Title
Single-cell multi-omic roadmap of human and murine pancreas development

Author
de la O, Sean Michael

Publication Date
2023

Single-cell multi-omic roadmap of human and murine pancreas development

by
Sean de la O

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Developmental and Stem Cell Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Todd Nystul
B55FB5CFA606452...

Todd Nystul
Chair

DocuSigned by:
Julie B. Sneddon
DocuSigned by:
Sarah Knox
DocuSigned by:
Faranak Fattahi
07D8B13C88524DE...

Julie B. Sneddon
Sarah Knox
Faranak Fattahi

Committee Members

Dedication

I would like to dedicate this dissertation to my parents, David and Evelyn de la O, and to my sister Tess de la O. To my parents, thank you for giving me the freedom to pursue my passion for science, even when it didn't make sense. To Tess, thank you for being the best sister anyone could ever want, and for being my diabetic buddy. It is my hope that one day that will change.

Acknowledgements

First, I would like to thank my Ph.D. mentors, Dr. Julie Sneddon and Dr. Sarah Knox. Sarah, thank you for teaching me to be bold, fiercely curious and to always keep asking questions. Julie, thank you for your kindness and compassion, and for giving me free reign in the lab even when I was entering into a completely new field and learning skillsets that were completely foreign to me. I could not have done any of this without both of your mentorships, and for that I am eternally grateful.

I would also like to thank my undergraduate mentor, Dr. Jenny Cappuccio. I'm sure it wasn't easy running a lab of undergraduate students with no research experience, but your passion for teaching both in the classroom and in the lab was what really sparked my interest in pursuing a Ph.D. and in pursuing research as a career. Thank you for always being calm and collected, even when it was obvious that I had no idea what I was doing; I will never forget my time in the Cappuccio lab. I would also like to thank Dr. Amy Sprowles for her steadfast belief in the scientific community at Humboldt State University. Without her direction and leadership in the HSU CIRM Bridges program, I would not be where I am today.

Thank you to my fellow Sneddon and Knox lab members, both past and present. They say that it takes a village, and it really does. I would not have been able to accomplish even half of what I did during my Ph.D. without your guidance and assistance. Thank you to all who contributed directly to the work in this thesis and thank you to all for your feedback and for listening to me try to explain some new computational analysis that I wanted to try at every lab meeting.

Thank you to my thesis committee members, Drs. Todd Nystul and Faranak Fattahi. Thank you to Dr. Nystul for serving as my committee chair and lending me an ear when I needed it. Thank you to Dr. Fattahi for always providing expert advice and feedback, especially after my project completely pivoted.

I would also like to thank the DSCB program and the friends that I made in graduate school along the way. I could not have dreamed of meeting such a welcoming group of people when I first joined, and you all made the first few years so much fun. To the upper years when I was a first year, thank you so much for all of the advice about navigating the peaks and valleys of graduate school, especially Ryan Boileau, Bryan Marsh and Ashley Libby. Thank you so much to Eliza Gaylord, the student turned teacher turned student again (and still manages to teach me so much), and Mariko Foecke, our DSCB adoptee, for your friendship, advice, and laughter all these years. Watching you two go from impressive rotation students to the even more stellar scientists you are now has been an absolute pleasure to witness. Ryan Samuel and Alex Duncan, thank you for the friendship, laughter and good times that we've had over the years, and for introducing me to Rosebud and Dee. To the class of 2017, thank you for being the amazing group of human beings that you are. I will never forget all of the memories we made during those first few years of graduate school. I would also like to thank my former roommate, quarantine buddy, and friend Steven Cincotta. Thank you for always lending me your ear when I had a million different scientific questions, for all of the Super Smash Brothers nights, wine, and your support over the years.

I would also like to thank all of my friends from back home, Humboldt, and my FOG rugby brothers who supported me during my PhD. My friends from Humboldt - Angel, Cameron, Andrew, Tim, Bobby and Yvette, and everyone else – thank you so much. My friends from back home – my best friends Chris and Thomas, Jose and Richard, WE DID IT! To my rugby family here in San Francisco – thank you for giving me the space to have fun and do other things outside of the lab. Coach Dany, Pablo, Tina and Ian, you all have made my experience on the SF rugby team amazing. To my teammates – Shawn, Ash, Stephen, Dylan and everyone else, thank you for being great friends these past few years.

I would also like to my wonderful girlfriend Emily and little Georgie girl. We started dating at probably one of my busiest times during my PhD and I'm sure that a lot of times I was

stressed and grumpy, but you always were there to support me and make me laugh. From coming in to lab on the weekend to help when I had a fever, to getting me hooked on your reality TV shows, and of course introducing me to your family in your parents and Georgie, thank you so much for all that you have done for me.

Last but not least, thank you to my family. To my siblings – Paul, Scott and Tess, for all of your support over the years, even when it probably didn't make sense what I was doing. And thank you to my parents, who have supported me all and have given me the freedom to do the things that I wanted in my life. This whole thing would not have been possible without you two, and I am so grateful for what you have done for me.

Contributions

All experiments and analyses for this dissertation were performed in the laboratories of Drs. Julie Sneddon and Sarah Knox at the University of California, San Francisco.

Chapter 2 of this dissertation is published as a manuscript in *Molecular Metabolism* with myself as first author. Co-authors on this manuscript include Xinkai Yao, Sean Chang, Dr. Zhe Liu, and Dr. Sneddon. This manuscript was written by myself and Dr. Sneddon. Xinkai Yao and Dr. Liu performed FACS on murine pancreas for single-cell sequencing. Sean Chang performed *in situ* hybridization for *in vivo* validation, and Dr. Liu performed qPCR for *Fev* mRNA expression. I performed all bioinformatics analyses.

Chapter 3 of this dissertation is unpublished. Dr. Liu and Dr. Daniel Wong prepared human fetal pancreata for single-cell sequencing. I performed the bulk of the computational analysis, with the assistance of Dr. Liu. Dr. Liu, Emily Chu, Sean Chang, and Peyton McDonnell performed immunostaining and *in situ* hybridization for single-cell validation. Drs. Kevin Yu and Aaron Tward assisted with computational analyses related to CellFindR results. Niko Eng and Dr. Yin Shen aided in snATAC-sequencing analysis. Drs. Anna Gloyn, Han Sun, and Seth Sharp performed type 1 and type 2 diabetes GWAS analyses. Dr. Wong generated the CRISPR/Cas9 edited FEV-KO hESC line with the assistance of Gabriel Peixoto. Sneha Rao, Dr. Liu, Emily Chu, Jacquelyn Bouza, and I performed directed differentiations of hESCs into beta-like cells. Dr. Liu, Dr. Wong, Emily Chu, and Sneha Rao performed flow cytometry analysis. I performed all remaining experiments and analyses in Chapter 2 and Chapter 3 and wrote this dissertation with input from Dr. Sneddon and my dissertation committee.

Single-cell multi-omic roadmap of human and murine pancreas development

Sean de la O

Abstract

The pancreas is a branched, lobular organ consisting of multiple cell types in two separate tissue compartments that work in concert to maintain proper food digestion and glucose homeostasis. The exocrine compartment synthesizes and secretes digestive enzymes that are shuttled to the intestine, while the endocrine compartment produces various hormones that enter the bloodstream and modulate systemic blood glucose levels. In this work, we sought to understand the diversity of cell types found in the human and mouse pancreas during fetal development. We start in Chapter 2 by performing single-nucleus Assay for Transposable Chromatin Sequencing (snATAC-Seq) on embryonic murine pancreas and construct gene regulatory networks of the developing endocrine and mesenchymal compartments. We uncover candidate TF regulators and downstream target genes of cell fate decisions across developmental time and confirm the expression of identified transcription factors. In Chapter 3, we identify novel and known cell types within the developing human fetal pancreas and their underlying transcriptomic profiles by utilizing single-cell RNA-Sequencing. We undercover significant cellular diversity in all cell types of the human fetal pancreas (endocrine, exocrine, mesenchymal, immune, neuronal and endothelial) and map the cell-cell interactions between these cell types. In the endocrine compartment, we identify novel endocrine progenitor cell types with varying differentiation potency and reconstruct their differentiation trajectories *in silico*. We then combine our transcriptional knowledge of human fetal pancreas development with chromatin landscape data through snATAC-Seq, allowing for a multi-omic analysis to construct gene regulatory networks and identification potential regulators of cell fate in the human fetal endocrine pancreas. With the chromatin landscape information from our snATAC-Seq data, we also identify development-enriched single nucleotide polymorphisms from genome

wide association studies of type 1 and type 2 diabetes. Additionally, we elucidate the cellular composition of *in vitro* stem cell-derived endocrine cells, identifying the transcriptional control of mis-differentiated populations. Lastly, we identify the transcription factor *FEV* as a regulator of beta cell differentiation, confirming *in silico* predictions. Together, we describe a novel in-depth transcriptional and epigenomic understanding of human and mouse pancreas cells, providing a rich database for the field and expanding upon our understanding of pancreas development.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1 Diabetes mellitus

Diabetes mellitus is a disease of the endocrine pancreas that leads to an impairment of glucose regulation within the affected individual. The most common forms of diabetes include type 1 and type 2 (T1D and T2D), representing the bulk of 537 million adults (aged 20-79 years old) worldwide affected in 2021 (The International Diabetes Federation; https://idf.org). In T1D and T2D the function of the endocrine beta cell, found within the Islets of Langerhans, is either lost or impaired. Pancreatic beta cells secrete insulin, a hormone that is essential for signaling to many cells throughout the body to transport glucose from the blood into the cells of the body to be utilized for energy production (1). Without insulin, glucose accumulates in the blood of the affected individual, leading to many complications such as neuropathy, retinopathy, nephropathy and, ultimately, death if not managed properly (2). T1D is a progressive autoimmune disease in which the immune system destroys the patient's own pancreatic beta cells, removing the sole source of endogenous insulin (3). T2D is commonly characterized by insulin resistance, in which the individual's peripheral tissues do not properly respond to insulin signaling, leading to blood glucose accumulation. This, in turn, increases the body's demand for insulin production from the pancreatic beta cells, leading to metabolic exhaustion of the beta cells which can no longer meet the body's demand for insulin (4). Current treatments for T1D and severe cases of T2D include constant monitoring of blood glucose levels and the administration of exogenous insulin to maintain euglycemia. This treatment, however, is far from perfect; only a fraction of patients achieve the target range of blood glucose levels long-term (5) and the constant monitoring of blood glucose, carbohydrate intake, exercise, etc. leads to a significant mental burden on the affected individual. Indeed, most, if not all, individuals develop diabetic burnout at some point (6), characterized by participation of self-destructive behaviors such as disregarding blood sugar levels and improperly estimating the required insulin intake for a meal instead of accurately counting the carbohydrate intake and calculating insulin

requirement. Diabetic burnout can therefore lead to hypo- and hyper-glycemic episodes and long-term complications (6). Thus, there is a significant need for new treatments and cures to alleviate the daily burdens of T1D and T2D.

1.2 Curative measures for diabetes mellitus

Outside of manual injections of insulin with each carbohydrate intake, there has been significant progress made in the production of so-called closed-loop insulin pumps, which regulate the administration of insulin to the individual by the crosstalk between the pump and a continuous glucose monitor, leading to clinically relevant improvements in glucose management compared to manual administration of insulin (7,8) These systems, however, can represent a significant financial burden on the affected individual, as well as interfere with lifestyle choices of affected patients, such as navigating the presence of tubing from insulin pumps. While these insulin pumps represent a significant step forward in regulating blood glucose levels in T1D patients, they still do not regulate blood glucose with the same precision that *bona fide* islet cells do, as their insulin delivery algorithms can still miscalculate the required amount of insulin and lead to hyper- and hypo-glycemic events.

Treatments not involving exogenous insulin involve the replacement of beta cells, either in the form of cadaveric islets or whole-pancreas transplantation. This cellular replacement therapy is typically performed on patients with “brittle” diabetes, in which the patient suffers from severe swings in in blood glucose levels and hypoglycemia unawareness (9). The first instance of successful insulin independence was achieved in 1966, in which whole pancreatic tissue was transplanted into two patients with T1D (10). The insulin independence was short-lived, however; one patient had evidence of insulin secretion for only six days, while the other was free from exogenous insulin therapy for four months. Additionally, these surgeries involved extensive transplantation of pancreatic, kidney and intestinal tissue. Therefore, while whole

pancreatic transplant became a viable option for curative measures for diabetes, there was still room for other improved transplantation methodologies.

Improved techniques for islet isolation (11,12), as well as the creation of more potent immunosuppressive agents, lead to the first successful islet transplantation study in 2000, achieved with the so-called Edmonton protocol (13). In this study, seven recipients with T1D underwent transplantation of isolated islets from non-diabetic cadaveric donors and were subsequently insulin independent. Since this landmark study, thousands of patients have undergone islet transplantations; current data from the Collaborative Islet Transplant Registry (CITR) report that, of all patients in their registry that received islets transplant alone (n = 1,108), 19.6% are insulin independent 0-24 years post first transplant, while 29.0% are still insulin dependent and 19.7% have no graft function (Collaborative Islet Transplant Registry; https://citregistry.org). Additionally, cadaveric islet transplantation was recently approved in the US under the name *Lantrida* (https://www.fda.gov/vaccines-blood-biologics/lantidra). These results point toward islet transplantation being a viable, albeit not perfect, option for effectively curing diabetes. One pressing issue of for islet transplantation, however, is the lack of adequate cellular mass for transplantation, as the demand for transplantable cadaveric islets far outweighs the donor supply. In the original Edmonton study, large islet mass (>11,000 islet equivalency (IEQ)/kg of body weight) was transplanted into the patients, equating to an average of 2-3 donors per recipient. The requirement of a large quantity of islet mass for infusion can be attributed to the loss of up to 60% of transplanted islet mass in the first few days following transplantation (14). Thus, a renewable cellular source of beta cells is needed, as well as improvement of transplant survival in the first few days following transplantation.

Significant research has been devoted to investigating methods for production of *de novo* pancreatic beta cells as a curative measure for T1D and T2D. The epithelial fraction of the adult pancreas consists of two distinct compartments that carry out different functions. The exocrine compartment, comprised of acinar and ductal cells, secretes and shuttles digestive

enzymes to the duodenum to aid in food digestion. The endocrine compartment, comprised of glucagon-expressing alpha cells, insulin-expressing beta cells, somatostatin-expressing delta cells, pancreatic polypeptide-expressing PP cells, and ghrelin-expressing epsilon cells, maintains blood glucose homeostasis. Recent research has investigated the possibility of transdifferentiating, or reprogramming, these non-beta epithelial cells into something resembling *bona fide* beta cells to replace lost beta cell mass. One study achieved viral-mediated over-expression of key transcription factors *Ngn3, Pdx1,* and *Mafa* in the pancreatic acinar tissue of adult mice, which reprogrammed these acinar cells into insulin-expressing beta-like cells (15). Numerous studies have also induced insulin expression in liver and intestinal tissue through virus-mediated expression of *Ngn3*, *Pdx1,* and *Mafa* (16,17). Of note, these studies resulted in only partial rescue of a diabetic phenotype, decreasing their possible efficacy of potential translatability in human diabetic patients. Another avenue to achieve trans-differentiation to beta-like cells would be to reprogram other endocrine cells such as pancreatic alpha cells. Alpha cells arise from the same endocrine progenitor cell as beta cells in mouse development and share many physiological and morphological features of beta cells, such as anatomical location in the pancreas, blood supply, glucose sensing, and hormone secretion machinery (18). Successful trans-differentiation of isolated human adult alpha cells has been shown through the overexpression of mouse *Pdx1, Mafa,* and *Nkx6.1* (19). These reprogrammed alpha cells secreted human insulin in presence of glucose, while curiously maintaining the expression of alpha cell specific genes *ARX* and *GCG* and rescued a diabetic phenotype in mouse models after transplantation. While trans-differentiation of human cells into pancreatic beta cells represents an attractive approach for *de novo* generation of beta cells, there are still issues that will need to be overcome for this approach to become a clinically viable option, with the most prominent likely being getting adequate reprogramming, and therefore beta cell mass, to reduce insulin dependence. Additional issues include general risks of viral-mediated gene therapies, such as virus-induced mutagenesis and oncogene activation (20).

Another promising approach for the *de novo* generation of beta cells is the differentiation of human pluripotent stem cells (hPSCs) towards the beta cell fate. hPSCs retain the ability to differentiate into any cell type within the body and therefore represent a near-unlimited source for the generation of pancreatic beta cells *in vitro*. Several directed differentiation protocols have been devised to generate insulin-secreted beta-like cells from hPSCs using stepwise differentiation platforms that aim to mimic endogenous development by recapitulating key cell stages through the carefully timed addition and withdrawal of defined combinations of signaling factors (21–25). These protocols, however, suffer from the production of non-endocrine cell types and a failure to match the transcriptional profiles and glucose responsiveness of primary adult human beta cells. This may be due, in part, to a relative lack of understanding about human endocrine development. Current protocols are based on knowledge of rodent development and therefore may be missing key regulatory pathways and lineage steps that are unique to human development. Gaining a deeper understanding of human endocrine development is therefore crucial for continued progress towards generating *in vitro*-derived beta-like cells that recapitulate *in vivo* development and functionality.

1.3 Murine pancreas epithelial development

Mouse pancreas development is broadly categorized into two “waves”, or transitions, during embryogenesis. The first wave begins at embryonic day (E) 8.5, when the gut tube endoderm and surrounding mesoderm interact with one another and induce the expression of *Pancreatic and duodenal homeobox 1* (*Pdx1*) at the mid-foregut junction (26–28). These *Pdx1*-expressing cells rapidly expand to form the dorsal pancreatic bud, which evaginates and pushes into the surrounding mesenchyme. *Pdx1* is required for this process, as lineage tracing (29) and gene ablation studies (30,31) have shown that it is essential for all downstream pancreatic lineages. The early pancreatic epithelium is comprised of multipotent pancreatic progenitor cells (MPCs) that eventually give rise to all exocrine and endocrine cells (32). Early differentiated

endocrine cells, consisting mainly of glucagon-expressing alpha or polyhormonal cells, are also present in the early pancreatic epithelium (33). These MPCs express *Pdx1* as well as other homeobox proteins such as *NK6 Homeobox 1* (*Nkx6-1*) and *NK2 Homeobox 2* (*Nkx2-2*) (34), while the early endocrine cells are derived from early *Neurogenin 3* (*Ngn3*)-expressing endocrine progenitors (35). At E11, microlumen formation occurs throughout the developing epithelium, which begin to coalesce into larger lumen structures at E11.5 to begin the initiation of branching morphogenesis (36,37), forming the branched epithelial plexus that will eventually give rise to the ductal network. This epithelial rearrangement leads to the compartmentalization and differentiation of MPCs into multipotent “tip” and bipotent “trunk” domains at E12. Tip cells are marked by the expression of *carboxypeptidase A1* (*Cpa1*) and *Pancreas Associated Transcription Factor 1a* (*Ptf1a*) and serve as progenitors for acinar, ductal, and endocrine cells in the pancreas (34). The trunk cells express *SRY-Box 9* (*Sox9*), *Nkx6-1*, and *HNF1 Homeobox B* (*HNF1β*) (38–40) and can give rise to ductal and endocrine cells (34).

The secondary wave of pancreas development, which begins at E13.5 and lasts until birth, is characterized by the differentiation of the progenitor pools into the three main cell types found in the adult pancreas epithelium: acinar, ductal, and endocrine cells. Tip cells begin to differentiate into protoacinar cells that not only express *Cpa1*, but also digestive enzyme genes such as amylase, elastase, and trypsinogen (41). These acinar cells further mature in late embryogenesis and are maintained through acinar proliferation in a subpopulation of adult acinar cells (42,43). Bipotent trunk cells differentiate to either mature ductal cells or *Ngn3*-expressing endocrine progenitor (EP) cells. This process is thought to be largely mediated by Notch-mediated lateral inhibition within the trunk network, in which Notch expression in trunk cells induces the expression of *Sox9*, which then directly activates the expression of *Ngn3* in a cell-autonomous manner (44). At high levels of Notch signaling, however, expression of the *Ngn3* repressor *Hes1* blocks endocrine differentiation. *Ngn3*-expressing EP cells themselves present Notch ligands to surrounding trunk cells and therefore repress the induction of an

endocrine cell fate in neighboring cells, resulting in a mosaic expression pattern of *Ngn3* and *Hes1* during development (45,46). These *Ngn3*-expressing EPs then give rise to the different subtypes of pancreatic endocrine cells in a process controlled by a complex gene regulatory network of activating and repressing transcription factors.

1.4 Heterogeneity within murine pancreatic endocrine progenitor cells

Induction of endocrine cell fate begins with the expression in bipotent progenitors within the developing trunk of *Ngn3* (29,47–50), whose expression is essential for the formation of all differentiated endocrine cell types (47,49,50). EP cells are thought to be unipotent and largely post-mitotic, with each EP cell giving rise to only one differentiated endocrine cell type (51). Although we have some understanding of the transcriptional machinery needed to differentiate an EP to a hormone-expressing endocrine cell, we do not yet fully understand the extrinsic and intrinsic factors that influence this important step in cell fate selection. *Ngn3*-expressing EPs may display differential competence across development time; in an over-expression system EPs derived from E8.5 embryonic pancreas predominantly produced *Gcg*+ alpha cells, while EPs acquired competence to make insulin-producing beta cells and pancreatic polypeptide (PP) cells at E11.5 and E12.5 (52). At E14.5 EPs become competent to produce somatostatin-producing delta cells, and drastically increased the propensity to produce beta cells. Recent work has described the differences in chromatin accessibility of mid- and late-stage EP cells, showing an increase in motif enrichment of beta cell-related transcription factors at E16.5 compared to E14.5 (53), indicating that late stage EPs have an increased capacity to bind key beta cell specific transcription factors and therefore may preferentially differentiate into beta cells.

With the advent of single-cell RNA-Sequencing (scRNA-Seq), a technique that allows the transcriptional profiling of hundreds to thousands of cells at single-cell resolution (54–56), our understanding of the heterogeneity among pancreatic EP and hormone-expressing cells has

increased dramatically (57). The application of scRNA-Seq to the developing pancreas by our laboratory and others has led to the discovery of multiple stages of EP populations that display developmental-time specific markers (53,58–62). These studies have described the emergence of “pre-alpha-first-cells” that arise at E9.5, maturing into first wave alpha cells by E11.5 – E13.5 (62) that display a specific gene signature that is different from second-wave alpha cells (60,62). Interestingly, multiple studies have also described heterogeneity in the EP pool beyond just *Ngn3*-expressing progenitor cells. Later progenitor EPs are marked by the expression of the transcription factor *Fev* (58), also known as *Pet1*, which was initially discovered as a regulator of serotonergic neuron cell fate in the developing brain (63). These Fev+ EPs express endocrine lineage genes such as *Chga, Isl1, Irx2*, and *Mafb*, indicating that they are downstream of *Ngn3*+ EPs but upstream of differentiated hormone-expressing cells, an observation that has been corroborated with genetic lineage tracing (58) and computational reconstruction of endocrine lineages *in silico* (53,58–62).

1.5 Transcriptional control of endocrine differentiation

Endocrine differentiation is a complex process requiring the temporal expression of key transcription factors which dictate the hormone-expressing endocrine cell into which each EP will differentiate. As stated above, the endocrine lineage is highly dependent on the expression of the pro-endocrine transcription factor *Ngn3*. Mice lacking Ngn3 display intact acinar and ductal compartments but succumb to diabetes shortly after birth (47), while overexpression of *Ngn3* induces endocrine differentiation in mouse embryonic stem cell-derived endodermal cells (64) and *xenopus* endoderm *in vivo* (65). *Ngn3* is detectable by E9.5, but virtually absent in adulthood (66). Direct targets of *Ngn3* include many transcription factors critical to endocrine development, including *BETA2/NeuroD* (67), *Paired Box 4* (*Pax4*) (68), and *Ngn3* itself (69). Other transcription factors expressed by EPs include *Regulatory Factor X3* (*Rfx3*) (70) and (*Rfx6*) (71,72), which, through scRNA-Seq studies, we now know are expressed by the Fev+

EPs downstream of Ngn3+ cells (58,60). *Pax4* and *Pax6* are also expressed by EPs and are essential for endocrine differentiation, with *Pax4* regulating beta and delta cell differentiation and *Pax6* being required for alpha and beta cells (73,74).

1.6 Human pancreas development

Human pancreas development is much less understood relative to murine development. Evidence of pancreatic tissue is first observed at Carnegie Stage 12 (CS12; ~29-31 days post conception (dpc)), where PDX1 can first be observed in the presumed pancreatic endoderm (75). These PDX1-expressing cells then begin to segregate into dorsal and ventral pancreatic buds by CS13 (~30-33 dpc), comprised of multipotent progenitors that rapidly expand and begin to form microlumens (76). Pancreatic SOX9+/GATA4+NKX6.1+ tip cells and SOX9+/NKX6.1+ trunk cells are present by CS19 (~45-47 dpc) (76). The onset of NEUROG3-expressing endocrine progenitors and endocrine commitment is observed at CS21 (~52-55 dpc) (77,78); interestingly, there is no evidence of the “primary” and “secondary” transitions observed in murine development (76).

Unlike in murine development, INS-expressing beta cells are the first predominant endocrine cell type observed in conjunction with NEUROG3 (79). The other hormone-expressing cells arise soon thereafter, and proto-islets can be observed by 10 weeks post conception (wpc). The architecture of adult islets is markedly different in humans than in mice; murine islets have a stereotypical alpha cell mantle and beta cell core, while human islet architecture is more interspersed and random (80). The cellular composition of mouse and human islets is also different, where beta cells make up 60-80% of the endocrine mass in the mouse pancreas, and 50-70% in humans (80). Human islets contain more alpha cells than mouse (20-40% vs. 10-20%, respectively), and the remaining endocrine cell types (delta, epsilon and PP cells) make up the remaining ~10% in humans and less than 5% in mouse islets. Whether these architecture and composition differences point to disparities in the

formation of the islets during development, however, is not well understood. NEUROG3 expression peaks during the first trimester and is largely absent by 35 wpc. It is likely that any increase in endocrine cell mass after this period is likely due to endocrine cell proliferation rather than differentiation of progenitors to hormone-expressing cells.

1.7 *In vitro* differentiation of human stem cells towards pancreatic endocrine cells

Over the past two decades, there has been great interest in the generation of functional human beta cells for cell-based therapies for T1D and T2D. The first report of human *in vitro* stem cell-derived beta-like cells occurred in 2001, when cells that secrete insulin in response to glucose were reported to arise spontaneously in human embryonic stem cell (hESC)-derived embryoid bodies (EBs) after the removal of the mouse embryonic fibroblast (MEF) feeder layer (81). While this study showed that *in vitro* generation of beta cells was indeed possible, there was still a need for controlled differentiation of beta cells to reach the efficiency needed for a cellular therapy for diabetes. The next advancement towards generating *in vitro*-derived beta cells was the generation of stem cell-derived pancreatic progenitor cells that had the capability of maturing into hormone expressing endocrine cells following transplantation into mice (82–84). These directed differentiation protocols were based on knowledge from classical mouse pancreas developmental studies to mimic Mother Nature, recapitulating the different cell types through which *in vivo* pancreatic cells must transition to reach the pancreatic progenitor stage. These protocols began with the induction of *FOXA2+/SOX17+* definitive endoderm lineage from hESCs through the addition of Activin A to culture media. Next, pancreatic progenitors were formed through the addition of retinoic acid, FGFs and hedgehog signaling inhibitors (84). These early differentiation protocols were carried out in two dimensional (2D) culture and in many studies the resulting cells were able to reverse diabetes in mice post-transplantation. These studies were followed by two landmark papers in which human embryonic stem cells were differentiated in either large-scale, three dimensional (3D) format (22) or in 2D and air

liquid interface (ALI) format (24) to generate glucose-responsive, monohormonal insulin-expressing cells that also expressed many of the key beta cell markers, as well as ultrastructural characteristics similar to primary adult beta cells such as mature insulin granules. These protocols utilized Alk5 inhibition, Triiodothyronine (T3) and gamma secretase inhibitor (XX or XXI) supplemented at the CHGA+ endocrine stages of the directed differentiations to achieve a more mature beta cell phenotype.

Since these early papers, many different directed differentiation protocols have been established with varying formats, uses, and efficacy in terms of functionality and ability to rescue diabetes in mice. These newer directed differentiation protocols modulate different aspects of beta cell differentiation, including reaggregating sorted cells (23,85), changes in culture methods, growth factors, small molecules, and cell lines used for differentiation (23,86,87). Additionally, in recent years studies have begun to focus on the generation of other pancreatic endocrine cell types such as alpha cells (88,89) in order to develop stem cell-derived pseudo-islets of defined proportions of the various endocrine cells in addition to beta cells.

While the field has made great strides in our ability to generate functional beta-like cells in past two decades, hurdles remain before these cells display all of the characteristics of *bona fide* adult beta cells. One such hurdle is the cellular mass needed per patient for a curative dose. Although millions of stem cells can be used as material for these directed differentiations, only about 30% of the cells in the end product resemble *bona fide* beta cells based on the co-expression of *NKX6.1*, a key transcription factor for *in vivo* beta cells, and insulin (22,23,87). The remaining end-stage cells are a mix of other endocrine cell types (i.e. GCG+ alpha cells, SST+ delta cells, polyhormonal cells, etc.), mis-differentiated cells that resemble endocrine cells of other tissues (23), or non-endocrine cells. Enrichment strategies have led to an increase in the percentage of beta-like cells (23,85) but are costly and labor intensive. Another hurdle for creating a viable cellular therapy from *in vitro*-derived beta cells is the functional immaturity that they exhibit when compared to endogenous adult beta cells. Although the beta-like cells do

exhibit insulin secretion when exposed to exogenous glucose during glucose-stimulated insulin secretion (GSIS) assays, they lack the stereotypical biphasic insulin secretion profile typically seen in adult beta cells, where readily made insulin granules are released during the first phase of insulin secretion and newly made insulin granules are produced and released during the second phase (90). Conversely, *in vitro* beta-like cells exhibit functional maturation after transplantation into rodents, so whether the functional maturation *in vitro* is “good enough” for *in vivo* transplantation remains to be seen. These two hurdles, namely, inefficient generation of pancreatic beta cells and functional immaturity, likely are multi-factorial and involve different contexts such as niche environment, transcriptional control, media conditions, among others.

1.8 Contribution to the field

This work elucidates cellular heterogeneity, transcriptional regulation, and chromatin accessibility of both human and mouse pancreas development. In the developing human pancreas, we use scRNA-Seq to investigate the heterogeneity of the different major cell types within the pancreas, uncovering more than 100 distinct cell types. We describe the transcriptional expression and cell-cell signaling of these cell types, with a focus on the mesenchymal and endothelial cells. Next, we focus on the endocrine cells of the human fetal pancreas, identifying novel heterogeneity in the endocrine progenitor pool and quantifying the proportional changes of this heterogenous population *in vivo*. With lineage reconstruction, we construct a map of human endocrine development and describe the transcriptional changes across differentiation. Furthermore, by integrating both scRNA-Seq and single-nucleus ATAC-Seq (snATAC-Seq), we construct a gene regulatory network (GRN) within endocrine cells and identify candidate TF regulators of human endocrine development and their downstream target chromatin regulatory regions and target genes. Lastly, we use our multi-omic datasets to understand the enrichment of T1D and T2D genome-wide association studies (GWAS) single nucleotide polymorphisms (SNPs) in fetal endocrine cells, compared to adult endocrine cells.

Together, our multi-omic analysis of human endocrine development will serve as a rich resource for further understanding not only endocrine development, but also the development of non-endocrine cell types.

Using our human fetal endocrine scRNA-Seq dataset as a roadmap, we also describe the similarities and differences of *in vivo* development to that of *in vitro* beta like cell development. This comparison describes the relatively similar lineage path taken by endocrine progenitors *in vitro* to that of *in vivo* development, except for the presence of so-called “stem cell derived enterochromaffin cells (sc-ECs)”. By comparing these sc-EC cells to both *in vivo* pancreas and intestinal endocrine development, we conclude that these sc-EC cells do in fact more resemble their *in vivo* intestinal counterpart and not a pancreatic pre-beta population as was previously described (91). Lastly, we use a FEV knockout (FEV KO) hESC line to query the functional role of the transcription factor FEV in governing beta cell development *in vitro*, showing a marked decrease in beta-like cells compared to wildtype (FEV WT) cells and confirming that FEV indeed plays a role in lineage allocation of beta cells.

In the developing murine pancreas, we use scRNA-Seq and snATAC-Seq to describe the multi-omic landscape of pancreatic epithelial and mesenchymal cells, constructing gene regulatory networks across developmental time and performing comparative analyses. These analyses identify both known and novel regulators of endocrine and mesenchymal cell fate, some of which we confirm *in vivo*. Our multi-omic profiling will serve as a rich resource for further understanding the emergence of cellular diversity and maturation that occurs across development time in both endocrine and mesenchymal cells of the developing murine pancreas.

Chapter 2: Single-cell chromatin accessibility of developing murine pancreas identifies cell state-specific gene regulatory programs

2.1 Introduction

Development of the mammalian pancreas requires the coordination of multiple cell lineages over time, culminating in the generation of a highly branched, mature organ consisting of both an exocrine and endocrine compartment. Specification of the murine pancreas begins at embryonic day (E) 8.5 with the expression of the transcription factor (TF) pancreatic and duodenal homeobox 1 (*Pdx1*) in a focal region of the endoderm-derived primitive foregut (92,93). These Pdx1(+) cells give rise to all of the epithelial lineages of the pancreas (duct, endocrine, and acinar) (29) and by E9 evaginate into the surrounding mesenchyme and begin to form a stratified epithelium. As branching morphogenesis progresses, regionalization of the epithelium results in the formation of both trunk and tip domains by E12.5. Cells located at the tip, marked by the expression of *Cpa1*, serve as multipotent progenitors that give rise to all three epithelial cell types until E13.5, at which point they undergo fate restriction to only give rise to acinar cells (29,34,44). Epithelial cells located in the trunk give rise to either ductal or endocrine lineages, a fate choice dependent on levels of Notch signaling (44).

Endocrine progenitor (EP) cells derive from a subset of ductal epithelial cells that experience lower levels of Notch and then activate expression of the TF neurogenin3 (*Neurog3*) (44). *Neurog3* expression marks early EP cells, which give rise to the main hormone-producing endocrine cells in the pancreas: alpha, beta, delta, and gamma (29,34,44,47). Gene knockout studies in mice revealed that the expression of a number of TFs that are critical for differentiation and maintenance of pancreatic endocrine lineages, such as paired box gene 4 (*Pax4*) and 6 (*Pax6*), neurogenic differentiation 1 (*Neurod1*), and LIM-homeodomain protein Islet 1 (*Isl1*), is dependent on *Neurog3* (47).

