over fetal β cells (Fig. 2*D*). These processes include vesicle-mediated transport and oxidation-reduction consistent with the idea that fetal β cells may not metabolize glucose or package insulin for secretion the same way adult β cells can (43).

These data represent a transcriptome-wide molecular characterization of human fetal and adult β cells and point to significant differences between mouse and human β-cell maturation. Further analysis of β cells at multiple time points during human and mouse development will further elucidate this species divergence.



	Gene Ontology Biological Process	Benjamini q value
Enriched in Adult β cells	Generation of precursor metabolites and energy	2.90E-07
	Oxidation reduction	0.001810
	Vesicle-mediated transport	0.004327
	Electron transport chain	0.028568
	Monosaccharide metabolic process	0.056043
Enriched in Fetal β cells	Cell morphogenesis	0.008113
	Cellular component morphogenesis	0.012270
	Cell projection organization	0.014508
	Wnt receptor signaling pathway	0.047234
	Cell projection morphogenesis	0.050197

Fig. 2. Human β-cell maturation. (*A*) FACS plots of human adult islets and human fetal pancreata sorted for INS$^+$ cells (APC$^+$). (*B*) Differentially expressed transcription factors between adult and fetal β cells. (*C*) Relative expression of UCN3 in mouse and human fetal and adult β cells. Expression normalized to fetal levels in each species. (*D*) Top five most significant (Benjamini q value) Gene Ontology biological processes relatively enriched in either adult or fetal β cells.

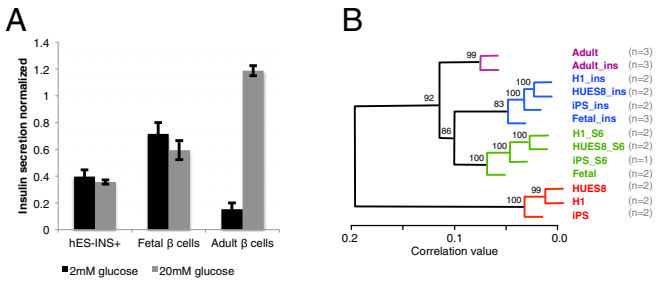
hPSC-Derived INS$^+$ Cells Resemble Human Fetal β Cells More Than Human Adult β Cells. Directed differentiation from hPSCs attempts to recreate human development in vitro. The extent to which this has been achieved remains unknown, because direct comparisons with human fetal cells have been difficult because of the scarcity of material and the inability to assess pure populations of β cells.

We first tested dispersed hPSC-INS$^+$ cells and fetal and adult β cells for glucose-stimulated insulin secretion (GSIS). In contrast to adult β cells, both hPSC-INS$^+$ cells and human fetal β cells have elevated basal glucose secretion and do not display a robust GSIS response (Fig. 3*A*). To transcriptionally compare hPSC-INS$^+$ cells and human fetal and adult β cells, we performed whole-genome expression analysis. Unsorted stage-6 cells and undifferentiated pluripotent stem cells were included as control groups. Hierarchical clustering across all genes identified four distinct groups of samples (Fig. 3*B*). hPSC-INS$^+$ cells clustered more closely with human β cells than with undifferentiated hPSCs. Notably, fetal β cells clustered with hPSC-derived INS$^+$ cells and not human adult β cells. The correlation between three biological replicates of adult β cells ($r^2 = 0.89 \pm 0.04$) was not significantly different from the correlation between fetal β cells and hPSC-INS$^+$ cells ($r^2 = 0.88 \pm 0.02$) ($P = 0.49$), indicating a high degree of similarity between the latter two cell types (Fig. 3*C*). Hierarchical clustering and correlation values based on 152 endocrine lineage genes confirmed this result (Fig. S3). Together, these data show that INS$^+$ cells derived from different hPSC lines are similar to fetal β cells and not adult β cells.

Transcriptional Differences Between hES-Derived INS$^+$ Cells and Human Adult β Cells. Given that the goal of hPSC-directed differentiation is to generate functional human β cells, we compared transcript expression by microarray and RNA-seq between hPSC-INS$^+$ cells and adult human β cells. We first focused our analysis on 152 endocrine lineage genes and observed that 26 of these genes were significantly different between adult β cells and hPSC-INS$^+$ cells (greater than threefold by microarray analysis, $P < 0.05$). RNA-seq analysis of two samples of human adult INS$^+$ cells and two samples of HUES8-INS$^+$ cells confirmed 24/26 genes as differentially expressed greater than threefold (Fig. 4*A*). For example, pancreatic and duodenal homeobox 1 (PDX1), v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog A (MAFA), and NK6 homeobox 1 (NKX6-1) are markers of adult human β cells (17) and in vivo-matured hPSC-INS$^+$ cells (22, 26). Using immunofluorescent staining, we confirmed the lack of expression of NKX6-1 and relatively heterogeneous expression of PDX1 and MAFA in hPSC-INS$^+$ cells (Fig. 4*B*).

We next focused on two categories of genes: those involved in endocrine subtype specification and those that might be relevant to GSIS. Factors involved in both endocrine specification [PDX1, NKX6-1, motor neuron and pancreas homeobox 1 (MNX1), forkhead box A1 (FOXA1), aristaless related homeobox (ARX), paired box 4 (PAX4), and iroquois homeobox 2 (IRX2)] and GSIS [MAFA; PAX4; GHRL; chromogranin B (CHGB); DPP4; potassium channel, subfamily K, member 1 (KCNK1); and potassium channel, subfamily K, member 3 (KCNK3)] were misexpressed in hPSC-INS$^+$ cells compared with adult β cells (Fig. 4*A*). PDX1, NKX6-1, and MNX1 are necessary for the determination of β-cell fate and have over threefold lower expression in hPSC-INS$^+$ than in human adult β cells (44–46). Conversely FOXA1, ARX, and IRX2, determinants of α cell fate, were misexpressed 4- to 40-fold higher in hPSC-INS$^+$ cells than adult β cells (47–49). MAFA, whose overexpression is sufficient to induce mature GSIS in immature P2 rat β cells (50), had significantly lower expression in hPSC-INS$^+$ cells compared with adult β cells. Conversely PAX4, whose prolonged expression was shown to blunt GSIS in mouse adult β cells (51), had significantly higher expression in hPSC-INS$^+$ cells than in adult β cells. Several other genes may be responsible

