Stem Cell Reports
Resource
OPEN ACCESS

Single-Cell Transcriptome Profiling of Mouse and hESC-Derived Pancreatic Progenitors

Nicole A.J. Krentz,[1,2,4,*] Michelle Y.Y. Lee,[1] Eric E. Xu,[1,2] Shannon L.J. Sproul,[1,2] Alexandra Maslova,[3] Shugo Sasaki,[1,2] and Francis C. Lynn[1,2,5,*]

[1]Diabetes Research Group, BC Children's Hospital Research Institute, Vancouver, BC V5Z 4H4, Canada
[2]Departments of Surgery and Cellular and Physiological Sciences, University of British Columbia, 950 28th Avenue West, Vancouver, BC V5Z4H4, Canada
[3]Graduate Program in Bioinformatics, University of British Columbia, 100-570 7th Avenue West, Vancouver, BC V5Z 4S6, Canada
[4]Twitter: @nkrentz
[5]Twitter: @nictitate
*Correspondence: nkrentz@alumni.ubc.ca (N.A.J.K.), francis.lynn@ubc.ca (F.C.L.)
https://doi.org/10.1016/j.stemcr.2018.11.008

SUMMARY

Human embryonic stem cells (hESCs) are a potential unlimited source of insulin-producing β cells for diabetes treatment. A greater understanding of how β cells form during embryonic development will improve current hESC differentiation protocols. All pancreatic endocrine cells, including β cells, are derived from Neurog3-expressing endocrine progenitors. This study characterizes the single-cell transcriptomes of 6,905 mouse embryonic day (E) 15.5 and 6,626 E18.5 pancreatic cells isolated from *Neurog3-Cre; Rosa26*$^{mT/mG}$ embryos, allowing for enrichment of endocrine progenitors (yellow; tdTomato + EGFP) and endocrine cells (green; EGFP). Using a *NEUROG3-2A-eGFP* CyT49 hESC reporter line (N5-5), 4,462 hESC-derived GFP+ cells were sequenced. Differential expression analysis revealed enrichment of markers that are consistent with progenitor, endocrine, or previously undescribed cell-state populations. This study characterizes the single-cell transcriptomes of mouse and hESC-derived endocrine progenitors and serves as a resource (https://lynnlab.shinyapps.io/embryonic_pancreas) for improving the formation of functional β-like cells from hESCs.

INTRODUCTION

Diabetes mellitus is a metabolic syndrome characterized by elevated blood glucose levels that result from reductions in insulin production or action. Insulin is produced by pancreatic β cells found within the endocrine islets of Langerhans. A potential treatment for diabetes is to replace insulin by transplantation of human embryonic stem cell (hESC)-derived β cells. Derivation of functional β cells from hESCs requires an in-depth understanding of how endocrine cells form during embryonic development.

During mouse and human pancreas development, pancreatic progenitors become restricted to the endocrine cell fate before differentiating to hormone-producing cells. This process involves many transcription factors (TFs) that drive the changes in gene expression necessary for endocrine cell genesis. Genetic loss-of-function studies have found roles for individual TFs in the formation of specific islet cell types. From this work, a map of the TF cascade that regulates the formation of endocrine cells, including the β cells, has emerged (Cano et al., 2014). However, our understanding of fate decisions during endocrine cell formation is based on studies of the whole population of progenitors, using technologies such as bulk RNA sequencing and often only in mouse cells. The gene expression of individual human and mouse cells during terminal differentiation is unknown.

A promising method to understand gene expression changes at single-cell resolution is single-cell RNA sequencing (scRNA-seq). Following the first publication in 2009 (Tang et al., 2009), commercial platforms and lower sequencing costs have made scRNA-seq a feasible technology for many biologists. Recently, several studies have investigated the single-cell transcriptome of healthy and type 2 diabetes human islets (Baron et al., 2016; Chu et al., 2016; Enge et al., 2017; Lawlor et al., 2017; Qiu et al., 2017; Segerstolpe et al., 2016; Wang et al., 2016; Xin et al., 2016; Zeng et al., 2017). From these studies, we have begun to appreciate the cell-type-specific gene expression changes that occur during diabetes progression, the differences between mouse and human islet cells, and the identity of islet and pancreatic cell types.

Several recent studies have started characterizing the single-cell transcriptomes of mouse and human progenitors during embryonic development, including mouse embryonic day (E) 13.5 (Stanescu et al., 2017), E14.5 (Byrnes et al., 2018), E16.5 (Scavuzzo et al., 2018), and E17.5 pancreas (Byrnes et al., 2018) and human fetal pancreas (Ramond et al., 2018). In this study, scRNA-seq was used to analyze 6,905 E15.5 pancreatic cells, 6,626 E18.5 pancreatic cells, and 4,462 hESC-derived endocrine progenitor (EP) cells. From these data, novel cell types were identified and comparisons between hESC-derived endocrine cells, mouse EP, and human islet cells were made. Characterization of these populations will aid efforts to generate an

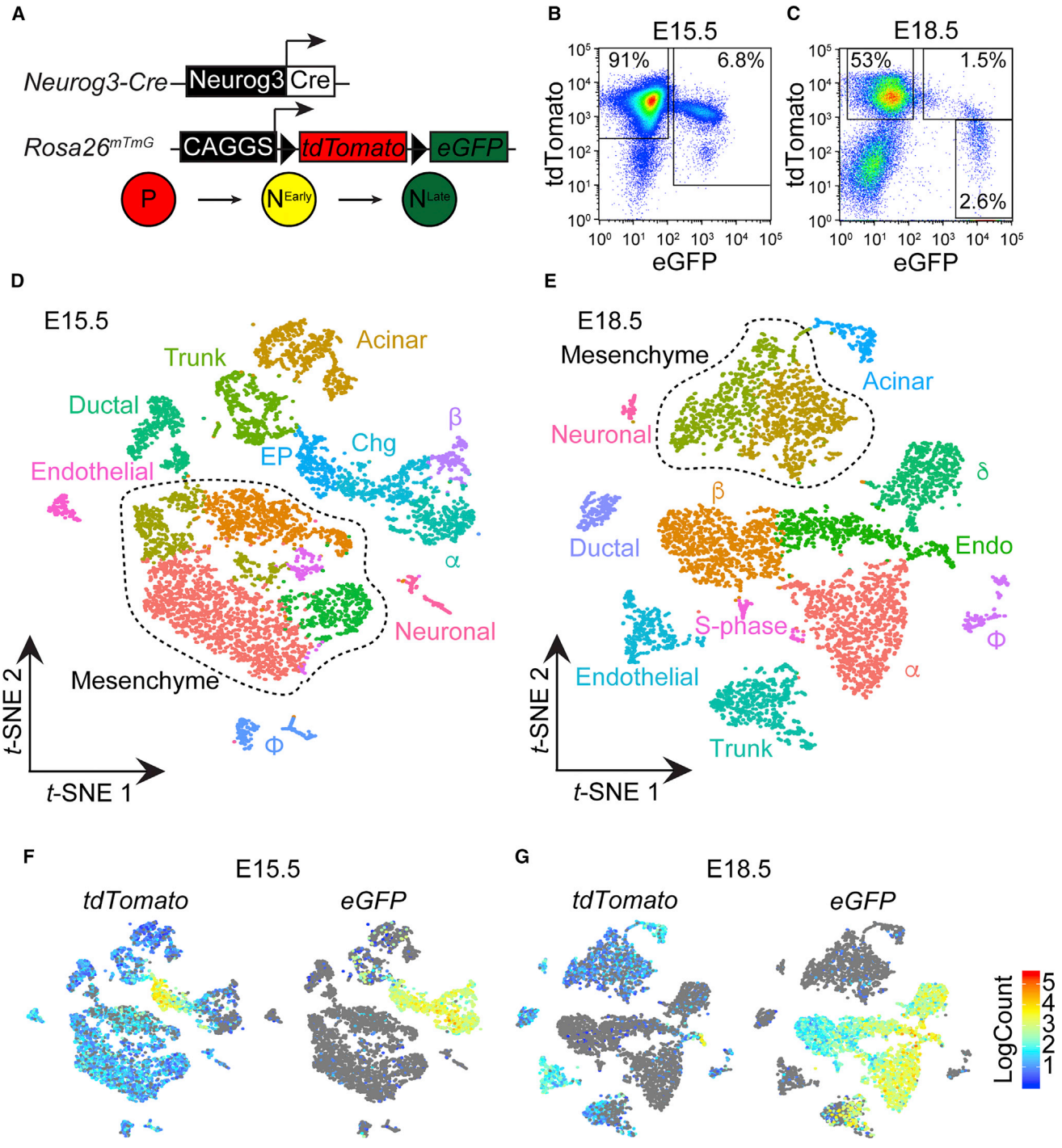


Figure 1. Cell Populations in E15.5 and E18.5 Mouse Pancreas

(A) Schematic overview of the two transgenic mouse lines used to isolate cell populations during pancreas development. Using this strategy, pancreatic progenitors (P; red) are tdTomato+, early Neurog3-lineage cells (N^{Early}; yellow) are tdTomato+ and eGFP+, and late Neurog3-lineage cells (N^{Late}; green) are eGFP+.

(B and C) FACS plot of (B) E15.5 and (C) E18.5 cells used for library generation.