Endocrine cell identity is specified and maintained by a complex network of TFs, many of which play dynamic roles throughout developmental time (94). For instance, early in development *Pdx1* is required for specification of pancreatic progenitors, but later in

development it is also important for the generation of beta cells and for the maintenance of beta cell identity (95,96). Along with *Pdx1* and *Pax4*, the TFs NK2 homeobox 2 (*Nkx2-2*) and NK6 homeobox 1 (*Nkx6-1*) are critical factors for beta cell differentiation, while aristaless related homeobox (*Arx*) is essential for alpha cell differentiation. *Arx* and *Pax4* play mutually opposing roles in the differentiation of alpha and beta cells, with *Arx* promoting the generation of alpha at the expense of beta and delta cells (97) and *Pax4* regulating the decision towards beta and delta at the expense of alpha and epsilon cell fate (97–99). When both *Arx* and *Pax4* are lost, delta cells persist but both alpha and beta cells are lost (98). Expression of *Nkx2-2* and *Nkx6-1* follows that of *Pdx1* in early pancreatic progenitors, then becomes progressively restricted to endocrine cells (100,101). Deletion of *Nkx2-2* results in a significant reduction of the four major endocrine cell types and an increase in ghrelin-producing epsilon cells (99–101). *Nkx6-1* functions downstream of *Nkx2-2* and is necessary for beta cell neogenesis through the maintenance and/or expansion of beta cell precursors following *Neurog3* expression but prior to the production of insulin, while later it is lost from developing alpha cells (100,102).

Extrinsic signals derived from non-epithelial cells are also important in guiding pancreatic organogenesis. Early pioneering work using pancreatic explants *ex vivo* showed that when E11 epithelial buds were cultured without their surrounding mesenchymal tissue, epithelial growth and differentiation were arrested (103). More recently, genetic ablation studies have demonstrated the requirement for pancreatic mesenchyme for expansion of the pool of early pancreatic progenitor cells early in development and for proliferation of differentiated cells later in development (104,105). Although the pancreatic mesenchyme is broadly appreciated as playing an important role in pancreatic organogenesis, however, it is still not well understood whether there exist biologically relevant sub-populations of mesenchyme with distinct lineages and/or functional roles.

Recent single-cell RNA-Sequencing (scRNA-Seq) studies have highlighted previously unappreciated levels of cellular heterogeneity among the epithelial cells of the developing

murine pancreas, particularly within the endocrine compartment (53,58–60,62). Although relatively less attention has been given to elucidating potential cellular heterogeneity within the mesenchymal compartment, evidence from scRNA-Seq and classical genetic lineage tracing experiments suggests that transcriptionally distinct mesenchymal cell types also exist during development (53,58,106). As a result of this body of work, we now have a greater understanding of the transcriptomic cues governing cell states across pancreatic development, but we still lack an understanding of the upstream epigenetic features that regulate cell fate decisions. Integration of gene expression data and chromatin accessibility data would permit identification of active transcription factor binding to accessible chromatin within a given cell type.

In recent years, Assay for Transposase-Accessible Chromatin followed by Sequencing (ATAC-Seq) has been developed to profile genome-wide chromatin accessibility for epigenetic analysis in a given cell type or tissue (107). This technique has been applied to sorted populations of endocrine cells from the murine pancreas to investigate the chromatin landscape of developing EP cells (53,108). These studies, however, lacked single-cell resolution to capture the chromatin states of the various subpopulations of developing endocrine cells that have been described (53,58–60,62). More recently, single-nucleus ATAC-Seq (snATAC-Seq) has emerged as a technology to provide insights into chromatin accessibility at single-cell resolution (109,110). snATAC-Seq has been used to profile the chromatin landscape of many developing tissue types and has revealed cell-type specific *cis*- and *trans*-regulatory elements governing gene expression and cell fate decisions (111–115). Furthermore, integration of scRNA- and snATAC-Seq data for multi-omic analysis permits refinement of expressed TFs to a further parsed subset that not only expressed but are also likely binding TF motifs in open regions of chromatin and actively regulating expression of downstream target genes.

Here, we generate snATAC-Seq data of developing murine pancreas and perform an integrated multi-omic analysis of both chromatin accessibility and RNA expression. We describe

at single-cell resolution the chromatin landscape of the developing epithelium at E14.5 and E17.5, stages at which the dynamic processes of expansion, differentiation, and morphogenesis are actively underway. We identify candidate TFs regulating transitions across the endocrine lineages and construct gene regulatory networks (GRNs) of active TFs binding to regulatory regions of downstream target genes. Additionally, we generate a snATAC-Seq dataset of developing pancreatic mesenchyme at E14.5, which to our knowledge represents the first ATAC-Seq dataset (bulk or single-nucleus) of this cell type. We believe that these datasets and analyses will serve as a valuable resource for the field of pancreatic biology in general and will contribute to our understanding of lineage plasticity among endocrine cell types. In addition, these data will serve as a reference as to which epigenetic states should be represented in the differentiation of stem cells to the pancreatic beta cell fate to best recapitulate *in vitro* the gene regulatory networks that are critical for progression along the beta cell lineage *in vivo*.

2.2 Results

2.2.1 Single-nucleus ATAC-Sequencing of the developing murine pancreas

To investigate chromatin accessibility in the developing pancreas, we aimed to capture a broad range of cell types, including both epithelial and non-epithelial populations. In addition, we were specifically interested in profiling endocrine progenitor (EP) cells but given their rare numbers we searched for a method to achieve enrichment of this population. We utilized ePet1-EYFP mice (referred to hereafter as eFev-EYFP, as the gene *Pet1* is also known as *Fev*), where EYFP expression is driven by a *Fev* enhancer (116,117). In previous work, we had identified *Fev* as a marker of an intermediate murine EP population downstream of the better-characterized Ngn3(+) population and upstream of differentiated, hormone-expressing endocrine cells (58).

As lineage reconstruction of scRNA-Seq data had revealed that this Fev-expressing EP population is likely the state at which endocrine lineage allocation occurs, we chose to focus on enrichment of Fev(+) cells. Although previous work with this eFev-EYFP mouse line had validated that EYFP expression faithfully reflected *Fev* expression in brain tissue, similar confirmation had not yet been performed in the pancreas (116). We performed dual *in situ* hybridization (ISH)/immunofluorescence (IF) staining of E14.5 eFev-EYFP pancreas tissue to evaluate the architecture of EYFP expression with respect to the expression of *Ngn3* and *Fev* transcripts, as well as Chromogranin A (Chga) protein, a marker of differentiated hormone-producing endocrine cells. Expression of *Ngn3* and *Fev* transcripts was mutually exclusive (**Figure 2.1a**), as expected from our previous work demonstrating by genetic lineage tracing and scRNA-Seq that Fev-expressing cells are downstream of an Ngn3(+) state (58). Also as expected, EYFP expression was only found in epithelial (E-cadherin(+)) cells, and mostly localized to ductal-like structures (**Figure 2.2a**). A significant fraction, but not all, of EYFP(+) cells were actively expressing *Fev* transcript (**Figure 2.1a**). In addition, we observed EYFP(+) cells also expressing Chga (**Figure 2.2a**). These data are consistent with a model of *Fev* expression in pancreatic EP cells in eFev-EYFP mice in which *Fev* transcript first begins to be expressed as *Ngn3* expression wanes, then expression of EYFP (under the control of the *Fev* enhancer) follows (**Figure 2.1b**). Persistence of EYFP in cells that no longer express *Fev* transcript likely reflects longer perdurance of EYFP fluorescent protein compared to *Fev* mRNA in these cells, similar to what has been observed for Ngn3-tdTomato (118), Ngn3-YFP (119), and Ngn3-EGFP transgenic mice (120).

We further validated the eFev-EYFP mouse line using fluorescence-activated cell sorting (FACS) and quantitative real-time polymerase chain reaction (qRT-PCR). Consistent with our IF staining, we observed little to no EYFP signal in cells that were negative for the epithelial marker EpCAM (**Figure 2.2b**). Within the population of cells positive for EpCam and negative for the mesenchymal marker CD140a, a bimodal distribution of EYFP signal was detected (**Figure**

2.2b). TaqMan qRT-PCR analysis revealed that the EYFP-low population had higher expression of *Fev* mRNA compared to EYFP-high cells (**Figure 2.2c**). The EYFP-low population thus likely corresponds to a stage in which EYFP expression is on the rise and *Fev* expression is still present, whereas the EYFP-high population likely represents a stage where EYFP has reached higher expression but *Fev* itself has begun to wane (**Figure 2.1b**). Thus, we selected this EYFP-low population, enriched for Fev(+) cells, for snATAC-Seq using the 10x Genomics platform (**Figure 2.1c**, **Figure 2.2d**). We included an additional two samples, biological replicates of whole pancreas tissue, to capture a broad range of cell types (**Figure 2.1c**).

Single cells were lysed to isolate nuclei, and chromatin was then subjected to the 10x Genomics pipeline and sequenced. The resulting dataset was analyzed with the computational package ArchR (121). First, the datasets were filtered to retain high-quality nuclei by thresholding on the number of unique nuclear fragments, as well as the transcription start site (TSS) enrichment score (see Methods). This step provides enrichment of cells displaying a high fraction of fragments that map to the TSS, versus other locations in the genome. Next, the datasets were subjected to doublet discrimination, resulting in a final dataset consisting of a combined total of 33,206 high-quality nuclei across the three samples. The data were then dimensionally reduced, clustered, and visualized in a 2D Uniform Manifold Approximation and Projection (UMAP) embedding (**Figure 2.1d**). The UMAP projection revealed that cells from both replicates of the whole pancreas contributed to each of the clusters within the dataset. As expected, cells from the eFev-EYFP(+) sample clustered only with endocrine cells from the two Whole Pancreas samples (**Figure 2.1d, inset**), reflecting successful enrichment of endocrine cells from the eFev-EYFP mouse line and effective integration of the three datasets. Each cluster was annotated as corresponding to a specific cell type found within the developing pancreas based on the gene score (accessibility of the gene promoter plus the gene body) of the following marker genes: *Col3a1* to mark mesenchymal cells, *Wt1* for mesothelial cells, *Cpa1*

and *Spp1 for* exocrine cells, *Chga* for endocrine cells, *Pecam1* for endothelial cells, *Rac2* for immune cells, and *Sox10* for neuronal cells (**Figure 2.1e, Supplemental file 2.1**).

2.2.2 Integration of single-cell transcriptional and chromatin accessibility data identifies epithelial heterogeneity in the developing murine pancreas

To reliably identify the heterogeneity of chromatin states within the epithelial cell types of the developing pancreas, we performed unconstrained integration of our snATAC-Seq data from all epithelial cells with E14.5 scRNA-Seq data previously published by our laboratory (58). First, we computationally isolated the epithelial cells from the scRNA-Seq dataset (13,093 epithelial cells total) and performed iterative sub-clustering with the computational package CellFindR (122) to identify biologically relevant cell types. Next, we correlated the gene expression profiles of each of the cells within this scRNA-Seq dataset with the gene scores of each of the cells within our snATAC-Seq dataset. After identifying correlated cell pairs between the two datasets, cells in the snATAC-Seq dataset were assigned the cell type label, as well as the gene expression profile, of the cognate cell from the scRNA-Seq data.

This integration resulted in a final epithelial snATAC-Seq dataset comprised of 16,250 nuclei representing 10 distinct cell types, including Acinar, Ductal, Spp1(+)/Neurog3(+) double positive EPs, Neurog3(+) single positive EPs, Fev(+)/Chgb(+) intermediate progenitors, and Pdx1(+)/Mafb(+) beta cell precursors, as well as Alpha, Beta, Delta and Epsilon cells (**Figure 2.3a**). As expected, the sorted EYFP(+) cells contributed highly to the endocrine but not the acinar or ductal compartments of the overall dataset (**Figure 2.3a, inset**). The relative proportions of these annotated cell types in the snATAC-Seq dataset roughly matched the proportions of the epithelial cells in the scRNA-Seq dataset (**Figure 2.3b**). Integration scores, a reflection of confidence in the assignment of cell identity, were highest among terminally differentiated cell types (e.g., exocrine and hormone-expressing endocrine cells) (**Figure 2.4a**), indicating less ambiguity in chromatin accessibility once cell fate is determined. Even in the

absence of integration with scRNA-Seq data, all cell types were identified when clustering on chromatin accessibility alone (**Figure 2.4b**).

Next, we confirmed the cell type annotations by assessing chromatin accessibility (gene score), as well as the transferred RNA expression from the integration (gene expression). We observed high concordance between chromatin accessibility and RNA expression for the marker genes defining our cell types (**Figure 2.3c**). Additionally, we observed cell-type specific chromatin accessibility of each marker gene locus (**Figure 2.4c**). To assess the reproducibility of called peaks across both biological samples, we subsetted the dataset by sample and re-called peaks on a subset of endocrine cells (see Methods). We detected 83,055 total peaks in the final endocrine peakset for the whole pancreas sample replicate 1, 119,809 total peaks for whole pancreas replicate 2, and 106,301 total peaks in the eFev-EYFP(+) sorted sample (**Figure 2.4d**). We observed an overlap of 71,169 peaks between the two whole pancreas peaksets (representing an 86% overlap for whole pancreas sample 1 and a 59% overlap for whole pancreas sample 2). The number of peaks common to all three samples was 64,082, which represents an overlap of 77% for whole pancreas 1, 53% for whole pancreas 2, and 60% for the the EYFP(+) sample (**Figure 2.4d**). When assaying differentially-accessible or -expressed genes, we observed far fewer significantly differentially accessible genes (n = 1,066) compared to differentially expressed (n = 4,567) (**Figure 2.3d, Supplemental file 2.1**). Among these differentially-accessible genes were top markers of each cluster identified by differential gene expression analysis of our scRNA-Seq dataset (**Supplemental file 2.1**). Taken together, these data confirm the existence of heterogeneous epithelial populations initially identified by scRNA-Seq, here by an orthogonal method.

2.2.3 Identification of candidate regulators of epithelial cell fate

To identify regulators of cell fate decisions in the developing pancreatic epithelium, we applied the peak calling algorithm MACS2 (123) to our dataset. We identified 232,056 peaks

across all epithelial clusters, with 63,097 peaks exhibiting differential accessibility across cell types (**Figure 2.5a**). Next, we assayed for TF motif enrichment in these differential peaks, identifying 404 enriched motifs. A number of TFs in the same family were deemed enriched due to the similarities in DNA binding motifs. For instance, TFs with enriched motifs included known regulators of pancreatic epithelial development, such as Sox family members (*Sox2, Sox4, Sox9*; Ductal), Hox family members (*Hoxb4, Hoxc4, Hoxa4*; Beta) and members of the Rfx family (*Rxf3* through *Rfx7*; Fev(+)/Chgb(+) and Pdx(+)/Mafb(+)) (**Supplemental file 2.2**). To distinguish among TFs with similar DNA motifs identified in a given cell type, we next identified significant TF motif deviations (calculated as deviation of motif enrichment in accessible peaks from the expected distribution based on the average across all cells) of each cell type using ChromVAR (124). The TFs from ChromVAR were then correlated with their gene expression profiles from the integrated RNA expression matrix, thereby identifying so-called "correlated TFs" that are both expressed and have significant motif deviation (**Figure 2.5b, Supplemental file 2.3**). By breaking this down further on a per-cluster basis, we then were able to observe the cell type-specific motif deviations and gene expression of the correlated TFs, narrowing the number of TFs with enriched motifs from 404 (**Figure 2.5a**) to 50 correlated TFs (**Figure 2.5b,c**). Correlated TFs included multiple members of the Fox family (*Foxk2* and *Foxo1*, Beta; *Foxp1*, Fev; *Foxa2*, Delta), as well as the Sox family (*Sox4*, Spp1(+)/Ngn3(+); *Sox9*, Ductal) (**Figure 2.5c**). The relationship between motif deviation and gene expression is further shown at single-cell resolution for several endocrine genes in **Figure 2.5d**. Interestingly, by observing not only motif deviation but also gene expression, we were able to determine that although both Ductal cells and Spp1(+)/Neurog3(+) EPs showed high motif deviation of *Sox4*, expression was significantly higher in the latter population (**Figure 2.5d**). This is in line with previously published work that shows that *Sox4* works with *Neurog3* to induce endocrine differentiation in the developing murine pancreas (125).

We next sought to understand the correlated TFs across the Alpha and Beta cell lineages. We first calculated the pseudotime values of cells along both trajectories (**Figure 2.5e**) and then applied the same motif deviation and gene expression correlation analysis for the genes and enriched motifs along these lineages (**Figure 2.5f, Supplemental file 2.3**). Across the Alpha lineage (including Ductal, Spp1(+)/Neurog3(+), Neurog3(+), Fev(+)/Chgb(+), and Alpha cells) we identified 34 correlated TFs. This included TFs in Ductal cells (*Nr4a1, Sox9*), progenitor cells (*Neurog3, Neurod2*), and Alpha cells (*Foxp1, Isl1, Mafb, Mafg*). For the Beta lineage, we identified 27 correlated TFs, including *Mnx1, Mafb, Mafg, Pdx1*, and *Foxo1*. In summary, the multi-layered approach taken here has further distilled the subset of TFs that likely play an important role in governing fate selection during endocrinogenesis.

2.2.4 Gene regulatory networks controlling epithelial cell fate

Our analyses thus far identified accessible chromatin and correlated TFs within the epithelial compartment of the developing endocrine pancreas. How and where these TFs bind and affect downstream target genes to govern cell fate decisions is not as well understood, however. To address this gap in knowledge, we next sought to construct a gene regulatory network (GRN) for Acinar, Ductal, and endocrine cells of the Alpha and Beta lineages (**Figure 2.6a**). We utilized the computational pipeline Integrated Regulatory Network Analysis (IReNA) v2 (113) (**Figure 2.6a, Figure 2.7a**), which combines both scRNA-Seq and snATAC-Seq data to predict TF binding of downstream target genes in a cell type-specific manner. First, we performed differential gene expression analysis on our scRNA-Seq dataset to identify genes enriched in each cell type (**Figure 2.7b, Supplemental file 2.4**). We then performed peak-to-gene linkage analysis in our integrated scRNA- and snATAC-Seq datasets, identifying accessible regions of chromatin (peaks) that are either positively or negatively significantly correlated with gene expression (genes) (**Figure 2.7c**). These peak-to-gene peaks were then further filtered and annotated as correlated accessible regions (CARs) belonging to one of three

categories: TSS (when the peak lies in the transcription start site (TSS) for the gene), gene body (when the distance between peak and TSS is less than 100 kb, and the peak-to-gene score calculated above is significant), or distal (when the peak is 100 kb upstream or downstream of the TSS of a correlated gene, and the peak-to-gene score is significant). We observed varying proportions of CAR categories among the cell types, with TSS representing the highest proportion, followed by roughly equivalent proportions of positive and negative CARs (**Figure 2.7d**). We next predicted the cell-type specific TF binding in these CARs by searching for TF DNA binding motifs in the CARs. Once identified, we then filtered the TFs by calculating their TF footprint score, retaining TFs with a score deemed significant by IReNA v2.

We observed the highest number of GRN TFs in the Fev(+)/Chgb(+) population (39 TFs), followed by Neurog3(+) (33 TFs) and Ductal (28 TFs) cells (**Figure 2.7e, Supplemental file 2.5**) Among our network of TFs and target gene-associated CARs, we binned these interactions as either activating or repressing by correlating the expression of each TF-target gene pair. Genes with a positive TF-target gene correlation were annotated as being activated by their given TF, while those with a negative TF-target gene correlation were annotated as being repressed. The Ductal and the Fev(+)/Chgb(+) populations had the highest number of regulations, followed by Pdx1(+)/Mafb(+) and Neurog3(+) (**Figure 2.7f**). The regulations among all the populations examined were relatively evenly split between activating and repressing. Lastly, from the GRN constructed above, we identified pairs of TFs that regulated one another; for each TF, we identified target genes that are also TFs and mapped these pairs as either activating or repressing depending on the correlation of gene expression of the target TF in the given cell type. This analysis permitted us to identify TFs that regulate the expression of other TFs in given cell types (**Figure 2.7g, Supplemental file 2.6**).

To examine the TFs comprising this epithelial GRN in more depth, we first focused on the hormone(+) populations within our dataset. We found that genes enriched in the Beta cell population are largely repressed in the Acinar, Ductal, and Spp1(+)/Neurog3(+) GRNs, then

gradually become activated as endocrine differentiation proceeds (**Figure 2.6c**). Within the Alpha cell population, Beta cell enriched genes are almost exclusively activated, consistent with previous studies investigating gene expression of individual TFs revealing that beta and alpha cells share common expression of genes needed for proper development and function (94). We then inquired within all of the GRNs defined for epithelial cell types, which TFs either activate or repress genes enriched in the Beta cell population. Among the top activating TFs were known regulators of Beta cell development, such as *Mafb, Neurod1,* and *Pdx1* (**Figure 2.6d**). Targets of these activating TFs identified by our GRN analysis included numerous genes, both known and novel (**Supplemental file 2.5**). Repressors of Beta cell enriched genes are largely contained within the Ductal GRN and include TFs such as *Tcf3*, *Tead2, Sox4,* and *Rest* (**Figure 2.6e**). As in Beta cell enriched genes, Alpha enriched genes follow a similar pattern of activation and repression across the epithelial GRN (**Figure 2.8a**). TFs activating Alpha cell enriched genes include Beta cell activating TFs such as *Mafb* and *Neurod1* (**Figure 2.6f**). TFs repressing Alpha cell enriched genes also included *Tcf3*, *Tead2, Sox4,* and *Rest* (**Figure 2.6g**), suggesting that these TFs repress global hormone(+) cell gene signatures. The six TFs that overlapped between the Alpha (18 TFs) and Beta (13 TFs) GRNs comprised known endocrine regulators *Pax6, Mafb, Neurod1,* and *Isl1*, as well as TFs less well studied in pancreas, *Zfp516* and *Meis2* (**Figure 2.6h**). TFs unique to the Beta GRN included known regulators of beta cell fate, such as *Nkx6-1*, *Pdx1,* and *Foxo1*, while those less well characterized included *Mlxipl*. We performed multiplexed immunofluorescence and *in situ* hybridization on sections of wildtype E14.5 pancreas tissue and validated the enrichment of *Mlxipl* expression in *Ins1*(+) Beta cells (**Figure 2.8b**). Examples of TFs unique to the Alpha GRN were *Irx2* and *Arx*, known regulators of Alpha cell fate. Less well characterized TFs included *Pbx1*, which is required for proper pancreas development (126), *Bbx, Peg3,* and *Etv1*. We also confirmed high enrichment of *Etv1* expression in Gcg(+) Alpha cells, as well as some expression in Fev(+) cells (**Figure 2.8c**).

Among the top activating TFs of Alpha cell enriched genes was *Etv1* (**Figure 2.6f**), reported to be a direct or indirect target of *Nkx2-2* (127). Within beta cells, failure to properly degrade *Etv1*, *Etv4*, and *Etv5* results in impaired insulin secretion (128). We took a closer look at the downstream targets of *Etv1* in the Alpha cell GRN and found that *Etv1* was predicted to activate 150 genes, and repress 130 genes, in the Alpha GRN (**Supplemental file 2.5**). When performing pathway analysis on these genes, we observed that *Etv1* activated pathways related to vesicle mediated transport, synaptic vesicle pathway, and membrane trafficking (**Figure 2.6i**). Other targets of *Etv1* included Alpha-cell related genes such as *Mafb* and *Irx2*, as well as endocrine hormones such as *Pyy* (**Supplemental file 2.5**). Pathways repressed by *Etv1* included those related to translation initiation, cytoplasmic ribosomal proteins, and cap-dependent translation initiation. When examining TF-TF interactions, we found that *Etv1* is repressed by *Sox9* and *Egr2* in the Ductal population and activated by *Stat3* in the Pdx1(+)/Mafb(+) population and by *Usf2* and *Vdr* in the Fev(+)/Chgb(+) population (**Figure 2.6j**).

In summary, the computational analyses described in this section have permitted the construction of a gene regulatory network of the acinar, ductal, and major endocrine lineages in the developing mouse pancreas. This workflow, which is dependent on integration of both chromatin accessibility and transcriptional data, has identified regulators of alpha and beta cell fate that can serve as the subjects of further experimental study.

2.2.5 Gene regulatory networks governing the initiation of endocrine differentiation

During mammalian pancreatic development, a subset of cells within the branching ductal epithelium activate the expression of the master regulator of endocrine differentiation, *Neurog3*. These rare Neurog3(+) cells represent the earliest known EP population, and considerable attention has been devoted to understanding this ductal to EP transition. In addition to *Neurog3*, which is required for mouse endocrine differentiation (47), numerous other TFs have been

identified that are also important for endocrinogenesis. Investigation of *NEUROG3* binding across the genome in human pluripotent stem cell (hPSC)-derived EP cells revealed widespread regulation of 138 TFs, some with known roles in endocrine development and others with unknown function (129). Further studies in human cells used inducible and knockout models in hPSC-derived endocrine cells to identify predicted targets of multiple endocrine TFs, including *NEUROG3*, *PDX1,* and *RFX6* (130). Generation of an Ngn3-timer fluorescent reporter mouse line that permitted the specific isolation of early Ngn3-expressing cells identified numerous putative direct targets of *Neurog3* in mouse EPs (131). These studies have highlighted the need for a broad, integrated analysis of all TFs and downstream targets that control the initiation of endocrine differentiation.

Towards this end, we began by investigating the GRN regulating the transition from a ductal to EP cell state. TFs in the Ductal GRN promoting the expression of Spp1(+)/Neurog3(+) EP genes include known regulators of endocrine cell fate, such as *Sox4* (125,132) (**Figure 2.9a**). Also identified was *Sox9*, which induces the expression of *Neurog3* and thus positively regulates the endocrine lineage before then being turned off during differentiation (44,132). Interestingly, *Tead2* and *Tcf3* activated the most genes enriched in the Spp1(+)/Neurog3(+) and Neurog3(+) EP populations (**Figure 2.9b**), indicating that these TFs are important initiators of an endocrine cell fate. The Yap/Tead signaling complex has previously been shown to activate multipotent progenitor cell enhancers and regulate epithelial outgrowth during human pancreatic development (133). *Tcf3*, also known as *E47*, has been shown in a human cell line to dimerize with *NEUROG3* to bind to the promoter region of the *INSM1* gene (134), which is required in mice to maintain mature beta cell function (135). TFs involved in the transition from a Spp1(+)/Neurog3(+) to a Neurog3(+) EP cell state include well known regulators of endocrine differentiation, such as *Nkx2-2*, *Pax4*, and *Neurod2* (**Figure 2.9a**). Other TFs with no known endocrine function included *Nhlh1*, whose expression was enriched in Neurog3+ cells by E14.5 scRNA-Seq and confirmed *in vivo* (**Figure 2.10a**). Major repressors of EP enriched genes in the

Ductal GRN include TFs such as *Rest* and *Hes1* (**Figure 2.9c**). The classification of *Hes1* as repressing is consistent with what is known about the role of Notch signaling in the initiation of EP cell fate (44,136). *Rest* is a master regulator of neurogenesis and has been previously described to inhibit direct reprogramming of pancreatic exocrine to endocrine cells by inhibiting the binding of *Pdx1* to key endocrine differentiation-related genes (137). In addition, loss of *Rest* results in increased generation of pancreatic endocrine cells during development (138,139). Consistent with previous studies, we confirmed expression of *Rest* in *Spp1*-expressing Ductal cells by *in vivo* staining of E14.5 pancreas and by scRNA-Seq (**Figure 2.10b**). Among TF-TF pairs identified in Ductal, Spp1(+)/Neurog3(+), and Neurog3(+) cells, most were classified as activating, with the exception of *Nfib* (Ductal) (**Figure 2.10c**). Taken together, our GRN analysis has identified novel candidate regulators, such as *Tcf3* and *Tead2*, of the ductal to EP cell state transition. These results expand upon our knowledge of this key developmental transition and serve as a resource for future studies.

We next focused on Fev(+)/Chgb(+) cells, as our previous work indicated that this cell state represents the bifurcation point at which the Alpha or Beta lineage is established (**Figure 2.6a**) (58). As expected, we observed that the Acinar and Ductal cell types largely repress genes that are enriched in the Fev(+)/Chgb(+) population (**Figure 2.9d**). These genes begin to be activated as an endocrine cell fate is established (Spp1(+)/Neurog3(+) and Neurog3(+) cell types) and are fully activated by the Fev(+)/Chgb(+) cell stage. Curiously, the Fev(+)/Chgb(+) enriched genes are not repressed in the Alpha and Beta cell types, suggesting that Alpha/Beta cell fate is due more to activation of key Alpha/Beta genes as opposed to the repression of progenitor-associated genes. Among the top TF activators of Fev(+)/Chgb(+) enriched genes (**Supplemental file 2.5**), we observed that known regulators of endocrine cell fate such as *Neurod1, Mafb*, and *Pax6* activated the most genes (**Figure 2.9e**). Activators also included less well described TFs, such as Rad21 and Peg3, as well as *Foxo1* and *Etv1* (**Figure 2.9e, Supplemental file 2.5**). Conversely, TF repressors of Fev(+)/Chgb(+) enriched genes included

the TFs *Tead2*, *Tcf3*, *Rest, Sox4*, and *Nfib,* among others (**Figure 2.9f**). Next, we constructed a network diagram of TF pairs that either activate or repress TFs enriched in the Fev(+)/Chgb(+) population (**Figure 2.9g**). Consistent with our observations in **Figure 2.9d**, TF-TF regulations were entirely repressive in the Acinar and Ductal cell states, and a mix of activating and repressing in the Spp1(+)/Neurog3(+) and Neurog3(+) cell states (**Figure 2.9g, Supplementary file 2.6**). In contrast, TF-TF regulations in the Alpha, Beta, and Pdx1(+)/Mafb(+) states were exclusively activating (**Figure 2.9g, Supplementary file 2.6**).

When comparing the GRNs among all progenitors and precursors, we identified 18 TFs unique to the Fev(+)/Chgb(+) population (**Figure 2.10d, Supplementary file 2.6**). Among these TFs identified within the Fev(+)/Chgb(+) GRN, *Rad 21* and *Mafb* were among the top activators of the transition from Fev(+)/Chgb(+) to either Pdx1(+)/Mafb(+), Alpha, or Beta cell states (**Figure 2.10e-g**). Among the top 10 TFs with the highest number of activating regulations across the transition from Fev(+)/Chgb(+) to Alpha cells was Vitamin D receptor (*Vdr*), whose expression has been linked to beta cell function and diabetes (**Figure 2.10f**) (140–142). As expected from **Figure 2.6f**, *Etv1* had a higher number of activating regulations for Alpha cell-enriched genes compared to Pdx1(+)/Mafb(+) or Beta populations (**Figure 2.10e-g**).

The gene *Fev* was initially described as a prototypical serotonergic transcription factor in the brain (143), then as a gene expressed in developing and adult mouse pancreatic islets (144). More recently, we found that *Fev* in the pancreas marks an intermediate progenitor of the mouse endocrine lineage (58). Our GRN analysis computed 83 genes activated and 57 genes repressed by *Fev* (**Supplemental file 2.5**). Pathway analysis of activated genes included pathways related to synaptic vesicle cycle, signaling by NRTKs, and membrane trafficking (**Figure 2.9h**). These data corroborate previous findings in which full body knockout of *Fev* resulted in decreased pancreatic insulin content, an impairment of insulin secretion, and consequently defects in glucose clearance (144). Pathway analysis of repressed genes included many pathways related to translation, such as translation initiation complex formation, cap-

dependent translation initiation, and eukaryotic translation initiation (**Figure 2.9h**). Downstream TF interactions of *Fev* were all activating and included the TFs *Foxa2*, *Usf2*, and *Nfe2l1* (**Figure 2.9i**).

Next, we sought to confirm the occupied peaks and downstream targets of *Fev* in our GRN analysis. Although ChIP-Seq data of *Fev* binding in the endocrine pancreas is not available, a previous study published ChIP-Seq data from *Fev*-expressing 5-HT(+) neurons in the E12.5-E15.5 developing hindbrain (117). Given the documented similarities in transcriptional networks between serotonergic neurons and pancreatic endocrine cells (144), we asked how much shared overlap there is between our gene-associated CARs in the pancreatic *Fev* GRN and serotonergic neuron ChIP-Seq peaks. When intersecting these peak sets, we observed 90 of the 270 (33%) *Fev* GRN CARs overlapping with peaks from the ChIP-Seq dataset (**Figure 2.11a**). Target genes of these gene-associated CARs included *Myo6*, *Mef2a*, and *Tmem30a* (deemed activated by *Fev* in our GRN analysis), as well as *Rrs1, Rpl41, and Cxadr* (deemed repressed by *Fev* in our GRN analysis). Non-overlapping activated gene-associated CARs included *Vamp3*, *Chga*, *Foxa2*, and *Klf7*. Repressed gene-associated CARs included *Notch2, Nr5ac,* and *Apex1.* The overlap of *Fev* GRN CAR and ChIP-Seq peak for *Mef2a* was visualized at the *Mef2a* gene locus in the Fev(+)/Chgb(+) and Pdx1(+)/Mafb(+) clusters (**Figure 2.11b**). Next, to confirm the potential gene targets of *Fev* in our GRN analysis, we interrogated published RNA-Seq data from wildtype and *Fev* knockout 5-HT(+) serotonergic neurons from E15.5 mouse hindbrains. We intersected the target genes of *Fev* in our GRN analysis with differentially expressed genes between *Fev* wildtype and knockout mice (**Figure 2.11c**) (117). We found that 38 of the targets of *Fev* in the embryonic pancreas were differentially expressed between the Fev wildtype and knockout embryonic serotonergic neurons, including genes such as *Celf1, Pcsk1, Mcm6, Chga, Dctn3,* and *Fabp5*.

Taken together, our data identify both known and novel regulators of pro-Alpha and pro-Beta cell fates that are active in the Fev(+)/Chgb(+) stage, the cell state that represents the

bifurcation point in the endocrine differentiation trajectory. Our analyses also yield a comprehensive view of potential targets of *Fev,* as well as insights regarding its function in regulating the machinery required for the production of endocrine hormone-containing vesicles. Lastly, our analysis has identified conserved *Fev*-occupied peaks and target genes between serotonergic neurons and Fev(+) endocrine progenitors, as well as potential tissue-specific regions of *Fev*-mediated regulation.

2.2.6 Characterization of chromatin accessibility and identification of GRNs within pancreatic mesenchymal cell types

Although proper development of the pancreatic epithelium depends on signals from the surrounding mesenchyme, the lineage and function of pancreatic mesenchymal cells remains vastly understudied. In previous work, we and others have used scRNA-seq to identify multiple transcriptionally distinct mesenchymal populations, including mesothelium, within the developing murine pancreas (53,58). Still, the upstream genetic regulators that maintain these distinct cell states are not defined. Data from snATAC-Seq of pancreatic mesenchymal cells would shed light on whether distinct states of chromatin accessibility correspond to transcriptionally distinct cell subpopulations and would reveal which TFs and binding sites are actively involved in controlling mesenchymal cell state.