DEVELOPMENTAL BIOLOGY



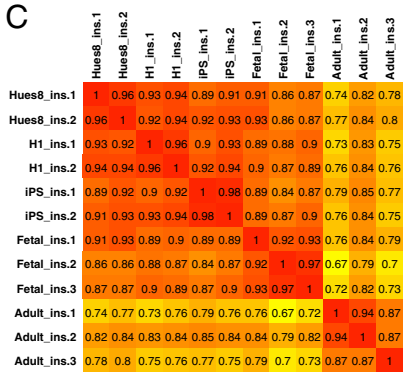


Fig. 3. hPSC-derived insulin-expressing cells resemble human fetal β cells. (*A*) Glucose stimulated insulin secretion of dispersed cells. In contrast to adult β cells, fetal β cells and hPSC-INS$^+$ cells both seem functionally immature, as indicated by increased basal glucose secretion and lack of glucose stimulation. (*B*) Hierarchical clustering based on microarray global gene expression across all genes indicated that hPSC-INS$^+$ cells cluster closely with human fetal and not adult β cells. Numbers in parentheses indicate biological replicates. Lengths in the dendrogram represent correlation distances. (*C*) r^2 values based on microarray data across all genes are shown. Each row and column represents one sample. r^2 values between biological replicates of adult β-cell samples (Adult_ins) are on average 0.89 ± 0.04. r^2 values between sorted hPSC-derived insulin$^+$ stage-6 cells and sorted fetal β cells are 0.88 ± 0.02. The biological variation between adult β cells is not statistically smaller than the variation between fetal β cells and hPSC-INS$^+$ stage-6 cells (P = 0.49). This indicates a high degree of similarity between hPSC-INS$^+$ cells and human fetal β cells. HUES8_ins, H1_ins, and iPS_ins are hPSC-derived MARIS-sorted stage-6 INS$^+$ cells; Fetal_ins are MARIS-sorted INS$^+$ cells from week-16 human fetal pancreata; Adult_ins are MARIS-sorted INS$^+$ cells from islet preparations of adult human pancreata.

for the lack of functional GSIS in hPSC-INS$^+$ cells: The presence of ghrelin suppresses GSIS (52), CHRB knock-out animals have reduced GSIS and elevated basal insulin secretion (53), PROX1 is associated with insulin secretion abnormalities (54), and lack of tandem pore domain potassium channels KCNK1 and KCNK3 may elevate resting membrane potential and cause hyperactivity and higher basal insulin secretion. Interestingly, with the exception of PAX4, no significant differences were observed in expression of endocrine subtype specification genes between fetal and adult β cells, whereas significant differences were apparent in the expression of the GSIS-relevant genes PAX4, CHGB, KCNK1, and KCNK3 (Fig. 4*C*).

This analysis suggests two challenges to producing functional human β cells from hPSC-INS$^+$ cells: (*i*) β-cell lineage commitment and (*ii*) functional maturation. Our observations support the hypothesis that hPSC-INS$^+$ cells resemble human fetal cells that are not fully committed to the β-cell lineage, as judged by transcription factor expression. The modulation of both endocrine lineage and GSIS genes may be critical for converting hPSC-INS$^+$ cells into a phenotype that more closely resembles adult human β cells.

We expanded the comparison between hPSC-INS$^+$ and adult β cells to the whole genome. Gene Ontology analysis of all differentially expressed genes ($P < 0.05$ by microarray) identified 22 statistically enriched (q < 0.05) biological processes in hPSC-INS$^+$ cells (Table S2). Among the 10 most differentially expressed processes, three involve cholesterol/sterol biosynthesis or metabolism, indicating that hPSC-INS$^+$ cells aberrantly express nonpancreatic, liver-specific genetic pathways (Fig. 4*D*).

A total of 755 genes were differentially expressed ($P < 0.05$, greater than threefold) by microarray analysis between hPSC-INS$^+$ and adult β cells; 583 genes were confirmed by RNA-seq (greater than threefold). We present a list of all differentially expressed transcription factors because they are of particular interest for their roles in modulating cell fates (Fig. 4*E*). Genes identified here are candidate markers that may distinguish hPSC-INS$^+$ cells from adult β cells, or targets to direct the conversion of hPSC-INS$^+$ cells into functional β cells.