(legend continued on next page)

unlimited source of insulin-producing β-like cells for diabetes treatment.

RESULTS

Strategy for Generating Single-Cell Transcriptomes of Embryonic Mouse Pancreas

Two transgenic mouse lines were used to isolate progenitor populations during pancreas development: *Neurog3-Cre* and *Rosa26*$^{mT/mG}$ (Figure 1A). In *Neurog3-Cre; Rosa26*$^{mT/mG}$ embryos, all cells are labeled with a membrane-targeted Tomato red fluorescent protein (tdTomato). Upon activation of the *Neurog3* promoter, Cre recombinase removes the floxed *tdTomato* cassette, resulting in expression of a membrane-targeted enhanced green fluorescent protein (eGFP). Cells that recently activated *Neurog3* express both tdTomato and eGFP (yellow; N^{Early}), while cells that are further along the endocrine cell lineage express eGFP only (green; N^{Late}) (Figure 1A) (Xu et al., 2015). This strategy was used to isolate by fluorescence-activated cell sorting (FACS) the three populations from the pancreas of one E15.5 and E18.5 embryo, and single-cell libraries were generated using a 10× Genomics Chromium single cell 3′ kit. In total, 7,502 E15.5 and 7,023 E18.5 single cells were sequenced at a depth of >50,000 reads per cell using Illumina NextSeq 500 (Table 1).

Identification of Cell Types in E15.5 and E18.5 Pancreas

At E15.5, 91% of pancreatic cells expressed tdTomato protein only (Figure 1B). As the yellow and green populations were less abundant (6.8%), these cells were pooled together and sequenced as one library (Figure 1B). At E18.5, the red (53%), yellow (1.5%), and green (2.6%) cells were used to generate three separate libraries (Figure 1C). To explore the cell types present within the pancreas, the sequenced red, yellow, and green cells at E15.5 and E18.5 were aggregated into single datasets using cellranger aggr and low-quality cells were excluded from analysis using Scater and Seurat (see Experimental Procedures for details). Following this, 6,905 E15.5 and 6,626 E18.5 cells were clustered using unsupervised *k*-means clustering and visualized using *t*-distributed stochastic neighbor embedding (*t*-SNE) (van der Maaten and Hinton, 2008). Cluster identity was inferred using differentially expressed genes (Tables S2 and S3 and Figure S1). At 15.5, there were 15 clusters representing mesenchyme cells (50.0%), acinar cells (8.8%), bipotent trunk progenitor cells (7.9%), ductal cells (6.6%), *Chga*-expressing immature endocrine cells (6.3%), α cells (6.3%), EPs (4.7%), macrophages (3.1%), β cells (2.6%), neuronal cells (1.9%), and endothelial cells (1.9%) (Figure 1D). Many of these clusters were also present at 18.5, including mesenchyme cells (22.8%), α cells (17.2%), β cells (14.6%), trunk cells (9.3%), endothelial cells (6.0%), acinar (3.1%), ductal (3.0%), macrophages (2.7%), and neuronal cells (1.1%) (Figure 1E). As we sequenced more yellow and green cells at E18.5 (3,950 versus 1,322 cells) (Table S1), three other endocrine populations were found: δ cells (9.4%), immature endocrine cells (Endo; 9.4%), and cells undergoing DNA replication (S-phase; 1.3%) (Figure 1E).

To verify the library identity of individual cells, the GRCm38 genome used for alignment was annotated to include the sequences for the *tdTomato* and *eGFP* transgenes. At E15.5 and E18.5, the trunk, acinar, ductal, mesenchymal, endothelial, neuronal, and macrophage cells expressed *tdTomato*, consistent with the non-endocrine lineage of these cell types (Figures 1F and 1G). At E15.5, the EP cells co-expressed both *tdTomato* and *eGFP* (Figure 1F), suggesting recent activation of *Neurog3*. In addition, a subset of Endo cells at E18.5 also co-expressed *tdTomato* and *eGFP*, consistent with a small EP population (Figure 1G). Interestingly, a subset of the trunk cells expressed *eGFP* at both E15.5 and E18.5 (Figures 1F and 1G), consistent with heterogeneous activation of *Neurog3* transcription in trunk cells. As all endocrine cells are derived from Neurog3+ progenitors (Gu et al., 2002), E15.5 and E18.5 endocrine cells expressed *eGFP* (Figures 1F and 1G). Taken together, scRNA-seq identified pancreatic cell populations and differentially expressed genes at E15.5 and E18.5.

Characterization of the Mouse Embryonic Endocrine Cell Transcriptome

To understand the transcriptional changes that occur during endocrine specification, the yellow and green cells were further characterized at E15.5. After filtering, 1,322 cells were analyzed using unsupervised *k*-means clustering and visualized using a *t*-SNE plot (Figure 2A). Eight clusters representing several cell populations were identified using the top ten differentially expressed genes (Table S4): α cells (20.3%), two EP populations (EP1, 19.1%, and EP2, 10.5%), β cells (18.8%), *Chga*-expressing endocrine cells (15.0%),

(D) Within the E15.5 pancreatic cells there were 15 clusters of 11 cell types: trunk, acinar, endocrine progenitor (EP), *Chg*-expressing endocrine (Chg), α cell (α), β cell (β), ductal, macrophage (ϕ), neuronal, endothelial, and five mesenchyme clusters.
(E) Within the E18.5 pancreatic cells there were 13 clusters of 12 cell types: trunk, acinar, ductal, endocrine cells (Endo), α cell (α), β cell (β), δ cell (δ), replicating cells (S-phase), macrophage (ϕ), neuronal, endothelial, and two mesenchyme clusters.
(F and G) Single-cell gene expression of *tdTomato* and *eGFP* at (F) E15.5 and (G) E18.5.

Table 1. Number of Targeted and Sequenced Cells in Six Single-Cell RNA-Sequencing Libraries

	Targeted Cell Number	Sequenced Cell Number
E15.5 red	6,000	6,010
E15.5 yellow and green	3,000	1,492
E18.5 red	5,000	2,846
E18.5 yellow	1,000	600
E18.5 green	5,000	3,577
S6D1 GFP	5,000	4,995

Endo cells (7.3%), trunk progenitor cells (4.8%), and ghrelin cells (ε; 4.3%).

As Neurog3+ progenitor cells exit the cell cycle during differentiation to endocrine cells (Desgraz and Herrera, 2009; Jensen et al., 2000; Miyatsuka et al., 2011), the cell cycle stage of individual cells at E15.5 was investigated. While the EP clusters included dividing cells, cells of the endocrine lineage mainly expressed G0/1 markers, consistent with cell cycle exit (Figure S2A). Interesting, the trunk population also contained many S- and G2/M-phase cells that expressed *tdTomato*, *eGFP*, and *Neurog3*, suggesting that these trunk cells recently activated the *Neurog3* promoter (Figures S2A–S2C) (Bechard et al., 2016).

To confirm the identity of the E15.5 yellow and green clusters, the expression of several genes was investigated. *Neurog3* was highly expressed in both EP clusters, while lower expression was also found in trunk, ε cells, and *Chg*-expressing cells (Figures 2B and S2C). *Neurod1*, a target of Neurog3, was expressed throughout the endocrine cell lineage except in the trunk cells (Figure S2C). Both *Ins1* and *Ins2* were expressed in the β cells while *Gcg* and *Ghrl* were specific to the α and ε cells, respectively (Figures 2C–2F and S2C).

To find cell-type-specific markers, the top ten differentially expressed genes in the EP1 (yellow), EP2 (blue), α (red), and β (green) cells were profiled (Figure S2D). The EP clusters expressed known marker genes such as *Neurog3*, *Pax4*, and *Sox4* along with previously undescribed genes, including *Midkine* (*Mdk*) and *Growth Arrest and DNA Damage Inducible Alpha* (*Gadd45a*) (Figure S2D). The α cells were enriched for expression of *Gcg* and previously proposed α cell markers *Slc38a5* (Stanescu et al., 2017) and *Transthyretin* (*Ttr*) (Westermark and Westermark, 2008). The β cell cluster expressed several β cell genes, including *Ins1*, *Ins2*, *Pdx1*, and *Iapp* (Figure S2D), along with the Neurod1 target *Nnat* (Chu and Tsai, 2005) and previously identified β cell marker *Ppp1r1a* (Martens et al., 2011).

We next aimed to characterize cells of the endocrine lineage at E18.5. To do this, cells from the E18 yellow and E18 green libraries were aggregated using cellranger aggr, resulting in 3,970 cells made up of 561 yellow and 3,409 green cells (Table S1). Visualizing these data using *t*-SNE revealed 11 clusters: two β cell populations (β1, 20%, and β2, 11.1%), α cells (21.0%), δ cells (15.0%), trunk cells (10.1%), ε cells (7.7%), *Chg*-expressing Endo cells (6.1%), S-phase cells (3.8%), stellate cells (2.7%), EPs (1.8%), and macrophages (0.6%) (Figure 2G). Yellow library cells were found in the trunk, EP, stellate, and macrophage cells and expressed *tdTomato* and *eGFP* (Figures S3A and S3B). Many of the same genes (*Spp1*, *Mt1*, and *Mt2*) were expressed in trunk cells at E15.5 and E18.5, while EP cells expressed *Neurog3*, *Gadd45a*, and *Pax4* (Figures S2E and S3C).