We integrated the snATAC-Seq data from the mesenchymal populations within our merged E14.5 whole pancreas dataset (replicates 1 and 2) with the age-matched (E14.5) scRNA-Seq data, using methods as described above. Clustering of the scRNA-Seq dataset identified six populations of mesenchymal cells, including one cluster enriched in the expression of *Gap43* (Gap43(+)), another cluster enriched in expression of *Sfrp2* (Sfrp2(+)), two clusters expressing chemokines (Cxcl12(+) and Cxcl13(+)), Vascular Smooth Muscle cells (VSM; Acta2(+)) and finally Mesothelium (Wt1+) (**Figure 2.12a**). Integration and cell label transfer classified all populations in our snATAC-Seq dataset (**Figure 2.12a,b**). Clustering on chromatin

accessibility alone, without integration with scRNA-Seq data, still resulted in the identification of similar clusters as in the integrated dataset (**Figure 2.13a**).

Once the mesenchymal clusters were identified, we next asked whether there were secreted factors uniquely produced by any of these mesenchymal subpopulations that may be signaling to any of the epithelial subpopulations identified above. We performed cell-cell communication analysis using CellChat (145) on the mesenchymal and epithelial compartments in our E14.5 scRNA-Seq dataset (**Figure 2.13b**). We found 33 cell-cell communication signaling pathways that were significantly active in the mesenchymal and epithelial datasets (**Supplemental file 2.7**), including communication based on ECM signaling (Mk, Ptn, Mpz, Laminin) and secreted factors (Bmp, Ephb, Ngf, Notch). When focusing on mesenchymal/mesothelial to epithelial signaling, we found that the ECM pathways Mk, Ptn and Collagen showed high relative strength for incoming signaling to Acinar and Ductal cells (**Figure 2.13b**), with high outgoing strength across multiple mesenchymal populations. Strong signaling to Beta cells included Grn, likely being received by the Cxcl12(+) mesenchymal population, which showed the strongest outgoing strength. Jam signaling, which is important in cell migration, was highest in outgoing signaling from Mesothelial cells, while highest in incoming signaling in the Acinar, Ductal, Spp1(+)/Neurog3(+), and Neurog3(+) populations. Taken together, these data elucidate the potential cell-cell communication between the developing mesenchymal, mesothelial, and epithelial cells of the embryonic pancreas.

Next, we identified differentially accessible peaks across all of the mesenchymal populations and found a total of 182,020 peaks, with 59,707 peaks displaying differential accessibility (**Supplemental file 2.2**). The majority of these differentially accessible peaks were enriched in the Gap43(+) and Mesothelial populations, with the remaining populations showing more modest numbers. Motif enrichment in these differentially accessible peaks identified 217 enriched motifs (**Supplemental file 2.2**). Correlation of gene expression and motif deviation

scores identified 34 correlated TFs (**Figure 2.12c,d**), including known regulators of mesenchymal cell fate, such as *Wt1*, *Twist2*, and *Hand2.*

We next assembled the mesenchymal GRN with IReNA v2 as described above. Among the mesenchymal cell types, the Gap43(+) and Mesothelial populations had the highest number of total regulations (**Figure 2.13c**) as well as the highest numbers of TFs in each of the GRNs (**Figure 2.13d**). In the TF-TF network, TFs active in the GRN of Mesothelium largely repressed the expression of Gap43(+)-associated TFs, whereas the Gap43(+) population activated Cxcl12(+)-associated TFs (**Figure 2.13e**). We next assessed which GRN TFs were either exclusive or shared among the mesenchymal populations (**Figure 2.13f**). The GRN in the Gap43(+) population contained 18 specific TFs, including genes involved in epithelial-mesenchymal transition (EMT) such as *Snai2, Twist2,* and *Zeb1*, suggesting that cells within the Gap43(+) population may be actively undergoing EMT processes during pancreatic development. Other Gap43(+) cell type-specific TFs included *Hoxb4* and *Prxx1*. TFs exclusive to the Cxcl12(+) population included members of the nuclear factor 1 (*Nfia* and *Nfic*), Ap-1 (*Junb* and *Jund*), and Klf families (*Klf6; Klf2* shared with the Gap43+ population). TFs specific to the Sfrp2(+) GRN included *Hoxb5, Meis2,* and *Bcl11a,* while *Pbx1* was shared with the Cxcl12(+) population, and *Barx1* was shared with the Mesothelial population.

We next focused on mesothelial cells, which consist of a monolayer of specialized cells that line the pleura and internal organs of adult tissues and serve numerous functions in the adult, including lubrication of tissue and immune surveillance (146). In the developing lung, lineage tracing studies have demonstrated that mesothelium also acts as a progenitor for certain specialized mesenchymal cell subtypes (147). Furthermore, previous work in our lab predicted the downstream lineages of mesothelial cells in the developing pancreas based on pseudotemporal ordering of scRNA-Seq data (58). Despite this, relatively little is known about how mesothelial cells are formed and maintained during pancreatic development. Within the snATAC-Seq dataset, we identified 20 TFs uniquely active within the GRN of the Mesothelial

population (**Figure 2.13f**). Top TFs activating Mesothelial-enriched genes included *Wt1,* which has previously been shown to be a master regulator of mesothelial formation (148), along with *Klf13,* whose role in mesothelial cell development and homeostasis is not well understood and thus warrants further study (**Figure 2.12e**). Top repressing TFs of Mesothelial-enriched genes included *Klf2*, *Ebf1,* and *Klf6* (**Figure 2.12f**). Pathway analysis of activated and repressed genes in the Mesothelium GRN identified pathways related to metabolism as activated, and ECM formation and deposition as repressed (**Figure 2.12g**).

Taken together, profiling of the chromatin accessibility within the cells of the developing mouse pancreatic mesenchyme has determined differentially accessible peaks among these populations and identified TFs potentially important in mesenchymal development through the use of motif enrichment and gene expression correlation analyses. Lastly, we have constructed a set of mesenchymal GRNs, identifying active TFs and their downstream target genes. These data will provide a resource for future work geared towards studying mesenchymal biology and understanding how this important but understudied non-epithelial population is maintained.

2.2.7 Mapping the gene regulatory networks of pancreatic epithelium across developmental time

Previous work from our laboratory (58) used scRNA-Seq to profile the murine pancreas across three developmental timepoints. Although this timecourse analysis revealed shifts in endocrine cell differentiation, proliferation, and gene expression programs across developmental time, the regulatory elements governing these changes remain poorly understood. To better characterize how gene regulatory networks within the pancreatic epithelium shift across developmental time, we performed snATAC-Seq on two independent samples of whole pancreas tissue at E17.5, each sample comprising at least n = 3 pancreata. After performing quality control and filtering steps (see Methods), we generated a merged dataset consisting of 78,669 high-quality nuclei comprised of the expected broad pancreatic cell

types (**Figure 2.15a**). Of these broad groups, 8,859 were labeled as endocrine based on the Gene Score for *Chga* (**Figure 2.14a, Figure 2.15a,b**). We then performed unconstrained integration of our E17.5 snATAC-Seq epithelial cells with E17.5 scRNA-Seq (58) as described above for E14.5. This integration resulted in a final epithelial snATAC-Seq dataset comprised of 8 distinct cell types, including Acinar, Ductal, Spp1(+)/Neurog3(+) double positive EPs, Neurog3(+) single positive EPs, Fev(+)/Chgb(+) intermediate progenitors, Alpha, Beta, and Delta cells. (**Figure 2.14a**). All epithelial cell populations annotated in the scRNA-Seq dataset were successfully captured in the snATAC-Seq dataset, although the relative proportions of some cell types varied, presumably due to cell isolation and nuclei capture (**Figure 2.14b**). We then identified correlated TFs across the hormone(+) lineages as detailed above. Across the Alpha lineage, we identified 39 correlated TFs, including Ductal TFs (*Nr5a2, Sox9*), progenitor cells (*Elf5, Neurog3, Foxa2*), and Alpha cells (*Rfx3, Rfx6, Arx, Isl1, Mafb, Mafg*) (**Figure 2.15c**). For the Beta lineage, we identified 32 correlated TFs, including *Arid3a, Pax6,* and *Pdx1* (**Figure 2.15d**). For the Delta lineage, we identified 34 correlated TFs, including *Rfx3* and *6, Lhx1, Pou3f4, Pax6, Arx* and *Pdx1* (**Figure 2.14c**).

Next, we sought to construct GRNs for each of the epithelial cell types in the E17.5 pancreas utilizing the IReNA v2 pipeline as detailed above for the E14.5 sample. After constructing the GRN, we observed the highest number of GRN TFs in the Neurog3(+) population (57), followed by the Fev(+)/Chgb(+) (50) and Ductal (44) populations (**Figure 2.15e**, **Supplemental file 2.4**). As in the E14.5 GRN, the Ductal, Fev(+)/Chgb(+), and Neurog3(+) populations had the highest number of regulations (**Figure 2.15f**). We observed that the Acinar population displayed a marked reduction in the number of regulations at E17.5 compared to E14.5 (**Figure 2.15c, Figure 2.7f**). Next, we focused on the hormone(+) cell types in our GRN analysis. The E17.5 Alpha cell population contained 23 total GRN TFs, 19 of which were unique to the Alpha cell population (**Figure 2.14d**). These Alpha GRN TFs included known regulators such as *Irx2, Arx*, and *Neurod1*, as well as unknown regulators such as *Bptf, Thap11*, and

Nfat5. When comparing Alpha GRN TFs at E17.5 (23) to those at E14.5 (18), we found that 7 TFs are unique to E17.5, including *Foxa2*, *Foxp1*, and *Bptf*, while two are unique to E14.5 (*Pbx1* and *Fbp1*) (**Figure 2.15g**). Alpha GRN TFs common between E14.5 and E17.5 (16) include key Alpha cell regulators such as *Irx2, Arx, Isl1*, and *Neurod1*, as well as *Etv1* and *Peg3*. The E17.5 Beta GRN contained six total TFs, two of which (*Insm1* and *Mlxipl)* are not found in the Alpha or Delta GRNs (**Figure 2.14d**). While a number of known regulators of beta cell fate such as *Nkx6-1, Mnx1, Mafg*, and *Isl1* are unique to the E14.5 Beta GRN, only *Banp* is unique to the E17.5 Beta GRN. Shared GRN TFs were *Pax6, Pdx1, Mafb, Insm1*, and *Mlxipl* (**Figure 2.15h**). Lastly, the E17.5 Delta GRN contained six TFs: known Delta regulator *Hhex*, along with *Pdx1, Zbtb20, Junb, Fos,* and *Klf7.* (**Figure 2.14d**). Top TFs regulating Delta cell-enriched genes included *Mafb*, *Neurod1*, *Klf7*, *Rad21,* and *Pax6* (**Figure 2.14e**), while TFs repressing Delta cell-enriched genes included *Tcf3, Tead2* and *Jun* (**Figure 2.14f**).

Lastly, we sought to understand how the targets of GRN TFs change across developmental time. We focused on activated target genes of *Fev* in the Fev(+)/Chgb(+) GRN and identified 91 targets in the E17.5 GRN compared to 83 in the E14.5 GRN (**Figure 2.14g, Supplemental file 2.3**). Targets unique to E14.5 included genes such as *Vamp3, Dctjn3, Mef2a, Klf10,* and *Ptpm2,* while shared genes included *Fam174a, Serinc1, Tmem30a, Chga,* and *Tmed8*. Targets unique to the E17.5 GRN included genes involved in pancreatic secretion, beta-catenin independent Wnt signaling, and signaling by NTRK1 (TRKA) (**Figure 2.14h**), such as *Pdx1*, *Itpr3*, *Rab6a*, *Gng12*, and *Pak3*. The predicted TF targets of *Fev* at E17.5 were *Nfix* (Spp1(+)/Neurog3(+))*, Nfib* (Ductal)*, Pdx1* (Beta)*, Usf2* (Fev(+)/Chga(+))*,* and *Ccnt2* (Fev(+)/Chga(+)) (**Figure 2.14i, Supplemental file 2.4**). Taken together, these data have permitted the construction of a gene regulatory network of E17.5 exocrine and endocrine cells in the developing mouse pancreas and permitted a direct comparison of the important TFs and target genes across these two timepoints to yield a more refined understanding of the regulatory dynamics of endocrine maturation across developmental time.

2.3 Discussion

Numerous studies have used scRNA-Seq to characterize developing mouse pancreas tissue, providing important insights into cellular heterogeneity and key transcriptional programs expressed in developing cell types (53,58–60,62). Still, these datasets lack information about which of the expressed TFs are active and binding, and about how the TFs are organized into regulatory networks. Profiling of the chromatin accessibility landscape at single-cell resolution has emerged as a powerful approach for generating new insights about regulatory programs governing development and cell fate decisions across multiple tissue types (113,114,149,150), and we have now extended this approach to developing mouse pancreas tissue. Given that we were particularly interested in interrogating mechanisms underlying endocrine lineage allocation, we utilized a genetic tool to achieve significant enrichment of EP cells. Previous work from our laboratory had identified the transcription factor *Fev* as a marker of a novel endocrine progenitor state, and lineage reconstruction analysis indicated that it is at this *Fev*(+) state that lineage allocation is executed (58). Here, we validated the use of an eFev-EYFP transgenic mouse line for enriching for *Fev*-expressing endocrine cells in the developing pancreas.

We have generated a comprehensive map of chromatin accessibility in the developing E14.5 and E17.5 murine pancreas, including enriched endocrine populations as well as non-endocrine cell types. Although previous studies have investigated chromatin accessibility in the developing pancreatic epithelium through bulk ATAC-Seq of sorted populations (53,108), to our knowledge our study represents the first to examine chromatin accessibility at true single-cell resolution. By integrating both scRNA-Seq and snATAC-Seq data, we successfully generated a refined list of correlated TFs that are not only expressed, but also likely binding to open regions of chromatin to control cell fate decisions across developmental time. The analyses performed in this study utilized separate, computationally integrated scRNA and snATAC Seq datasets; in the future, performing multi-omic profiling of the transcriptional and chromatin states within the

same cell may lend even greater confidence in the GRN reconstruction and identification of TFs governing cell fate.

Here, we have constructed cell-type specific GRNs describing active TFs and their putative target genes through the binding of *cis*-regulatory regions. Future work could leverage the datasets presented here to catalog in greater depth the gene-associated CARs to identify specific *cis*-regulatory elements operant in specific subtypes of cells on a gene-by-gene basis in a manner reported in other tissues, such as the human retina (151). An advantage of our single cell level data is that it permits identification of not only cell type-specific, but even cell subtype-specific differences in active regulatory elements. Our analysis identified several known regulators of endocrine cell fate, such as TFs *Pdx1* and *Nkx6-1* in beta cells and *Arx* in alpha cells, as well as identified novel candidate TFs, such as *Mlxipl* in beta cells and *Pbx1* and *Peg3* in alpha cells. Furthermore, we identified a number of potential developmental timepoint-specific GRN TFs that may drive changes in proliferation and maturation that occur in endocrine cells as development progresses (53,58,60). Although here our focus within the epithelial compartment was on the endocrine lineages, our dataset also provides a rich resource for future interrogation of gene regulatory networks controlling acinar and ductal cell fates. Identification of these networks will inform efforts underway at generating stem cell-derived exocrine cells *in vitro* for studies aimed at understanding exocrine cell physiology and modeling of diseases such as cystic fibrosis, pancreatitis, and pancreatic cancer (152,153).

Traditional single-gene studies, along with genomic studies, have led to the identification of numerous TFs that play a functional role in regulating pancreatic endocrine differentiation. Although some individual TF-TF interaction pairs have been identified through these methods, the field still lacks an understanding of how these TFs are broadly arranged in regulatory networks across cell types and developmental stages. Our analysis permitted the creation of a TF-TF regulatory network, identifying TFs that control cell fate decisions through the binding and regulation of other important TFs. The assembly of these networks has identified both known

and novel TF-TF interacting pairs whose associations can be experimentally validated in future studies using tools such as Chromatin Immunoprecipitation Sequencing (ChIP-Seq) for confirmation of binding to specific DNA regions. Furthermore, CRISPR-mediated gene editing can be used to assess the downstream consequences of loss of individual candidate TFs on cell fate outcomes.

One transcription factor that we focused on in this work is the gene *Fev*. In a beta cell line, *Fev* has been demonstrated to bind not only to serotonergic genes, reflective of common transcriptional cascades that drive the differentiation of both serotonergic neurons and of beta cells (154), but also to a conserved insulin gene regulatory element (144). To investigate ubiquitous versus tissue-specific regions bound by Fev, we compared target peaks of *Fev* from our GRN analysis to *Fev* ChIP-Seq data obtained from developing serotonergic neurons and found some overlap of *Fev*-occupied chromatin between the two datasets. To probe downstream targets of Fev, we assessed the intersection of target genes of Fev in our embryonic pancreatic endocrine GRN with genes differentially expressed between Fev WT and knockout embryonic serotonergic neurons. This analysis identified candidate genes directly regulated by Fev in both tissue types. Generation of ChIP-Seq data from embryonic pancreas will permit further validation of our GRN analysis in a tissue-specific manner.

In contrast to the pancreatic epithelium, the cellular composition and transcriptional features of the pancreatic mesenchyme have been less well described. We and others have applied scRNA-Seq to mesenchymal tissue to identify transcriptionally distinct sub-populations (53,58) and infer lineage relationships among some of these cell subtypes (*51*). In addition, functional heterogeneity among pancreatic mesenchymal cells has begun to be explored. For instance, one study reported that expression of *Pbx1* in a subset of Nkx2-5+ mesenchymal cells defines an anatomically specialized, pro-endocrine niche (106). Which genes, including TFs, govern the acquisition of mesenchymal subpopulation identity, however, is poorly understood. Our work begins to investigate novel TFs regulating mesenchymal cell fate and will serve as an

important resource for understanding mesenchymal development and function. Our dataset identifies which gene networks should be activated in order to generate not only organ-specific mesenchyme, but mesenchymal subtypes from pluripotent stem cell sources (154). Generation of stem cell-derived mesenchymal subtypes would then permit their co-culture with stem cell-derived endocrine cells in order to recapitulate *in vitro* the endogenous signaling events between mesenchyme and endocrine cells during development. In addition, the CellChat analysis presented here provides a complementary set of data by identifying candidate secreted mediators of cell-cell signaling between mesenchymal and endocrine sub-populations.

The map of chromatin accessibility generated here not only provides deeper understanding of fundamental mechanisms underlying genetic control of developmental programs, but also holds relevance to the translational goals of beta cell regeneration and cell replacement therapy. For instance, our comprehensive characterization of chromatin state across endocrine development provides insights into the lineage plasticity observed among endocrine cells (94), and future work can leverage information about active endocrine cell type-specific GRNs to improve strategies for trans-differentiation of non-beta endocrine cells to the beta cell fate. Furthermore, the generation of functionally mature beta cells from hPSCs remains a strong focus of cell replacement therapeutic strategies for patients with diabetes, and such *in vitro* protocols would benefit from an improved understanding of the dynamics in chromatin accessibility across endocrine development *in vivo*. Our dataset identifies which GRNs should be modulated *in vitro* to better approximate *in vivo* development. For instance, it will be interesting to benchmark recently published multi-omic datasets of hPSCs undergoing differentiation to a beta cell fate (91,155) against our multi-omic dataset generated here to evaluate the fidelity of cells generated *in vitro* to their *in vivo* counterparts.

2.4 Materials and Methods

Animal studies

All mouse procedures were approved by the University of California, San Francisco (UCSF) Institutional Animal Care and Use Committee (IACUC). Mice were housed in a 12-hour light-dark cycle in a controlled temperature climate. Noon of the day of a vaginal plug was considered embryonic day (E)0.5.

eFev-EYFP (ePet1-EYFP) mice were kindly donated by Dr. Evan Deneris, and have been previously described (116,117). Mice were maintained on a C57BL/6J background. Wildtype C57BL/6J mice used for breeding and for the whole pancreas samples were obtained from the Jackson Laboratory. Genotyping of eFev-EYFP mice was conducted on tail DNA, with forward primer TGCGATGGGAAGATAAGAGGGG and reverse primer GAAGTTCACCTTGATGCCGTTC.

Histology, immunofluorescence, and imaging

E14.5 pancreata were dissected in ice cold 1x PBS, then fixed in 4% paraformaldehyde (PFA) overnight at 4 C. After washing three times in 1x PBS, tissues were preserved in 30% sucrose in PBS at 4 C overnight and then embedded in Optimal Cutting Temperature (O.C.T.) compound (Tissue-Tek) and flash frozen prior to sectioning at 10 μm thickness.

For immunofluorescence staining, cryosections were washed 3 times in 1x PBS, permeabilized in 0.5% triton X-100 in PBS (PBT) for 10 minutes at room temperature (RT), and then blocked with 5% normal donkey serum (NDS) in 0.1% PBT for 1 hour. Sections were stained overnight at 4 C using primary antibodies against GFP (1:500, Abcam Cat. ab13970), Chga (1:250, Abcam Cat. ab15160), or E-cadherin (1:100, BD Transduction Laboratories Cat. 610182). The next day, sections were washed three times in 1x PBS and then incubated with species-specific Alexa 488-, 555-, or 647-conjugated secondary antibodies and DAPI in 5%

NDS in 0.1% PBT for 1 hour at RT. Sections were washed three times in 1x PBS and covered in Fluoromount-G mounting solution (SouthernBiotech, Cat. 0100-01).

Images were captured with an SP8 Leica confocal laser scanning microscope. Maximum intensity Z-projections were then prepared using Image J software (156).

In situ hybridization

Multiplexed *in situ* hybridization/immunofluorescence was performed with RNAscope technology using probes purchased from Advanced Cell Diagnostics, Inc. Probes against mouse *Fev* (Cat. 413241), *Ngn3* (Cat. 422401), *Mlxipl* (Cat. 558141), *Etv1* (Cat. 557891), *Nhlh1* (Cat. 585751), *Spp1* (Cat. 435191), and *Rest* (Cat. 316251) were used according to the manufacturer's instructions for the RNAscope multiplex fluorescent detection V2 kit (Advanced Cell Diagnostics, Inc., Cat. 323110). 10 µm thick cryosections were brought to RT, washed with PBS to remove O.C.T., and treated with hydrogen peroxide and proteinase III. Tissue was hybridized with the probe mixture for 2 hours at 40 C. Hybridization signals were amplified via sequential hybridization of amplifier AMP1, AMP2, and AMP3 and label probes Opal 570 (1:1500, PerkinElmer, Cat. FP1488001KT), Opal 650 (1:1500, PerkinElmer, Cat. FP1496001KT), and Opal 690 (1:1500, PerkinElmer, Cat. FP1497001KT).

Following signal amplification of the target probes, sections were incubated in 1x blocking buffer for 1 hour at RT, followed by staining with primary antibodies against GFP (1:500, Abcam Cat. ab13970), Epcam (1:100, BD Biosciences, Cat. 552370), Ins (1:100, Abcam Cat. ab7842), or Gcg (1:250, Cell Signaling Technology Cat. 2760S). The next day, sections were washed three times with 1x PBS and then incubated with species-specific Alexa 488- or Alexa 555-secondary antibodies and DAPI in 5% NDS in 0.1% PBT for 1 hour at RT. Sections were then washed three times in 1x PBS, mounted with ProLong Gold Antifade Mountant (Invitrogen, Cat. P36930) and stored at 4 C prior to imaging. Optical sectioning images were

taken with a Leica confocal laser scanning SP8 microscope equipped with white light sources. 10 steps X 1 mm thickness Z-sections were captured for each imaging area.

Dissociation and sorting of murine pancreas tissue for quantitative RT-PCR

E14.5 pancreata were dissected from embryos of pregnant eFev-EYFP dams and kept in separate wells of a 96-well plate. EYFP fluorescence was assessed under a microscope to confirm the genotype of each pancreas. Pancreata with EYFP fluorescence (EYFP(+)) were then pooled together and transferred to a 1.5 ml microcentrifuge tube, then dissociated into single cells by incubating with 250 ul of TrypLE Express dissociation reagent (Gibco, Cat. 12604013)) at 37 C for 20 minutes, with pipet trituration at 5-minute intervals. Dissociation was neutralized with FACS buffer (10% FBS + 2 mM EDTA in phenol-red free HBSS), and the single-cell suspensions were passed through 30 μm cell strainers.

Cells were stained with SYTOX Blue dead cell stain (Invitrogen, Cat. S34857) to remove dead cells, then with a PE-conjugated antibody against mesenchymal marker CD140a (1:50; eBioscience Cat. 12-1401-81) and an APC-conjugated antibody against epithelial marker CD326/EpCam (1:50; eBioscience Cat. 17-5791-82) at 4 C for 30 minutes. Stained cells were washed twice in FACS buffer and sorted using a BD FACSAria II cell sorter (BD Biosciences). After size selection to remove debris and doublets and sorting on SYTOX Blue negative (live) events, cells were further subgated on CD140a(-)/CD326(+) (epithelial) cells and then on EYFP fluorescence.

RNA was extracted from EYFP(-), EYFP-low, and EYFP-high sorted cells with the RNeasy Mini Kit (Qiagen, Cat. 74106). Reverse transcription was performed with the PrimeScript High Fidelity RT-PCR Kit (Takara, Cat. R022A). RT-PCR was run on a 7900HT Fast RT-PCR instrument (Applied Biosystems) with Taqman probes for *Fev* (assay ID: Mm00462220_m1, Cat. 4331182) and *GAPDH* (assay ID: Mm99999915_g1, Cat. 4331182) in

triplicate. Data were normalized to *GAPDH*. Error bars represent standard error of the mean (SEM).

Dissociation and sorting of murine pancreas tissue for snATAC-Seq

For the whole pancreas samples, C57BL/6J embryonic pancreata (n=10 from 3 litters for E14.5 replicate 1 and n=16 from 3 litters for E14.5 replicate 2; n=7 from 1 litter for E17.5 replicate 1 and n=3 from 1 litter for E17.5 replicate 2) were pooled into a single 1.5 ml microcentrifuge tube per replicate, per timepoint. Biological replicate experiments for each timepoint were performed on different days. For the eFev-EYFP sample, E14.5 pancreata (n=15) were dissected from embryos of two pregnant eFev-EYFP dams, and each pancreas was kept in a separate well of a 96-well plate. EYFP fluorescence was assessed under a microscope to confirm the genotype of each pancreas. E14.5 pancreata with EYFP fluorescence (EYFP(+)) (n=5) were then pooled together, pancreata without EYFP fluorescence (EYFP(-)) (n=10) were pooled together as a negative control, and each sample was transferred to a separate 1.5 ml microcentrifuge tube.

The E14.5 whole pancreas, E17.5 whole pancreas, E14.5 EYFP(+), and E14.5 EYFP(-) samples were dissociated into single cells by incubating with 250-350 ul per sample of TrypLE Express dissociation reagent (Gibco, Cat. 12604013)) at 37 C for 20 minutes, with pipet trituration at 5-minute intervals. Dissociation was neutralized with FACS buffer (10% FBS + 2 mM EDTA in phenol-red free HBSS) and the single-cell suspensions were passed through 37 μm cell strainers.

All samples were stained with SYTOX Blue dead cell stain (Invitrogen, Cat. S34857) to remove dead cells. Cells were washed twice with FACS buffer and sorted using a BD FACSAria II cell sorter (BD Biosciences). After size selection to remove debris and doublets, all cells were sorted on SYTOX Blue negative (live) events, and the E14.5 EYFP(+) and EYFP(-) samples were further sub-gated on EYFP fluorescence. Live cells from whole pancreas and live EYFP(+)

cells from the EYFP(+) sample were collected into separate tubes containing 1x FACS buffer and immediately subjected to extraction of nuclei as described below.

Extraction of nuclei

All buffers (e.g., 0.1x lysis buffer, lysis dilution buffer, and wash buffer) were freshly prepared according to the 10x Genomics Demonstrated protocol (CG000212 RevC) and maintained at 4 C. Nuclei were isolated from whole pancreas (25,000 cells for E14.5 replicate 1 and 50,000 cells for E14.5 replicate 2; 1,000,000 cells for E17.5 replicate 1 and 500,000 cells for E17.5 replicate 2) or from Fev-high (EYFP(+)) cells (25,000 cells for E14.5) using the demonstrated protocol. Sorted cells were added to a 2 ml microcentrifuge tube and centrifuged at 500 rcf for 5 minutes at 4 C. All supernatant was removed without disrupting the cell pellet. 100 ul chilled 0.1x lysis buffer was then added and pipetted 5 times to fully mix the buffer with the cells, then incubated for 3 minutes on ice to achieve full cell lysis. 1 ml chilled wash buffer was added to the lysed cells to terminate the lysis. Lysed cells were centrifuged at 500 rcf for 5 minutes at 4 C, and supernatant was gently removed. Nuclei were resuspended in 50 ul wash buffer, transferred to a 200 ul tube, and spun down and resuspended in 10 ul 1x Nuclei buffer (10x Genomics, Part Number 2000153). 2 ul of the suspension was loaded onto a hemacytometer to determine the concentration of nuclei and simultaneously assess nucleus quality. High-quality nuclei from the whole pancreas and eFev-EYFP samples were then used for downstream library construction and sequencing.

snATAC-seq capture, library construction, and sequencing

Input nuclei were subjected to transposition, partitioning, and library construction using 10x Genomics Chromium Next GEM Single Cell ATAC Reagent Kit v1.1 Chemistry, according to the manufacturer's instructions. An Agilent Fragment Analyzer was used for assessing the

fragment distribution of both the whole pancreas and eFev-EYFP libraries, which were run on the Illumina NovaSeq 6000 platform.

Clustering of murine scRNA-Seq data

For clustering of murine scRNA-Seq data for integration with our snATAC-Seq data, we applied the clustering algorithm CellFindR (122) to our previously published scRNA-Seq dataset of developing murine pancreas tissue (58). 10x Genomics outputs of E14.5 and E17.5 pancreata were downloaded from the Gene Expression Omnibus (GEO) (GSE101099; samples GSM3140916, GSM3140919, GSM3140920, GSM2699157, GSM3140917, and GSM3140918), and analyzed with Seurat v3.2.3. Seurat objects were created from each 10x output with Read10x() and CreateSeuratObject() and filtered to retain high quality cells (nFeature_RNA > 1250 and percent.mt < 7 for GSM3140916; nFeature_RNA > 1500 and percent.mt < 5 for GSM3140919 and GSM3140920; nFeature_RNA > 500 and < 3000 and percent.mt < 5 for GSM2699157; nFeature_RNA > 500 and < 5000 and percent.mt < 5 for GSM3140917 and GSM3140918). The datasets were then normalized and variable features calculated with NormalizeData() and FindVariableFeatures(), respectively. The samples were then integrated using Seurat's standard batch correction method (157) with SelectIntegrationFeatures(), FindIntegrationAnchors() and IntegrateData(). The integrated object was then scaled with ScaleData() and principal component analysis (PCA) performed with RunPCA(). UMAP dimensional reduction was calculated with RunUMAP() with dims = 1:30. Neighbors were found in the dataset with FindNeighbors() with dims = 1:30 and clustering performed with FindClusters(), resolution = 0.2 for E14.5 and 0.4 for E17.5. Next, broad cell types were manually annotated based on expression of known marker genes (i.e. *Col3a1* for mesenchyme) and used for subsequent iterative sub-clustering.

For iterative subclustering with CellFindR, each broad cell type (Mesenchyme, Mesothelium, Exocrine, and Endocrine) was subsetted individually. PCA, Neighbors and UMAP

were recalculated as described above (Mesenchyme: dims 1:15; Mesothelium: dims 1:15 for E14.5 and 1:10 for E17.5; Exocrine: dims 1:10 for E14.5 and 1:15 for E17.5; and Endocrine: dims 1:10) and the first clustering resolution calculated with find_res() from CellFindR. Iterative subclustering was then performed with sub_clustering(). Subclusters that displayed characteristics of doublets (expressing markers of more than one broad group e.g., *Col3a1+*/*Cpa1+* acinar cells) or low quality (e.g. clustering based on high mitochondrial gene content) were manually removed.

snATAC-Seq analysis

FASTQ files were generated from raw sequencing reads using the bcl2fastq function from Illumina. BAM files and single-cell accessibility counts were generated using the cellranger-atac count function from Cell Ranger software (version 1.0.1 for the E14.5 datasets, version 2.1.0 for the E17.5 datasets). Reference genome used was Mus musculus assembly mm10, annotation gencode.vM17.basic. Files processed with Cell Ranger ATAC were then analyzed using ArchR (version 1.0.1) (121).

Unless otherwise noted, parameters and function calls were kept the same between the E14.5 and E17.5 datasets. First, ArchR Arrow files were created with the ArrowFiles() function with default settings. For 14.5, an ArchR project was then created using both whole pancreas and EYFP(+) sorted cells with ArchRProject(), resulting in an initial dataset consisting of 46,453 nuclei. For E17.5, an ArchR project was created using the whole pancreas sample, resulting in an initial dataset of 113,071 nuclei. Each project was then filtered for high quality nuclei (TSS enrichment >= 10 and number of fragments >= 3,000) (39,845 nuclei retained in E14.5; 102,644 retained in E17.5), and doublets removed with addDoubletScores() and filterDoublets() (35,823 nuclei retained in E14.5; 94,175 retained in E17.5). Next, iterative LSI was performed on both E14.5 and E17.5 datasets with the addIterativeLSI() function, with clustering parameters of resolution = 0.2, sampleCells = 10,000, n.start = 10, varFeatures = 25000. Clustering was

performed with addClusters() with resolution = 0.1 and method = “Seurat”. UMAP dimensional reduction was performed with addUMAP(), with minDist = 0.5. Clusters were manually annotated based on the Gene Score of known marker genes with addImputeWeights() and then visualized by UMAP. For E14.5, cluster 9 was removed as no discernable markers for the known pancreatic cell types were detected, resulting in a final dataset consisting of 33,206 total cells with a median number of 14,756 fragments per cell and a median TSS enrichment score of 14.914. For E17.5, each Broad Group was subsetted and re-clustered at higher resolution in order to separate out potential doublet clusters or empty droplets that were not removed with the doublet filtering detailed above. Doublet/empty droplet clusters were removed from the Mesenchymal, Epithelial and Endothelial Broad Groups, resulting in a final dataset consisting of 78,669 nuclei with a median number of 15,965 fragments per cell and a median TSS enrichment score of 21.037.

For epithelial analysis, epithelial (exocrine and endocrine) nuclei were subsetted based on accessibility of known marker genes (*Cpa1, Spp1, Chga*). Iterative LSI was recalculated with iterations = 4 for E14.5 and = 2 for E17.5, resolutions = 0.1, 0.2, 0.4, and 0.5 for E14.5 and = 0.5 for E17.5, sampleCells = 5,000 for E14.5 and 10,000 for E17.5, n.start = 10, and varFeatures = 15000 for E14.5 and 25,000 for E17.5. Clustering was performed with resolution = 0.9 for E14.5 and = 1.2 for E17.5, and UMAP recalculated with minDist = 0.5. Clustered epithelial cells from the scRNA-Seq data described above were used for unconstrained integration with addGeneIntegrationMatrix(). Chromatin accessibility peaks were then called with Macs2 via ArchR with addGroupCoverages(), addReproduciblePeakSet() and addPeakMatrix(). Marker peaks within the epithelial compartment were calculated with getMarkerFeatures() using the “PeakMatrix”. For motif analysis within marker peaks, motif annotations were added with addMotifAnnotations() with the “cisbp” motif set and then calculated with peakAnnoEnrichment(). ChromVAR (124) analysis was performed with addBgdPeaks() and addDeviationsMatrix(). Correlated transcription factors were correlated

between the “GeneIntegrationMatrix” (RNA expression from the unconstrained integration) and “MotifMatrix” (ChromVAR motif deviations) with correlateMatrices(), keeping TFs with a correlation > 0.5, padj < 0.01 and max delta greater than 0.5 of the upper quartile.