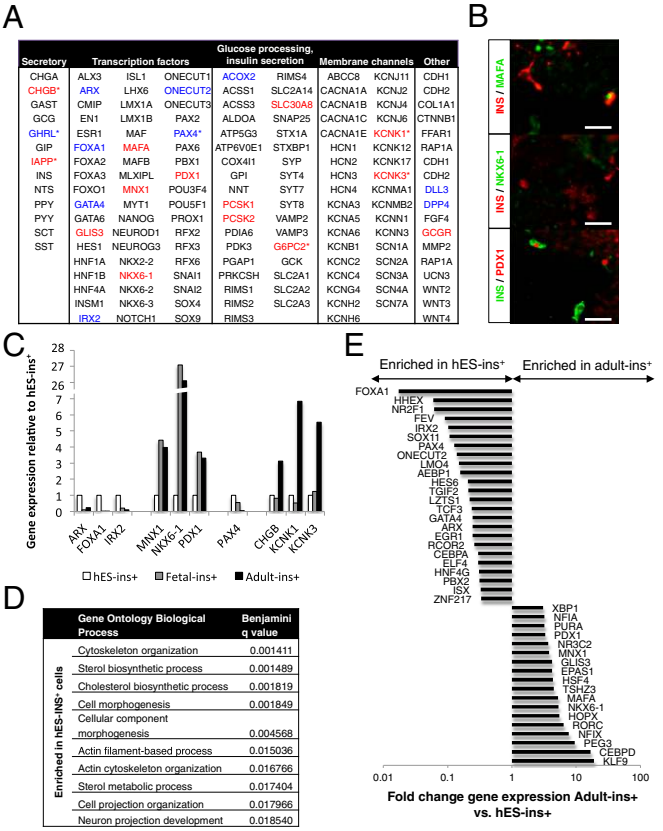


Fig. 4. Differential expression between human β cells and hPSC-derived insulin-expressing cells. (*A*) The list of 152 pancreatic lineage genes is colored for genes that are differentially overexpressed in adult β cells (red) or hPSC-INS$^+$ cells (blue). Differential gene expression was calculated based on microarray data between human adult β cells and hPSC-INS$^+$ cells (greater than threefold change, $P < 0.05$) and confirmed by RNA-seq between HUES8-INS$^+$ cells and adult β cells (greater than threefold change). Asterisk indicates genes that are also differentially expressed between fetal and adult β cells. (*B*) Immunofluorescence. hPSC-INS$^+$ cells with PDX1, NKX6-1, and MAFA. (Scale bars, 100 μm.) (*C*) Relative gene expression of selected differentially expressed genes normalized to expression in hES-INS$^+$ cells. ARX, FOXA1, IRX2, MNX1, NKX6-1, and PDF are expressed at similar levels between fetal and adult β cells but differentially expressed in hES-INS$^+$ cells. PAX4, CHGB, KCNK1, and KCNK3 are expressed at similar levels between hES-INS$^+$ cells and fetal β cells but differentially expressed in adult β cells. (*D*) Top 10 most significant (Benjamini q value) gene ontology biological processes enriched hPSC-INS$^+$ cells over adult β cells. (*E*) A list of 42 transcription factors that were differentially expressed based on microarray data between human adult β cells and hPSC-derived INS$^+$ cells (greater than threefold change, $P <$ 0.05) and confirmed by RNA-seq between HUES8-INS$^+$ cells and adult β cells (greater than threefold change).

Discussion

Here we compare the transcriptome of hPSC-derived pancreatic cells produced in vitro to human fetal and adult β cells, as well as test for the variation that may result from using different stem cell lines. These transcriptional analyses were made possible by antibody staining (for insulin) followed by cell sorting so that relatively pure populations of cells could be compared.

One conclusion from these results is that there is a high degree of similarity between INS^+ cells derived from three different pluripotent stem cell lines. The degree of correlation between INS^+ cells, derived after more than 20 d of directed differentiation, resembles the degree of correlation between different undifferentiated pluripotent stem cells. Furthermore, correlation between INS^+ cells was similar to the degree of correlation between human β-cell samples obtained from two different donors. The high degree of similarity we observe between pancreatic cells derived from different hPSC lines presents an important proof-of-principle observation for hPSC-directed differentiation.

Data from the analysis of the insulin-GFP knock-in hESC-reporter line points to a potentially high degree of similarity between our hPSC-INS^+ cells and those produced by other laboratories, using different cell lines and variations of differentiation protocols (28). Characterization of sorted hPSC-INS^+ cells from additional cell lines and differentiation protocols is necessary to confirm this observation.

In the absence of tools to study human development, model organisms have informed our understanding of human development and biology, although the degree to which human development resembles the development of other organisms has not been rigorously investigated. Using MARIS, we have compared the transcriptional profile of human fetal and adult β cells. This analysis allows for a second conclusion, namely, a list of genes that are differentially expressed during human β-cell maturation. These genetic signatures can be used as markers for generating functional human β cells. Moreover, the results indicate that gene expression changes during human β-cell maturation may not resemble changes in mouse β-cell gene expression during the late prenatal and early postnatal period. A possible explanation is that we are comparing two different stages of β-cell maturation. For example, the gene expression of week-16 human β cells could resemble the gene expression of early embryonic mouse β cells and not late prenatal mouse β cells. Alternatively, our data may be a result of intrinsic developmental differences between mouse and human. Further study of multiple stages during human and mouse development using RNA-seq in addition to microarrays is necessary to determine the degree of similarity between mouse and human β-cell maturation. However, despite gene expression analysis, it is not possible to carry out human in vivo lineage tracing studies, and therefore the lineage relationship between human fetal β cells and adult β cells remains unknown.

Whether hPSC-directed differentiation protocols produce cells with gene expression patterns that that are immature or fetal, instead of adult, is a question of interest (27). These assessments are generally made based on the expression of a handful of fetal-specific genes (such as the expression of other hormones in addition to insulin), or the absence of a number of adult markers (such as MAFA). Because of the limited number of genes analyzed in these earlier studies, the degree to which cells derived in vitro actually resemble true human fetal cells remained unknown. Our genome-wide expression comparison of hPSC-INS^+ cells, from three different pluripotent stem cell lines, with human fetal and adult β cells points to differentiated hPSC-INS^+ cells being most like fetal cells. Although there are many differences in gene expression between hPSC-INS^+ cells, human fetal β cells and human adult β cells (suggesting that no two cell types are fully equivalent), this analysis showed close clustering of hPSC-INS^+ cells with human fetal β cells and not human adult β cells. hPSC-INS^+ cells and fetal β cells were no more different from each other than the biological replicates of human adult β cells. This result was confirmed by correlation analysis based on 152 pancreatic lineage genes.

Our work does not address the heterogeneity of INS^+ cells or the possibility that there is a smaller subset of hPSC-INS^+ cells and fetal β cells that closely resemble adult β cells. Further study of hPSC-INS^+ cells and fetal INS^+ cells sorted into smaller subsets based on expression of other markers (such as known hormones or adult β-cell transcription factors) would be needed to address these questions.

A thorough transcriptional analysis of in vivo-matured hPSC-INS^+ cells suggests that they are more similar to adult islets than to unsorted in vitro-derived hPSCs that were differentiated into INS^+ cells (33). Determining the degree of similarity between sorted in vivo-matured hPSC-INS^+ cells and fetal and adult β cells would be of great interest to better understand the state of in vivo-derived hPSC-INS^+ cells. Additionally, it would be interesting to analyze using MARIS recently described glucose-responsive hPSC-derived insulin-expressing cells (55) to determine their relationship to adult β cells.

The data reported here point to at least three classes of genes that are differentially expressed between INS^+ cells produced from stem cells in vitro and bona fide adult human β cells. The first class of genes regulates the lineage commitment of hPSC-INS^+ cells toward β cells and away from other pancreatic endocrine cell types. The second class of genes, which are similarly expressed in human fetal β cells and hPSC-INS^+ cells, and differentially expressed in adult β cells, may be responsible for functional β-cell maturation to achieve normal glucose response. The third class, which is uniquely expressed in hPSC-INS^+ cells, are sterol biosynthesis/metabolism genes typically expressed in liver instead of pancreatic cells and may represent inappropriate gene expression arising from the differentiation protocols. The challenge for the field can be defined as finding ways to manipulate the expression of these multiple genes so that one can reproducibly prepare large numbers of fully functional human β cells.