The endocrine cells were found within the E18.5 green cell library and expressed *eGFP* only (Figures S3A and S3B). Both *Ins1* and *Ins2* were highly expressed in the two β cell populations, β1 and β2, as well as in the S-phase cells (Figures 2H, 2I, and S3C), consistent with the start of the wave of replication that is required for the acquisition of β cell mass (Georgia and Bhushan, 2004). Differential expression analyses of the β cell populations reveal cluster-specific differences in gene expression (Table S5). The top ten differentially expressed genes in the β1 cluster included markers of mature cells, including *Ins1*, *Ins2*, *G6pc2*, and *Slc2a2*, while the β2 cluster was a *Nnat*-expressing immature cell state (Figure S3E). The *Chg*-expressing Endo cell state also expressed *Nnat* along with the progenitor cell marker *Neurod1* (Figure S3E). The S-phase cells expressed high levels of *Ins1*, *Ins2*, and *Gcg*, suggesting this cluster represents a mixture of α and β cells (Figure S3C) undergoing DNA replication and mitosis (Figure S3E). As expected, *Sst* expression was upregulated in the δ cell population (Figure 2J), *Ghrl* was upregulated in the ε cell cluster (Figure 2L), and *Gcg* was highly expressed in the α cell cluster (Figure 2M). Although none of the clusters showed specific upregulation of *Ppy* at this stage, some α cells and δ cells expressed *Ppy* (Figure 2K).

Pseudotime Analysis of Endocrine Lineage Using RNA Velocity

To understand the developmental cell lineage and resultant changes in gene expression, RNA velocity was used to order the E15.5 and E18.5 endocrine cells in pseudotime (Manno et al., 2018). RNA velocity uses unspliced and spliced mRNA to estimate the rate of change between cell states. Using the E15.5 aggr library (from Figure 1), a reduced dataset of 2,869 cells was generated using the trunk, acinar, ductal, EP, Chg, α, and β cells and visualized using a *t*-SNE plot (Figure 3A). Based on the ratio of unspliced/spliced transcripts (Figure S4), a velocity was assigned, and the average vector field was represented by an arrow indicating the direction of differentiation. Within the ductal

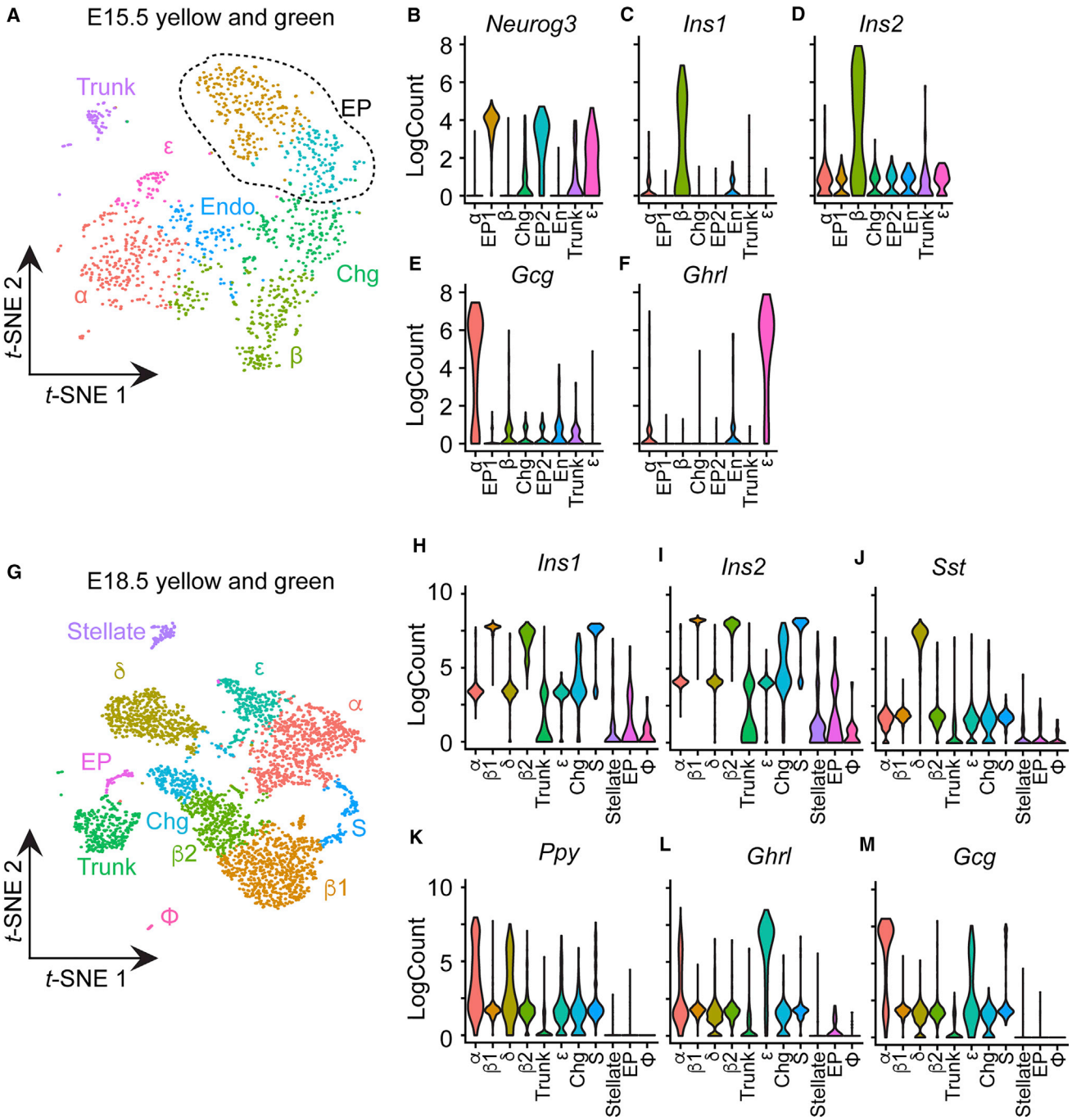


Figure 2. Cell Populations in Mouse Endocrine-Lineage Cells at E15.5 and E18.5

(A) *t*-SNE plot of eight clusters in E15.5 yellow and green cells. These clusters were identified as trunk cells (Trunk; 4.8%), two endocrine progenitor clusters (EP1 yellow, 19.1%, and EP2 blue, 10.5%), *Chga*-expressing immature endocrine cells (Chg; 15.0%), endocrine cells (Endo; 7.3%), α cells (α; 20.3%), β cells (β; 15.0%), and ghrelin cells (ε; 4.3%).

(B–F) Single-cell gene expression of (B) *Neurog3*, (C) *Ins1*, (D) *Ins2*, (E) *Gcg*, and (F) *Ghrl* across cell clusters.

(G) *t*-SNE plot of 11 cell clusters from E18.5 yellow and green cells: trunk (10.1%), endocrine progenitor (EP; 1.8%), *Chg*-expressing (Chg; 6.1%), α cell (α; 21.0%), two β cell (β1, 20%, and β2, 11.2%), δ cell (δ; 15.0%), Ghrl cell (7.7%), S-phase cell (S; 3.8%), stellate (2.7%), and macrophage (ϕ; 0.6%).

(H–M) Expression of endocrine hormones (H) *Ins1*, (I) *Ins2*, (J) *Sst*, (K) *Ppy*, (L) *Ghrl*, and (M) *Gcg* across clusters.