For pseudotime lineage calculations, we manually imputed the cell states for each cell lineage (Alpha and Beta for E14.5; Alpha, Beta and Delta for E17.5) and computed the pseudotime values with addTrajectory().

To compare peaks across the E14.5 whole pancreas and EYFP(+) sorted samples, we first subsetted the epithelial dataset to retain all cell types that had sufficient cell numbers to call peaks after being split by sample. This included the following clusters: Neurog3(+), Fev(+)/Chgb(+), Pdx1(+)/Mafb(+), Alpha, and Beta cells. We then split this subsetted dataset according to the sample from which they were derived and then re-called peaks with addGroupCoverages(), addReproduciblePeakSet() and addPeakMatrix(). We extracted these peaks with getPeakset() for each sample and assessed overlap with findOverlaps() from the R package IRanges (https://github.com/Bioconductor/IRanges).

For mesenchymal and mesothelial analysis in the E14.5 datasets, nuclei were subsetted based on accessibility of known marker genes (*Col3a1, Wt1*). Iterative LSI was recalculated with iterations = 2, resolution = 0.5, sampleCells = 10,000, n.start = 10, and varFeatures = 25000. Clustering was performed with resolution = 0.3, and UMAP recalculated with minDist = 1.5. Clustered mesenchymal and mesothelial cells from the scRNA-Seq data described above were used for unconstrained integration. Peak calling, marker peak identification, motif analysis, ChromVAR analysis and transcription factor correlation were performed as described above.

Gene regulatory network analysis

Gene regulatory networks (GRNs) were constructed as described by Lyu et al., 2021 (113) (https://github.com/Pinlyu3/IReNA-v2). Candidate cluster-enriched genes were calculated with the scRNA-Seq dataset of epithelial or mesenchymal cells with Seurat’s FindAllMarkers()

with min.pct = 0.1 and logfc.threshold = 0.25. DEGs were retained with an average logFC > 0 and padj < 0.01. DEGs were then mapped to specific cell types with the IReNA v2 function Process_DEGs_to_Celltypes().

Peak-to-gene linkage was performed with ArchRs addPeak2GeneLinks() function for both epithelium and mesenchyme using the “GeneIntegrationMatrix” (integrated RNA Seq counts). Peak-to-gene links were then extracted with the IReNA v2 function Get_p2g_fun().

To identify potential *cis*-regulatory elements for each candidate gene, called correlated accessible regions (CARs), we separated the peak-to-gene links into three categories: TSS (when the peak lies within the transcription start site (TSS) for the gene), gene body (when the distance between peak and TSS is less than 100kb, and the peak-to-gene score calculated above is significant), or distal (when the peak is 100kb upstream or downstream of the TSS of correlated gene, and the peak-to-gene score is significant). These peak-to-gene links were then filtered to only include genes in the DEG list calculated above with Selection_peaks_to_one().

Next, we predicted the cell-type specific transcription factors binding in these CARs. We first took the snATAC fragments for each dataset (whole pancreas and Fev-high) and then extracted the fragments for each cell type. We converted these fragment lists to .BAM files and corrected the Tn5 insertion bias with TOBIAS (158) ATACorrect with default parameters except --read_shift 0 0. We then converted the TOBIAS output bigwig files to GRanges with the IReNA v2 function Check_normalized_Signal(). Next, we calculated TF binding motifs in our peaks with motifmatcher (https://github.com/GreenleafLab/motifmatchr), filtering calculated TFs out from the motif analysis if they were not enriched in each cell type by the DEG analysis. Next, we calculated the NC (average bias-corrected Tn5 signal in the center of the motif), NL and NR (average bias-corrected Tn5 signal in the left and right flanking regions of the motif) scores with Calculate_footprint_celltypes() and filtered TFs with a score of NC < -0.1 and NL > 0.1 and NR > 0.1.

Next, we used MAGIC (Mining Algorithm for GenetIc Controllers) (159) to compute correlation between TF and target gene gene expression. We retained the top and bottom 2.5% of correlations for our downstream analysis.

Lastly, we constructed the cell-type specific GRNs. We combined the peak-target links from our third step with the cell-type specific TF-peak links from our fourth step with Reg_one_cells_RPC_MG(). We then classified these interactions as either activating or repressing with our TF-target gene interactions calculated above with Add_Cor_to_GRN_network_and_Filter(). We then identified feedback TF-TF pairs in our constructed GRN with FoundFeedBackPairs_new() and Process_the_Feedback_res().

Confirmation of *Fev* GRN peaks and targets

Peaks from *Fev* ChIP-Seq data detected in E12.5-E15.5 hindbrain were obtained from GEO (accession number GSE74315), from the supplementary file labeled "GSE74315_mycPet-1_ChIP_peaks.mm10.bed.gz." *Fev*-occupied CARs from the Fev(+)/Chgb(+) GRN were taken written in .bed format. These peaks were then intersected using the bedtools software (https://github.com/arq5x/bedtools2) with default settings. Differentially expressed genes between *Fev* wildtype and knockout E15.5 hindbrains were downloaded from GEO (accession number GSE74315) under the supplementary file labeled "GSE74315_WT_vs_Pet1KO_5-HT_neuron_E15.5C_Clontech_cuffdiff.txt.gz."

Data Availability

Raw and processed data for the single-nucleus ATAC-Sequencing analyses reported in this manuscript have been deposited to the GEO database under accession number GSE210569. The authors declare that all data supporting the findings of this study are available within the

article and its supplementary information files, or from the corresponding author upon reasonable request.

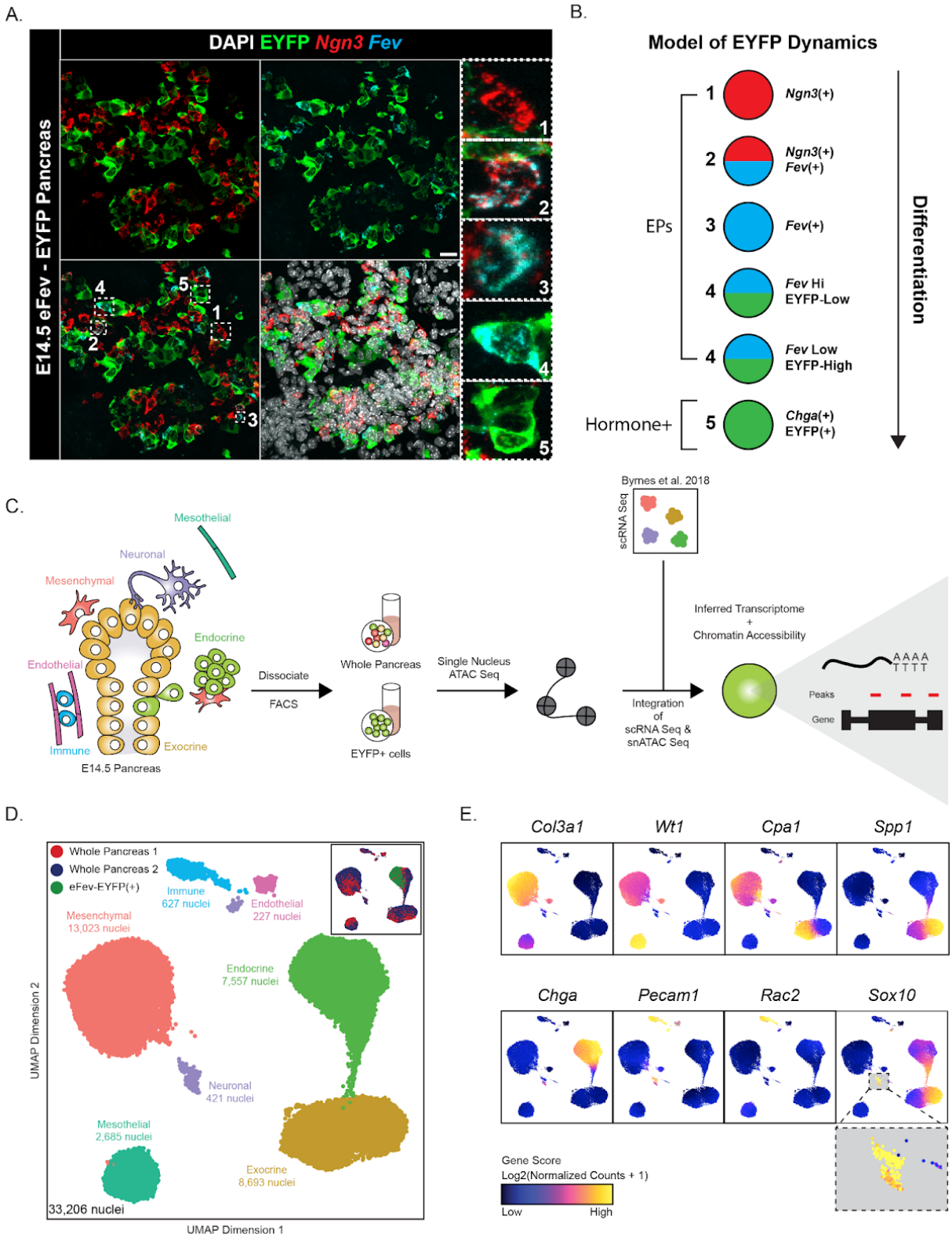
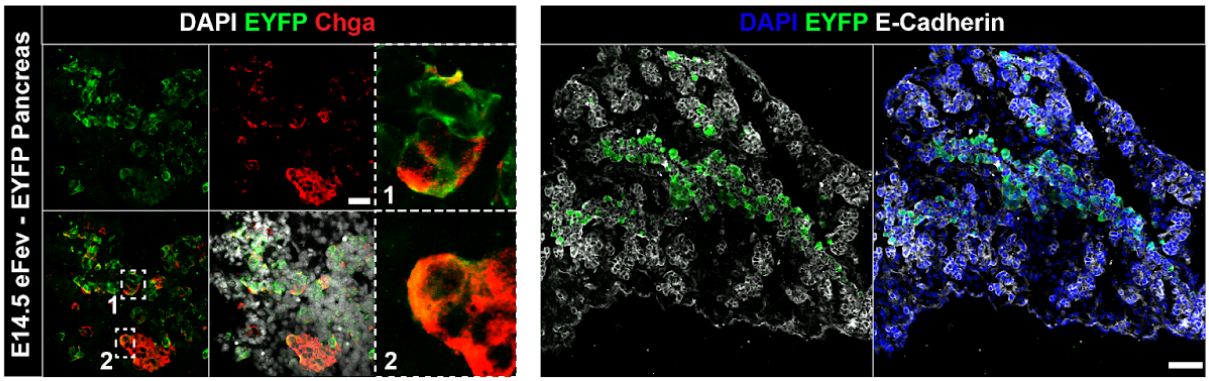

A.
DAPI EYFP Ngn3 Fev
E14.5 eFev - EYFP Pancreas
B.
Model of EYFP Dynamics
1 Ngn3(+)
2 Ngn3(+) Fev(+)
EPs
3 Fev(+)
4 Fev Hi EYFP-Low
4 Fev Low EYFP-High
Hormone+
5 Chga(+) EYFP(+)
Differentiation
C.
Mesothelial
Neuronal
Mesenchymal
Endocrine
Endothelial
Immune
Exocrine
E14.5 Pancreas
Dissociate
FACS
Whole Pancreas
EYFP+ cells
Single Nucleus ATAC Seq
Byrnes et al. 2018
scRNA Seq
Inferred Transcriptome + Chromatin Accessibility
Integration of scRNA Seq & snATAC Seq
Peaks
Gene
D.
Whole Pancreas 1
Whole Pancreas 2
eFev-EYFP(+)
Immune 627 nuclei
Endothelial 227 nuclei
Mesenchymal 13,023 nuclei
Endocrine 7,557 nuclei
Neuronal 421 nuclei
Mesothelial 2,685 nuclei
Exocrine 8,693 nuclei
33,206 nuclei
UMAP Dimension 2
UMAP Dimension 1
E.
Col3a1
Wt1
Cpa1
Spp1
Chga
Pecam1
Rac2
Sox10
Gene Score
Log2(Normalized Counts + 1)
Low
High

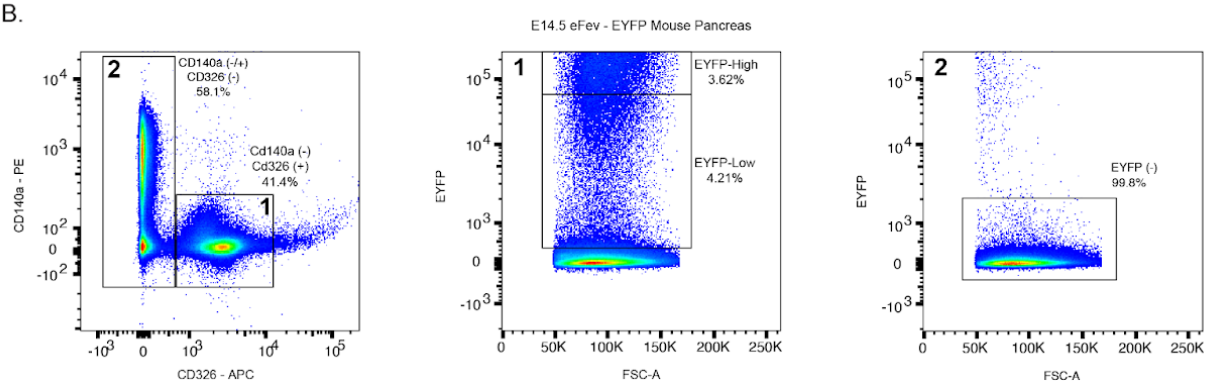
Figure 2.1: Single-nucleus ATAC-Seq of developing murine pancreas.

(**A**) Multiplexed immunofluorescence and *in situ* hybridization staining of E14.5 pancreas from eFev-EYFP transgenic mouse embryos. *Neurog3* transcript is shown in red and marks early endocrine progenitors (EPs), *Fev* transcript is shown in cyan and marks intermediate EPs, and EYFP protein is shown in green. DAPI marks nuclei in white. Five selected regions of interest (ROIs) are outlined by dashed white rectangles on the merged image and shown at higher magnification to the right. These ROIs highlight examples of cells that are 1) Ngn3(+)/Fev(-)/EYFP(-); 2) Ngn3(+)/Fev(+)/EYFP(-); 3) Ngn3(-)/Fev(+)/EYFP(-); 4) Ngn3(-)/Fev(+)/EYFP(+); and 5) Ngn3(-)/Fev(-)/EYFP(+). Scale bar is 20 μm. (**B**) Model of Fev and EYFP expression in EPs undergoing differentiation to a hormone-producing, Chromogranin A (Chga)-expressing state in eFev-EYFP reporter mice. Each circle represents a cell state across endocrine differentiation. (**C**) Overview of experimental approach for generating single-nucleus ATAC-Seq (snATAC-Seq) data. To enrich for Fev-high EPs, E14.5 eFev-EYFP murine pancreas was dissociated, and the resulting single cell suspension was subjected to FACS to enrich for EYFP(+) epithelial cells ("EYFP+ Cells"). In parallel, E14.5 pancreata from control (C57BL/6J) embryos were dissociated and subjected to FACS to isolate all live cells ("Whole Pancreas") to profile a broad spectrum of cell types, including non-epithelial cells. After subjecting samples to snATAC-Seq, data were then integrated with previously-published single-cell RNA-Sequencing (scRNA-Seq) datasets of E14.5 murine pancreas previously published by our laboratory (58). (**D**) Uniform Manifold Approximation and Projection (UMAP) visualization of merged snATAC-Seq datasets from both Whole Pancreas (biological replicates 1 and 2) and EYFP(+) samples, comprising a total of 33,206 nuclei. Each dot represents a single cell, and each cell is colored according to cell type. Contribution of each sample (Whole Pancreas and eFev-EYFP(+)) to the total dataset is depicted in the inset, with the eFev-EYFP(+) sample contributing only to the endocrine cluster, as expected. (**E**) Feature plots depicting the Gene Scores (accessibility of the gene promoter plus the gene body) for some of the marker genes used to annotate the cell types in panel D. Inset for Sox10 shows Sox10-expressing neuronal cells at higher magnification.

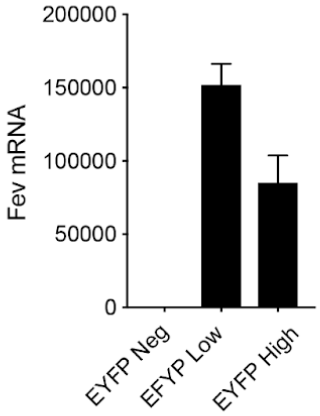
A.


E14.5 eFev - EYFP Pancreas
DAPI EYFP Chga
DAPI EYFP E-Cadherin
1
2
1
2

B.


E14.5 eFev - EYFP Mouse Pancreas
2
CD140a (-/+)
CD326 (-)
58.1%
Cd140a (-)
Cd326 (+)
41.4%
1
CD140a - PE
CD326 - APC
1
EYFP-High
3.62%
EYFP-Low
4.21%
EYFP
FSC-A
2
EYFP (-)
99.8%
EYFP
FSC-A

C.


Fev mRNA
200000
150000
100000
50000
0
EYFP Neg
EFYP Low
EYFP High

D.

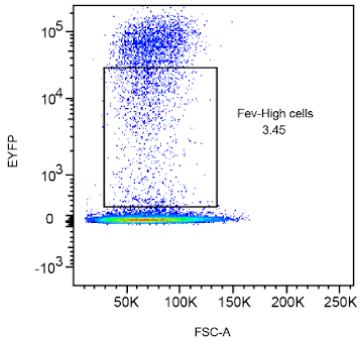

Fev-High cells
3.45
EYFP
FSC-A

Figure 2.2: Strategy for enrichment of Fev-expressing pancreatic endocrine progenitors.

(**A**) Immunofluorescence staining of E14.5 pancreas from eFev-EYFP transgenic mouse embryos. In the left panel, Chga marks differentiated endocrine cells in red, and green reflects EYFP expression. Nuclei are stained with DAPI in white. In the right panel, E-cadherin marks the cell membranes of pancreatic epithelial cells in white, and green reflects EYFP expression. Nuclei are stained with DAPI in blue. Scale bars iare 20 μm (left panels) and 60 μm (right panels). (**B**) FACS plots depicting the gating strategy used to assess EYFP expression in eFev-EYFP(+) embryos. Cells were dissociated from E14.5 pancreata and stained with antibodies against EpCAM (an epithelial marker also known as CD326) and CD140a (a mesenchymal marker also known as PDGFRa). Two populations were sorted: EpCAM(+)/CD140a(-) epithelial cells (gate 1) and EpCAM(-) non-epithelial cells, which comprise both a CD140a(+) mesenchymal and CD140a(-) population (collectively captured by gate 2). Flow analysis of EYFP expression revealed both an EYFP high and an EYFP low population within the EpCAM(+)/CD140a(-) epithelial population, and an absence of EYFP expression in EpCAM(-) non-epithelial cells. (**C**) Quantitative RT-PCR (Taqman) analysis of sorted EYFP(-), EYFP-low, and EYFP-high cells confirmed that EYFP efficiently reflected *Fev* expression in the embryonic pancreas, with no *Fev* expression detected in EYFP(-) cells. Somewhat higher expression of *Fev* was detected in the EYFP-low vs. -high population. (**D**) FACS plot depicting the gating strategy used for isolating Fev-High (EYFP-low) cells for single nucleus ATAC-Seq (snATAC-Seq).

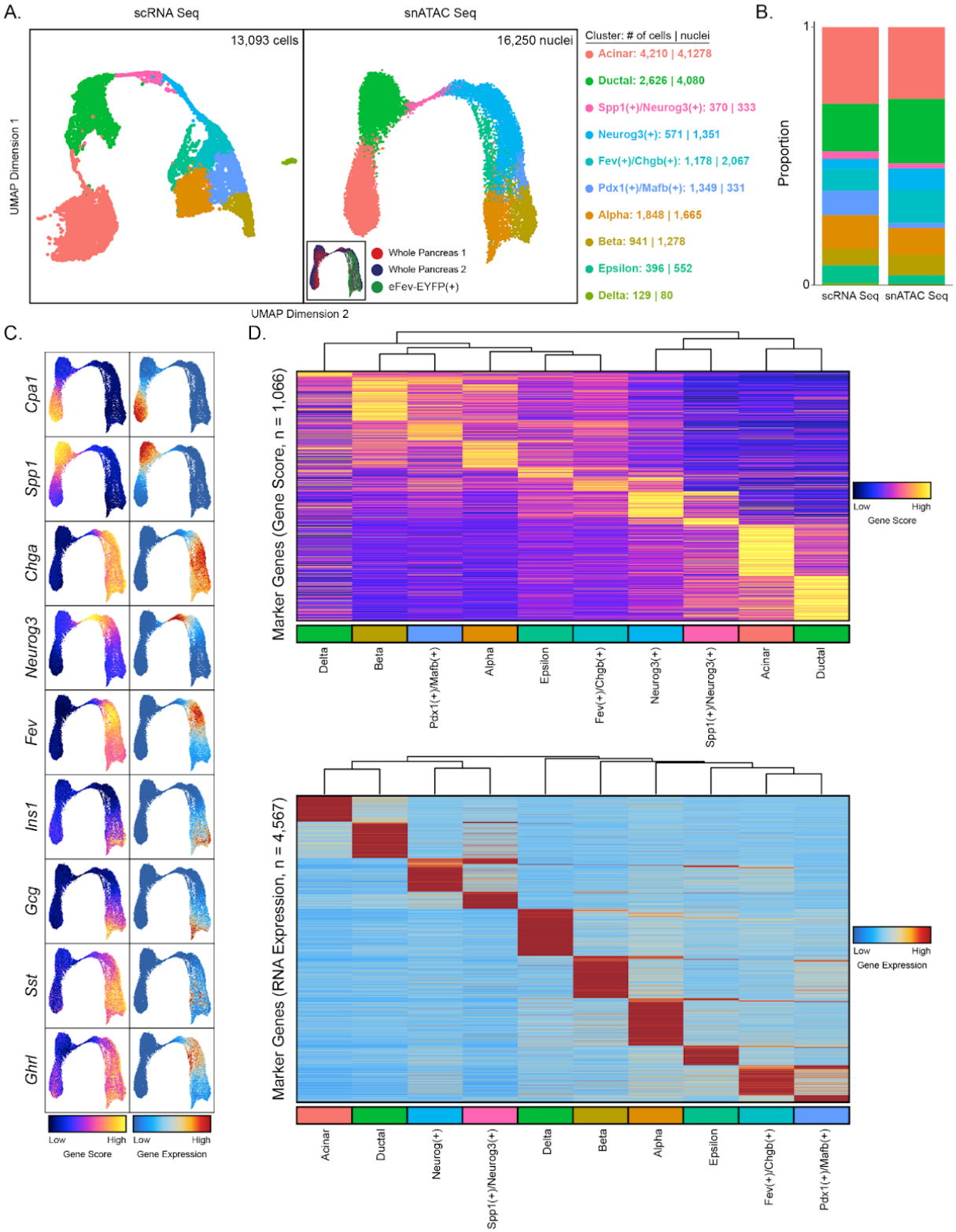

A.
scRNA Seq
snATAC Seq
13,093 cells
16,250 nuclei
UMAP Dimension 1
UMAP Dimension 2
Whole Pancreas 1
Whole Pancreas 2
eFev-EYFP(+)
Cluster: # of cells | nuclei
Acinar: 4,210 | 4,1278
Ductal: 2,626 | 4,080
Spp1(+)/Neurog3(+): 370 | 333
Neurog3(+): 571 | 1,351
Fev(+)/Chgb(+): 1,178 | 2,067
Pdx1(+)/Mafb(+): 1,349 | 331
Alpha: 1,848 | 1,665
Beta: 941 | 1,278
Epsilon: 396 | 552
Delta: 129 | 80
B.
Proportion
1
0
scRNA Seq
snATAC Seq
C.
Cpa1
Spp1
Chga
Neurog3
Fev
Ins1
Gcg
Sst
Ghrl
Low
High
Gene Score
Low
High
Gene Expression
D.
Marker Genes (Gene Score, n = 1,066)
Low
High
Gene Score
Delta
Beta
Pdx1(+)/Mafb(+)
Alpha
Epsilon
Fev(+)/Chgb(+)
Neurog3(+)
Spp1(+)/Neurog3(+)
Acinar
Ductal
Marker Genes (RNA Expression, n = 4,567)
Low
High
Gene Expression
Acinar
Ductal
Neurog(+)
Spp1(+)/Neurog3(+)
Delta
Beta
Alpha
Epsilon
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)

Figure 2.3: Integration of single-cell transcriptional and chromatin accessibility data identifies epithelial heterogeneity in the developing murine pancreas.

(**A**) UMAP plots enabling visualization of scRNA-Seq (left) and snATAC-Seq (right) data for all epithelial cells in the E14.5 pancreas. Numbers of cells or nuclei are depicted on the right, along with cell type annotations. The scRNA-Seq dataset was previously published by our group (58). **(B**) Bar graph depicting the proportion of all each cell type as a fraction of all epithelial cells in the scRNA-Seq and snATAC-Seq datasets. Colors match the cell types in (**A**). (**C**) Feature plots showing chromatin accessibility (Gene Score; left) or Gene Expression (right) of genes that mark each epithelial cell type. *Cpa1*, *Spp1*, *Chga*, *Neurog3*, *Fev*, *Ins1*, *Gcg*, *Sst*, and *Ghrl* mark acinar, ductal, pan-differentiated endocrine, early endocrine progenitor (EP), intermediate EP, beta, alpha, delta, and epsilon cells, respectively. **(D**) Heatmaps depicting genes that are differentially accessible (gene score; top heatmap; 1,066 genes) or differentially expressed (gene integration matrix; bottom heatmap; 4,567 genes) across the epithelial clusters. Genes listed were selected from the set of genes determined to be differentially expressed among scRNA-Seq clusters in panel (**A**).

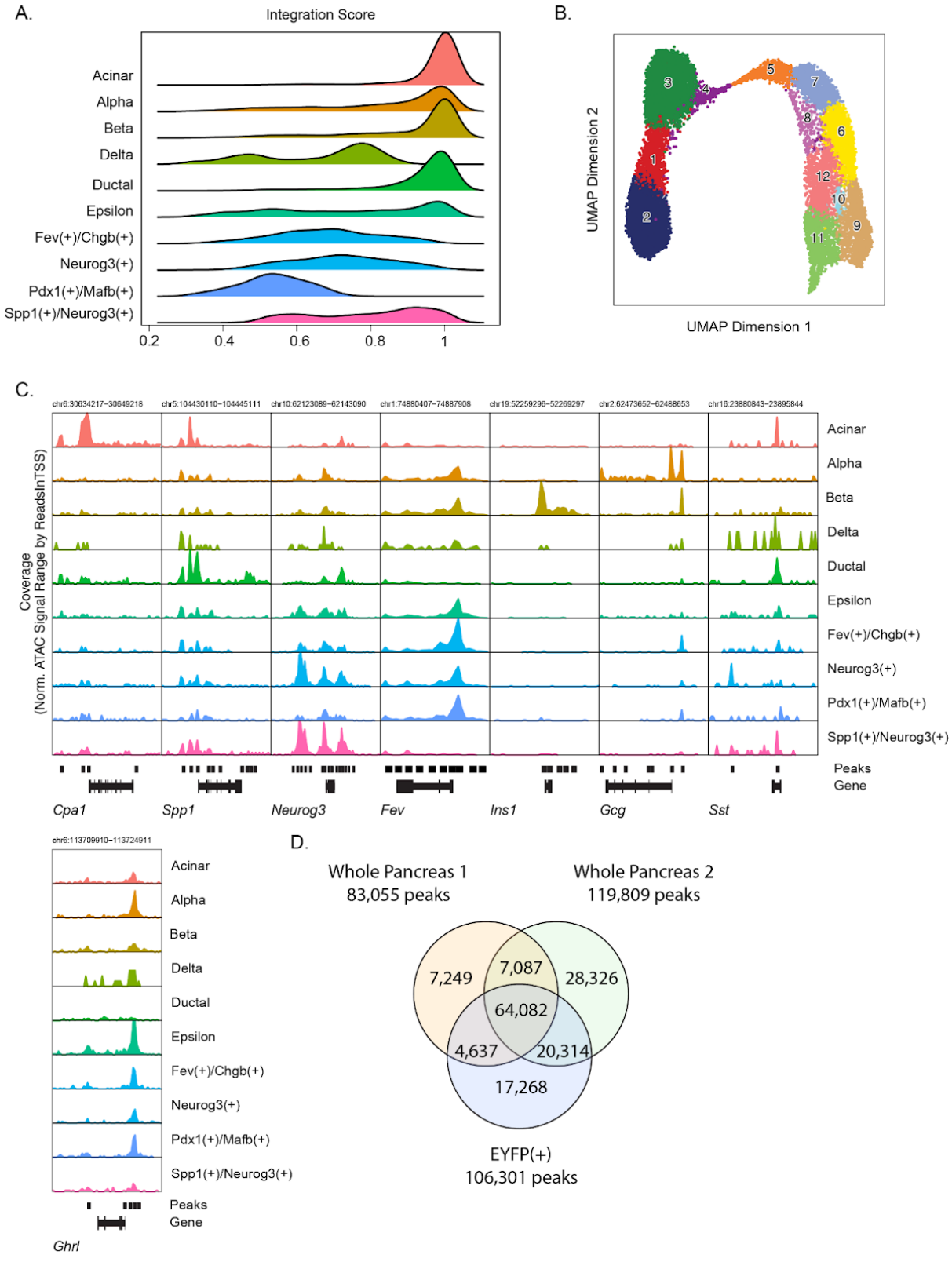

A.
Integration Score
Acinar
Alpha
Beta
Delta
Ductal
Epsilon
Fev(+)/Chgb(+)
Neurog3(+)
Pdx1(+)/Mafb(+)
Spp1(+)/Neurog3(+)
0.2
0.4
0.6
0.8
1
B.
UMAP Dimension 2
UMAP Dimension 1
C.
Coverage (Norm. ATAC Signal Range by ReadsInTSS)
Peaks
Gene
Cpa1
Spp1
Neurog3
Fev
Ins1
Gcg
Sst
Ghrl
D.
Whole Pancreas 1
83,055 peaks
Whole Pancreas 2
119,809 peaks
7,249
7,087
28,326
64,082
4,637
20,314
17,268
EYFP(+)
106,301 peaks

Figure 2.4: Confirmation of cell type assignments.

(**A**) Ridge plots depicting scores from the integration of scRNA-Seq and snATAC-Seq data for each cell type identified in **Figure 2.3a**. A higher integration score indicates higher confidence in the assignment of cell identity. (**B**) UMAP visualization of snATAC-Seq clustering of the epithelial compartment, here without integration with scRNA-Seq data. All clusters in **Figure 2.3a** are present even when clustered based on chromatin accessibility alone. (**C**) Track plots of cell type-specific marker genes. Each row indicates a cell type, and each column is a marker gene. The x-axis represents position along the chromosome, which is labeled at the top of each plot. The y-axis represents normalized ATAC signal aggregated across all cells within a population. Regions identified as peaks are depicted as bars above the gene body. (**D**) Venn diagram showing overlap of peaks called in a subset of endocrine cells (Neurog3(+), Fev(+)/Chgb(+), Pdx1(+)/Mafb(+), Alpha, and Beta cells) in the EYFP(+) sample and the whole pancreas samples (replicates 1 and 2).

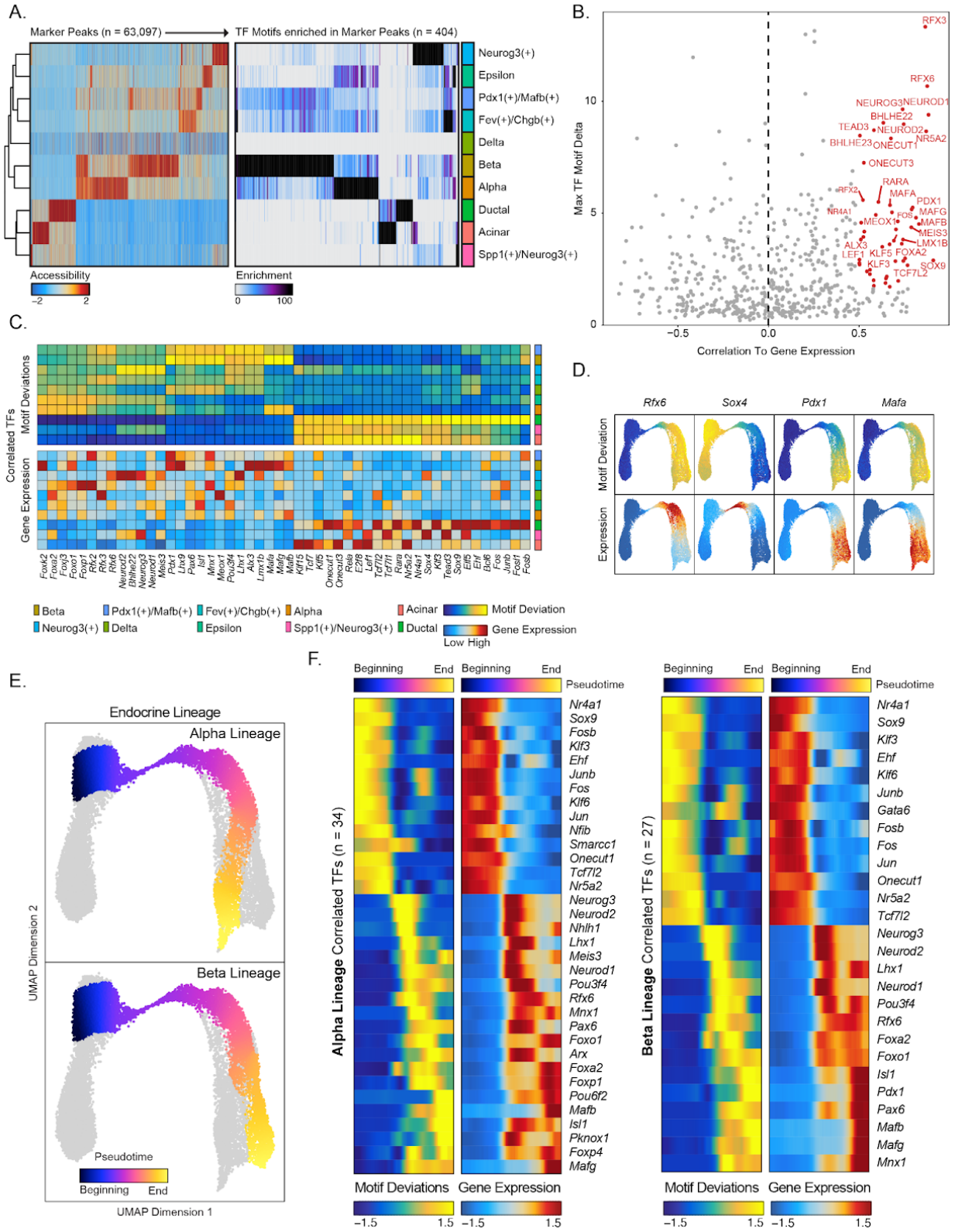

A.
Marker Peaks (n = 63,097)
TF Motifs enriched in Marker Peaks (n = 404)
Neurog3(+)
Epsilon
Pdx1(+)/Mafb(+)
Fev(+)/Chgb(+)
Delta
Beta
Alpha
Ductal
Acinar
Spp1(+)/Neurog3(+)
Accessibility
−2
2
Enrichment
0
100
B.
Max TF Motif Delta
Correlation To Gene Expression
RFX3
RFX6
NEUROG3
NEUROD1
BHLHE22
TEAD3
NEUROD2
NR5A2
BHLHE23
ONECUT1
ONECUT3
RARA
RFX2
MAFA
PDX1
NR4A1
FOS
MAFG
MEOX1
MAFB
MEIS3
ALX3
LMX1B
LEF1
KLF5
FOXA2
KLF3
SOX9
TCF7L2
C.
Correlated TFs
Motif Deviations
Gene Expression
Beta
Pdx1(+)/Mafb(+)
Fev(+)/Chgb(+)
Alpha
Acinar
Motif Deviation
Neurog3(+)
Delta
Epsilon
Spp1(+)/Neurog3(+)
Ductal
Gene Expression
Low High
D.
Rfx6
Sox4
Pdx1
Mafa
Motif Deviation
Expression
E.
Endocrine Lineage
Alpha Lineage
Beta Lineage
Pseudotime
Beginning
End
UMAP Dimension 2
UMAP Dimension 1
F.
Beginning
End
Pseudotime
Alpha Lineage Correlated TFs (n = 34)
Beta Lineage Correlated TFs (n = 27)
Motif Deviations
Gene Expression
−1.5
1.5

Figure 2.5: Identification of candidate correlated transcription factors governing pancreatic epithelial cell fate.