Experimental Procedures

MARIS Staining and FACS. hPSC-derived cells and human islets were dispersed to a single-cell suspension using TrypLE Express (Invitrogen). Human fetal pancreata were mechanically dispersed in the presence of 1 mg/mL Dispase (Roche) and 1 mM Collagenase P (Roche). All cells were passed through a 40-μm filter and washed with PBS at least twice. Cells were fixed and stained according to the MARIS. The list of primary and secondary antibodies used is provided in Table S3. Cells were sorted on the FACSAria (BD Biosciences) using FACSDiva software. Gates were set with reference to negative controls. The sorting speed was adjusted to ensure sorting efficiency above 90%. Cells were collected in tubes that were coated with a small amount of Sort buffer.

Global Gene Expression Analysis (Microarray). Using the Illumina TotalPrep RNA Amplification kit (Ambion), double-stranded cDNA was generated following reverse transcription from 100 ng of total RNA. In vitro transcription overnight with biotin-labeled nucleotides created amplified mRNA (cRNA), which was concentrated by vacuum centrifugation at 30 °C; 750 ng cRNA per sample was then hybridized to Human HT-12 Expression BeadChips (Illumina) using the Whole- Genome Expression Direct Hybridization kit (Illumina). Finally, chips were scanned on the Illumina Beadstation 500. The chip annotation manifest was version 4, revision 1. For differential expression analysis and the generation of gene lists for functional annotation and pathway analysis, microarray data were processed in GenomeStudio (V2011.1; Illumina). Raw data were adjusted by background subtraction and rank-invariant normalization. Before calculating fold change, an offset of 20 was added to all probe set means to eliminate negative signals. The *P* values for differences between mean signals were calculated in GenomeStudio by *t* test and corrected for multiple hypotheses testing by the Benjamini–Hochberg method in combination with the Illumina custom false discovery rate model.

Global Gene Expression Analysis (RNA-seq). Isolated RNA was obtained from two biological replicates of HUES8-derived INS^+ cells and human adult β cells, as well as one replicate of live and processed stage-6 cells. Samples were poly-A–purified and converted to cDNA libraries using the Illumina TruSeq protocol and prepared into Illumina libraries using the Beckman Coulter Genomics SPRIworks system using custom adapters; 6-nt 3′ barcodes were added during PCR enrichment and the resulting fragments were evaluated using Agilent BioAnalyzer 2100. Samples were multiplexed two-per-lane for sequencing using the Illumina HiSEq 2000 platform with paired-end read lengths of 80 nt, resulting in 68 million to 112 million paired reads per sample, and an average biological fragment length of 168–179 nt. Reads were aligned to the human genome (GRCh37/hg19) using STAR (version 2.2.0c) guided by GENCODE gene annotations (version 14) (56). RNA-seq FPKM (fragments per kilobase of exon per million fragments) gene enrichment was determined using maximum likelihood by Cuffdiff (57, 58) (version 2.0.2) and visualized using CummeRbund (57). Transcript differential expression was calculated by Cuffdiff using the default negative binomial model, with significant hits also confirmed using the count-based technique DESEq (59).

ACKNOWLEDGMENTS. The authors thank Dena Cohen for helpful discussion and Eliza Scadden and Priya Makhijani for technical help. S.H. is a Biological and Biomedical Sciences Ph.D. student supported by the Sternlicht Director's Fund Fellowship. J.R.M. was supported by a Harvard Stem Cell Institute (HSCI) Training Grant. D.A.M. is an investigator of the Howard Hughes Medical Institute (HHMI). This project was supported in part by the Helmsley Charitable Trust, the Harvard Stem Cell Institute, and the National Institutes of Health Grants 2U01DK07247307, RL1DK081184, and 1U01HL10040804.