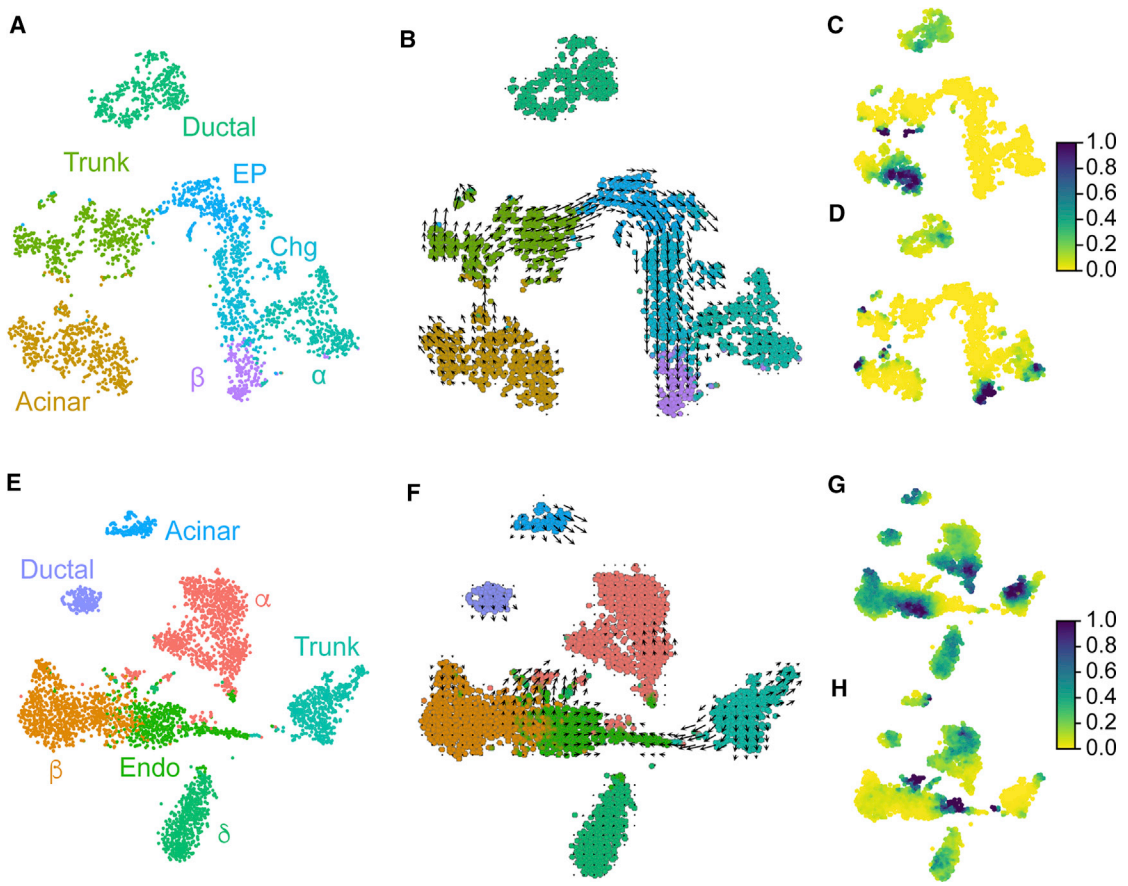


Figure 3. Pseudotime Analysis of E15.5 and E18.5 Endocrine Cells Using RNA Velocity

(A) *t*-SNE projection of E15.5 populations used for RNA velocity analysis: acinar, ductal, trunk, endocrine progenitor (EP), *Chg*-expressing (Chg), α, and β cells.

(B) Velocity field projected onto *t*-SNE plot. Arrows represent average velocity.

(C and D) The root (C) and endpoint (D) of differentiation was identified using a Markov random-walk model.

(E) *t*-SNE projection of E18.5 populations used for RNA velocity analysis: acinar, ductal, trunk, immature endocrine (Endo), α, β, and δ cells.

(F) Velocity field projected onto *t*-SNE plot. Arrows represent average velocity.

(G and H) The root (G) and endpoint (H) of differentiation was identified using a Markov random-walk model.

population, there was little progression toward other cell fates, suggesting terminal differentiation (Figure 3B). Interestingly, a subset of acinar cells showed progression toward the trunk cell fate (Figure 3B). Within the trunk cluster, there were three root points determined using a Markov random-walk model (Figure 3C). From these roots, the cells transitioned through the EP state toward α and β cells (Figure 3B). This was also confirmed by the endpoints of differentiation (Figure 3D). The same method was used to determine pseudotime of trunk, acinar, ductal, Endo, α, β, and δ cells at E18.5 (from Figure 1). For this analysis, a reduced dataset of 4,263 cells was visualized using *t*-SNE (Figure 3E). Unlike at E15.5, cells at E18.5 had less overall velocity, suggesting more terminally differentiated cells (Figure 3F). Within the Endo population there was a single differentiation root (Figure 3G) that bifurcates toward two cell types, the α cells and a subpopulation of Endo cells. This is consistent with a population of EPs that serve as a root for differentiation and suggests that they either differentiate toward an endocrine cell lineage or remain as Endo cells. Together, these data highlight the heterogeneity of cell fate decisions during mouse pancreas development.

Characterization of Endocrine Cell Population at E18.5

To further investigate the heterogeneity of embryonic endocrine cells, the E18.5 green library was studied. Following filtering, 3,382 cells were visualized using a *t*-SNE plot and the following cell types were identified based on gene expression (Table S6): two β cell populations