(A) Heatmaps depicting enriched marker peaks (left; n = 63,097) and transcription factor (TF) motifs enriched in marker peaks (right; n = 404) for each epithelial cell type. (**B**) Dot plot shows so-called “correlated” TFs (those with high correlation between motif deviation score and gene expression) in all epithelial cells. (**C**) Heatmaps revealing cell type-specific motif deviation scores (top) and gene expression values (bottom) of positive TFs identified in (**B**). (**D**) Feature plots displaying motif deviation (top) and gene expression (bottom) of selected positively correlated TFs at single-cell resolution. (**E**) Pseudotemporal ordering of epithelial cells along the Alpha and Beta lineages based on chromatin accessibility. (**F**) Heatmaps depicting positively correlated TFs across pseudotime (from left to right) for the Alpha lineage (left heatmaps; n = 34 correlated TFs) and Beta lineage (right heatmaps; n = 27 correlated TFs).

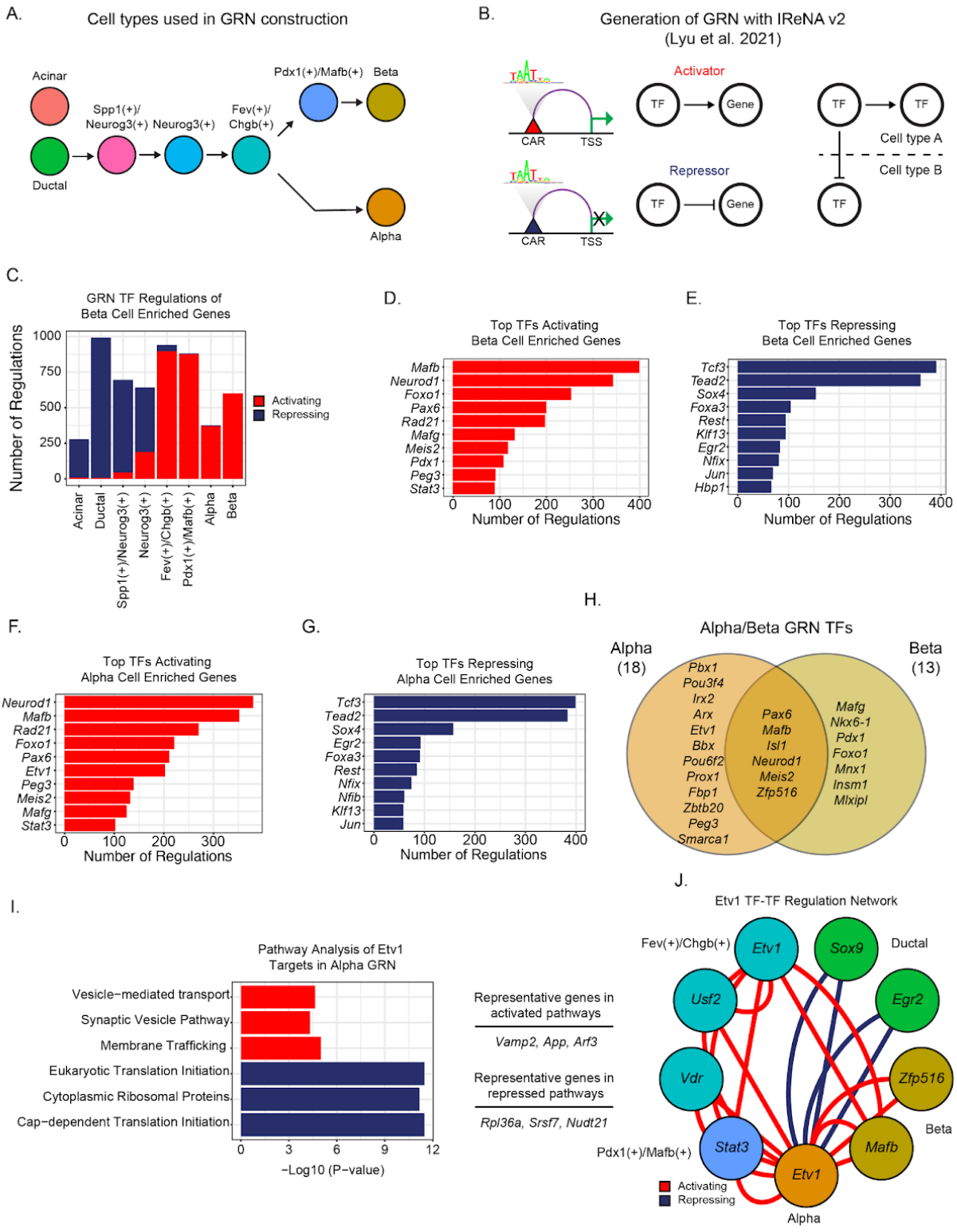

A.
Cell types used in GRN construction
Acinar
Ductal
Spp1(+)/
Neurog3(+)
Neurog3(+)
Fev(+)/
Chgb(+)
Pdx1(+)/Mafb(+)
Beta
Alpha
B.
Generation of GRN with IReNA v2
(Lyu et al. 2021)
Activator
Repressor
CAR
TSS
TF
Gene
Cell type A
Cell type B
C.
GRN TF Regulations of
Beta Cell Enriched Genes
Number of Regulations
Activating
Repressing
Acinar
Ductal
Spp1(+)/Neurog3(+)
Neurog3(+)
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)
Alpha
Beta
D.
Top TFs Activating
Beta Cell Enriched Genes
Mafb
Neurod1
Foxo1
Pax6
Rad21
Mafg
Meis2
Pdx1
Peg3
Stat3
Number of Regulations
E.
Top TFs Repressing
Beta Cell Enriched Genes
Tcf3
Tead2
Sox4
Foxa3
Rest
Klf13
Egr2
Nfix
Jun
Hbp1
Number of Regulations
F.
Top TFs Activating
Alpha Cell Enriched Genes
Neurod1
Mafb
Rad21
Foxo1
Pax6
Etv1
Peg3
Meis2
Mafg
Stat3
Number of Regulations
G.
Top TFs Repressing
Alpha Cell Enriched Genes
Tcf3
Tead2
Sox4
Egr2
Foxa3
Rest
Nfix
Nfib
Klf13
Jun
Number of Regulations
H.
Alpha/Beta GRN TFs
Alpha
(18)
Beta
(13)
Pbx1
Pou3f4
Irx2
Arx
Etv1
Bbx
Pou6f2
Prox1
Fbp1
Zbtb20
Peg3
Smarca1
Pax6
Mafb
Isl1
Neurod1
Meis2
Zfp516
Mafg
Nkx6-1
Pdx1
Foxo1
Mnx1
Insm1
Mlxipl
I.
Pathway Analysis of Etv1
Targets in Alpha GRN
Vesicle−mediated transport
Synaptic Vesicle Pathway
Membrane Trafficking
Eukaryotic Translation Initiation
Cytoplasmic Ribosomal Proteins
Cap−dependent Translation Initiation
−Log10 (P−value)
Representative genes in
activated pathways
Vamp2, App, Arf3
Representative genes in
repressed pathways
Rpl36a, Srsf7, Nudt21
J.
Etv1 TF-TF Regulation Network
Fev(+)/Chgb(+)
Etv1
Sox9
Ductal
Usf2
Egr2
Vdr
Zfp516
Beta
Pdx1(+)/Mafb(+)
Stat3
Mafb
Etv1
Activating
Repressing
Alpha

Figure 2.6: Mapping the gene regulatory networks active in the Alpha and Beta cell populations.

(**A**) Cell populations used as input for IReNA. Arrows denote known lineage relationships. (**B**) Schematic depicting application of the Integrated Regulatory Network Analysis (IReNA) v2 pipeline to identify gene regulatory networks (GRNs) within specific cell types. The GRN refers to active, cell type-specific transcription factors (TFs) and their target genes. (**C**) Bar graph indicating the number of downstream target genes for TFs enriched in the Beta cell type. Activating (red bars) and repressing (blue bars) refers to positive or negative correlation between gene expression levels of the TF and the target gene. (**D-E**) Bar graphs showing the top 10 TFs with the highest number of activating (**D**) or repressing (**E**) regulations of target genes enriched in the Beta cell type. (**F-G**) Bar graphs showing the top 10 TFs with the highest number of activating (**F**) or repressing (**G**) regulations of target genes enriched in the Alpha cell type. (**H**) Venn diagram depicting the overlap between Alpha and Beta GRN TFs. (**I**) Bar graph depicting top most significantly (p-value < 0.01) enriched pathways of genes activated (red bars) or repressed (blue bars) by the TF *Etv1* in the Alpha cell GRN. (**J**) Network diagram representing regulations between *Etv1* and interacting TFs. Each TF is represented by a circle (node) that is colored by the cell type in which that TF is active in the GRN. Activating regulations are depicted by red lines, while repressing are depicted by blue lines.

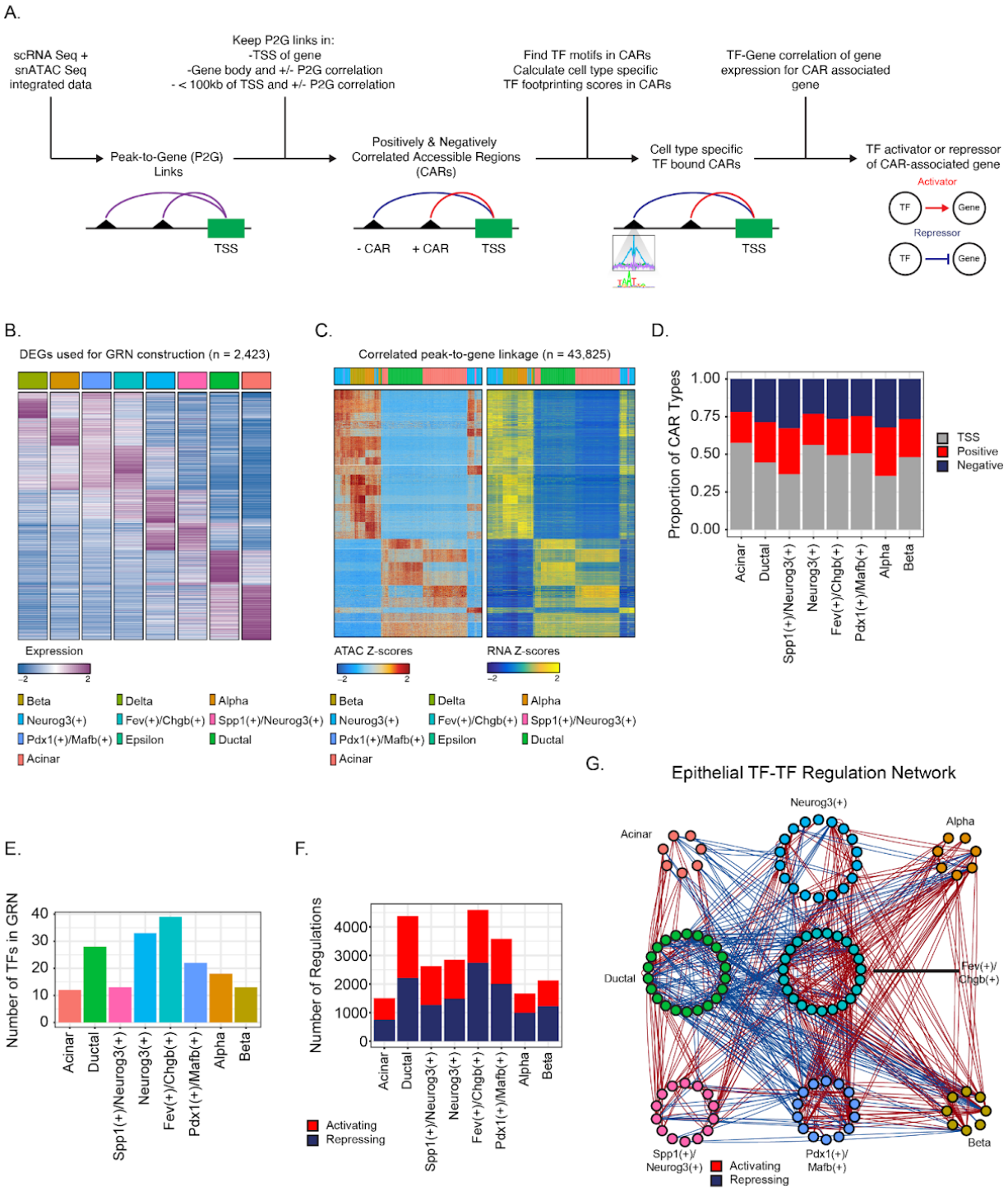
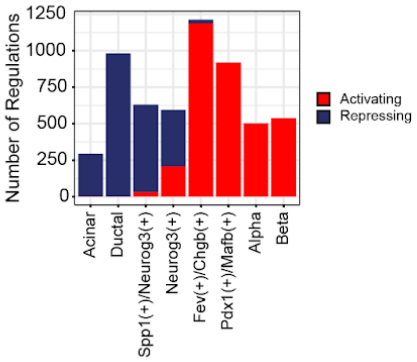

A.
scRNA Seq + snATAC Seq integrated data
Peak-to-Gene (P2G) Links
Keep P2G links in:
-TSS of gene
-Gene body and +/- P2G correlation
- < 100kb of TSS and +/- P2G correlation
Positively & Negatively Correlated Accessible Regions (CARs)
Find TF motifs in CARs
Calculate cell type specific TF footprinting scores in CARs
Cell type specific TF bound CARs
TF-Gene correlation of gene expression for CAR associated gene
TF activator or repressor of CAR-associated gene
Activator
Repressor
TF
Gene
TSS
- CAR
+ CAR
B.
DEGs used for GRN construction (n = 2,423)
Expression
C.
Correlated peak-to-gene linkage (n = 43,825)
ATAC Z-scores
RNA Z-scores
Beta
Delta
Alpha
Neurog3(+)
Fev(+)/Chgb(+)
Spp1(+)/Neurog3(+)
Pdx1(+)/Mafb(+)
Epsilon
Ductal
Acinar
D.
Proportion of CAR Types
TSS
Positive
Negative
Acinar
Ductal
Spp1(+)/Neurog3(+)
Neurog3(+)
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)
Alpha
Beta
E.
Number of TFs in GRN
F.
Number of Regulations
Activating
Repressing
G.
Epithelial TF-TF Regulation Network
Neurog3(+)
Alpha
Acinar
Ductal
Fev(+)/ Chgb(+)
Beta
Spp1(+)/ Neurog3(+)
Pdx1(+)/ Mafb(+)
Activating
Repressing

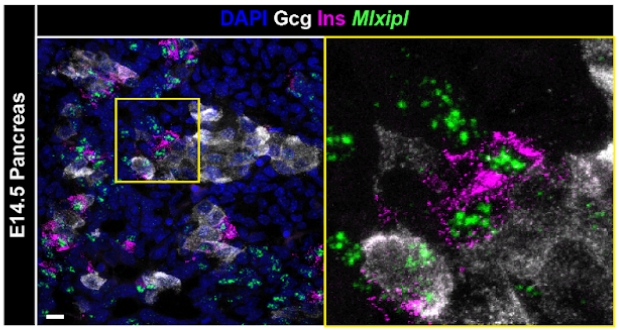
Figure 2.7: Gene regulatory network map of the developing pancreatic epithelium.

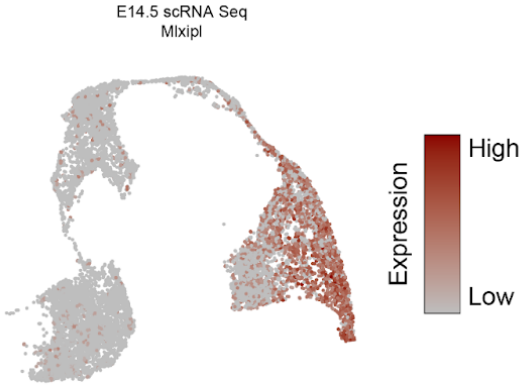
(**A**) Diagram depicting the IReNA v2 pipeline. (**B**) Heatmap showing the top differentially expressed genes (DEGs) used as input for GRN construction. Each column represents the average expression of a given DEG within each of the epithelial cell populations shown in **Figure 2.3a**. (**C**) Heatmaps showing ATAC accessibility (left) and RNA expression (right) of identified peak-to-gene (P2G) links used as input for GRN construction. Columns represent single cells and rows represent accessible peaks (left) and their correlated genes (right). (**D**) Bar graph cataloging correlated accessible regions (CARs), broken down according to CAR type, for each cell population. The three CAR types include those that lie within the transcription start site (TSS) of a gene (gray), as well as those that are positively (red) or negatively (blue) correlated with their linked gene. (**E**) Histogram showing the number of TFs in the GRN of each of the epithelial cell types. (**F**) Bar graph representing the proportion of all TF-gene interactions that are activating vs. repressing. (**G**) Bidirectional network diagram depicting all TF-TF interactions among all cell types shown in **Figure 2.3a**. TFs are denoted by each node, which is colored by the cell type in which the TF is active in the GRN. Activating regulations are depicted by red lines, while repressing are depicted by blue lines.

A.

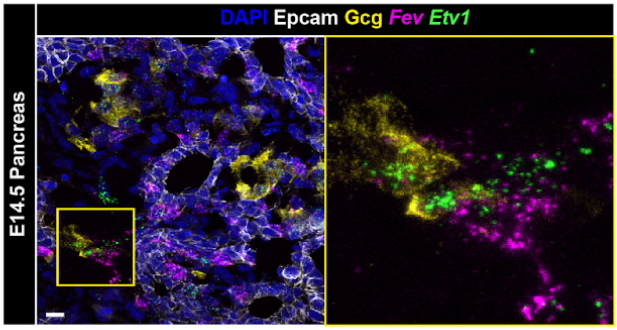

Number of Regulations
1250
1000
750
500
250
0
Activating
Repressing
Acinar
Ductal
Spp1(+)/Neurog3(+)
Neurog3(+)
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)
Alpha
Beta

B.


DAPI Gcg Ins Mlxipl
E14.5 Pancreas


E14.5 scRNA Seq
Mlxipl
Expression
High
Low

C.


DAPI Epcam Gcg Fev Etv1
E14.5 Pancreas

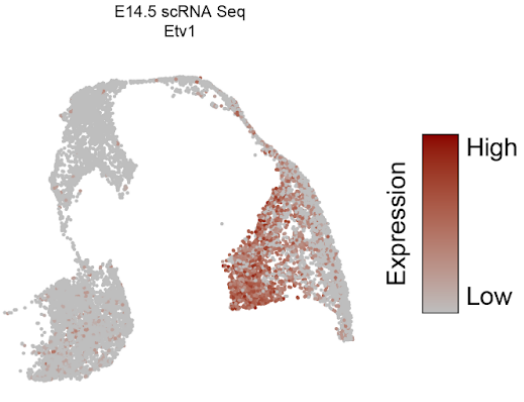

E14.5 scRNA Seq
Etv1
Expression
High
Low

Figure 2.8: Confirmation *in vivo* of cell type specific GRN TF expression in the developing pancreas.

(**A**) Bar graph indicating the number of downstream target genes for TFs enriched in the Alpha cell type. Activating (red bars) and repressing (blue bars) refers to positive or negative correlation between gene expression levels of the TF and the target gene. (**B-C**) Multiplexed immunofluorescence and *in situ* hybridization staining of WT E14.5 pancreas tissue. (**B**) Ins-expressing Beta cells are shown in magenta, and Gcg-expressing Alpha Cells are shown in white. *Mlxipl* transcript expression is shown in green. DAPI marks nuclei in blue. Expression of *Mlxipl* across all epithelial cells, as measured by scRNA-Seq, is depicted in the feature plot to the right. Cluster annotation is shown in **Figure 2.3a**. (**C**) Alpha cells expressing Gcg protein are shown in yellow. *Etv1* transcript expression is shown in green, and *Fev*-expressing progenitor cells are shown in magenta. DAPI marks nuclei in blue and Epcam marks epithelial cells in white. Expression of *Etv1* as measured by scRNA-Seq is depicted in the feature plot to the right. All scale bars are 20 μm.

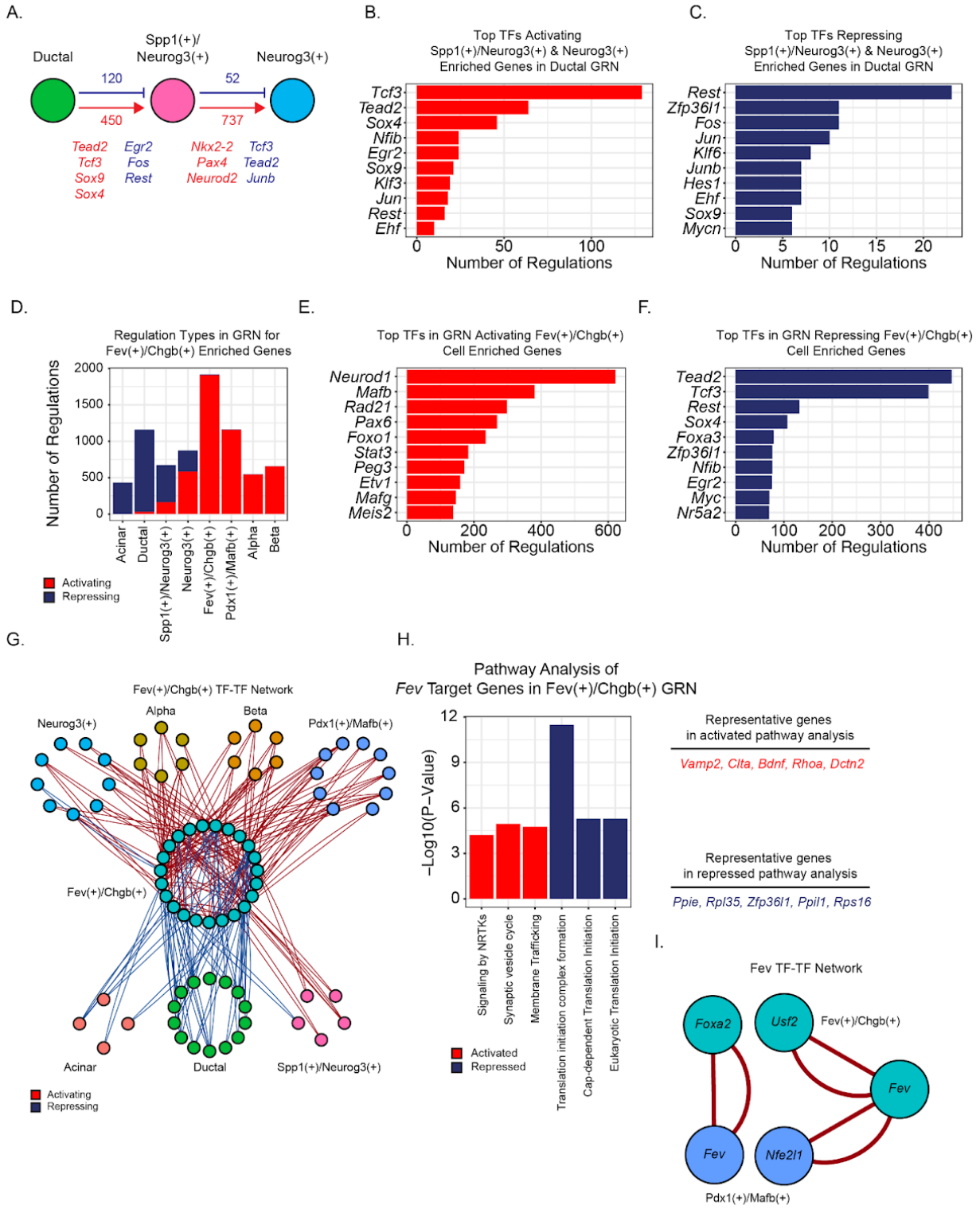

A.
Ductal
Spp1(+)/
Neurog3(+)
Neurog3(+)
120
450
52
737
Tead2
Tcf3
Sox9
Sox4
Egr2
Fos
Rest
Nkx2-2
Pax4
Neurod2
Tcf3
Tead2
Junb
B.
Top TFs Activating
Spp1(+)/Neurog3(+) & Neurog3(+)
Enriched Genes in Ductal GRN
Tcf3
Tead2
Sox4
Nfib
Egr2
Sox9
Klf3
Jun
Rest
Ehf
0
50
100
Number of Regulations
C.
Top TFs Repressing
Spp1(+)/Neurog3(+) & Neurog3(+)
Enriched Genes in Ductal GRN
Rest
Zfp36l1
Fos
Jun
Klf6
Junb
Hes1
Ehf
Sox9
Mycn
0
5
10
15
20
Number of Regulations
D.
Regulation Types in GRN for
Fev(+)/Chgb(+) Enriched Genes
Number of Regulations
2000
1500
1000
500
0
Acinar
Ductal
Spp1(+)/Neurog3(+)
Neurog3(+)
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)
Alpha
Beta
Activating
Repressing
E.
Top TFs in GRN Activating Fev(+)/Chgb(+)
Cell Enriched Genes
Neurod1
Mafb
Rad21
Pax6
Foxo1
Stat3
Peg3
Etv1
Mafg
Meis2
0
200
400
600
Number of Regulations
F.
Top TFs in GRN Repressing Fev(+)/Chgb(+)
Cell Enriched Genes
Tead2
Tcf3
Rest
Sox4
Foxa3
Zfp36l1
Nfib
Egr2
Myc
Nr5a2
0
100
200
300
400
Number of Regulations
G.
Fev(+)/Chgb(+) TF-TF Network
Neurog3(+)
Alpha
Beta
Pdx1(+)/Mafb(+)
Fev(+)/Chgb(+)
Acinar
Ductal
Spp1(+)/Neurog3(+)
Activating
Repressing
H.
Pathway Analysis of
Fev Target Genes in Fev(+)/Chgb(+) GRN
−Log10(P−Value)
12
9
6
3
0
Signaling by NRTKs
Synaptic vesicle cycle
Membrane Trafficking
Translation initiation complex formation
Cap-dependent Translation Initiation
Eukaryotic Translation Initiation
Activated
Repressed
Representative genes
in activated pathway analysis
Vamp2, Clta, Bdnf, Rhoa, Dctn2
Representative genes
in repressed pathway analysis
Ppie, Rpl35, Zfp36l1, Ppil1, Rps16
I.
Fev TF-TF Network
Foxa2
Usf2
Fev(+)/Chgb(+)
Fev
Fev
Nfe2l1
Pdx1(+)/Mafb(+)

Figure 2.9: Mapping the gene regulatory networks active in the *Fev*-expressing pancreatic endocrine population.

(**A**) Cartoon depicting the number of activating (red) and repressing (blue) regulations between transcription factors (TFs) in Ductal, Spp1(+)/Neurog3(+), and Neurog3(+) populations, with representative TFs indicated. (**B-C**) Bar graphs depicting the top 10 TFs in the Ductal GRN with the highest number of activating (**B**) or repressing (**C**) regulations of target genes that are enriched in the Spp1(+)/Neurog3(+) and Neurog3(+) cell types. (**D**) Bar graph depicting the number of TF regulations of target genes enriched in the Fev(+)/Chgb(+) cell type. Activating (red bars) and repressing (blue bars) refers to positive or negative correlation, respectively, of gene expression between the TF and target gene. (**E-F**) Bar graphs depicting the top 10 transcription factors with the highest number of activating (**E**) or repressing (**F**) interactions of target genes enriched in the Fev(+)/Chgb(+) cell type. (**G**) Network diagram depicting all TF-TF regulations between TFs enriched in the Fev(+)/Chgb(+) GRN and all other GRNs. TFs are denoted by each node, which is colored by the cell type in which the TF is found. Each activating regulation is depicted by a red line, while each repressing regulation is depicted by a blue line. (**H**) Bar graph depicting top significant (p-value < 0.01) pathways of genes activated (red bars) or repressed (blue bars) by *Fev* in the GRN analysis. (**I**) Network diagram depicting the TFs activated (red lines) by *Fev* in the GRN analysis. Each node (TF) is colored according to the cell type in which it is expressed.

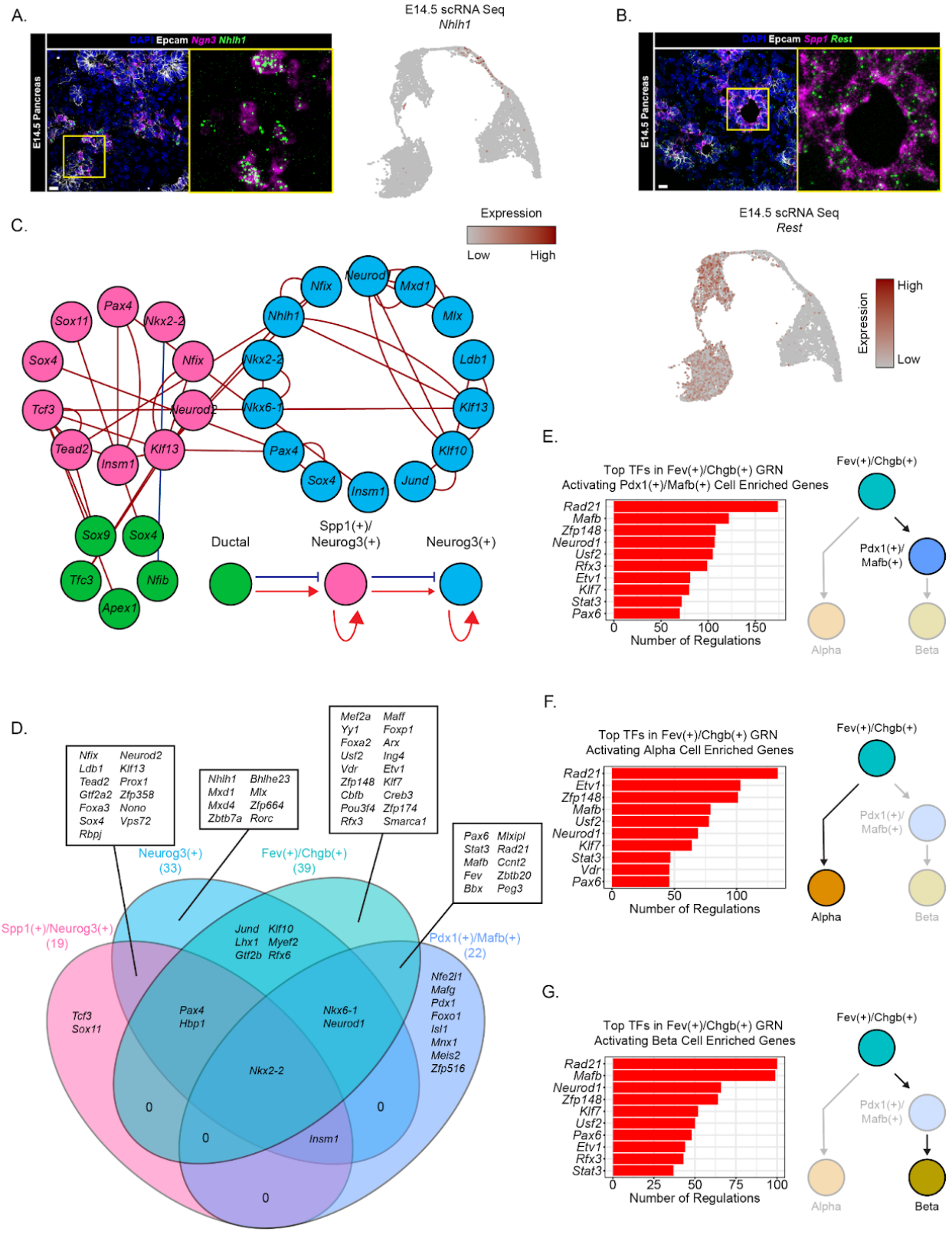
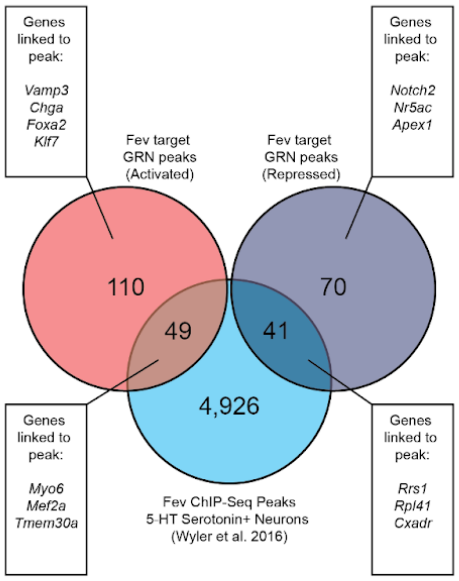

A.
DAPI Epcam Ngn3 Nhlh1
E14.5 Pancreas
E14.5 scRNA Seq
Nhlh1
B.
DAPI Epcam Spp1 Rest
E14.5 Pancreas
C.
Expression
Low
High
Sox11
Pax4
Nkx2-2
Sox4
Nfix
Tcf3
Neurod2
Tead2
Insm1
Klf13
Sox9
Sox4
Tfc3
Nfib
Apex1
Nfix
Neurod2
Mxd1
Nhlh1
Mlx
Nkx2-2
Ldb1
Nkx6-1
Klf13
Pax4
Klf10
Sox4
Insm1
Jund
Ductal
Spp1(+)/
Neurog3(+)
Neurog3(+)
E14.5 scRNA Seq
Rest
Expression
High
Low
D.
Nfix Neurod2
Ldb1 Klf13
Tead2 Prox1
Gtf2a2 Zfp358
Foxa3 Nono
Sox4 Vps72
Rbpj
Nhlh1 Bhlhe23
Mxd1 Mlx
Mxd4 Zfp664
Zbtb7a Rorc
Mef2a Maff
Yy1 Foxp1
Foxa2 Arx
Usf2 Ing4
Vdr Etv1
Zfp148 Klf7
Cbfb Creb3
Pou3f4 Zfp174
Rfx3 Smarca1
Pax6 Mlxipl
Stat3 Rad21
Mafb Ccnt2
Fev Zbtb20
Bbx Peg3
Neurog3(+)
(33)
Fev(+)/Chgb(+)
(39)
Spp1(+)/Neurog3(+)
(19)
Pdx1(+)/Mafb(+)
(22)
Jund Klf10
Lhx1 Myef2
Gtf2b Rfx6
Tcf3
Sox11
Pax4
Hbp1
Nkx6-1
Neurod1
Nkx2-2
Insm1
Nfe2l1
Mafg
Pdx1
Foxo1
Isl1
Mnx1
Meis2
Zfp516
0
0
0
0
E.
Top TFs in Fev(+)/Chgb(+) GRN
Activating Pdx1(+)/Mafb(+) Cell Enriched Genes
Rad21
Mafb
Zfp148
Neurod1
Usf2
Rfx3
Etv1
Klf7
Stat3
Pax6
0
50
100
150
Number of Regulations
Fev(+)/Chgb(+)
Pdx1(+)/
Mafb(+)
Alpha
Beta
F.
Top TFs in Fev(+)/Chgb(+) GRN
Activating Alpha Cell Enriched Genes
Rad21
Etv1
Zfp148
Mafb
Usf2
Neurod1
Klf7
Stat3
Vdr
Pax6
0
50
100
Number of Regulations
Fev(+)/Chgb(+)
Pdx1(+)/
Mafb(+)
Alpha
Beta
G.
Top TFs in Fev(+)/Chgb(+) GRN
Activating Beta Cell Enriched Genes
Rad21
Mafb
Neurod1
Zfp148
Klf7
Usf2
Pax6
Etv1
Rfx3
Stat3
0
25
50
75
100
Number of Regulations
Fev(+)/Chgb(+)
Pdx1(+)/
Mafb(+)
Alpha
Beta

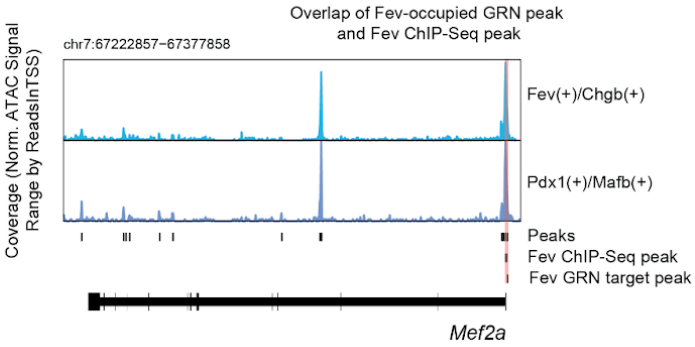
Figure 2.10: Gene regulatory networks active in the endocrine progenitor populations.

