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

Siniša Hrvatina, Charles W. O'Donnellab, Francis Denga, Jeffrey R. Millmana, Felicia Walton Pagliuc, Philip Dilorioa, Alireza Rezania, David K. Giffordb, and Douglas A. Meltonab

*Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute and †Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138; ‡Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; ‡Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA 01605; and ‡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 these variations 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 casts doubt on 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 glucagon-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 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.


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

1To 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 iP5-17b 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. 1A). 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. 1B).

The RNA isolated from hPSC-INS cells 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. 1C). 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. 1D).

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 based on these endocrine lineage genes confirmed a high degree of similarity between hESC-derived and hiPSC-derived insulin cells (Fig. S2A and B). Together, these data suggest that INS cells derived from different hPSC lines are very similar.

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. S2C) (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. 2A).

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 \( (P < 0.05, \text{fold change} > 3) \) (Fig. 2A). 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. 2C). 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. 2D). 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.

**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, \( P < 0.05 \)). 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), aristatess related homeobox (ARX), paired box 4 (PAX4), and irquiso homeobox 2 (IRX2)] and GSIS [MAFA; PAX4; GHR; chronomabrin 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. 4A). 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

---

**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.
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 protein channels KCN3 and KCN1K 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. 4C).

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, there involve cholesterol/sterol biosynthesis or metabolism, indicating that hPSC-INS+ cells aberrantly express nonpancreatic, liver-specific genetic pathways (Fig. 4D).

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. 4E). 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.
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 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 remains 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-differentiated 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 pancreas 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 FACS Diva 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-Gene 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, following the protocol of live and preserved stage-6 cells. Samples were poly-A–purified and converted to DNA 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 pair-end reads lengths of 80 nt resulting in 68 million total 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 2.8.2) and visualized using cufflinks (59). 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 discussions and Eliza Scadden and Priya Mahkijani for technical help. S.H. is a Biologist at 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 1U01HL1004084.