1. Takahashi K, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5):861–872.
2. Thomson JA, et al. (1998) Embryonic stem cell lines derived from human blastocysts. *Science* 282(5391):1145–1147.
3. Yu J, et al. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318(5858):1917–1920.
4. Park I-H, et al. (2008) Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 451(7175):141–146.
5. Colman A, Dreesen O (2009) Pluripotent stem cells and disease modeling. *Cell Stem Cell* 5(3):244–247.
6. Daley GQ (2010) Stem cells: Roadmap to the clinic. *J Clin Invest* 120(1):8–10.
7. Rubin LL (2008) Stem cells and drug discovery: The beginning of a new era? *Cell* 132(4):549–552.
8. Chin MH, et al. (2009) Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* 5(1):111–123.
9. Doi A, et al. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* 41(12):1350–1353.
10. Hu B-Y, et al. (2010) Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc Natl Acad Sci USA* 107(9):4335–4340.
11. Osafune K, et al. (2008) Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat Biotechnol* 26(3):313–315.
12. Bock C, et al. (2011) Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 144(3):439–452.
13. Chetty S, et al. (2013) A simple tool to improve pluripotent stem cell differentiation. *Nat Methods* 10(6):553–556.
14. Nostro MC, et al. (2011) Stage-specific signaling through TGFβ family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. *Development* 138(5):861–871.
15. Ginis I, et al. (2004) Differences between human and mouse embryonic stem cells. *Dev Biol* 269(2):360–380.
16. Odom DT, et al. (2007) Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* 39(6):730–732.
17. Riedel MJ, et al. (2012) Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas. *Diabetologia* 55(2):372–381.
18. Polak M, Bouchareb-Banaei L, Scharfmann R, Czernichow P (2000) Early pattern of differentiation in the human pancreas. *Diabetes* 49(2):225–232.
19. De Krijger RR, et al. (1992) The midgestational human fetal pancreas contains cells coexpressing islet hormones. *Dev Biol* 153(2):368–375.
20. Hayek A, Beattie GM (1997) Experimental transplantation of human fetal and adult pancreatic islets. *J Clin Endocrinol Metab* 82(8):2471–2475.
21. Beattie GM, Butler C, Hayek A (1994) Morphology and function of cultured human fetal pancreatic cells transplanted into athymic mice: A longitudinal study. *Cell Transplant* 3(5):421–425.
22. Rezania A, et al. (2012) Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice. *Diabetes* 61(8):2016–2029.
23. Maehr R, et al. (2009) Generation of pluripotent stem cells from patients with type 1 diabetes. *Proc Natl Acad Sci USA* 106(37):15768–15773.
24. Chen S, et al. (2009) A small molecule that directs differentiation of human ESCs into the pancreatic lineage. *Nat Chem Biol* 5(4):258–265.
25. Kelly OG, et al. (2011) Cell-surface markers for the isolation of pancreatic cell types derived from human embryonic stem cells. *Nat Biotechnol* 29(8):750–756.
26. Kroon E, et al. (2008) Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat Biotechnol* 26(4):443–452.
27. D'Amour KA, et al. (2006) Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol* 24(11):1392–1401.
28. Basford CL, et al. (2012) The functional and molecular characterisation of human embryonic stem cell-derived insulin-positive cells compared with adult pancreatic beta cells. *Diabetologia* 55(2):358–371.
29. Micallef SJ, et al. (2012) INS(GFP/w) human embryonic stem cells facilitate isolation of in vitro derived insulin-producing cells. *Diabetologia* 55(3):694–706.
30. Jiang J, et al. (2007) Generation of insulin-producing islet-like clusters from human embryonic stem cells. *Stem Cells* 25(8):1940–1953.
31. Tateishi K, et al. (2008) Generation of insulin-secreting islet-like clusters from human skin fibroblasts. *J Biol Chem* 283(46):31601–31607.
32. Zhang D, et al. (2009) Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells. *Cell Res* 19(4):429–438.
33. Xie R, et al. (2013) Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* 12(2):224–237.
34. Morán I, et al. (2012) Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab* 16(4):435–448.
35. Dorrell C, et al. (2011) Transcriptomes of the major human pancreatic cell types. *Diabetologia* 54(11):2832–2844.
36. Hrvatin S, Deng F, O'Donnell CW, Gifford DK, Melton DA (2014) MARIS: Method for analyzing RNA following intracellular sorting. PLoS ONE, 10.1371/journal.pone.0089459.
37. Bonal C, Herrera PL (2008) Genes controlling pancreas ontogeny. *Int J Dev Biol* 52(7):823–835.
38. Blum B, et al. (2012) Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. *Nat Biotechnol* 30(3):261–264.
39. Bonett RM, Hu F, Bagamasbad P, Denver RJ (2009) Stressor and glucocorticoid-dependent induction of the immediate early gene kruppel-like factor 9: Implications for neural development and plasticity. *Endocrinology* 150(4):1757–1765.
40. Lavado A, Lagutin OV, Chow LM, Baker SJ, Oliver G (2010) Prox1 is required for granule cell maturation and intermediate progenitor maintenance during brain neurogenesis. *PLoS Biol* 8(8):e1000460.
41. Paz H, Lynch MR, Bogue CW, Gasson JC (2010) The homeobox gene Hhex regulates the earliest stages of definitive hematopoiesis. *Blood* 116(8):1254–1262.
42. Steele-Perkins G, et al. (2005) The transcription factor gene Nfib is essential for both lung maturation and brain development. *Mol Cell Biol* 25(2):685–698.
43. Weinhaus AJ, et al. (2003) Insulin secretagogues, but not glucose, stimulate an increase in [Ca2+]i in the fetal human and porcine beta-cell. *J Clin Endocrinol Metab* 88(6):2753–2759.
44. Dalgin G, et al. (2011) Zebrafish mnx1 controls cell fate choice in the developing endocrine pancreas. *Development* 138(21):4597–4608.
45. Sander M, et al. (2000) Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. *Development* 127(24):5533–5540.
46. Yang Y-P, Thorel F, Boyer DF, Herrera PL, Wright CV (2011) Context-specific α- to-β-cell reprogramming by forced Pdx1 expression. *Genes Dev* 25(16):1680–1685.
47. Collombat P, et al. (2003) Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev* 17(20):2591–2603.
48. Petri A, et al. (2006) The effect of neurogenin3 deficiency on pancreatic gene expression in embryonic mice. *J Mol Endocrinol* 37(2):301–316.
49. Shih DQ, Navas MA, Kuwajima S, Duncan SA, Stoffel M (1999) Impaired glucose homeostasis and neonatal mortality in hepatocyte nuclear factor 3alpha-deficient mice. *Proc Natl Acad Sci USA* 96(18):10152–10157.
50. Aguayo-Mazzucato C, et al. (2011) Mafa expression enhances glucose-responsive insulin secretion in neonatal rat beta cells. *Diabetologia* 54(3):583–593.
51. Brun T, et al. (2004) The diabetes-linked transcription factor PAX4 promotes beta-cell proliferation and survival in rat and human islets. *J Cell Biol* 167(6):1123–1135.
52. Gelling RW, et al. (2003) Lower blood glucose, hyperglucagonemia, and pancreatic alpha cell hyperplasia in glucagon receptor knockout mice. *Proc Natl Acad Sci USA* 100(3):1438–1443.
53. Obermüller S, et al. (2010) Defective secretion of islet hormones in chromogranin-B deficient mice. *PLoS ONE* 5(1):e8936.
54. Ingelsson E, et al.; MAGIC investigators (2010) Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes* 59(5):1266–1275.
55. Cheng X, et al. (2012) Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell* 10(4):371–384.
56. Harrow J, et al. (2012) GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 22(9):1760–1774.
57. Trapnell C, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578.
58. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
59. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.

Supporting Information

Hrvatin et al. 10.1073/pnas.1400709111

SI Experimental Procedures

Directed Differentiation. Human pluripotent stem cells (hPSCs) were routinely cultured on human embryonic stem cell (hESC)-certified matrigel (BD Biosciences) in mTeSR medium (Stemcell Technologies). Cells were passaged at the ratio of 1:6–1:20 every 4–7 d using TrypLE Express (Invitrogen). Two different basal media were used during differentiation. Basal media 1 (BM-1) contained MCDB-131 (Invitrogen) supplemented to 10 mM glucose, 1× GlutaMAX (Gibco, Life Technologies), 2.35 g/L $NaHCO_3$, and 0.1% reagent-grade BSA (Proliant). Basal media 2 (BM-2) contained MCDB-131 (Invitrogen) supplemented to 8 mM glucose, 1× GlutaMAX (Gibco, Life Technologies), 2.93 g/L $NaHCO_3$, 2% (wt/wt) reagent-grade BSA (Proliant), 1:200 ITS-X (Invitrogen), and 44 mg/L vitamin C (Sigma–Aldrich).