(**A-B**) Multiplexed immunofluorescence and *in situ* hybridization staining of WT E14.5 pancreas tissue. **(A)** Endocrine progenitors expressing *Ngn3* transcript are shown in magenta. *Nhlh1* transcript expression is shown in green, and epithelial cells are marked by Epcam in white. DAPI marks nuclei in blue. Expression of *Nhlh1* across epithelial cells, as assessed by scRNA-Seq, is depicted in the feature plot to the right. Cluster annotation is shown in **Figure 2.3a.** (**B**) Ductal cells expressing *Spp1* transcript are shown in magenta. *Rest* transcript expression is shown in green, and epithelial cells are marked by Epcam in white. DAPI marks nuclei in blue. Expression of *Rest* across epithelial cells, as assessed by scRNA-Seq, is depicted in the feature plot below. Scale bars in (A) and (B) are 20 μm. (**C**) Network diagram depicting the TFs activated (red lines) or repressed (blue lines) among Ductal (green), Spp1(+)/Neurog3(+) (pink), and Neurog3(+) (blue) cell types. Nodes represent TFs and are grouped and colored according to the cell type in which they are active. The directionality of regulations is modeled in the diagram at the bottom. (**D**) Venn diagram depicting TFs that are shared vs. unique within the GRN constructed from all four EP populations. (**E-G**) Bar graphs depicting the top 10 TFs within the Fev(+)/Chgb(+) GRN with the highest number of activating regulations of target genes enriched in the (**E**) Pdx1(+)/Mafb(+) (**F**) Alpha or (**G**) Beta cell types.

A.


Genes linked to peak:
Vamp3
Chga
Foxa2
Klf7
Fev target GRN peaks (Activated)
Fev target GRN peaks (Repressed)
Genes linked to peak:
Notch2
Nr5ac
Apex1
110
49
70
41
4,926
Genes linked to peak:
Myo6
Mef2a
Tmem30a
Fev ChIP-Seq Peaks
5-HT Serotonin+ Neurons
(Wyler et al. 2016)
Genes linked to peak:
Rrs1
Rpl41
Cxadr

B.


Overlap of Fev-occupied GRN peak and Fev ChIP-Seq peak
chr7:67222857–67377858
Coverage (Norm. ATAC Signal Range by ReadsInTSS)
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)
Peaks
Fev ChIP-Seq peak
Fev GRN target peak
Mef2a

C.

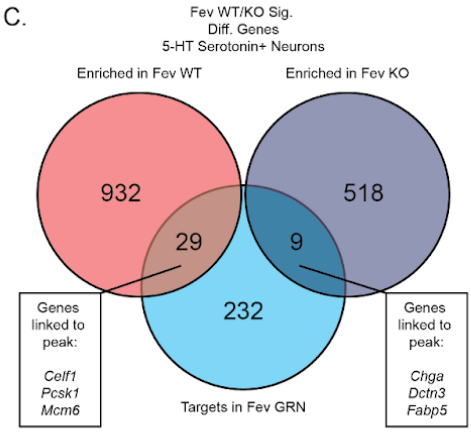

Fev WT/KO Sig. Diff. Genes
5-HT Serotonin+ Neurons
Enriched in Fev WT
Enriched in Fev KO
932
29
518
9
232
Genes linked to peak:
Celf1
Pcsk1
Mcm6
Targets in Fev GRN
Genes linked to peak:
Chga
Dctn3
Fabp5

Figure 2.11: Confirmation of Fev GRN targets.

(**A**) Venn diagram of intersected peaks from GRN targets of Fev in the pancreatic Fev(+)/Chgb(+) population and Fev ChIP-Seq data from E12.5 to E15.5 mouse hindbrains (35). (**B**) Track plot displaying accessibility of the *Mef2a* locus in Fev(+)/Chgb(+) cells. Fev GRN and Fev ChIP-Seq peaks are highlighted. Overlap of peak sets is denoted by the red box. (**C**) Venn diagram showing intersection of Fev targets and significantly differentially expressed genes in wildtype (WT) and Fev knockout (KO) E15.5 mouse hindbrains (117).

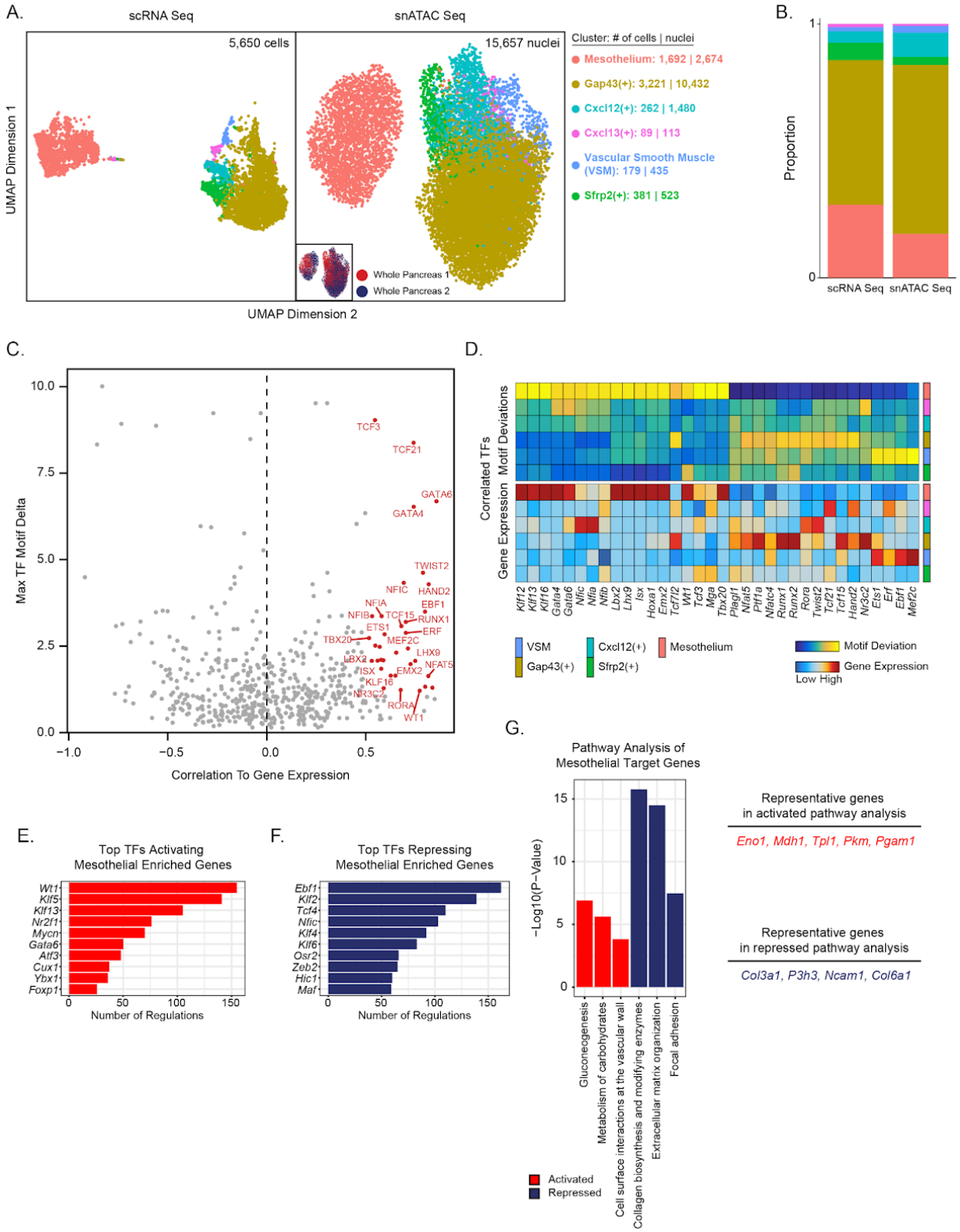

A.
scRNA Seq
snATAC Seq
5,650 cells
15,657 nuclei
UMAP Dimension 1
UMAP Dimension 2
Cluster: # of cells | nuclei
Mesothelium: 1,692 | 2,674
Gap43(+): 3,221 | 10,432
Cxcl12(+): 262 | 1,480
Cxcl13(+): 89 | 113
Vascular Smooth Muscle (VSM): 179 | 435
Sfrp2(+): 381 | 523
Whole Pancreas 1
Whole Pancreas 2
B.
Proportion
scRNA Seq
snATAC Seq
C.
Max TF Motif Delta
Correlation To Gene Expression
TCF3
TCF21
GATA6
GATA4
TWIST2
NFIC
HAND2
NFIA
EBF1
NFIB
TCF15
RUNX1
ETS1
ERF
TBX20
MEF2C
LBX2
LHX9
ISX
EMX2
NFAT5
KLF16
NR3C2
RORA
WT1
D.
Correlated TFs
Motif Deviations
Gene Expression
VSM
Gap43(+)
Cxcl12(+)
Sfrp2(+)
Mesothelium
Motif Deviation
Gene Expression
Low High
E.
Top TFs Activating Mesothelial Enriched Genes
Number of Regulations
F.
Top TFs Repressing Mesothelial Enriched Genes
Number of Regulations
G.
Pathway Analysis of Mesothelial Target Genes
−Log10(P−Value)
Gluconeogenesis
Metabolism of carbohydrates
Cell surface interactions at the vascular wall
Collagen biosynthesis and modifying enzymes
Extracellular matrix organization
Focal adhesion
Activated
Repressed
Representative genes in activated pathway analysis
Eno1, Mdh1, Tpl1, Pkm, Pgam1
Representative genes in repressed pathway analysis
Col3a1, P3h3, Ncam1, Col6a1

Figure 2.12: Chromatin accessibility and gene regulatory network in the developing pancreatic mesenchyme.

(**A**) UMAP plots enable visualization of scRNA-Seq (left) and snATAC-Seq (right) data for all mesenchymal cells in the E14.5 pancreas. Numbers of cells/nuclei are depicted on the right, along with cell type annotations. The scRNA-Seq dataset is from our previously published work (58). (**B**) Bar graph depicts the proportion of each cell type in the scRNA-Seq and snATAC-Seq datasets. Colors match the cell types in (**A**). (**C**) Dot plot showing correlated transcription factors (as determined by correlation between motif deviation score and gene expression) in mesenchymal and mesothelial cells. (**D**) Heatmaps reveal cell type-specific motif deviation scores (top) and gene expression values (bottom) of positive transcription factors identified in (**C**). (**E, F**) Bar graph depicting the top TFs activating (**E**) and repressing (**F**) genes enriched in the mesothelial population. (**G**) Bar graph depicting top significant (p-value < 0.01) pathways of genes activated (red bars) or repressed (blue bars) by mesothelial cells in the GRN analysis. VSM, vascular smooth muscle.

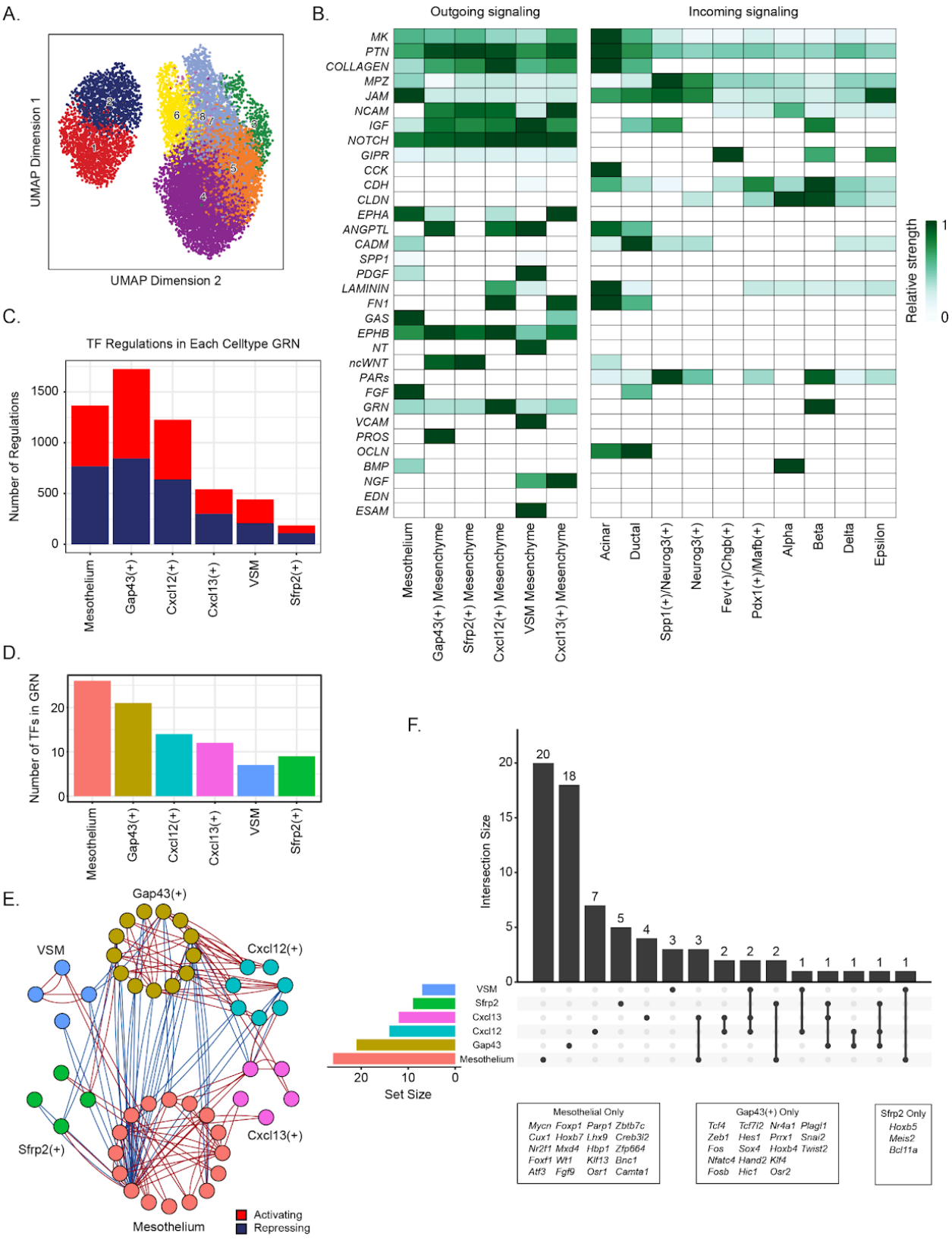

A.
UMAP Dimension 1
UMAP Dimension 2
B.
Outgoing signaling
Incoming signaling
MK
PTN
COLLAGEN
MPZ
JAM
NCAM
IGF
NOTCH
GIPR
CCK
CDH
CLDN
EPHA
ANGPTL
CADM
SPP1
PDGF
LAMININ
FN1
GAS
EPHB
NT
ncWNT
PARs
FGF
GRN
VCAM
PROS
OCLN
BMP
NGF
EDN
ESAM
Mesothelium
Gap43(+) Mesenchyme
Sfrp2(+) Mesenchyme
Cxcl12(+) Mesenchyme
VSM Mesenchyme
Cxcl13(+) Mesenchyme
Acinar
Ductal
Spp1(+)/Neurog3(+)
Neurog3(+)
Fev(+)/Chgb(+)
Pdx1(+)/Mafb(+)
Alpha
Beta
Delta
Epsilon
Relative strength
1
0
C.
TF Regulations in Each Celltype GRN
Number of Regulations
1500
1000
500
0
Mesothelium
Gap43(+)
Cxcl12(+)
Cxcl13(+)
VSM
Sfrp2(+)
D.
Number of TFs in GRN
20
10
0
Mesothelium
Gap43(+)
Cxcl12(+)
Cxcl13(+)
VSM
Sfrp2(+)
E.
Gap43(+)
VSM
Cxcl12(+)
Cxcl13(+)
Sfrp2(+)
Mesothelium
Activating
Repressing
F.
Intersection Size
20
18
7
5
4
3
3
2
2
2
1
1
1
1
1
VSM
Sfrp2
Cxcl13
Cxcl12
Gap43
Mesothelium
20
10
0
Set Size
Mesothelial Only
Mycn Foxp1 Parp1 Zbtb7c
Cux1 Hoxb7 Lhx9 Creb3l2
Nr2f1 Mxd4 Hbp1 Zfp664
Foxf1 Wt1 Klf13 Bnc1
Atf3 Fgf9 Osr1 Camta1
Gap43(+) Only
Tcf4 Tcf7l2 Nr4a1 Plagl1
Zeb1 Hes1 Prrx1 Snai2
Fos Sox4 Hoxb4 Twist2
Nfatc4 Hand2 Klf4
Fosb Hic1 Osr2
Sfrp2 Only
Hoxb5
Meis2
Bcl11a

Figure 2.13: Gene regulatory network in the developing pancreatic mesenchyme.

(**A**) UMAP visualization of snATAC-Seq clustering of mesenchymal cells, here without integration with scRNA-Seq data. All clusters in **Figure 2.12a** are present even when clustered based on chromatin accessibility alone. (**B**) CellChat analysis depicting cell-cell communication between mesenchyme and epithelium at E14.5. Heatmaps depict outgoing signals derived from mesenchymal cells (left) and incoming signals to epithelial cells (right). Cell populations are represented across columns, and signaling factors are represented across rows. (**C**) Bar graph depicting the proportion of all TF-gene interactions that are activating (red) vs. repressing (blue). (**D**) Histogram showing the total number of GRN TFs identified within each cell population. (**E**) Network diagram showing all TF-TF interactions among all cell types shown in **Figure 2.12a**. Each TF is denoted by a node, which is colored by the cell type in which the TF is active. Activating regulations are depicted by red lines, while repressing are depicted by blue lines. (**F**) UpSet plot depicting shared and exclusive Mesenchymal TFs within the GRN. Numbers above each bar designate the number of TFs within that set. VSM, vascular smooth muscle.

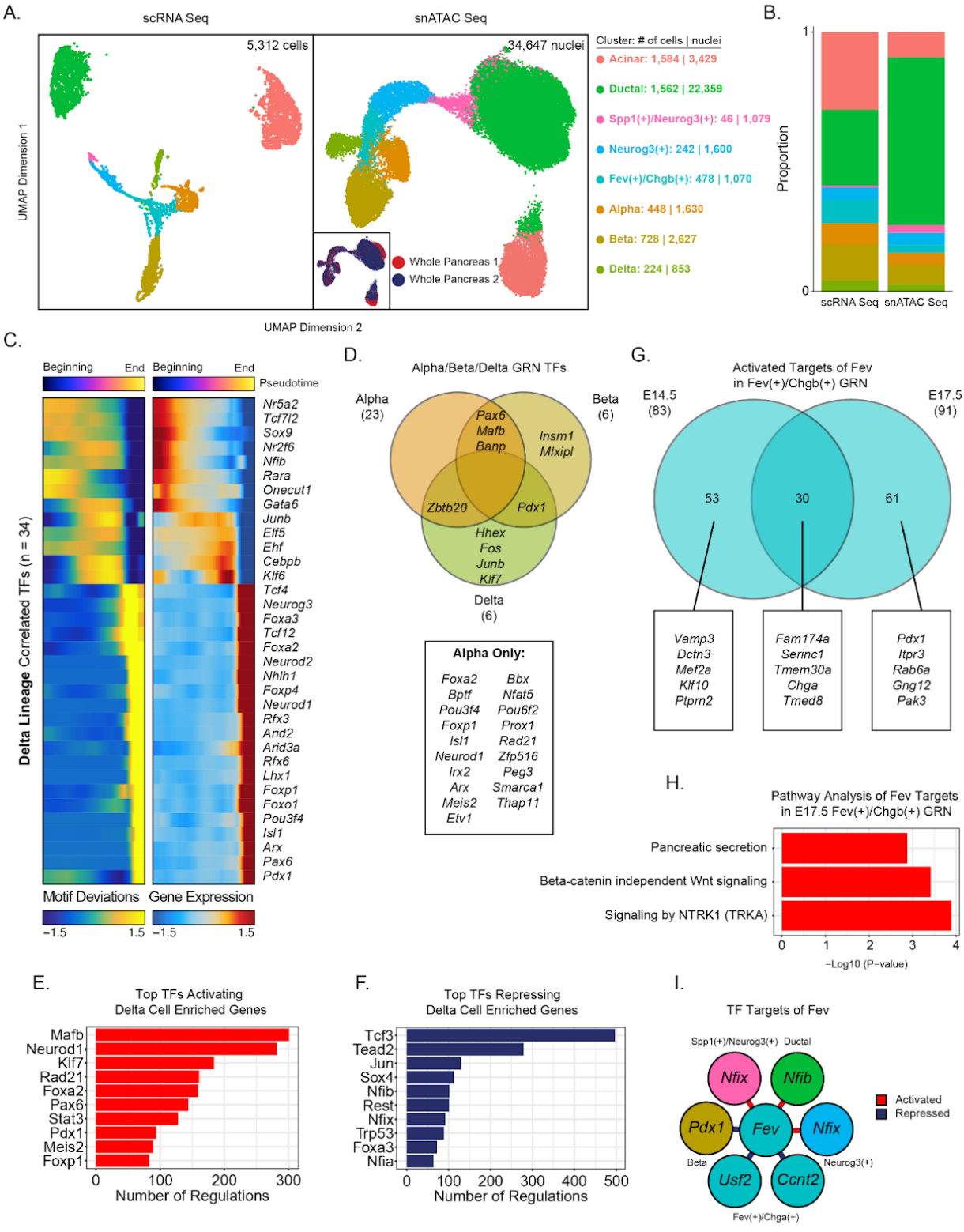

A.
scRNA Seq
snATAC Seq
5,312 cells
34,647 nuclei
UMAP Dimension 1
UMAP Dimension 2
Whole Pancreas 1
Whole Pancreas 2
Cluster: # of cells | nuclei
Acinar: 1,584 | 3,429
Ductal: 1,562 | 22,359
Spp1(+)/Neurog3(+): 46 | 1,079
Neurog3(+): 242 | 1,600
Fev(+)/Chgb(+): 478 | 1,070
Alpha: 448 | 1,630
Beta: 728 | 2,627
Delta: 224 | 853
B.
Proportion
scRNA Seq
snATAC Seq
C.
Beginning
End
Pseudotime
Delta Lineage Correlated TFs (n = 34)
Nr5a2
Tcf7l2
Sox9
Nr2f6
Nfib
Rara
Onecut1
Gata6
Junb
Elf5
Ehf
Cebpb
Klf6
Tcf4
Neurog3
Foxa3
Tcf12
Foxa2
Neurod2
Nhlh1
Foxp4
Neurod1
Rfx3
Arid2
Arid3a
Rfx6
Lhx1
Foxp1
Foxo1
Pou3f4
Isl1
Arx
Pax6
Pdx1
Motif Deviations
Gene Expression
−1.5
1.5
D.
Alpha/Beta/Delta GRN TFs
Alpha
(23)
Beta
(6)
Delta
(6)
Pax6
Mafb
Banp
Insm1
Mlxipl
Zbtb20
Pdx1
Hhex
Fos
Junb
Klf7
Alpha Only:
Foxa2
Bptf
Pou3f4
Foxp1
Isl1
Neurod1
Irx2
Arx
Meis2
Etv1
Bbx
Nfat5
Pou6f2
Prox1
Rad21
Zfp516
Peg3
Smarca1
Thap11
G.
Activated Targets of Fev
in Fev(+)/Chgb(+) GRN
E14.5
(83)
E17.5
(91)
53
30
61
Vamp3
Dctn3
Mef2a
Klf10
Ptprn2
Fam174a
Serinc1
Tmem30a
Chga
Tmed8
Pdx1
Itpr3
Rab6a
Gng12
Pak3
H.
Pathway Analysis of Fev Targets
in E17.5 Fev(+)/Chgb(+) GRN
Pancreatic secretion
Beta-catenin independent Wnt signaling
Signaling by NTRK1 (TRKA)
−Log10 (P−value)
E.
Top TFs Activating
Delta Cell Enriched Genes
Mafb
Neurod1
Klf7
Rad21
Foxa2
Pax6
Stat3
Pdx1
Meis2
Foxp1
Number of Regulations
F.
Top TFs Repressing
Delta Cell Enriched Genes
Tcf3
Tead2
Jun
Sox4
Nfib
Rest
Nfix
Trp53
Foxa3
Nfia
Number of Regulations
I.
TF Targets of Fev
Spp1(+)/Neurog3(+)
Ductal
Nfix
Nfib
Pdx1
Fev
Nfix
Beta
Neurog3(+)
Usf2
Ccnt2
Fev(+)/Chga(+)
Activated
Repressed

Figure 2.14: Mapping the gene regulatory networks of pancreatic epithelium across developmental time.

(**A**) UMAP plots enable visualization of scRNA-Seq (left) and snATAC-Seq (right) data for all epithelial cells in the E17.5 mouse pancreas. Numbers of cells or nuclei are depicted to the right, along with cell type annotations. The scATAC-Seq dataset is comprised of two independent biological replicates (n = 7 pancreata from 1 litter for replicate 1; n = 3 pancreata from 1 litter for replicate 2) of whole pancreas. Contribution of each sample (replicate 1 and replicate 2) to the merged dataset is depicted in the inset. The scRNA-Seq dataset is from our previously published work (58). (**B**) Bar graph depicts the proportion of each cell type within the scRNA-Seq and snATAC-Seq datasets. Colors match the cell types in (**A**). (**C**) Heatmaps depicting positively correlated TFs across pseudotime (from left to right) for the Delta lineage (Ductal to Spp1(+)/Neurog3(+), to Neurog3(+), to Fev(+)/Chgb(+), to Delta). (**D**) Venn diagram depicting the overlap in TFs among Alpha, Beta, and Delta GRNs. (**E-F**) Bar graphs showing the top 10 TFs with the highest number of activating (**E**) or repressing (**F**) regulations of target genes enriched in the Delta cell type. (**G**) Venn diagram depicting both the common and the timepoint-specific activated target genes of Fev in the E14.5 and E17.5 Fev(+)/Chgb(+) GRNs. Selected TFs within each group are listed in the boxes at the bottom. (**H**) Bar graph depicting pathways significantly (p-value < 0.01) enriched among target genes activated by Fev that are unique to the E17.5 Fev(+)/Chgb(+) GRN. (**I**) Network diagram representing regulations between Fev and interacting TFs. Each TF is represented by a circle (node) that is colored by the cell type in which that TF is active in the GRN. Activating regulations are depicted by red lines, while repressing are depicted by blue lines.

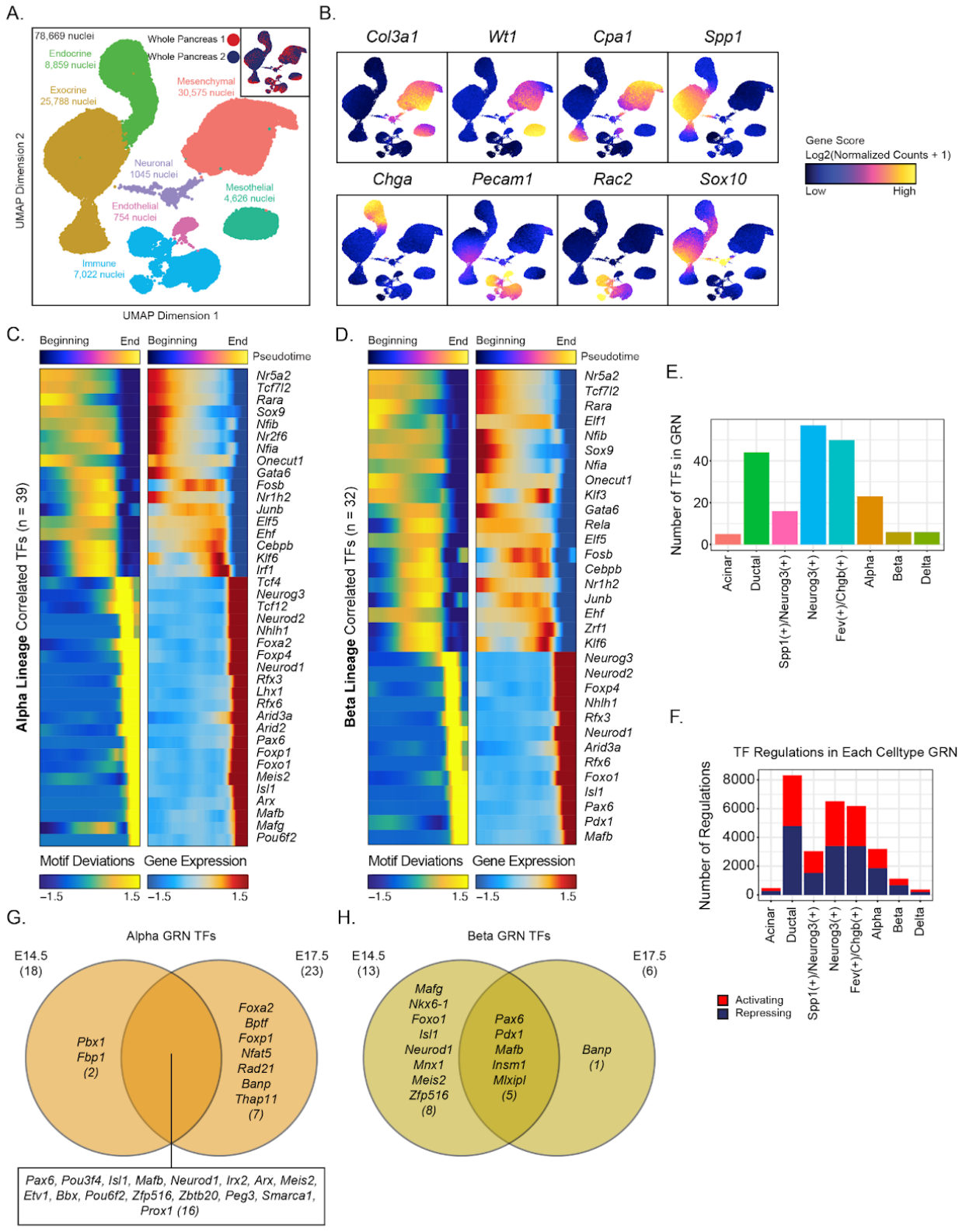

A.
78,669 nuclei
Whole Pancreas 1
Whole Pancreas 2
Endocrine
8,859 nuclei
Exocrine
25,788 nuclei
Mesenchymal
30,575 nuclei
Neuronal
1045 nuclei
Mesothelial
4,626 nuclei
Endothelial
754 nuclei
Immune
7,022 nuclei
UMAP Dimension 2
UMAP Dimension 1
B.
Col3a1
Wt1
Cpa1
Spp1
Chga
Pecam1
Rac2
Sox10
Gene Score
Log2(Normalized Counts + 1)
Low
High
C.
Beginning
End
Pseudotime
Alpha Lineage Correlated TFs (n = 39)
Nr5a2
Tcf7l2
Rara
Sox9
Nfib
Nr2f6
Nfia
Onecut1
Gata6
Fosb
Nr1h2
Junb
Elf5
Ehf
Cebpb
Klf6
Irf1
Tcf4
Neurog3
Tcf12
Neurod2
Nhlh1
Foxa2
Foxp4
Neurod1
Rfx3
Lhx1
Rfx6
Arid3a
Arid2
Pax6
Foxp1
Foxo1
Meis2
Isl1
Arx
Mafb
Mafg
Pou6f2
Motif Deviations
Gene Expression
−1.5
1.5
D.
Beta Lineage Correlated TFs (n = 32)
Nr5a2
Tcf7l2
Rara
Elf1
Nfib
Sox9
Nfia
Onecut1
Klf3
Gata6
Rela
Elf5
Fosb
Cebpb
Nr1h2
Junb
Ehf
Zrf1
Klf6
Neurog3
Neurod2
Foxp4
Nhlh1
Rfx3
Neurod1
Arid3a
Rfx6
Foxo1
Isl1
Pax6
Pdx1
Mafb
E.
Number of TFs in GRN
0
20
40
Acinar
Ductal
Spp1(+)/Neurog3(+)
Neurog3(+)
Fev(+)/Chgb(+)
Alpha
Beta
Delta
F.
TF Regulations in Each Celltype GRN
Number of Regulations
0
2000
4000
6000
8000
Activating
Repressing
G.
Alpha GRN TFs
E14.5
(18)
E17.5
(23)
Pbx1
Fbp1
(2)
Foxa2
Bptf
Foxp1
Nfat5
Rad21
Banp
Thap11
(7)
Pax6, Pou3f4, Isl1, Mafb, Neurod1, Irx2, Arx, Meis2, Etv1, Bbx, Pou6f2, Zfp516, Zbtb20, Peg3, Smarca1, Prox1 (16)
H.
Beta GRN TFs
E14.5
(13)
E17.5
(6)
Mafg
Nkx6-1
Foxo1
Isl1
Neurod1
Mnx1
Meis2
Zfp516
(8)
Pax6
Pdx1
Mafb
Insm1
Mlxipl
(5)
Banp
(1)

Figure 2.15: Gene regulatory analysis of the developing E17.5 pancreatic epithelium.