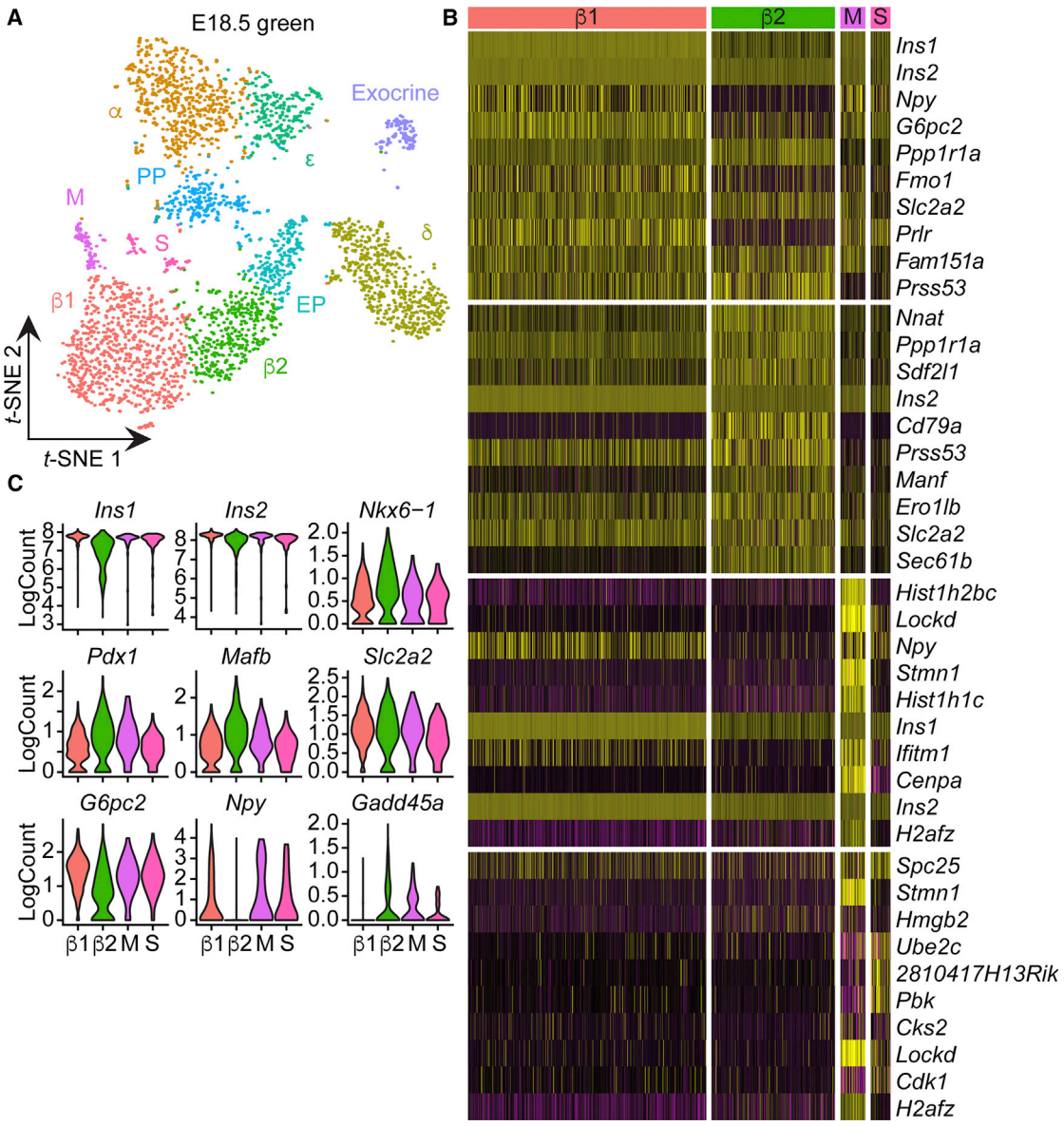


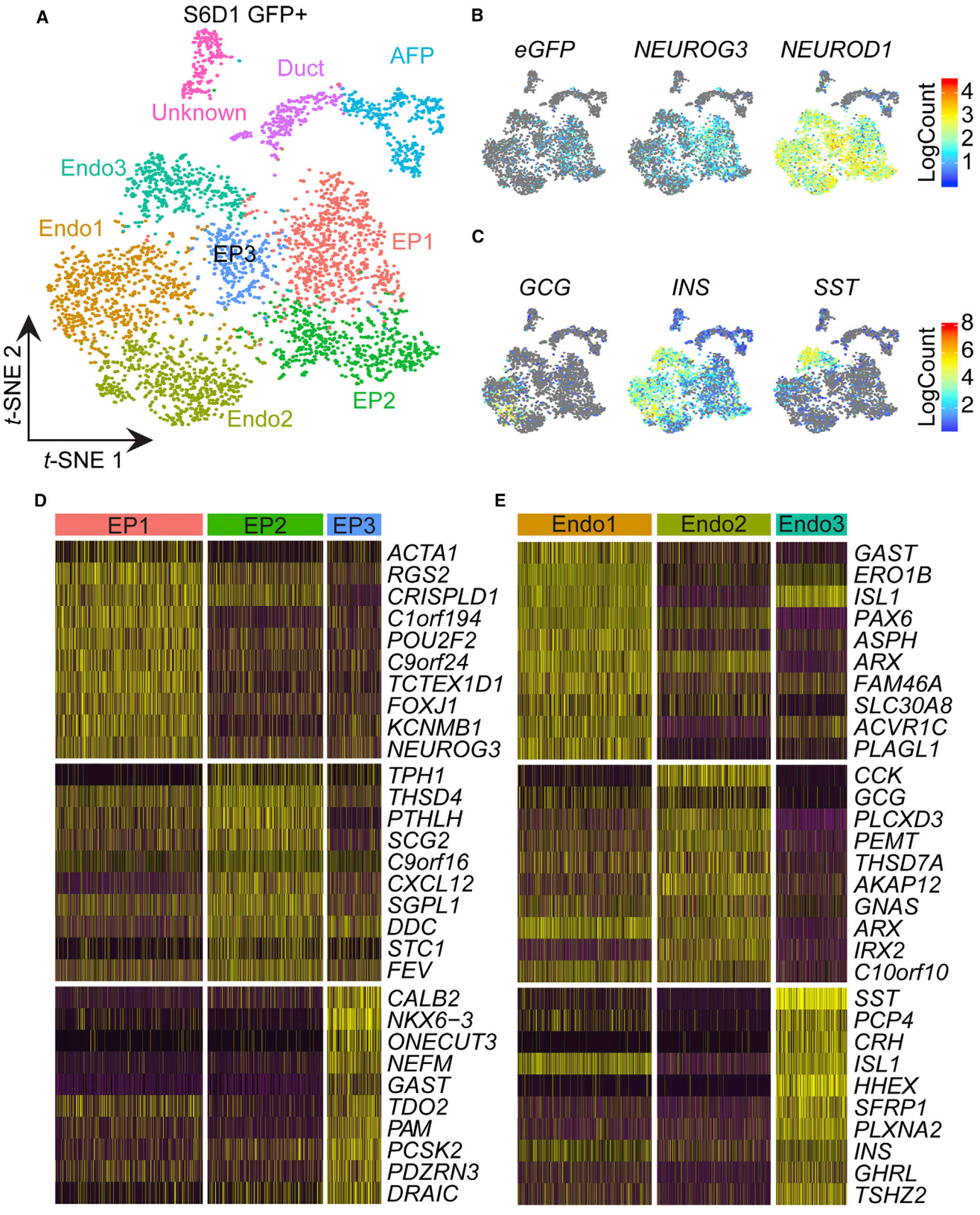
Figure 4. Characterization of Endocrine Cells in E18.5 Green Cells

(A) *t*-SNE plot of 10 cell clusters from E18.5 green cells: endocrine progenitors (EP; 7.0%), α cells (α; 18.2%), two β-cell clusters (β1, 23.4%, and β2, 11.8%), δ cells (δ; 16.7%), Ghrl-cells (ε; 8.0%), pancreatic polypeptide cells (PP; 6.9%), S-phase cells (S; 2.0%), mitotic cells (M; 2.4%), and exocrine cells (3.6%).

(B) Top ten differentially expressed genes in the β cell clusters: β1 (red), β2 (green), M (purple), and S (pink).

(C) Violin plots of average gene expression of *Ins1*, *Ins2*, *Nkx6-1*, *Pdx1*, *Mafb*, *Slc2a2*, *G6pc2*, *Npy*, and *Gadd45a* in β1 (red), β2 (green), M (purple), and S (pink).

(β1, 23.4%, and β2, 11.8%), α cells (18.2%), δ cells (16.7%), ε cells (8.0%), pancreatic polypeptide cells (PP; 6.9%), exocrine cells (3.6%), S-phase cells (2.4%), and mitotic cells (2.0%) (Figure 4A). The cell cycle phase of individual cells in the E18.5 green library was determined. As expected, the S-phase cluster contained both S- and G2/M-phase cells, while the M cluster had G2/M-phase cells (Figure S5A). To confirm the cell identities, the expression of endocrine hormones was determined in single cells. Both *Ins1* and *Ins2* were found in the cells of β1, β2, S, and M clusters, while *Gcg* expression was specific to the α cell population (Figure S5B). The δ cell cluster expressed *Sst*, the ε cells expressed *Ghrl*, and PP cells contained *Ppy* transcripts (Figures S5B and S5C).



(legend on next page)

To understand the heterogeneity within the β cell populations the expression of the top ten genes for β1 (pink), β2 (green), S-phase (pink), and M-phase cells (purple) was determined (Figure 4B). Most cells were in the β1 cluster; expressed genes involved in glucose metabolism, including *Slc2a2* and *G6pc2*; and expressed high levels of *Ins1* and *Ins2* (Figure 4B and S5B). The top ten differentially expressed genes of the β2 cluster included the progenitor markers *Nnat* and *Ppp1r1a*, suggesting a less mature β cell population (Figure 4B). The other *Ins*-expressing cells are located within the S- and M-phase clusters. The S cluster expressed genes specific to the S phase, such as *Cdk1*, suggesting these cells represent a small (2.0%) population of β cells undergoing DNA replication (Figures 4B, S5A, and S5B). In the M cluster (2.4%), the cells expressed the histone genes *Hist1h2bc* and *Hist1h1c*, suggesting that these cells are undergoing mitosis (Figures 4B, S5A, and S5B).

Previous studies in adult β cells suggest proliferation is accompanied by a decrease in the function and maturation of β cells (Puri et al., 2018; Szabat et al., 2016). To understand if a similar process occurs during mouse β cell development, we profiled the expression of several β cell maturity and progenitor markers in the β1, β2, S-phase, and M-phase populations of cells (Figure 4C). The β2 cluster contained a subset of cells with lower *Ins1*, *Ins2*, *G6pc2*, and *Npy*, consistent with a less functional cell state (Figure 4C). In addition, this cluster showed an upregulation of genes expressed in immature progenitor cells: *Nkx6-1*, *Pdx1*, *Mafb*, and *Gadd45a* (Figure 4C). Interestingly, the cells of the S and M cluster exhibit a gene expression profile similar to that of the β1 cluster, suggesting that in the embryonic state proliferation does not reduce maturation (Figure 4C).

Single-Cell Transcriptome of NEUROG3 Lineage during hESC Differentiation

To profile the transcriptome of human EPs, a CyT49 *NEUROG3-2A-eGFP* hESC reporter line (N5-5) was used (Krentz et al., 2017). N5-5 cells were differentiated using the protocol of Rezania et al. (Rezania et al., 2014) and collected for scRNA-seq before the transition to stage 6 (S6D1), when the number of GFP+ cells peaks (Figure S6A). The differentiation that was used for library generation had 23.9% GFP+ cells at S6D1 (Figure S6B). After filtering, 4,462 GFP+ cells were visualized using *t*-SNE, revealing nine clusters (Figure 5A). These clusters can be classified as five cell types based on gene expression: EPs (39.1%), polyhormonal endocrine (Endo; 42.0%), duct (6.0%), liver (8.2%), and an unknown cell type (4.7%) (Figure 5A).

Further examination of the expression of endocrine-specific genes supported these cell classifications. Although all cells expressed GFP protein, only a few cells localized to the EP clusters expressed *eGFP* and *NEUROG3* transcripts (Figure 5B). *NEUROD1*, a direct target of NEUROG3 (Huang et al., 2000), was widely expressed throughout the EP and endocrine clusters (Figure 5B). To understand how the three EP clusters differ, the top ten genes that are specific for each individual cluster were determined. The largest EP cluster (EP1) expressed *NEUROG3*, while EP2 contained genes that are associated with serotonin production, including *TPH1* and *FEV* (Ohta et al., 2011) (Figure 5D). In EP3, the *GAST* gene was upregulated, consistent with previous reports of *GAST* induction in INS+ cells during hESC differentiation (Suissa et al., 2013) (Figure 5D).

The endocrine cell population was made up of hormone+ cells, many of which co-expressed multiple hormones, including *GCG*, *SST*, and *INS* (Figure 5C), consistent with a previous study characterizing hESC-derived endocrine cells using single-cell qPCR (Petersen et al., 2017). At the protein level, most cells were single-hormone positive, but a small proportion of double-hormone-positive cells was detected (Figure S6C). Of the three hormones, *INS* was the most abundantly expressed and can be detected in EP and endocrine cells (Figure 5C). Differential gene expression analysis between the three endocrine clusters revealed an enrichment of β cell genes *ERO1B* (Zito et al., 2010) and *SLC30A8* (Davidson et al., 2014) in Endo1, suggesting these cells were differentiating β cells (Figure 5D). The Endo2 cluster appeared fated toward the α cell based on expression of *GCG*, *PEMT* (Segerstolpe et al., 2016), and *IRX2*, while the expression of *SST* and *HHEX* in Endo3 is suggestive of the δ cell fate (Figure 5D).

Comparison of hESC-Derived Endocrine Cells to Mouse and Human Endocrine Cells

To understand the "developmental age" of hESC-derived endocrine cells, the single-cell transcriptomes of S6D1 GFP+ cells were compared with mouse endocrine cells

Figure 5. Characterization of hESC-Derived Endocrine Cells

(A) *t*-SNE plot of nine cell clusters from *NEUROG3-eGFP* endocrine progenitor cells.

(B) Single-cell gene expression of *eGFP*, and progenitor cell markers *NEUROG3* and *NEUROD1*.

(C) Single-cell gene expression of the hormones *GCG*, *INS*, and *SST*.

(D) Top ten differentially expressed genes in the endocrine progenitor (EP) cell clusters: EP1 (red; 17.5%), EP2 (green; 14.4%), and EP3 (blue; 9.7%).