To initiate differentiation the cells were dissociated using TrypLE Express to single cells and seeded at 150,000 cells/cm^2 onto 1:30 dilution of growth factor reduced matrigel (BD Biosciences) in DMEM/F12 in mTeSR media with 10 μM Y27632 (StemGent). Two days following seeding the differentiation was started.

Day-1 (stage-1.1) cells were exposed to BM-1 supplemented with 3 μM CHIR-99021 (Stemgent) + 100ng/mL rhActivinA (R&D Systems). Days 2–3 (stage 1.2): BM-1 + 100 ng/mL rhActivinA (R&D Systems). Days 4–5 (stage 2): BM-1 + 50 ng/mL FGF7 (Peprotech). Days 6–9 (stage 3): BM-2 + 50 ng/mL FGF7 (Peprotech) + 2 μM RA (Sigma) + 0.25 μM SANT-1 (Sigma) + 20 ng/mL rhActivinA (R&D Systems) + 100 ng/mL rhNoggin (R&D Systems). Days 10–12 (stage 4): BM-2 + 100 ng/mL rhNoggin (R&D Systems) + 0.25 μM SANT-1 (Sigma) + 100 nM PDBu (EMD Biosciences). Days 13–15 (stage 5): BM-2 + 100 ng/mL rhNoggin (R&D Systems) + 1 μM Alk5 inhibitor (Axxora). From day 15 onward cells were kept in BM-2 media awaiting analysis (stage 6).

RNA Quality Analysis. After isolation, RNA quantity was evaluated using a NanoDrop 2000 (NanoDrop Technologies). RNA quality was assessed by microcapillary electrophoresis on the Bioanalyzer 2100 (Agilent) using RNA Nano 6000 or RNA Pico 6000 chips, depending on the RNA concentration. Agilent 2100 Expert software was used to visualize the electropherograms and calculate the RNA integrity number, a standardized categorization of total RNA quality on a scale of 1 (worst) to 10 (best) (1).

Quantitative RT-PCR. CDNA (cDNA) was made from 4 ng unamplified total RNA with random hexamer priming using the High Capacity cDNA Reverse Transcription with RNase Inhibitor kit (Applied Biosystems). One fourth of the volume of cDNA was used for each TaqMan quantitative RT-PCR (qRT-PCR). Reactions contained transcript-specific TaqMan probes and Fast Universal PCR Master Mix with no AmpErase UNG (Applied Biosystems). The following transcript-specific TaqMan probes were used: β-actin (Hs99999903_m1), insulin (Hs00355773_m1), glucagon (Hs00174967_m1), somatostatin (Hs00356144_m1), pancreatic polypeptide (Hs00358111_g1), and ghrelin (Hs01074053_m1). Reactions were run on an Applied Biosystems 7900HT Fast Real-Time PCR System with default settings. Detection thresholds were automatically computed by SDS 2.3 software (Applied Biosystems). Threshold data were analyzed in DataAssist 3.0 (Applied Biosystems) using the comparative Ct (ΔΔCt) relative quantitation method, using β-actin as the endogenous control.

Microarray Expression Clustering. Gene-level microarray expression values were generated by GenomeStudio, using rank-invariant normalization with background subtraction, and analyzed using the R package (2). Per-sample and per-condition (averaging gene levels) hierarchical clustering was performed based on Pearson correlation and dendogram visualizations were drawn. pvclust's multiscale bootstrap resampling was used (with 500,000 iterations) to estimate "approximately unbiased" (AU) *P* values indicating the significance of each subcluster choice in the hierarchy given the underlying data.

DAVID Gene Ontology. The representation of Gene Ontology (GO) biological process terms within the lists of differentially expressed genes was assessed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID; http://david.abcc.ncifcrf.gov) Bioinformatics Resource. The *P* values were corrected for multiple hypotheses testing by the Benjamini–Hochberg false discovery rate procedure and expressed as a q value. Terms were ranked by q value.

Immunofluorescent Staining of Adherent Cells. Adherent cells were washed three times for 5 min in PBS and fixed with 4% paraformaldehyde for 30 min at room temperature. Cells were then washed three times for 5 min in PBS and blocked with 10% (vol/vol) donkey serum (Jackson Immunoresearch) in PBS/0.3% Triton X-100. Primary antibodies were incubated overnight at 4 °C. Secondary antibodies were incubated for 1 h at room temperature.

The list of primary and secondary antibodies used is provided in Table S3.

Glucose-Stimulated Insulin Secretion. Approximately 10^5 dispersed hPSC-derived S6 cells or dispersed fetal cells were plated per well of a 96-well plate and allowed to attach overnight. Alternatively, ∼5,000 dispersed islet cells were plated amid 1×10^5 hESCs (for cell–cell contact and attachment). Cells were then washed two times in PBS. Cells were incubated for 1 h in Kreb's buffer with no glucose (128 mM NaCl, 5 mM KCl, 2.7 mM CaCl2, 1.2 mM MgCl2, 1 mM Na_2HPO_4, 1.2 mM KH_2PO_4, 5 mM NaHCO3, 10 mM Hepes, and 0.1% BSA) at 37 °C, 5% CO_2, and then incubated for 60 min in Kreb's buffer with either (*i*) 2 mM (low) glucose, (*ii*) 20 mM (high) glucose, or (*iii*) 2 mM glucose with 30 mM KCl. Supernatant fractions after each exposure to glucose were collected and stored at −80 °C until analysis. Insulin concentrations were measured using the Mouse Ultrasensitive Insulin ELISA kit (80-INSMSU-E01; Alpco), which cross-reacts >100% with human insulin. Concentrations were calculated from cubic spline interpolation of a standard curve and normalized to input cell number.