(**A**) Uniform Manifold Approximation and Projection (UMAP) visualization of merged E17.5 snATAC-Seq dataset, comprising a total of 78,669 nuclei. Contribution of each sample (whole pancreas replicates 1 and 2) to the merged dataset is depicted in the inset. (**B**) Feature plots depicting the Gene Scores (accessibility of the gene promoter plus the gene body) for a subset of the marker genes used to annotate the cell types in panel **A.** (**C-D**) Heatmaps depicting positively correlated TFs across pseudotime (from left to right) for the Alpha lineage (**C**) and Beta lineage (**D**). (**E**) Histogram showing the total number of TFs identified within each cell population in the E17.5 GRN. (**F**) Bar graph representing the proportion of all TF-gene interactions that are activating (red bars) vs. repressing (blue bars). (**G-H**) Venn diagrams depicting the overlap and exclusivity of TFs in the E14.5 and E17.5 (**G**) Alpha and (**H**) Beta GRNs.

Chapter 3: Single-cell multi-omic roadmap of human fetal pancreas development

3.1 Introduction

Diabetes mellitus is one of the most common endocrine disorders worldwide, affecting hundreds of millions of individuals across the globe (160). Type 1 diabetes (T1D) is a disease of the endocrine pancreas characterized by immune-mediated destruction of insulin-producing beta cells. Beta cell replacement therapy holds great promise for eliminating the need for exogenous insulin delivery and effectively curing the disease (161,162). Several protocols have been devised to generate insulin-secreting beta-like cells from human pluripotent stem cells (hPSCs) using stepwise differentiation platforms that aim to mimic endogenous development by recapitulating key cell stages through the carefully timed addition and withdrawal of defined combinations of signaling factors (22–25,163,164). These protocols suffer, however, from the production of non-endocrine cell types and a failure to match the transcriptional profiles and glucose responsiveness of primary adult human islets. This may be due to a relative lack of understanding about human endocrine development *in vivo*, as current protocols are based on knowledge of rodent development and may therefore be missing key regulatory pathways and lineage steps unique to human development. Indeed, multiple studies have identified discrepancies between mouse and human pancreatic islets, including structural (165), transcriptomic (166), and metabolic (167) differences. Therefore, gaining a deeper understanding of human endocrine development is crucial for continued progress towards generating *in vitro*-derived beta-like cells that recapitulate endogenous function.

In mice, all pancreatic epithelial cell types, including both exocrine and endocrine, are generated from a domain of the gut tube that begins to express the transcription factor *Pdx1* around embryonic day (E) 8.5 (93). The initial pancreatic bud then branches extensively, forming the tip and trunk regions of the finger-like projections that comprise the ductal epithelium as it expands and begins regional specification (168). The hormone-expressing endocrine cells, including insulin-producing beta cells, derive from endocrine progenitor (EP)

cells that activate the expression of the transcription factor *Neurog3* in a subset of ductal trunk cells (29,169). During human development, *NEUROG3* is also a presumed marker of EP cells (169). The expression of *NEUROG3* in humans begins as early as 8 weeks post conception (w), peaks at 10-12 w, and then gradually decreases to very low levels by 35 w (169). While in murine pancreatic development the expression of Neurog3 in EPs occurs in two distinct waves, in human development it has been reported to occur in a single wave, further highlighting differences between human and rodent pancreas development (35,76).

Heterotypic interactions between epithelial and non-epithelial cells are broadly important for mammalian development, including in the pancreas, where they regulate expansion of pancreatic epithelial progenitors as well as their subsequent differentiation (103–105,170). In particular, islet development depends on complex interactions between endocrine cells and multiple other cell types, including neurons and endothelial cells (170–172). To fully understand human endocrine pancreas development *in vivo* and then successfully mimic the process *in vitro* will require a comprehensive catalog of not only the endocrine cells, but also the non-epithelial pancreatic cell types, as well as signaling pathways through which they act.

Murine single-cell RNA-Sequencing (scRNA-Seq) studies performed by our laboratory and others have uncovered EP subtypes downstream of the *Neurog3*-expressing population, and some have cataloged the cell heterogeneity within other non-endocrine compartments as well[25–27]. Single-cell studies of human pancreas have begun to reveal heterogeneity of human endocrine cells. For instance, scRNA-Seq has been applied to cells generated by beta cell differentiation protocols *in vitro* and to adult human islets, where previously unappreciated levels of cellular heterogeneity were described (23,166,173–178). A study of human pancreas tissue at 9 w used single-cell qPCR on a small number of sorted cells to detect the expression of 96 prospectively defined developmental genes and described a putative EP population as well as an early differentiated endocrine cell population (179). More recently, a paper focusing on early pancreatic epithelial progenitors reported potential pathways that may be broadly mediating

interactions between mesenchymal and epithelial cells; however, the number of endocrine cells investigated was limited (178). Work focusing on human fetal endocrinogenesis identified putative EP cell states *in silico*, although these states remain to be confirmed *in vivo* and there was no investigation of how endocrine cells interact with other cell types in the developing pancreas (177). Thus, a comprehensive characterization of the full panoply of both epithelial and non-epithelial cell types in the developing human pancreas is still lacking. Importantly, the field also still lacks an understanding at the single-cell level of how cell composition and lineage trajectories of *in vitro* stem cell-derived beta cells compare to those of endogenous developing human cells.

In the quest to characterize the relevant cell states of endocrine differentiation *in vivo* that need to be recapitulated *in vitro* for cell replacement therapy approaches, the field also lacks an understanding of the epigenetic mechanisms by which those endocrine cell states are established and maintained. A study investigating *in vitro* generation of pancreatic progenitors identified heterogeneity in global chromatin accessibility depending on which *in vitro* differentiation protocol was used, highlighting the need for an *in vivo* comparator against which epigenetic data from *in vitro* differentiation platforms can be benchmarked (180). Recent studies using single-nucleus Assay for Transposase-Accessible Chromatin with Sequencing (snATAC-Seq) have provided evidence of distinct epigenetic states across endocrine subclusters in the adult islet (181). These studies also localized type 1 and type 2 diabetes-associated genetic risk variants to regions of accessible chromatin in adult islet cells and predicted their regulatory function by interpreting their co-accessibility with target genes (181–183). Although a large number of diabetes genetic risk variants have been discovered in Genome-Wide Association Studies (GWAS) (184–186), it is not clear which are operant during development, thereby exhibiting regulatory functions in a cell-specific and/or developmental stage-specific manner.

In this study, we utilize large-scale scRNA-Seq to generate a comprehensive atlas of human fetal pancreas tissue ranging from 8 to 20 w. We describe previously unappreciated

levels of cell heterogeneity within the endothelial, mesenchymal, exocrine, neuronal, immune, and endocrine lineages of the human pancreas and identify putative signaling pathways active among these lineages. Within the endocrine lineage, we identify four novel progenitor cell types, confirm their existence in independent tissue samples, and reconstruct their lineage trajectories *in silico*. By performing snATAC-Seq analysis on 12 w human fetal pancreatic tissue and performing a multi-omic analysis integrating snATAC-Seq and scRNA-Seq data, we provide novel insights into regulatory landscapes of single cells in the developing human endocrine pancreas. We also leverage the snATAC-Seq data to identify a potential developmental-specific role of multiple diabetes GWAS risk alleles. In addition, we compare the molecular profiles and cellular trajectories of endogenous human fetal pancreatic endocrine cells to those generated *in vitro* from human stem cells. Through genome editing of hPSCs, we identify the transcription factor *FEV* as a regulator of human endocrine differentiation. This study will serve broadly as a resource for the field of pancreatic development and will also provide the foundation for future improvements in therapeutic strategies for generating replacement beta cells that more closely resemble their *in vivo* counterparts both in identity and function.

3.2 Results

3.2.1 Interrogating the developing human pancreas at single-cell resolution

To characterize the cellular heterogeneity within the human fetal pancreas, we performed droplet-based scRNA-Seq on eight independent biological samples ranging from 8 to 20 w, a window of human pancreatic development encompassing specification of endocrine progenitors (EPs), differentiation into hormone-producing cell types, and islet morphogenesis (76). Each tissue sample was dissociated and subjected to red blood cell lysis to deplete erythrocytes. Resulting single-cell suspensions were then used directly (“Total Pancreas”) or subjected to magnetic bead-based selection to either enrich for EPCAM+ cells (“Epithelial(+)”)

or deplete CD45+ cells (“Immune(-)”), and then prepared for sequencing using the 10x Chromium Single-Cell Gene Expression v3 platform (**Figure 3.1a, Supplemental file 3.1**). Following sequencing and quality control processing (Methods), the resulting data were computationally merged into a final dataset consisting of 114,873 cells, each of which was then classified as belonging to a “Broad Group" based on the expression of established marker genes: Mesenchymal (*COL3A1*), Immune (*RAC2*), Exocrine (*CPA1*), Endothelial (*PECAM1*), Neuronal (*SOX10*), Endocrine (*CHGA*), and Proliferating (*TOP2A*) (**Figure 3.1b**). Individual biological samples showed varying degrees of contribution to all Broad Groups, with experimental enrichment or depletion affecting the overall contribution as expected. For instance, samples subjected to positive selection for epithelial cells showed enrichment of endocrine and exocrine populations and relatively fewer cells classified as mesenchymal, immune, or endothelial (**Figure 3.2a,b**). Importantly, we observed high concordance of technical and biological replicates based on Pearson correlation of Epithelial cells (**Figure 3.2c**).

To further investigate the cellular heterogeneity within each Broad Group, we next applied the clustering algorithm CellFindR (122,187) to our dataset (see Methods). CellFindR iteratively increases Louvain clustering resolution based on the condition that each cluster expresses a minimum of 10 genes with greater than 2-fold expression in comparison to all other clusters. Once this condition is broken, principal components are recalculated and each cluster is further sub-clustered following the same condition, creating tiers of clusters with the nomenclature of the initial cluster represented by an integer and subsequent sub-clusters followed by a period and an integer. CellFindR defines cell populations that are biologically relevant and generates a clustering map with multi-tier hierarchy. Within the merged datasets, CellFindR identified a total of 103 clusters in the developing human pancreas, including 15 mesenchymal, 8 exocrine, 8 endocrine, 8 endothelial, 13 neuronal, 31 immune, and 20 proliferating clusters. This highlights the striking and previously undescribed cell heterogeneity in the developing human pancreas (**Figure 3.1c-h**).

3.2.2 Cellular heterogeneity and cell-cell communication across non-endocrine lineages of the fetal pancreas

The importance of mesenchyme in guiding pancreatic organogenesis has been demonstrated through mechanical removal and recombination experiments, as well as genetic ablation studies (103,105,188). Recent work has shed light on the functional heterogeneity within murine pancreatic mesenchyme, where an Nkx2.5+ mesenchymal sub-population has been reported to establish a pro-endocrine niche during mouse pancreatic development (106). The full panoply of cell subtypes within the human fetal pancreatic mesenchyme and their roles in heterotypic cellular signaling, however, remain unknown. In this study, we employed CellFindR to identify 15 sub-clusters of mesenchymal cells (**Figure 3.1d, Figure 3.3a**), including known cell types such as vascular smooth muscle (Immature and Mature VSM, clusters 6.0 and 6.1, respectively), Pericytes (cluster 2) and Mesothelial cells (cluster 7), annotated based on differential gene expression analysis (**Figure 3.3b**, **Supplemental file 3.1**). In addition, several novel populations were identified that express modulators of WNT signaling, including genes encoding Secreted Frizzled Related Protein 1 (*SFRP1*) (SFRP1^{hi}CEBPD+, cluster 1.0; SFRP1^{hi}, cluster 1.1) and *SFRP2* (SFRP2+, cluster 0). We also discovered a heterogenous population of cells enriched for expression of C-C Motif Chemokine Ligand 2 (*CCL2*): CCL2^{hi}PRRX1+ (cluster 3.0), and CCL2^{hi}CCL21+ (cluster 3.2). In contrast to the other mesenchymal populations, the relative proportions of mature VSM (0.35% of total mesenchymal cells at 8 w; 2.9% at 10 w, 2.04% at 12 w) and CCL2^{hi}CCL21+ (0.18% of total mesenchymal cells at 8 w; 2% at 10 and 12 w) populations increased only by 10 w (**Figure 3.3c**), suggesting that these populations may expand later in development compared to the rest of the mesenchymal compartment.

We next validated the existence of novel mesenchymal subpopulations by performing multiplexed *in situ* hybridization (ISH) and immunofluorescence (IF) staining for *SFRP1*, *SFRP2,*

and *CCL21* transcripts, along with the broad mesenchymal marker protein VIMENTIN (VIM) on tissue at developmental stage 18 w (**Figure 3.3d**). *SFRP1*, *SFRP2*, and *CCL21* colocalized with VIM, confirming their classification as mesenchymal genes. *SFRP1* was detected in a broad population of VIM+ mesenchymal cells, whereas *SFRP2* was expressed in a restricted subset of VIM+ cells that were also found to be positive for *SFRP1* (**Figure 3.3d**), confirming the predictions of the scRNA-Seq data (**Figure 3.4a**). In addition, *CCL21* expression overlapped with a subset of VIM+ mesenchyme (**Figure 3.3d**). Collectively, these data confirmed the *in vivo* existence of heterogeneous mesenchymal populations inferred by CellFindR.

To understand how mesenchymal cell subtypes may be communicating with one another via cell-cell signaling, we employed the computational package CellChat, which infers cellular communication within complex scRNA-Seq datasets (145). Signaling pathways scored as significantly active between mesenchymal subtypes included the FGF, EGF, and COLLAGEN pathways (**Supplemental file 3.2**). We focused on the PDGF signaling pathway, as it has been shown to be important for mesenchymal development in multiple other organs, including metanephric organs and the gastrointestinal tract (189,190). We found that the immature and mature VSM cell populations scored highest as producers of PDGF ligands ("Senders"), while multiple clusters were predicted "Receiver" populations (**Figure 3.3e**). In particular, the PDGFA-PDGFRB ligand-receptor pair was predicted to have the highest relative contribution to mesenchymal PDGF signaling, followed by PDGFA-PDGFRA (**Figure 3.3f**). To better understand the contribution of each ligand-receptor pair, we analyzed the dominant Senders and Receivers for each pair separately. For the PDGFA-PDGFRB ligand-receptor pair, the dominant sources of PDGFA ligand were predicted to be the immature and mature VSM clusters, and the dominant receiver through the PDGFRB receptor was predicted to be pericytes (**Figure 3.3g**). This result is similar to PDGF signaling in the retina, where pericytes are recruited to developing endothelium through PDGF signaling to aid in formation of the blood-retinal barrier[51]. When analyzing the PDGFA-PDGFRA ligand receptor pair, we found that the

dominant receivers of the PDGFA ligand through the PDGFRA receptor were predicted to be the SFRP-expressing clusters (SFRP1^{hi}CEBPD+, SFRP1^{hi}, and SFRP2+) (**Figure 3.3h**). The effects of PDGFA-PDGFRA signaling in these populations are unknown and warrant further study.

Endothelial cell-derived signals are essential for proper formation of the murine pancreas (171,191). The murine pancreas is surrounded by vasculature as early as E10.5, with arterial and venous specification occurring at E11 (192). In human fetal pancreas tissue, CD31+ blood vessels are present as early as 7 w (193). Previous studies demonstrated phenotypic (194) and functional (192,195) heterogeneity within pancreatic endothelial populations. However, transcriptional heterogeneity among endothelial cells in the human fetal pancreas has not yet been investigated at the single cell level. Thus, we next focused on the endothelial compartment of our single-cell dataset to better understand its potential signaling role in pancreatic organogenesis. CellFindR identified eight clusters within the endothelial Broad Group, with five main subtypes (**Figure 3.1e**). These included RGCC+ capillaries (cluster 0), NR2F2+/ACKR1+ venous cells (cluster 2), PRND+/IGFBP3+ angiogenic tip cells (cluster 1), proliferating cells (clusters 4.0 and 4.1), and a heterogeneous population of GJA5+ arterial cells (GJA5+/HPGD+; cluster 3.0; GJA5+; cluster 3.1; GJA5+/SERPINE2^{hi}, cluster 3.2) (**Figure 3.3i,j, Figure 3.4b, Supplemental file 3.1**). We detected no major shifts in prevalence of these populations across the developmental timepoints within our dataset (**Figure 3.4c**), suggesting that pancreatic endothelial specialization is already established at 8 w.

To validate our CellFindR inferences, we performed ISH for arterial marker *GJA5* (196), capillary/venous marker *PLVAP* (197), and venous marker *ACKR1,* along with antibody staining for pan-endothelial marker CD31, in 18 w fetal tissue (**Figure 3.3k,l**). *In vivo* analysis revealed mutually exclusive GJA5+ arterial and ACKR1+ venous blood vessels that co-localized with CD31 (**Figure 3.3k**). We also observed PLVAP+/CD31+ capillary/venous cells that did not

colocalize with *GJA5* probe (**Figure 3.3l**). Taken together, these data establish endothelial cellular heterogeneity in the developing fetal pancreas.

To investigate cellular communication among pancreatic endothelial populations, we again applied CellChat analysis, which identified the NOTCH signaling pathway as among the most significant (**Supplemental file 3.2**). While NOTCH signaling has previously been shown to be a critical regulator of endothelial specification (198) its role in fetal pancreatic endothelial development and maturation has yet to be described. We found that Arterial populations 1, 2, and 3 scored the highest as “Senders” of NOTCH ligands, while the Arterial_1 and Angiogenic Tip Cell populations scored the highest as “Receivers” (**Figure 3.3m**). When assessing the contribution of each ligand-receptor pair, we observed that the JAGGED1 (JAG1)-NOTCH4 pair was predicted to make the highest relative contribution to NOTCH signaling within our endothelial dataset (**Figure 3.3n**). NOTCH ligand JAG1 has been implicated as a pro-angiogenic molecule that can counteract the anti-angiogenic effects of Delta-Like Canonical Notch Ligand 4 (DLL4)-NOTCH signaling in mice (199). Within our dataset, the arterial populations were scored highest as “Senders” of the JAG1 ligand, with both the Angiogenic Tip cell and Arterial_1 population predicted as “Receivers” (**Figure 3.3o**). These data suggest that pancreatic arterial cells may maintain angiogenic tip cell fate through NOTCH signaling mediated by the pro-angiogenic molecule JAG1.

Heterogeneity was also discovered within the remaining Broad Groups. In the exocrine compartment, for instance, we observed *CFTR*-expressing ductal cells (clusters 2.0, 2.1.0 and 2.1.1; **Supplemental file 3.1**) (200). In addition, three sub-clusters of acinar cells appeared to represent a continuum of maturation states, characterized by varying degrees of expression of digestive enzymes (clusters 1, 3.0, and 3.1) (**Figure 3.4d**, **Supplemental file 3.1**). We annotated Exocrine cluster 0.0 as Pre-Acinar and 0.1 as Pre-Ductal cells, based on their displaying gene expression profiles most highly correlated to acinar and ductal cells, respectively (**Figure 3.4e**), but not markers characteristic of mature acinar or ductal populations

(i.e *PRSS1* and *CELA3*, Acinar; *CFTR*, Ductal) (**Supplemental file 3.1**). These may represent tip and trunk cells that eventually give rise to the differentiated exocrine tissue (34). In the Neuronal Broad Group, we identified clusters representing myelinating Schwann Cells (clusters 0.0, 0.1, and 0.2), as well as peripheral nerve subtypes expressing various neurotransmitters such as *VIP, NPY,* and *NOS1* (clusters 1.0.0, 1.0.1, 1.1.0, 1.1.1 and 1.2) and proliferating neuronal cells (clusters 2.0.0, 2.0.1, 2.1.0, 2.1.1, 2.2) (**Supplemental file 3.1**; **Figure 3.4f**). In the Immune Broad Group, we identified 16 lymphoid lineage populations, including T, B, and NK cells. We also identified 15 myeloid lineage populations, including neutrophils, monocytes, macrophages, and mast cells (**Figure 3.4g**)**.** The proportion of myeloid and lymphoid lineage cells remained relatively stable from 8 to 12 w, while a substantial increase in the proportion of lymphoid cells was observed at 16 w, possibly due to the infiltration of B cells from the blood (**Figure 3.4h**).

Next, we deployed CellChat to interrogate the incoming and outgoing signaling pathways active among all subtypes of each Broad Group (**Figure 3.4i,j**). The Midkine (MK), pleiotrophin (PTN), and RESISTIN pathways were the top three most highly scored both for incoming and outgoing signaling (**Supplemental file 3.2**). Previous studies in mice revealed the fundamental function of VEGF signaling in islet vascularization and vessel architecture (192,195). CellChat predicted that in the developing human fetal pancreas, VEGF signaling was strictly sensed by endothelial cells (receivers) and secreted by epithelial cells. VEGF signaling cross talk was also observed between pancreatic mesenchyme (sender), epithelium (sender), and endothelial cells (receiver), suggesting that mesenchymal and epithelial cells promote vascular modeling through the secretion of VEGF in the fetal human pancreas, paralleling murine development. In addition, mesenchymal cells were found to produce a variety of paracrine factors, including WNT and FGF ligands (**Figure 3.4j**), which are critical for endothelial and epithelial development (201). CXCL secreted by cells in the endothelial Broad Group and the $CCL2^{hi}$CCL19+ mesenchymal population was predicted to act on a wide range of immune cell types, including monocytes,

macrophages, and DCs (**Figure 3.4i,j**). Taken together, these predicted coordinated signaling interactions across cell types highlight the significance of cell-to-cell communication in human fetal pancreas organogenesis and provide a framework for future studies investigating heterotypic cellular signaling in the human fetal pancreas.

3.2.3 Discovery and characterization of novel human endocrine progenitor cell populations

Mapping the cellular and molecular landscape across human endocrine development is a critical step for improving stem cell-derived therapies for diabetes. CellFindR permitted the identification of four hormone-expressing endocrine clusters in the human fetal pancreas, distinguished by the expression of *INS* in beta cells (cluster 0; 6,700 cells), *GCG* in alpha cells (cluster 1; 3,388 cells), *SST* in delta cells (cluster 3; 1,554 cells), and *GHRL* in epsilon cells (cluster 4; 909 cells) (**Figure 3.5a,b**). The fifth major pancreatic endocrine cell type, *PPY*-expressing PP cells, were not detected as a separate cluster, although *PPY* expression was observed in alpha cells (**Figure 3.5b**). We also observed overlapping expression of genes encoding hormone markers in some of the hormone+ cells (**Figure 3.5b**), likely representing polyhormonal cells that have been previously reported to exist in human endocrine development (202,203). Indeed, multiplexed IF staining of 9 w human fetal pancreas tissue for hormones C-PEP, GCG, and SST confirmed that a fraction of endocrine cells are bi- or even poly-hormonal at the protein level (**Figure 3.6a**).

In addition to the expected hormone+ populations, CellFindR also annotated a remaining fifth endocrine cluster, cluster 2, which we classified as a putative endocrine progenitor (EP) cell population based on its specific expression of *NEUROG3* and absence of hormone expression (**Figure 3.5a,b**). CellFindR further sub-clustered cluster 2 into four sub-clusters: 2.0.0 (453 cells), 2.0.1 (450 cells), 2.1.0 (421 cells), and 2.1.1 (356 cells), resulting in a final endocrine

dataset comprised of eight clusters (**Figure 3.1h**, **Figure 3.5a**). These data reveal heterogeneity within the progenitor pool of the human pancreatic endocrine compartment.

Next, we tested the hypothesis that the four subclusters of cluster 2 indeed represent biologically distinct EP states with varying degrees of potency. Lineage inference with RNA splicing was not successful in our endocrine dataset, reflecting recognized issues with RNA Velocity in multiple datasets (204). Instead, we used the computational package CellRank (205), in combination with pseudotemporal ordering, to analyze the endocrine cells from all developmental timepoints. First, we used Slingshot (206) to assign pseudotime values to each cell (**Figure 3.5c**). We set cluster 2.1.0 as the root cluster (beginning of pseudotime), as it has the highest expression of the EP marker *NEUROG3* (**Figure 3.5b**). We then used these pseudotime values as input for CellRank, creating a directed transition matrix (**Figure 3.5d**). This matrix was then decomposed into macrostates (summarized individual gene expression profiles) that represent the inferred initial state (which a cell is unlikely to revisit after leaving) and terminal state (which a cell is unlikely to leave after visiting) (**Figure 3.6b**). Finally, the transition matrix was used to calculate the fate probability of each cell, computing the likelihood that a given cell will transition to each of the terminal macrostates (**Figure 3.5e,f**).

In our analysis we calculated 5 macrostates: 4 representing terminal cell types (Alpha, Beta, Delta, and Epsilon) and the EP cluster 2.1.0 (**Figure 3.6b**). As the transition matrix and absorption probabilities indicated that cluster 2.1.0 served as a common progenitor for the Alpha, Beta and Epsilon populations (**Figure 3.5e,f**), we henceforth refer to this cluster as the Common Endocrine Progenitor (cEP) population. Cluster 2.1.1 (henceforth referred to as the Pre-Alpha/Pre-Epsilon population) followed the cEP cluster in the transition matrix (**Figure 3.5d**) and served as a bifurcation point, being biased in absorption probabilities to either the Alpha or Epsilon clusters (**Figure 3.5e,f**). Along an alternative branch, cEP cells were followed in the transition matrix by cluster 2.0.1, which expressed the highest levels of the gene *FEV* (**Figure 3.5g**), a transcription factor we and others previously identified as a marker of a novel EP cell

state during murine endocrine development (58,60,62). This highly *FEV*-expressing cluster, henceforth referred to as the FEV High population, shared high absorption probabilities with cluster 2.0.0 (henceforth referred to as the Pre-Beta population) for a Beta cell fate (**Figure 3.5e,f**). Taken together, these results predict the lineage trajectories for the Alpha, Beta, and Epsilon lineages, while the Delta lineage remains ambiguous. The lineage trajectories in human endocrine development predicted here are distinct from those reported in mouse and human development (58,60,62,177).

To assess how each of these distinct endocrine cell states varied across developmental time, we quantified their population dynamics across developmental time. Each of the four EP populations was present in every biological scRNA-Seq sample, across all timepoints sampled (**Figure 3.6c**). The relative proportion of each of the progenitor populations was highest at 8 w (cEP, 12.8%; Pre-Alpha/Pre-Epsilon, 11.0%; FEV High, 12.4%; Pre-Beta, 10.2% of all endocrine cells), then decreased as developmental time progressed to reach their lowest levels at 20 w (cEP, 2.2%; Pre-Alpha/Pre-Epsilon, 1.6%; FEV High, 1.8%; Pre-Beta, 2.1% of total endocrine cells) (**Figure 3.5h**), consistent with their characterization as progenitor populations. In contrast, the proportion of beta cells in the developing pancreas steadily increased across developmental time (24% of all endocrine cells at 8 w; 60% at 20 w). The proportion of alpha, delta and epsilon cells remained relatively stable from 10 to 20 w at proportions approximating those reported in human adult islets (207). These data are consistent with a model whereby the pool of more differentiated endocrine cells increases over developmental time at the expense of a dwindling progenitor pool. These results further support the hypothesis that the four endocrine cell subclusters represent novel endocrine progenitor populations in the human fetal pancreas.

3.2.4 Transcriptional regulation of human pancreatic endocrine development

To identify the transcriptional features that distinguish each endocrine cell population, we performed a series of analyses. First, we conducted differential gene expression analysis across

all eight endocrine clusters by comparing each cluster against all other clusters, resulting in a total of 858 genes with a log2 fold-change (log2FC) value of at least 0.5 (**Figure 3.6d, Supplemental file 3.1**). Of these 858 differentially expressed genes (DEGs), the majority were most highly expressed by the EP populations (643 genes). Pre-Beta (n = 73 genes) and FEV High (n = 163 genes) clusters expressed genes associated with neuronal development (*TUB1A1, NNAT* by Pre-Beta; *FEV* by FEV High) (63,208,209), while the cEP cluster expressed genes (n = 239) associated with mRNA processing and chromatin remodeling (*SRSF3, RBMX*) (210,211), potentially representing dynamic priming of gene expression needed for endocrine differentiation. Pre-Alpha/Pre-Epsilon cells also expressed genes (n = 168) associated with neuronal development and RNA processing (**Supplemental file 3.1**). Pathway analysis of the differentially expressed genes corroborated these findings, revealing that enriched pathways included those annotated as being involved in neuronal development (Pre-Alpha/Pre-Epsilon) and endocrine function (Pre-Beta) (**Figure 3.6e**).

Among hormone-producing cells, DEGs expressed by the Beta Cell population (n = 27 genes) included genes poorly characterized in beta cells (*LMO2*, *ASPH*), as well as known markers of beta cell identity (*INS)* (212) and function (*IAPP, SLC30A8, PCSK1*) (213–215) (**Figure 3.5i, Supplemental file 3.1**). Alpha Cells (n = 37 genes) expressed both genes known to be expressed in alpha cells, (*IRX2*, *TTR*), as well as those whose function in alpha cells is not well characterized, such as *EDN3*, *SPINT2,* and *CDNK1C* (**Supplemental file 3.1**). Delta Cells (n = 23 genes) highly expressed markers of delta cell identity (*SST*) and genes encoding peptides with known endocrine function (*NPW, CRH*) (216,217) (**Figure 3.5i, Supplemental file 3.1**). Of the four hormone-expressing cell types, Epsilon Cells had the highest number of DEGs detected (n = 127 genes), including genes associated with epsilon cell identity and function (*GHRL*), and those involved in cellular signaling (*FGF12, FGF1)* (**Figure 3.5i, Supplemental file 3.1**). The enriched pathways in the hormone-expressing populations included Insulin Processing (Beta Cells), Retinoic Acid Signaling (Alpha Cells) and ATF-2 Transcription Factor

Network signaling (Epsilon Cells) (**Figure 3.6e**). Together, these data describe transcriptional programs and enriched signaling pathways specifically active within each human pancreatic endocrine population.

We next set out to focus specifically on differentially-expressed transcription factors (TFs), as they are critical regulators of cell fate determination (76,218). We identified 108 TFs that were differentially expressed across the endocrine lineage (**Figure 3.6d, Supplemental file 3.1**). cEP cells displayed the highest expression of *NEUROG3* targets *NKX2-2*, *NEUROD1*, and *INSM1* (134,219,220), as well as TFs involved in Hippo-YAP signaling (*TEAD2*) and Notch signaling (*RBPJ* and *HES6*), consistent with previous evidence in mice that Notch activity is critical for maintenance of endocrine progenitor cell fate (221) (**Figure 3.6d**). The reciprocally inhibitory interactions between *Pax4* and *Arx* promote the acquisition of beta cell or alpha cell fate during murine endocrine development (97). Accordingly, we found that the FEV High and Pre-Beta EP populations expressed high levels of *PAX4*, whereas the Pre-Alpha/Pre-Epsilon EP cluster showed elevated levels of *ARX* expression during human pancreas development. Expression of the TF *NKX6.1*, a crucial regulator of beta cell fate (102), was enriched in the FEV High and Pre-Beta populations. Moreover, the highest expression of beta cell regulators *PDX1* (222), *MNX1* (223), and *PAX6* (224) appeared in the Pre-Beta but not in the FEV High EP cells, suggesting that these genes may play a functional role in beta cell fate restriction. We also identified TFs that were specifically expressed in a specific EP cell type, such as *TOX3* (cEP), *SIM1* (Pre-Beta), and *POU2F2* (Pre-Alpha/Pre-Epsilon) (**Figure 3.6d**). These results provide a rich dataset of TFs that warrant further study in developing human endocrine cells.

We next aimed to compare the gene expression profiles among the four sub-clustered EP populations to identify which genes had driven their distinction by CellFindR. Comparison of the cEP vs. Pre-Alpha/Pre-Epsilon clusters uncovered higher expression in cEP cells of *SOX4,* which cooperates with *Neurog3* to regulate endocrine induction in the murine pancreas (125) (**Figure 3.6f**). In contrast, Pre-Alpha/Pre-Epsilon cells more highly expressed genes associated

with differentiated endocrine cells, such as *ARX*, *ISL1*, and *GHRL*. When comparing cEPs vs. FEV High cells, we observed higher expression of *SFRP1, PROX1* and *CITED2* in the cEP cluster (**Figure 3.6g**). Fev High cells expressed genes associated with hormone secretion, such as *PCSK1N, PCSK1* (215) and *KCNK17* (225). Pairwise comparison of Pre-Beta vs. FEV High cells revealed higher expression of genes associated with beta cell maturation and function (*MAFB, PCSK1N, GNAS*) in the Pre-Beta cluster, while the FEV High cluster showed higher expression of genes *FEV* and *HES6* (**Figure 3.6h**). The results from these pairwise comparisons corroborate our hypothesis, based on lineage reconstruction, that the EP populations represent distinct cell states that are pre-committed to one or more hormone-producing cell fates. Taken together, these data have enabled the construction of a model of human endocrine lineage specification and the identification of novel genes governing cell fate decisions during human fetal development.

3.2.5 Confirmation of novel endocrine progenitor populations *in vivo*

To confirm the findings of EP cell heterogeneity that had emerged from the scRNA-Seq data, we first set out to identify genes or combinations of genes that could serve as specific markers of each of the four EP populations. Manual curation of top differentially expressed genes across EP cells identified Sushi Domain Containing 2 (*SUSD2),* LIM Homeobox Transcription Factor 1 Beta *(LMX1B)*, Peripherin (*PRPH*), and Aristaless-Related Homeobox *(ARX)* as highly enriched in the cEP, FEV High, Pre-Beta, and Pre-Alpha/Pre-Epsilon EP populations, respectively **(Figure 3.7a,b).** Among these four genes, *SUSD2* has been broadly described to label *NEUROG3*-expressing EP cells in the developing human pancreas (226,227). *ARX* is an important regulator of alpha cell fate (228) that has not been described to mark human Pre-Alpha/Pre-Epsilon EP cells. *LMX1B* and *PRPH* are novel markers labeling sequential progenitor states during human beta cell development.