(E) Top ten differentially expressed genes in endocrine cell clusters: Endo1 (orange; 17.3%), Endo2 (green; 15.0%), and Endo3 (blue; 9.7%).

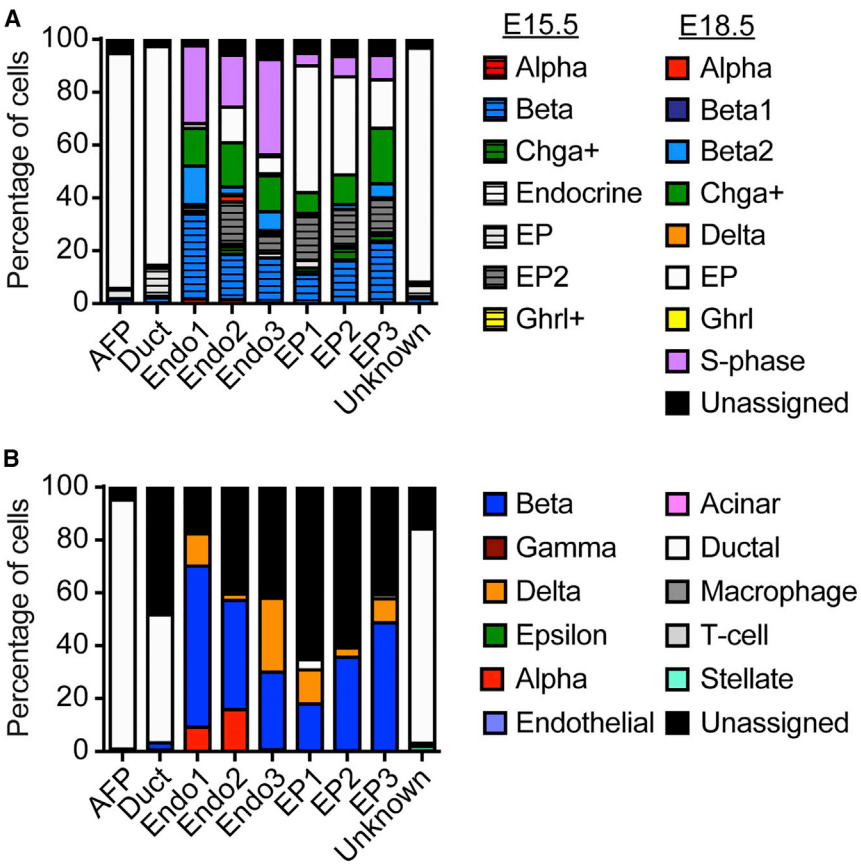


Figure 6. Comparison of hESC-Derived Endocrine Cells with Mouse and Human Endocrine Cells

Proportion of cells (%) within hESC-derived clusters that map to (A) E15.5 or E18.5 mouse endocrine cells or (B) human islet cells

using scmap (Kiselev et al., 2018). This method allows comparison of multiple datasets by projecting a cell onto a reference dataset, inferring cellular identity based on what reference cell type it transcriptionally resembles. First, individual hESC-derived cells were mapped to E15.5 and E18.5 mouse endocrine cell types based on transcriptional similarity. Of the total number of hESC-derived EP cluster cells, 55.2% resembled mouse EPs and 14.9% resembled Endo cells (Figure 6A). Interestingly, 92.9% of all hESC-derived α-fetoprotein (AFP)-expressing cells and 93.7% of all hESC-derived duct cells mapped to mouse EP cells, suggesting that these populations represent an EP-like cell (Figure 6A). Of the endocrine cluster cells, 32.4% mapped with mouse embryonic β cells, suggesting that most hESC-derived endocrine cells were fated to the β cell lineage.

To understand how hESC-derived cells compare with adult human endocrine cells, we next compared hESC-derived cells with human islet cells (Baron et al., 2016). As expected, most cells of the three EP clusters were unassigned, as EPs are found only in embryonic pancreas (Figure 6B). However, the cells that do map within the EP clusters resemble β cells, δ cells, and ductal cells (Figure 6B). Within the Endo clusters, 61.1%, 41.3%, and 29.3% of Endo1, Endo2, and Endo3 mapped to the β cell lineage, respectively (Figure 6B). Consistent with the expression of *GCG* and *SST* (Figure 5C), 9.6% of Endo1 and 16.3% of Endo2 cells resemble α cells, while 27.9% of Endo3 cells map to δ cells (Figure 6B). Interestingly, the three non-endocrine hESC-derived cell types, AFP, duct, and unknown, mapped to ductal cells in the adult dataset (Figure 6B). Taken together, 74.3% of assigned cells mapped to an endocrine lineage, with the vast majority mapping to the human β cells, suggesting the current differentiation protocols generate endocrine cells that share transcriptomic similarity to human β cells (Petersen et al., 2017).

DISCUSSION

scRNA-seq allows for the discovery of cell types and cell-state-specific genes, and the appreciation of cellular heterogeneity within a population. Here, scRNA-seq was used to generate a resource of single-cell transcriptomes from 6,905 E15.5 embryonic pancreatic cells, 6,626 E18.5 embryonic pancreatic cells, and 4,462 hESC-derived *NEUROG3-2A-eGFP* cells. Several unique observations were made, including genes that may regulate endocrine cell formation and previously unidentified populations of

cells generated in hESC differentiations. These data are publicly available online (https://lynnlab.shinyapps.io/embryonic_pancreas) and will serve as a single-cell gene expression resource for mouse embryonic pancreas and hESC-derived endocrine cells.

Comparison of E15.5 and E18.5 EP cells resulted in a list of potential markers, including two genes, *Btg2* and *Gadd45a*, that are involved in neural development. Pro-neural proteins induce expression of *Gadd45g* (de la Calle-Mustienes et al., 2002; Huang et al., 2010), resulting in Gadd45-dependent cell cycle exit by upregulation of the cell cycle inhibitor *Cdkn1a* (Hildesheim and Fornace, 2002) and direct interaction with Cdk1/CyclinB (Zhan et al., 1999). While Gadd45 proteins are implicated in pancreatic cancer (Hildesheim and Fornace, 2002), their role in pancreas development has not been investigated. The expression of *Gadd45a* in Neurog3+ EP cells suggests that it may play a role, along with Cdkn1a, in regulating cell cycle exit.

Btg2, also known as *Tis21*, is a negative regulator of the cell cycle that inhibits transcription of *CyclinD1*, preventing the G1-S transition (Canzoniere et al., 2004; Guardavaccaro, 2000). Deletion of *Btg2* in the adult dentate gyrus shortens G1 length in progenitor cells and prevents their terminal differentiation (Farioli-Vecchioli et al., 2009). This is thought to be caused in part by the direct binding of Btg2 to the *Id3* promoter. Id proteins bind class A basic-helix-loop-helix (bHLH) proteins, which are obligate heterodimerization partners of class B bHLH TFs like Neurog3. By sequestering class A bHLH proteins and preventing their association with pro-neural bHLH TFs, Id acts to prevent terminal differentiation (Lyden et al., 1999). Btg2 may also act to inhibit *Id3* transcription in the pancreas, allowing for the activation of pro-endocrine genes, including *Neurog3*.

The liver, like the pancreas, is derived from the foregut endoderm. The region of the endoderm that gives rise to the liver can also form the ventral pancreas (Tremblay and Zaret, 2005). One of the mechanisms that controls the decision between liver and pancreas is the secretion of fibroblast growth factors by the cardiac mesoderm, which permits the formation of liver while preventing ventral pancreas formation (Zaret et al., 2008). The similar developmental origins of the pancreas and the liver make the unintended generation of liver cells during hESC differentiations toward pancreas likely. However, finding liver cells downstream of *NEUROG3* is surprising. It is possible that, like in the mouse, a small population of hESC-derived pancreatic endoderm cells have low transcription of *NEUROG3* that is not sufficient to induce the endocrine lineage (Bechard et al., 2016). Alternatively, this may also be due to precocious activation of *NEUROG3* during early stages of the differentiation before cells are competent to become endocrine cells (Russ et al., 2015). Using a *NEUROG3* lineage-tracing hESC line, it would be interesting to investigate the plasticity of cells that activate *NEUROG3* transcription. Our previous studies suggest that NEUROG3 protein in hESC differentiations is hyperphosphorylated (N.A.J.K. and F.C.L., unpublished data), resulting in rapid degradation. Efforts to stabilize NEUROG3 protein may prevent the unintended formation of other endodermal cell types, including liver cells.

In conclusion, the single-cell transcriptomes of mouse pancreatic progenitors, EPs, and endocrine cells at E15.5 and E18.5 as well as *NEUROG3*-expressing cells derived from hESCs were characterized. These data are a resource for developmental biologists interested in studying heterogeneity in the developing mouse pancreas and for stem cell researchers aiming to improve current differentiation protocols for generating β-like cells.

EXPERIMENTAL PROCEDURES

Animals

Mice were housed on a 12-hr light-dark cycle in a climate-controlled environment according to protocols approved by the University of British Columbia Animal Care Committee. *Rosa26*$^{mT/mG}$ (Stock No. 007576) (Muzumdar et al., 2007) and *Neurog3-Cre* (Stock No. 005667) mice were purchased from The Jackson Laboratory.