1. Schroeder A, et al. (2006) The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 7:3.

2. Suzuki R, Shimodaira H (2006) Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540–1542.

Stage 1		Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
MCDB-131 0.1% BSA 10mM Glucose	MCDB-131 0.1% BSA 10mM Glucose	MCDB-131 0.1% BSA 10mM Glucose	MCDB-131 2% BSA 8mM Glucose	MCDB-131 2% BSA 8mM Glucose	MCDB-131 2% BSA 8mM Glucose	MCDB-131 2% BSA 8mM Glucose
ActivinA, CHIR99021	ActivinA	FGF7, Vitamin C	FGF7, ActivinA, SANT1, RA, Noggin, ITS-X, Vitamin C	SANT1, PdBu, Noggin, ITS-X, Vitamin C	Noggin, Alk5, ITS-X, Vitamin C	ITS-X, Vitamin C
1d	3d	2d	4d	3d	3d	3d+

Fig. S1. Directed differentiation protocol and RNA quality. Stepwise differentiation from hESCs to pancreatic endocrine cells. DE, definitive endoderm; EN, endocrine cells; EP, endocrine progenitor; PP, pancreatic progenitor. Table contains reagents used during each stage of directed differentiation.

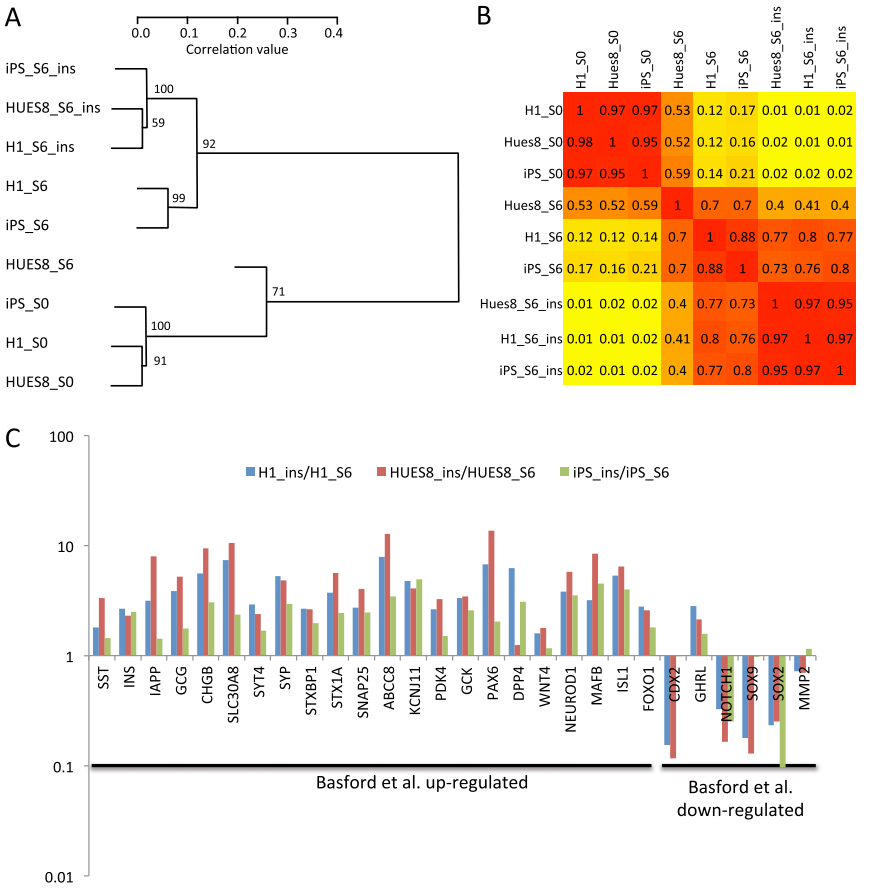


Fig. S2. Correlation, clustering, and gene expression of cells derived by directed differentiation. (*A*) Hierarchical clustering based on microarray data across 152 pancreatic lineage genes is shown. Lengths in the dendrogram represent correlation value. AU *P* values are displayed. Insulin-expressing (INS^+) cells from different cell lines form a statistically significant cluster. (*B*) r^2 values between groups of samples. (*C*) Comparable gene expression. Basford et al. (1) identified 28 genes that were differentially expressed between hESC-derived INS^+ cells and insulin-negative (INS^-) cells. Here shown is fold change in gene expression between hPSC-INS^+ and unsorted stage-6 cells for all 28 genes; 27 of the 28 genes had the same pattern of gene expression. Ghrelin (GHRL) was down-regulated in Basford et al. (1) but up-regulated in all cell lines in our study.

1. Basford CL, et al. (2012) The functional and molecular characterisation of human embryonic stem cell-derived insulin-positive cells compared with adult pancreatic beta cells. *Diabetologia* 55(2):358–371.