Preparing Cells for Single-Cell RNA Sequencing

For mouse studies, *Neurog3-Cre; Rosa26*$^{mT/mG}$ embryos were collected on E15.5 and E18.5 and dissected on ice. To generate single cells, embryonic pancreases were incubated in 2 mL of pre-warmed 37°C 0.25% trypsin with mild agitation for 8 or 20 min for E15.5 and E18.5 pancreases, respectively. To stop digestion, 1 mL of cold fetal bovine serum (FBS) and 2 mL of cold PBS were added and mixed by inversion, followed by filtering through a 40-μm nylon filter. Cells were then centrifuged at 4°C for 5 min at 200 × *g*. After the supernatant was aspirated, the cells were resuspended in cold 2% FBS in PBS, placed on ice, and immediately sorted into tdTomato+, tdTomato+ eGFP+ (yellow), and eGFP+ fractions using a Beckman Coulter MoFlo Astrios (Mississauga, ON, Canada) in 20% FBS in PBS.

For hESC studies, N5-5 cells were differentiated and collected following 72 hr in stage 5 (S6D1) (see Supplemental Experimental Procedures for details). Cells were washed once with PBS before adding 500 μL of Accutase per well of a 12-well plate. Following 5 min at 37°C, 500 μL of 2% BSA MCDB medium was added to each well and the cells were transferred to a 15-mL conical tube. Cells were centrifuged for 5 min at 200 × *g*, washed once with PBS, and resuspended in 350 μL of ice-cold PBS. GFP+ cells were sorted into stage 5 medium with 10 μM Y-27632 dihydrochloride using a Beckman Coulter MoFlo Astrios. scRNA-seq libraries were generated with 10× Genomics Chromium pipeline (see Supplemental Experimental Procedures for details).

Data Analyses

Following sequencing, data were analyzed using publicly available software programs and R pipelines. First, cellranger mkfastq (10× Genomics) was used to generate FASTQ files from the raw sequencing data. Next, cellranger count aligned FASTQ files to reference mouse (GRCm38) and human genomes (GRCh38) and generated single-cell gene counts. To determine counts for the lineage-tracing transgenes, the mouse and human reference genomes were annotated to include the *tdTomato* and/or *eGFP* sequences. Cellranger aggr was used to combine data from multiple libraries to generate E15.5 aggr (E15.5 red and E15.5 green and yellow), E18.5 aggr (E18.5 red, E18.5 yellow, and E18.5 green), and E18.5 endocrine (E18.5 yellow and E18.5 green), using the normalization parameter set to "mapped" to normalize sequencing depth across libraries.

Two additional R pipelines were used to filter out cells that did not meet the quality control standard. Scater (https://bioconductor.org/packages/release/bioc/html/scater.html) was used to discard cells: (1) with fewer than 1,000 transcripts, (2) with counts (transcripts/cell) greater than 3 median absolute deviation (MAD) away from the median (removing potential doublets and debris), (3) with genes (genes/cell) greater than 3 MAD away from the median (to remove low-abundance genes or genes with high dropout rate), and (4) with mitochondrial DNA transcripts/total transcripts greater than 3 MAD away from the median (McCarthy et al., 2017). The runPCA automatic filtering algorithm was used to filter out genes and cells based on principal-component analysis (PCA). Finally, a gene must be expressed in at least three cells to be included in the downstream analysis.

This quality control dataset was then analyzed using the Seurat v.2.0 pipeline (http://satijalab.org/seurat/), another R toolkit for single-cell transcriptomics (Butler et al., 2018). Seurat was used to remove common sources of variation, including number of genes (each cell must express a minimum of 500 genes), number of counts (each gene must be expressed in a minimum of three cells), and cell cycle phase. The Seurat object was log normalized with a default size factor of 10,000, and data were scaled while regressing out cell-cycle-dependent changes in gene expression. To determine variable genes, mean expression was used with the LogVMR dispersion function and imputed into PCA. The first 15 principal components were used to cluster based on a shared nearest-neighbor modularity optimization-based clustering algorithm with default parameters and resolution of 0.6. *t*-SNE embedding using the first 15 principal components was used to visualize data in reduced dimensions. To identify differentially expressed genes in each cluster, the Seurat function FindAllMarkers was used. For a gene to be differentially expressed in a cluster it must be expressed by at least 10% of cells, have a log-fold change greater than 0.25, and reach statistical significance of an adjusted $p < 0.05$ as determined by the Wilcox test.

Human-Mouse Comparison

scmap was used to analyze the similarity between hESC-derived cells and mouse embryonic pancreas or human islet cells (Kiselev et al., 2018). For the comparison with embryonic mouse cells, endocrine cells from the E15.5 yellow and green and E18.5 yellow and green libraries were used. For the comparison with human islets, four human islet libraries were used (Baron et al., 2016). Scater was used to normalize gene expression by library size and perform a log2 transformation of the data, and libraries were subset to include only the expression of shared genes across the datasets. The top 500 genes for each dataset on the linear model of log(expression) versus log(dropout rate) were selected using selectFeatures. The mean expression, or centroid, of the selected genes was calculated for each cluster in the reference datasets using indexCluster with default parameters. The comparison between a query dataset (hESC-derived) to each reference library was carried out by scmapCluster with a minimum similarity threshold of 0.5.

Pseudotime Analysis

Pseudotime analysis was performed on the post-Scater and Seurat dataset using RNA Velocity (Manno et al., 2018). First, the expression of spliced and unspliced transcripts of each gene was determined. Cells in the bottom 0.5% of the total unspliced transcript count were removed. Next, the following filtering was performed to discard genes: (1) with fewer than 40 total spliced counts across all cells, (2) with fewer than 30 cells expressing the spliced variant, (3) with fewer than 25 total unspliced counts across all cells, (4) with fewer than 20 cells expressing the unspliced variant, or (5) with an average spliced variant expression of less than 0.08 or average unspliced variant expression of 0.01 in at least one cluster. Genes were then ranked using the coefficient of variation versus mean metric, selecting the top 3,000 genes as features. The velocity estimation was performed in Python using the gene relative estimation method. Gamma fit was performed using cell nearest-neighbor pooling with $k = 500$. The resulting velocity estimates were projected onto the *t*-SNE embedding obtained in Seurat.

ACCESSION NUMBERS

Sequencing data are available under GEO Accession GSE120522.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, seven tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.stemcr.2018.11.008.

AUTHOR CONTRIBUTIONS

Conceptualization, N.A.J.K., E.E.X, S.S., and F.C.L.; Methodology and Investigation, N.A.J.K., M.L., E.E.X., S.L.J.S., A.M., S.S., and F.C.L. Writing, N.A.J.K. and F.C.L.; Funding Acquisition, F.C.L.

ACKNOWLEDGMENTS

This work was supported by operating grants to F.C.L.: Stem Cell Network (DT3 and DT4), Canadian Foundation for Innovation (#33644). Salary (F.C.L.) was supported by the Michael Smith Foundation for Health Research (MSFHR; #5238 BIOM), the Canadian Diabetes Association, and the BC Children's Hospital Research Institute. Fellowship support was provided by the CIHR-BC Transplantation Trainee Program, the BC Children's

Hospital Research Institute (N.A.J.K., M.L., E.E.X., S.L.J.S.), University of British Columbia (N.A.J.K., M.L.), the National Science and Engineering Research Council of Canada (N.A.J.K., A.M.), the JDRF (S.S.), and the MSFHR (S.S.). We thank members of the Lynn lab for technical support, discussion, and critical reading of the manuscript. We thank Sara Mostafavi for technical advice. Training on the 10× Genomics platform was provided by Jens Durruthy (10× Genomics).

Received: April 18, 2018
Revised: November 9, 2018
Accepted: November 12, 2018
Published: December 11, 2018

REFERENCES

Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst. *3*, 346–360.e4.

Bechard, M.E., Bankaitis, E.D., Hipkens, S.B., Ustione, A., Piston, D.W., Yang, Y.-P., Magnuson, M.A., and Wright, C.V.E. (2016). Precommitment low-level Neurog3 expression defines a long-lived mitotic endocrine-biased progenitor pool that drives production of endocrine-committed cells. Genes Dev. *30*, 1852–1865.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Byrnes, L.E., Wong, D.M., Subramaniam, M., Meyer, N.P., Gilchrist, C.L., Knox, S.M., Tward, A.D., Ye, C.J., and Sneddon, J.B. (2018). Lineage dynamics of murine pancreatic development at single-cell resolution. Nat. Commun. *9*, 3922.

Cano, D.A., Soria, B., Martin, F., and Rojas, A. (2014). Transcriptional control of mammalian pancreas organogenesis. Cell. Mol. Life Sci. *71*, 2383–2402.

Canzoniere, D., Farioli-Vecchioli, S., Conti, F., Ciotti, M.T., Tata, A.M., Augusti-Tocco, G., Mattei, E., Lakshmana, M.K., Krizhanovsky, V., Reeves, S.A., et al. (2004). Dual control of neurogenesis by PC3 through cell cycle inhibition and induction of Math1. J. Neurosci. *24*, 3355–3369.

Chu, K., and Tsai, M.J. (2005). Neuronatin, a downstream target of BETA2/NeuroD1 in the pancreas, is involved in glucose-mediated insulin secretion. Diabetes *54*, 1064–1073.