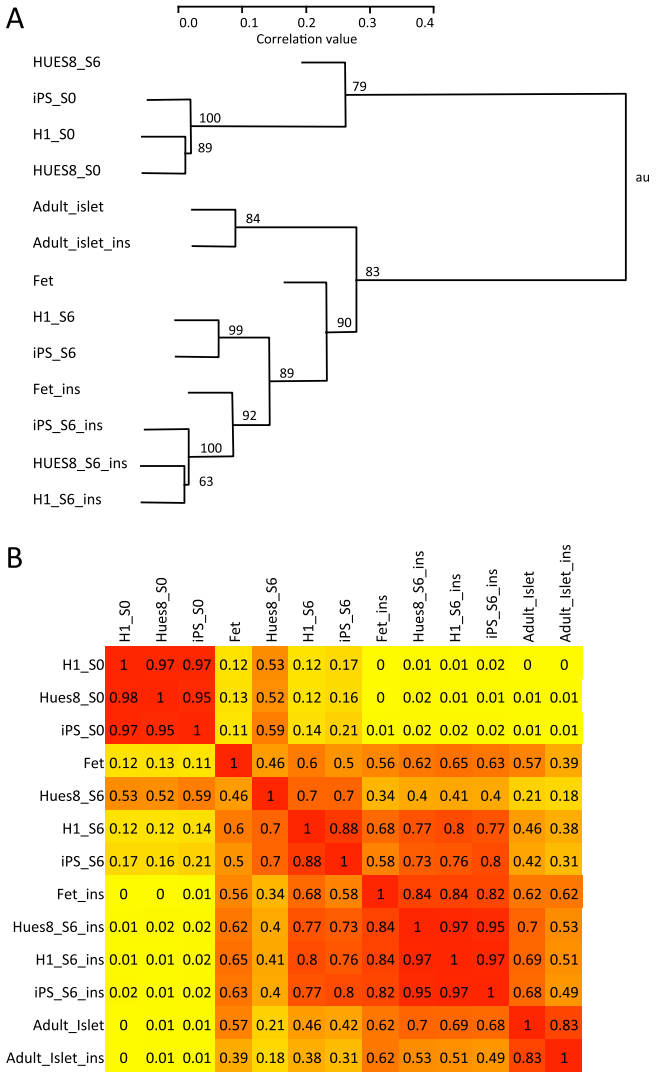


Fig. S3. Correlation and clustering of hPSC-INS$^+$ cells and fetal and adult β cells. (*A*) Hierarchical clustering based on microarray data across 152 pancreatic lineage genes is shown. Lengths in the dendrogram represent correlation value. AU *P* values are displayed. INS$^+$ cells from different cell lines form a statistically significant cluster. (*B*) r^2 values and hierarchical clustering based on microarray data across 152 pancreatic lineage genes is shown.

Table S1. Gene list

Secretory	Transcription factors			Glucose processing, insulin secretion		Membrane channels		Other
CHGA	ALX3	ISL1	ONECUT1	ACOX2	RIMS4	ABCC8	KCNJ11	CDH1
CHGB	ARX	LHX6	ONECUT2	ACSS1	SLC2A14	CACNA1A	KCNJ2	CDH2
GAST	CMIP	LMX1A	ONECUT3	ACSS3	SLC30A8	CACNA1B	KCNJ4	COL1A1
GCG	EN1	LMX1B	PAX2	ALDOA	SNAP25	CACNA1C	KCNJ6	CTNNB1
GHRL	ESR1	MAF	PAX4	ATP5G3	STX1A	CACNA1E	KCNK1	FFAR1
GIP	FOXA1	MAFA	PAX6	ATP6V0E1	STXBP1	HCN1	KCNK12	RAP1A
IAPP	FOXA2	MAFB	PBX1	COX4I1	SYP	HCN2	KCNK17	CDH1
INS	FOXA3	MLXIPL	PDX1	GPI	SYT4	HCN3	KCNK3	CDH2
NTS	FOXO1	MNX1	POU3F4	NNT	SYT7	HCN4	KCNMA1	DLL3
PPY	GATA4	MYT1	POU5F1	PCSK1	SYT8	KCNA3	KCNMB2	DPP4
PYY	GATA6	NANOG	PROX1	PCSK2	VAMP2	KCNA5	KCNN1	FGF4
SCT	GLIS3	NEUROD1	RFX2	PDIA6	VAMP3	KCNA6	KCNN3	GCGR
SST	HES1	NEUROG3	RFX3	PDK3	G6PC2	KCNB1	SCN1A	MMP2
	HNF1A	NKX2-2	RFX6	PGAP1	GCK	KCNC2	SCN2A	RAP1A
	HNF1B	NKX6-1	SNAI1	PRKCSH	SLC2A1	KCNC4	SCN3A	UCN3
	HNF4A	NKX6-2	SNAI2	RIMS1	SLC2A2	KCNG4	SCN4A	WNT2
	INSM1	NKX6-3	SOX4	RIMS2	SLC2A3	KCNH2	SCN7A	WNT3
	IRX2	NOTCH1	SOX9	RIMS3		KCNH6		WNT4

Pancreatic lineage genes (n = 152) were selected based on published literature for their relevance in pancreatic development or β-cell function.

Table S2. GO biological pathways enriched in hPSC-INS$^+$ cells

GO biological process	Benjamini q value
Cytoskeleton organization	0.0014108
Sterol biosynthetic process	0.0014891
Cholesterol biosynthetic process	0.0018188
Cell morphogenesis	0.0018485
Cellular component morphogenesis	0.0045675
Actin filament-based process	0.0150359
Actin cytoskeleton organization	0.0167663
Sterol metabolic process	0.0174035
Cell projection organization	0.0179661
Neuron projection development	0.0185401
Cholesterol metabolic process	0.0187551
Neuron development	0.0199759
Cell part morphogenesis	0.0201410
Neuron projection morphogenesis	0.0213534
Cell projection morphogenesis	0.0228888
Lipid transport	0.0264480
Lipid localization	0.0286305
Isoprenoid biosynthetic process	0.0295110
Neuron differentiation	0.0301086
Lipoprotein particle clearance	0.0440558
Response to organic substance	0.0444088
Steroid biosynthetic process	0.0456816

Genes more highly expressed in hPSC-INS$^+$ than in adult β cells ($P < 0.05$ by microarray) were analyzed using the DAVID. Twenty-two statistically enriched (Benjamini corrected $q < 0.05$) GO terms referencing biological processes are shown. GO terms are ranked by q value.

Table S3. Primary and secondary antibodies used in the study

Antigen	Species	Manufacturer and item no.	Dilution for MARIS*	Dilution for planar immunofluorescence
Insulin	Guinea pig	Dako, A0564	1:1,000	
Glucagon	Mouse	Abcam, ab82270	1:200	
Somatostatin	Goat	Santa Cruz Biotechnology, sc-7819	1:2,000	
C-peptide	Rat	DSHB, GN-ID4		1:500
Glucagon	Rabbit	Cell Signaling Technology, 2760		1:500
PDX1	Goat	R&D, AF2419		1:500
NKX6-1	Mouse	DSHB, F55A12		1:500
MAFA	Rabbit	Abcam, ab-26405		1:1,000
ARX	Goat	Santa Cruz Biotechnology, sc-48843		1:100
IgG (mouse, rat, rabbit, guinea pig, goat)	Donkey	Invitrogen, Alexa-488, Alexa-594 conjugated	1:500	1:500
IgG (guinea pig)	Donkey	Jackson Immunoresearch, Alexa-647, 706–605-148	1:500	1:500

*Method for Analyzing RNA following Intracellular Sorting.