Chu, L.F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D.T., Choi, J., Kendziorski, C., Stewart, R., and Thomson, J.A. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. *17*, 173.

Davidson, H.W., Wenzlau, J.M., and O'Brien, R.M. (2014). Zinc transporter 8 (ZnT8) and beta cell function. Trends Endocrinol. Metab. *25*, 415–424.

de la Calle-Mustienes, E., Glavic, A., Modolell, J., and Luis Gomez-Skarmeta, J. (2002). Xiro homeoproteins coordinate cell cycle exit and primary neuron formation by upregulating neuronal-fate repressors and downregulating the cell-cycle inhibitor XGadd45-y. Mech. Dev. *119*, 69–80.

Desgraz, R., and Herrera, P.L. (2009). Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. Development *136*, 3567–3574.

Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., and Quake, S.R. (2017). Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. Cell *171*, 321–330.e14.

Farioli-Vecchioli, S., Saraulli, D., Costanzi, M., Leonardi, L., Cina, I., Micheli, L., Nutini, M., Longone, P., Oh, S.P., Cestari, V., et al. (2009). Impaired terminal differentiation of hippocampal granule neurons and defective contextual memory in PC3/Tis21 knockout mice. PLoS One *4*, e8339.

Georgia, S., and Bhushan, A. (2004). Beta cell replication is the primary mechanism for maintaining postnatal beta cell mass. J. Clin. Invest. *114*, 963–968.

Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development *129*, 2447–2457.

Guardavaccaro, D.C., Corrente, G., Covone, F., Micheli, L., D'Agnano, I., Starace, G., Caruso, M., and Tirone, F. (2000). Arrest of G1-S progression by the p53-inducible gene PC3 is Rb dependent and relies on the inhibition of cyclin D1 transcription. Mol. Cell. Biol. *20*, 1797–1815.

Hildesheim, J., and Fornace, A.J., Jr. (2002). Gadd45a: an elusive yet attractive candidate gene in pancreatic cancer. Clin. Cancer Res. *8*, 2475–2479.

Huang, H.P., Liu, M., El-Hodiri, H.M., Chu, K., Jamrich, M., and Tsai, M.J. (2000). Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. Mol. Cell. Biol. *20*, 3292–3307.

Huang, H.S., Kubish, G.M., Redmond, T.M., Turner, D.L., Thompson, R.C., Murphy, G.G., and Uhler, M.D. (2010). Direct transcriptional induction of Gadd45gamma by Ascl1 during neuronal differentiation. Mol. Cell. Neurosci. *44*, 282–296.

Jensen, J., Heller, R.S., Funder-Nielsen, T., Pedersen, E.E., Lindsell, C., Weinmaster, G., Madsen, O.D., and Serup, P. (2000). Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. Diabetes *49*, 163–176.

Kiselev, V.Y., Yiu, A., and Hemberg, M. (2018). scmap: projection of single-cell RNA-seq data across data sets. Nat. Methods *15*, 359–362.

Krentz, N.A.J., van Hoof, D., Li, Z., Watanabe, A., Tang, M., Nian, C., German, M.S., and Lynn, F.C. (2017). Phosphorylation of NEUROG3 links endocrine differentiation to the cell cycle in pancreatic progenitors. Dev. Cell *41*, 129–142.e6.

Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., Kycia, I., Robson, P., and Stitzel, M.L. (2017). Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res. *27*, 208–222.

Lyden, D., Young, A.Z., Zagzag, D., Yan, W., Gerald, W., O'Reilly, R., Bader, B.L., Hynes, R.O., Zhuang, Y., Manova, K., et al. (1999). Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. Nature *401*, 670–677.

Manno, G.L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lonnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Martens, G.A., Jiang, L., Hellemans, K.H., Stange, G., Heimberg, H., Nielsen, F.C., Sand, O., Van Helden, J., Van Lommel, L., Schuit, F., et al. (2011). Clusters of conserved beta cell marker genes for assessment of beta cell phenotype. PLoS One *6*, e24134.

McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

Miyatsuka, T., Kosaka, Y., Kim, H., and German, M.S. (2011). Neurogenin3 inhibits proliferation in endocrine progenitors by inducing Cdkn1a. Proc. Natl. Acad. Sci. U S A *108*, 185–190.

Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593–605.

Ohta, Y., Kosaka, Y., Kishimoto, N., Wang, J., Smith, S.B., Honig, G., Kim, H., Gasa, R.M., Neubauer, N., Liou, A., et al. (2011). Convergence of the insulin and serotonin programs in the pancreatic beta-cell. Diabetes *60*, 3208–3216.

Petersen, M.B.K., Azad, A., Ingvorsen, C., Hess, K., Hansson, M., Grapin-Botton, A., and Honore, C. (2017). Single-cell gene expression analysis of a human ESC model of pancreatic endocrine development reveals different paths to beta-cell differentiation. Stem Cell Reports *9*, 1246–1261.

Puri, S., Roy, N., Russ, H.A., Leonhardt, L., French, E.K., Roy, R., Bengtsson, H., Scott, D.K., Stewart, A.F., and Hebrok, M. (2018). Replication confers beta cell immaturity. Nat. Commun. *9*, 485.

Qiu, W.L., Zhang, Y.W., Feng, Y., Li, L.C., Yang, L., and Xu, C.R. (2017). Deciphering pancreatic islet beta cell and alpha cell maturation pathways and characteristic features at the single-cell level. Cell Metab. *25*, 1194–1205.e4.

Ramond, C., Beydag-Tasoz, B.S., Azad, A., van de Bunt, M., Petersen, M.B.K., Beer, N.L., Glaser, N., Berthault, C., Gloyn, A.L., Hansson, M., et al. (2018). Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. Development *145*, dev165480.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Russ, H.A., Parent, A.V., Ringler, J.J., Hennings, T.G., Nair, G.G., Shveygert, M., Guo, T., Puri, S., Haataja, L., Cirulli, V., et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J. *34*, 1759–1772.

Scavuzzo, M.A., Hill, M.C., Chmielowiec, J., Yang, D., Teaw, J., Sheng, K., Kong, Y., Bettini, M., Zong, C., Martin, J.F., et al. (2018). Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. Nat. Commun. *9*, 3356.

Segerstolpe, A., Palasantza, A., Eliasson, P., Andersson, E.M., Andreasson, A.C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M.K., et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab. *24*, 593–607.

Stanescu, D.E., Yu, R., Won, K.J., and Stoffers, D.A. (2017). Single cell transcriptomic profiling of mouse pancreatic progenitors. Physiol. Genomics *49*, 105–114.

Suissa, Y., Magenheim, J., Stolovich-Rain, M., Hija, A., Collombat, P., Mansouri, A., Sussel, L., Sosa-Pineda, B., McCracken, K., Wells, J.M., et al. (2013). Gastrin: a distinct fate of neurogenin3 positive progenitor cells in the embryonic pancreas. PLoS One *8*, e70397.

Szabat, M., Page, M.M., Panzhinskiy, E., Skovso, S., Mojibian, M., Fernandez-Tajes, J., Bruin, J.E., Bround, M.J., Lee, J.T., Xu, E.E., et al. (2016). Reduced insulin production relieves endoplasmic reticulum stress and induces beta cell proliferation. Cell Metab. *23*, 179–193.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tremblay, K.D., and Zaret, K.S. (2005). Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. Dev. Biol. *280*, 87–99.

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

Wang, Y.J., Schug, J., Won, K.J., Liu, C., Naji, A., Avrahami, D., Golson, M.L., and Kaestner, K.H. (2016). Single-cell transcriptomics of the human endocrine pancreas. Diabetes *65*, 3028–3038.

Westermark, G.T., and Westermark, P. (2008). Transthyretin and amyloid in the islets of Langerhans in type-2 diabetes. Exp. Diabetes Res. *2008*, 429274.

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

Xu, E.E., Krentz, N.A., Tan, S., Chow, S.Z., Tang, M., Nian, C., and Lynn, F.C. (2015). SOX4 cooperates with neurogenin 3 to regulate endocrine pancreas formation in mouse models. Diabetologia *58*, 1013–1023.

Zaret, K.S.W., Watts, J., Xu, J., Wandzioch, E., Smale, S.T., and Sekiya, T. (2008). Pioneer factors, genetic competence, and inductive signaling: programming liver and pancreas progenitors form the endoderm. Cold Spring Harb. Symp. Quant. Biol. *73*, 119–126.

Zeng, C., Mulas, F., Sui, Y., Guan, T., Miller, N., Tan, Y., Liu, F., Jin, W., Carrano, A.C., Huising, M.O., et al. (2017). Pseudotemporal ordering of single cells reveals metabolic control of postnatal beta cell proliferation. Cell Metab. *25*, 1160–1175 e11.

Zhan, Q., Antinore, M.J., Wang, X.W., Carrier, F., Smith, M.L., Harris, C.C., and Fornace, A.J., Jr. (1999). Association with Cdc2 and inhibition of Cdc2/Cyclin B1 kinase activity by the p53-regulated protein Gadd45. Oncogene *18*, 2892–2900.

Zito, E., Chin, K.T., Blais, J., Harding, H.P., and Ron, D. (2010). ERO1-beta, a pancreas-specific disulfide oxidase, promotes insulin biogenesis and glucose homeostasis. J. Cell Biol. *188*, 821–832.