Once markers of each of the four novel EP subtypes had been identified, we next validated the existence of each EP cell state *in vivo* by performing multiplexed ISH/IF staining of sections of independent human fetal pancreas samples at 8, 12, and 18 w. ISH was performed for each putative marker gene alongside the pan-EP marker *NEUROG3* and combined with IF staining for the pan-differentiated endocrine cell marker CHGA. When staining tissues to validate the presence of the cEP population, we found five unique cell states based on their expression of *SUSD2*, *NEUROG3*, and/or CHGA (**Figure 3.7c,d**). As predicted by our scRNA-Seq analysis, we detected putative cEP cells, characterized as *SUSD2*$^{+}$*NEUROG3*$^{+}$CHGA^{-} (23.7% at 8 w, 9.8% at 12 w, 2.1% at 18 w) (**Figure 3.7c,d**). The same experimental and quantification approaches were adopted to validate the existence of the other three EP types. Thus, we observed the presence of *LMX1B*$^{+}$*NEUROG3*$^{+}$ (putative FEV High EP cells; 20.2% at 8 w, 16.7% at 12 w, 3.6% at 18 w), *PRPH*$^{+}$*NEUROG3*$^{+}$ (putative Pre-Beta EP cells; 12.6% at 8 w, 9.3% at 12 w, 2.5% at 18 w), and *ARX*$^{+}$*NEUROG3*$^{+}$ (putative Pre-Alpha/Pre-Epsilon EP cells; 2.2% at 8 w, 4.9% at 12 w, 1.1% at 18 w) cells that were also negative for CHGA (**Figure 3.7e-j**). All four EP populations were detected in nine independent biological samples of pancreas tissue.

To further confirm the existence of four distinct populations, we next performed multiplexed ISH staining at 8 w to simultaneously detect markers of all four EP cell types; we found that the expression of genes predicted to specifically mark each EP population indeed showed mutual exclusivity *in situ* (**Figure 3.8a**). Moreover, the relative prevalence of all four EP populations decreased over developmental time (**Figure 3.7d,f,h,j**), consistent with the characterization of these populations as progenitor populations. Collectively, these results confirmed the presence of the novel EP subtypes in the developing human pancreas as predicted by computational analysis of the scRNA-Seq data.

Our CellFindR inferences identified the TF *FEV* as a marker for both FEV High and Pre-Beta EP populations (**Figure 3.5g**). Given that we previously had identified *Fev* as a marker of

EP cells in the murine pancreas (58,62), we assessed the dynamics of *FEV* expression in the human fetal pancreas. Consistent with our scRNA-Seq data, the percentage of $FEV^{+}NEUROG3^{+}$ double positive cells (marking FEV High or Pre-Beta EP cells) decreased over developmental time (13.4% at 8 w, 5.9% at 12 w, 2.8% at 18 w) (**Figure 3.8b**). We also observed a significant rise in $FEV^{+}CHGA^{+}$ double positive cells in 18 w tissue compared to 12 w and in 12 w tissue compared to 8w **(Figure 3.8b)**. Further investigation revealed that *FEV* expression localized to GCG-expressing alpha cells, but not INS-expressing beta cells, at 18 w **(Figure 3.8c).** This observation agrees with previously published reports that in adult human islets, *FEV* is exclusively expressed in the alpha cells (173,174,229). Along the beta cell lineage trajectory, *FEV* expression was detected in the FEV High and Pre-Beta populations but absent in differentiated beta cells themselves. In contrast, along the alpha cell lineage, *FEV* was expressed in differentiated alpha cells themselves but not in their progenitors **(Figure 3.7a,b, Figure 3.8d)**. These results demonstrate regulation of *FEV* expression across human endocrine development and suggest a dynamic and lineage-specific role in regulating alpha vs. beta cell fate and/or function.

3.2.6 Single-nucleus ATAC-Seq of human fetal endocrine cells reveals dynamic chromatin accessibility and gene regulatory networks

Recent advances in single-cell technologies have allowed for the integration of multi-omic single-cell data, leading to new insights into developmental biology (111,230,231). We set out to gain an understanding of the epigenetic mechanisms upstream of gene expression that are important in governing cell identity during endocrine development. To this end, we performed snATAC-Seq on human fetal pancreas using the 10x Genomics Chromium Next GEM Single Cell ATAC v1.1 platform. To increase the resolution for endocrine cell types, we enriched for EpCAM+ epithelial cells in 12-week fetal pancreas, a particularly active time of cell expansion and diversification (**Figure 3.9a**). Filtering, dimensional reduction, initial clustering,

and analysis steps were performed with the R package ArchR (121), resulting in a final dataset comprising 5,949 nuclei with a median of 31,158 fragments captured per nucleus (**Figure 3.10a**). Among these nuclei, 1,737 were classified as belonging to endocrine cells (**Figure 3.9a, Figure 3.10a**) based on the Gene Score Matrix (accessibility of gene body plus promoter) of *CHGA*.

Next, we integrated our endocrine snATAC-Seq and scRNA-Seq datasets to perform multi-omic analysis in the same cell types. Each cell in the snATAC-Seq dataset was correlated with its most similar counterpart in the scRNA-Seq dataset by correlating the Gene Score Matrix with the scRNA-Seq expression matrix (RNA transcript counts) on a per-cell basis (**Figure 3.9a**). Once these highly correlated pairs were found, the snATAC-Seq data from each cell were associated with the corresponding cell type label and RNA expression matrix. Of note, transfer of a cell label was not forced if the inferred gene score from snATAC-Seq data did not correlate with gene expression in any of the cells within the scRNA-Seq dataset. Our integration analyses identified eight endocrine populations in the snATAC-Seq dataset, including the four newly identified EP populations (cEP, FEV High, Pre-Beta, and Pre-Alpha/Pre-Epsilon) and four hormone-expressing populations (**Figure 3.10b**). The Gene Score of the EP marker genes identified by scRNA-Seq analysis showed high concordance with their corresponding RNA expression levels (**Figure 3.10c**), confirming the existence of EP cell states using single-nucleus chromatin accessibility analysis. Due to the low numbers of nuclei within each individual EP subpopulation, we merged all four of the EP populations into one single cluster, resulting in a final dataset of 5 clusters, consisting of Alpha (373 nuclei), Beta (608 nuclei), Delta (273 nuclei), Epsilon (160 nuclei), and pooled Endocrine Progenitor (EP) cells (323 nuclei) (**Figure 3.9b,c**). These data provide further confirmation of the discovery by scRNA-Seq of four EP cell states, using an orthogonal method of snATAC-Seq.

We next sought to understand the gene regulatory networks (GRN) that govern human endocrine cell fate decisions. We applied the computational package Single Cell rEgulatory

Network Inference and Clustering plus (SCENIC+) (232) to our endocrine dataset. SCENIC+ uses either same-cell or paired scRNA-Seq and snATAC-Seq multi-omic data to infer TF binding to target regions of DNA to affect expression of downstream target genes. The set of TFs with their corresponding regions of DNA binding and target genes are referred to as eRegulons and can be classified as either activating (where the TF binds to target regions of DNA and induces gene expression of the target genes; denoted as "_+_") or repressing (where the TF binds to target regions of DNA and represses the expression of the target genes; denoted as "_-_").

To avoid variability in RNA expression across developmental time, we used just the endocrine scRNA-Seq dataset from the 12 w EPCAM sorted sample, in combination with the 12 w EPCAM sorted snATAC-Seq dataset, as input for analysis by SCENIC+. First, we identified regions of accessible chromatin in each of the endocrine cell types using MACS2 (123), identifying 197,522 consensus peaks. We then used this peak set, along with the associated gene expression data from the scRNA-Seq dataset, to create a GRN of the endocrine cells using SCENIC+ (see Methods for details). The result was the identification of 217 TFs with predicted regulatory function and a total of 76,579 regulatory interactions between TFs, target chromatin, and target genes (**Supplemental file 3.3**). Of these, 66,031 were deemed as activating and 10,548 as repressive. We then performed dimensional reduction on the eRegulon target regions and target genes (**Figure 3.10d**). Next, we identified cell-type specific activating and repressing eRegulons by calculating the Regulon Specificity Score (RSS) for each eRegulon. First, we identified high confidence eRegulons by correlating the region (peak)-based and gene-based regulons. These eRegulons were then scored for each cell type by the their Regulon Specificity Score (RSS). This analysis identified 64 cell type-specific activating eRegulons and 29 repressive eRegulons (**Figure 3.9d**). Known activators in the EP cell type included TFs such as *NEUROG3*, *PAX4*, *INSM1*, *NEUROD1/2* and *FEV*, while unknown TFs included numerous zinc finger nucleases (i.e., *ZNF219*, *ZNF300*, *ZNF431*), *TCF3, LMX1B,* and

NHLH1 (**Figure 3.9d, Figure 3.10e**). Targets of *NEUROG3* included TFs such as *PAX4*, *LMX1B*, *NKX2.2*, *SOX4*, *RFX3*, and *NEUROG3* itself (**Supplemental file 3.3**), all of which are consistent with recent CUT&RUN data performed in human stem cell-derived pancreatic EP cells (129). There were several specific eRegulons for the beta population, including known TFs such as *GLIS3*, *MAF* family members (*MAF*, *MAFA*, *MAFB*, *MAFG*), *NKX6-1*, and *PDX1*, as well as unknown TFs such as *NR3C1*, *ARID3A*, *ETS2*, and *LHX4* (**Figure 3.9d**). Epsilon eRegulon TFs included *ETV1, JUNB, OVOL2*, and *PBX1*, while Delta eRegulon TFs included *HHEX, ISL1, NKX6-3*, and *FOXO1*. Interestingly, there were no eRegulons that were deemed specific to the alpha population.

Next, we sought to identify eRegulons that display cooperativity in their binding to shared regions of chromatin. We queried the overlap in target regions for the top TFs per cell type based on the RSS calculated above and identified TFs that target the same open chromatin. TFs that display cooperativity were mainly present in the EPs (*NEUROG3, NEUROD2, NHLH1*) and Beta cells (*MAFA, MAFB, LHX4, MNX1*) (**Figure 3.10f,g**). Together, these data represent a rich resource for the field of pancreas endocrine biology for querying not only TFs regulating endocrine development, but also their downstream target chromatin regions and target genes.

3.2.7 Identification of development-specific diabetes GWAS risk loci

Gene discovery efforts for monogenic forms of diabetes attest to the importance of TFs involved in pancreatic development for normal glucose homeostasis (233). T2D is a complex disease with multiple associated genetic risk loci identified through GWAS (>400 distinct signals identified (184,234), the majority of which are located in non-coding regions of the genome with a presumed regulatory function (184). This stands in contrast to T1D, which has far fewer loci identified by GWAS (78 distinct signals (235)), the majority of which exert their effect through the immune system (235). There is compelling evidence from both human physiology (233,236) and epigenomics (237) that pancreatic islets are a key tissue mediating a large proportion of the

genetic risk for T2D. Since gene regulation is highly context-specific, we hypothesized that defects in islet cell development could emerge earlier in the cell lineage in progenitor cells and that some of the regions of chromatin more accessible during development would overlap with T2D, and potentially T1D, risk loci. Therefore, a comparison between human fetal endocrine and adult islet cells would present a unique opportunity to identify signals specific to endocrine development.

We compared chromatin accessibility in endocrine cells from our 12 w fetal pancreas snATAC-Seq dataset with snATAC-Seq data from adult human islets (183), focusing on differences between (a) adult vs. fetal beta cells and (b) adult hormone-positive cells vs. fetal EPs. We identified 146,589 differentially accessible peaks between adult and fetal beta cells, with the majority of peaks accessible in fetal beta cells (129,937 peaks) compared to adult (16,652 peaks) (**Supplemental file 3.4**). Next, we investigated whether T1D (186) and T2D (184) risk alleles from published fine-mapped credible sets at each locus (cumulative posterior probability of association ≥ 99%) were differentially accessible in fetal vs. adult beta cells. First, we used fGWAS (238) to calculate the overall enrichment of T1D and T2D single nucleotide polymorphisms (SNPs) within the peaks of each cell type in both the fetal and adult datasets (**Figure 3.9e**). Consistent with previous studies, we observed strong enrichment of T2D SNPs in both fetal and adult endocrine cells. Conversely, T1D SNPs showed weaker enrichment in fetal and adult endocrine cells, corroborating the idea that most T1D SNPs exert their effect in non-endocrine cells (235). We thus focused our follow up investigations on T2D SNPs in fetal vs. adult endocrine cells.

When comparing differential chromatin accessibility in fetal vs. adult beta cells, we identified 28 loci (GoShifter score > 0.2) that were enriched within the differential peaks in fetal beta cells. Among these loci most highly ranked by GoShifter were known endocrine regulatory genes, including genes involved in monogenic diabetes such as *HNF1A*, *HNF1B*, *GCK*, and *NEUROG3* (**Figure 3.9f, Supplemental file 3.4**). Other top ranked loci included genes reported

to regulate beta cell mass, such as *CCND2* (239) (rs3217792 and rs76895963) and *KCNQ1/CDKN1C* (rs2237897) (240), which is in line with the hypothesis that developmentally determined beta cell mass may predispose individuals to T2D (241). We also identified loci containing genes not annotated to have a functional role in endocrine development, such as *LRFN2*, a gene involved in neurite outgrowth in the brain (242) met, where fine-mapping has previously resolved the casual variant to a single SNP (184) (**Figure 3.9f**). Our analysis also identified variants at the *PROX1* locus (**Figure 3.9g**), where we have previously demonstrated tissue-specific effects (in liver vs. islets) of T2D-associated variants on expression (234,243). Our findings here now expand the possibility that these tissue-specific effects extend to temporal effects on expression of this TF during islet cell development (**Figure 3.9f,g**). A similar analysis comparing adult hormone-positive cells (11,242 differentially accessible peaks) vs. fetal EPs (98,334 differentially accessible peaks) identified significant peaks at 23 T2D loci (**Figure 3.11a**), including *WDR72*, a gene involved in endocytic vesicle trafficking mediated enamel mineralization (244) that has no reported function in endocrine development or function. As expected, based on the fGWAS results (**Figure 3.9e**), we observed less enrichment of T1D SNPs by GoShifter within the differentially accessible chromatin of fetal beta cells compared to the T2D SNPs (**Figure 3.11b**). Among the top ranked risk variants were genes with reported function in beta cells, such as *GLIS3*, *INS-IGF2*, and *MEG3* (**Figure 3.11c**), as well as *TYK2*, which has recently been reported to play a role in human fetal beta cell development (245). Together, these results provide a framework for identifying cell type- and developmental stage-specific T1D and T2D genetic risk loci, thus generating mechanistic insights into the regulatory mechanisms of diabetes-associated SNPs in the context of human pancreas development.

3.2.8 Benchmarking *in vitro* stem cell-derived endocrine cells against *in vivo* human fetal development

Our analyses thus far permitted the construction of a cellular and transcriptional roadmap of human fetal endocrine differentiation, elucidating cellular heterogeneity, inferred lineage relationships, and candidate cell fate regulators. This presented an opportunity to benchmark protocols for differentiating human stem cells into beta-like cells (BLCs) *in vitro* against endogenous human fetal development *in vivo*. Current *in vitro* protocols entail the differentiation of hPSCs through a step-wise process, first to definitive endoderm, then to primitive gut tube, then the pancreatic progenitor stage, followed by an endocrine progenitor stage, and finally terminating in insulin-expressing BLCs (**Figure 3.12a**) (22,163,246). Although tremendous progress has been made in refining these directed differentiation protocols, hurdles still remain to generating beta cells that are fully functionally mature and that are not plagued by the generation of mis-differentiated cell types (162).

A recent scRNA-Seq study profiled the cellular composition and transcriptomic landscape of hPSCs undergoing directed differentiation to BLCs, specifically at the EP (Stage 5) and beta-like cell (Stage 6) stages (23). This work identified nine stem cell (sc)-derived endocrine populations that arise during the seven days of culture at the EP stage (Stage 5): NEUROG3+ progenitors (sc_Neurog3 Early, Mid, and Late), differentiated cell types (sc_Alpha, sc_Beta, and sc_SST_HHEX (Delta)), as well as three populations that presumably represent mis-differentiated cell types (sc_Enterochromaffin (sc_EC), sc_Phox2a, and sc_FEV_High_ISL1_Low). The sc_EC population was annotated as enterochromaffin-like due to some transcriptional similarities to endogenous enterochromaffin cells found in the intestine (23). More recently, it has been postulated that sc_EC cells are transcriptionally similar to human fetal EP cells and therefore may actually represent a beta cell progenitor population *in vitro* (91). We set out to resolve these discordant models, and to more broadly determine how

analogous the cellular populations generated *in vitro* are to those we observed during human fetal endocrine pancreas development (**Figure 3.5a**). We performed a comparative analysis between cells undergoing endocrinogenesis *in vitro* (during Stage 5 of the directed differentiation) and cells in our human fetal endocrine dataset. We also incorporated scRNA-Seq data from a recently published study of human fetal intestinal enteroendocrine development (247). We utilized the supervised cell type classifier scPred (248) to train a prediction model using our human fetal endocrine scRNA-Seq dataset, as well as various intestinal cell types, including enteroendocrine progenitors (Intestine Progenitor (NEUROG3+)), enteroendocrine (EE) cells (Intestine EEs) and differentiated enterochromaffin cells (Intestine EC (TAC1+) and Intestine EC (NPW+)) (**Figure 3.12b**). The classifier annotated the sc_Beta and sc_Alpha populations found at Stage 5 as largely beta (64%) or alpha (71%) in identity, respectively, while the sc_SST_HHEX population was annotated as a mixture of alpha, beta, epsilon, and delta cells, with the highest proportion (38%) of cells annotated as delta (**Figure 3.12c,d**). scPred mainly classified the sc_Neurog3_Early cluster as either Intestine Progenitor (NEUROG3+) (45%) or the cEP human fetal pancreatic population (39%), while the sc_Neurog3_Mid cluster was classified as a mixture of mostly pancreatic cEP (48%) and FEV High (30%) progenitors and sc_Neurog3_Late was largely classified as either the FEV High (32%) or Pre-Beta (36%) (**Figure 3.12c,d**). These results indicate that the transcriptional profiles of EP cell types found *in vitro* are largely similar to those of EP cell types found during endogenous human fetal development, with the exception of the sc_Neurog3_Early, which resembles both early pancreatic and intestinal EPs.

We next delved into sc_EC cells in more detail. We found that the sc_EC population was largely classified as either pancreatic Pre-Beta progenitor or Intestinal EC (TAC1+) cells (37% and 36%, respectively) (**Figure 3.12e, Figure 3.13a**). No sc_EC cell was classified as the second EC subtype found in fetal intestinal development, Intestine EC (NPW+). We did observe that sc_EC cells began to more closely resemble bona fide fetal intestine EC cells as *in vitro*

differentiation progressed through the days of Stage 5 (**Figure 3.12e**). Genes involved in the synthesis and secretion of serotonin, the main function of EC cells, were enriched in the sc_EC and intestinal EC cells, but not the Pre-Beta EP or fetal Beta cells (**Figure 3.12f**). Since sc_EC cells indeed resemble endogenous intestinal EC counterparts as time in culture continues, they are likely not *in vitro* beta cell progenitors, but may arise during *in vitro* differentiation due to the similarities between intestinal and pancreatic endocrine development (249). Taken together, our work demonstrates key similarities between *in vitro*-derived human EPs and BLCs to their endogenous human fetal counterparts and identifies the generation of cell types not found in normal pancreatic development.

When re-examining *in vitro* lineage relationships within the framework of the trained cell type classifier, we observed that beta cell differentiation *in vitro* occurs in a manner largely similar to endogenous human fetal development (**Figure 3.12g**). Veres *et al.* showed that Stage 5 sc_Neurog3_Late endocrine cells likely give rise not only to sc_Beta cells, but also to sc_EC cells. As the sc_EC population constitutes a large portion of the cells produced at the completion of the directed differentiation protocol (**Figure 3.12c, Figure 3.13b**), understanding transcriptional mechanisms that regulate their formation would aid in driving the progenitors at the previous EP stage to differentiate into beta cells as opposed to this undesired population. To identify such mechanisms, we performed differential gene expression analysis among the fetal pancreatic beta cells and immediate progenitors (Beta and Pre-Beta), *in vitro* sc_EC cells, and fetal *TAC1*+ intestinal EC cells (**Figure 3.12h, Supplemental file 3.5**). This analysis resulted in 1,607 DEGs with at least 0.5 log2FC in expression among all populations, 93 of which are TFs. TFs enriched in the sc_EC population included *MNX1*, while those enriched in the fetal *TAC1*+ intestinal EC population included *CDX2*, *PITX2,* and *HOPX*. The fetal Pre-Beta cluster showed the highest expression levels of beta cell-related TFs, such as *NKX2-2 a*nd *NKX6-2* (**Figure 3.12h**). The fetal beta population was enriched for *MAFA* and *PLAGL1*. In future work, inducing the expression of beta cell lineage genes in the terminal progenitor population *in vitro*

(sc_Neurog3_Late) may reduce or prevent the formation of these undesired, mis-differentiated cell types and improve overall efficiency of BLC generation *in vitro*.

We next aimed to determine the transcriptional differences between hormone-producing endocrine cells generated *in vitro* vs. *in vivo* by comparing *in vitro*-derived endocrine cells at the final stage of differentiation, Stage 6, to their fetal counterparts. Data on Stage 6 cells were generated by Veres *et al.* by sampling cells weekly from the same differentiation flask (week 0 – week 5). Our fetal cell type classifier classified the sc_Beta and sc_Alpha cell types largely as their fetal counterparts, and sc_EC cells were again classified as intestinal EC cells (**Figure 3.13c,d**). Correlation analysis revealed that at Stage 6, *in vitro*-derived BLCs undergo maturation and more closely resemble their *in vivo* counterparts as time in culture proceeds (**Figure 3.13e**). Differential gene expression analysis between *in vitro* and endogenous beta cells revealed enrichment of genes *MEG3*, *PLAGL1*, *INS*, and *MAFA* in the endogenous vs *in vitro* beta cells (**Figure 3.13f, Supplemental file 3.5**). ISH staining against *MEG3,* which is a long non-coding RNA implicated in murine beta cell maturation (250,251), confirmed higher expression in fetal beta cells compared to *in vitro* BLCs (**Figure 3.13g**). Although *in vitro*-derived beta cells showed higher expression of beta cell maturation genes *HOPX* and *IAPP* (252,253), they also showed higher expression of beta cell progenitor markers such as *FEV* and *PAX4* (58), as well as *DDC,* which is involved in catalyzing 5-HTP to serotonin (**Figure 3.13f**). These results highlight the value of our dataset comprising intermediate progenitor populations at single-cell resolution in enabling the more precise mapping of *in vitro*-derived populations to their endogenous counterparts.

Next, we compared *in vitro*-derived alpha cells to their fetal counterparts and found high concordance between the two populations, irrespective of the duration in culture (correlation coefficient > 0.88 at all timepoints) (**Figure 3.13e**). Differential gene expression analysis between the two alpha cell populations again revealed the enrichment of neuronal-associated genes in the *in vitro* condition, but we observed similar expression of key alpha cell-related

genes such as *ARX, IRX2, GCG, TTR,* and *FOXA2* between the *in vitro* and endogenous alpha cells (**Figure 3.13h, Supplemental file 3.5**). These results suggest that *in vitro*-derived alpha cells closely resemble fetal alpha cells in expression of key alpha cell fate regulators, despite the fact that the differentiation protocol was optimized for the generation of beta cells.

3.2.9 A functional role for *FEV* in regulating human endocrine differentiation

Our analysis of the developing human endocrine pancreas identified *FEV* as a marker of two of the EP populations within the beta cell lineage (FEV High and Pre-Beta) (**Figure 3.5g, Figure 3.8d**). Previous genetic ablation studies in the developing murine pancreas demonstrated that global loss of the transcription factor *Fev* leads to reduced insulin production and secretion, as well as impaired glucose tolerance in adult mice (144). The role of *FEV* in regulating human pancreas endocrine development, however, is still unknown. We set out to interrogate whether *FEV* is simply a marker of pre-beta cell populations or it itself has a functional role in endocrine development in human cells. First, we queried our SCENIC+ results to determine which cell types showed enrichment of activating and repressing eRegulons for FEV. The activating FEV eRegulon (304 target genes, 402 target regions) showed enrichment in the EP population, while the repressing eRegulon (282 target genes, 394 target regions) showed enrichment in the non-EP cell types (**Figure 3.14a**), suggesting that *FEV* largely activates EP-related genes and represses genes in hormone+ cells. FEV also directly targets many TFs with known function in endocrine development, including *PAX4, NEUROG3, INSM1* (Activating), as well as *MEIS3, HHEX* and *FOXO1* (Repressing) (**Figure 3.14b**). Activators of *FEV* expression include *NEUROG3,* which has been shown to directly bind to the *Fev* promoter in a mouse beta cell line (144), *NKX2-2*, and *FEV* itself (**Supplemental file 3.3**). Repressors of *FEV* include *MEIS3, HHEX*, and *PHF1*.

Our GRN data suggests that *FEV* may play a direct role in endocrine development by binding to key TFs. To evaluate whether our *in vitro* stem cell differentiation platform could

provide experimental validation of the function of *FEV* in human endocrine development, we first asked whether the dynamics of *FEV* expression *in vitro* recapitulated those *in vivo*. *FEV* mRNA expression was first detected during the pancreatic endocrine progenitor stage (Stage 4) and peaked during the endocrine progenitor stage (Stage 5) (**Figure 3.15a**) during the differentiation. As in human fetal beta cell differentiation, *FEV* expression co-localized with *NEUROG3* at the endocrine progenitor stage (**Figure 3.15b**). These results indicated that FEV is expressed in human EP cells *in vitro* and that the stem cell differentiation platform could be used to evaluate the function of *FEV* in human EPs.

To determine the consequences of loss of *FEV* during differentiation to the beta cell lineage, we established a clonally-derived *FEV* knockout (FEV-KO) hESC line using the CRISPR-Cas9 system (**Figure 3.14c, Figure 3.15c,d**). We subjected this line, alongside a clonally-derived WT (non-edited) control line, to directed differentiation to the beta cell lineage. As measured by flow cytometry, loss of *FEV* did not lead to any significant change in the maintenance of pluripotency (NANOG/OCT3/4 co-expression) or in the efficiency of generating target cells at stages 1 through 5 of the differentiation, including the proportion of SOX17+/FOXA2+ cells at the end of Stage 1, of PDX1+ cells at the end of Stage 3, or of PDX1+/NKX6.1+ cells at the end of Stage 5 (**Figure 3.15e**). However, flow cytometric analysis revealed a significant reduction in the number of early BLCs as measured by co-expression of C-PEPTIDE (C-PEP, a proxy for INS) and NKX6.1 (a key TF marking beta cells) in FEV-KO cells (7.9 ± 2.3%; Mean ± SEM) vs. WT cells (22.7 ± 8.5%; Mean ± SEM) (**Figure 3.14d**). These results are consistent with *in silico* TF perturbation analysis using the computational package CellOracle (254), which uses scRNA-Seq and snATAC-Seq data to build GRNs among defined cell types. This GRN is then used in conjunction with pseudotemporal ordering to predict the effect of TF knockdown on the expression of downstream target genes and on trajectory outcomes. Simulated KO of key endocrine transcription factors *MAFB* and *ARX* in our human fetal scRNA-Seq/snATAC-Seq datasets resulted in expected simulated differentiation

outcomes based on experimental knowledge of murine endocrine development (255–258) (**Figure 3.15f**). Analysis of *in silico* KO of *FEV* by CellOracle predicted an inhibition of the beta lineage, confirming our *in vitro FEV* KO results. Taken together, these results suggest that *FEV* expression is not required for pluripotency, specification of definitive endoderm or pancreatic progenitors, or induction to the endocrine lineage but plays a critical role later in beta cell fate specification.

To characterize the transcriptional changes in cells upon loss of *FEV*, we performed bulk RNA-Sequencing (RNA-Seq) on FEV-KO and WT cells at the BLC stage (Stage 6, day 10) from three independent, paired differentiations. As expected, we observed very few genes (< 10 per timepoint) differentially expressed between WT and KO cells at the end of Stages 1 and 4 (**data not shown**), timepoints before peak *FEV* expression. In contrast, differential gene expression analysis of WT and KO cells at the BLC stage identified 1,837 genes with at least 0.5 log2FC change in expression (964 genes in WT, 873 genes in KO) (**Figure 3.14e, Supplemental file 3.6**). Pathway analysis of all DEGs identified enrichment of pathways related to beta cell function, such as insulin secretion and calcium signaling in WT cells, while KO cells displayed enriched pathways such as cell cycle and extracellular matrix organization (**Figure 3.14f**). Among the DEGs were known direct targets of *Fev* in the murine pancreas and brain, including *SLC6A4, GCK, LMX1B*, and *DDC* (117,144), confirming the efficacy of our CRISPR-generated KO line (**Figure 3.15g**).

To determine whether loss of *FEV* affects the expression of regulators of beta cell fate, we analyzed the intersection between genes identified by bulk RNA-Seq as enriched in either the WT or KO beta-like cells, and genes identified by scRNA-Seq as differentially expressed across the beta lineage populations during development. This analysis revealed that ablation of *FEV* resulted in down-regulation of fate regulators enriched in FEV High and Pre-Beta cells, including *NKX2.2*, *NEUROD1*, and *PAX6* (**Figure 3.14g,h**), as well as regulators of beta cell maturation and function such as *INS, MNX1,* and *IAPP* (**Figure 3.14g,h**). To verify these

findings, we performed IF staining of hormone markers on the WT and FEV-KO cell clusters at the BLC stage and observed a significant decrease in the proportion of SST+ or C-PEP+ hormonal cells in KO vs. WT cells. The ratios of SST+ C-PEP+ and GCG+ C-PEP+ bi-hormonal cells were also significantly reduced in the KO condition (**Figure 3.14i**). Taken together, these data indicate that *FEV is* not simply a marker of pre-beta EP cells but indeed also plays a role in regulating endocrine cell specification in the developing human pancreas.

3.3 Discussion

In this study, we have comprehensively characterized the transcriptome of human fetal pancreas at six developmental timepoints, ranging from 8 to 20 w, at single-cell resolution. We have identified previously unappreciated levels of heterogeneity within the various pancreatic cell types, including 15 mesenchymal, 8 exocrine, 8 endocrine, 8 endothelial, 13 neuronal, 31 immune, and 20 proliferating clusters. This resource provides a cellular and gene regulatory roadmap of early human fetal pancreas organogenesis and lays the groundwork for the interrogation of the functional significance of each of these cell types. We confirmed the presence of representative subpopulations in the endothelial and mesenchymal lineages by *in situ* hybridization in fetal tissue. Future studies are warranted to validate other newly described populations and assess their presence or absence at earlier or later developmental timepoints not covered by this study.

By computational inference we have observed active cell-cell communication between Broad Groups in the developing human pancreas. In particular, analysis using CellChat identified NOTCH-JAG and PDGF-PDGFR as potentially functional ligand-receptor pairs within the endothelial lineage and the mesenchymal lineage, respectively. Given these data and previous evidence that non-epithelial cells such as endothelial, neuronal, and mesenchymal play important roles in guiding the development of the murine pancreatic epithelium (104,106,171,172,188), the signaling pathways identified here await future functional

confirmation by methods such as genetic or small molecule-mediated loss- and gain-of-function experiments in human pancreatic tissue *ex vivo* or using hPSC-derived pancreatic cells *in vitro*.

Our detailed investigation of the human fetal endocrine compartment identified four novel progenitor populations that each express unique marker genes. Multiplexed tissue staining confirmed the existence of these distinct populations across multiple developmental timepoints and in independent biological samples *in vivo*. Recently, an independent study identified endocrine subpopulations, several of which (termed EP2, EP3, EP4) appear analogous to the ones we have described here, providing additional evidence of their existence *in vivo* (177). Hierarchical clustering of biological samples based on gene expression showed that epithelial cells from samples of similar developmental time points clustered more closely to one another, despite some samples having been subjected to positive or negative cell selection (**Figure 3.2c**). In addition, the high concordance among our *in silico* replicates gave us high confidence that these represent bona fide cellular populations in tissue (**Figure 3.2c**). The CellFindR clustering algorithm was critical in the discovery of these EP populations, as initial analyses utilizing standard clustering methods failed to distinguish these subpopulations from one another and instead annotated all four as belonging to a single EP cluster. Of note, no evidence of heterogeneity was found within each differentiated hormone-producing endocrine cell type, a topic that is debated in the adult human pancreas (166,259,260). Given the sensitivity of the clustering algorithm CellFindR in identifying cellular heterogeneity, it is likely that these heterogeneous populations do not exist in the fetal endocrine pancreas during the developmental timepoints covered by this study, although it is possible that they were too rare to detect or that they only arise later in development. Additionally, by adopting a variety of computational methods, we have inferred the gene regulatory networks and differentially expressed genes among endocrine populations. These insights serve to significantly improve our understanding of human pancreatic endocrine development at the cellular and molecular levels.

The endocrine lineage predictions constructed in this study provide a contrasting account of endocrine differentiation in human cells compared to murine development. Unlike in the mouse pancreas, where multipotent intermediate progenitors give rise to all endocrine cell types (58,60,62), our data presented here predict that three of the four human EP populations act as fate-committed progenitors that are either uni- or bi-potent. This differentiation potency is reflected in the genes expressed in each population, as endocrine cell-related genes gradually increase in expression as the EPs begin to become more fate restricted. The alpha and beta lineage predictions presented here are consistent with a recent study that utilized mitochondrial genome variants within adult alpha and beta cells as endogenous lineage tracing markers at single-cell resolution and concluded that human alpha and beta cells arise from separate progenitor populations (261). Prediction of the delta lineage presented here, however, is less clear. In particular, our analysis was unable to determine which of the progenitors directly give rise to the Delta cell cluster. **(Figure 3.5e,f, Figure 3.7b)**. Of note, our lineage analysis of human endocrine differentiation differs from a recently published study using single-cell sequencing and lineage reconstruction of fetal endocrine cells (177). Future studies at early developmental timepoints might resolve this issue by increasing the cell number of EPs captured, increasing the chances of detecting potentially rare and/or transient, pre-delta progenitor cells. In the future, applying state-of-the-art methods for lineage barcoding to human pancreas tissue *ex vivo* could represent an exciting approach for experimental validation of predictions generated *in silico.*

Identification of the effector transcripts or genes through which disease-associated variants influence risk is a critical first step towards biological inference and thus clinical translation. The context-specificity of gene regulation presents an additional challenge. Single-cell resolution of both gene expression and chromatin accessibility in human adult islets has recently demonstrated the importance of cell state, cell type, and the potential of co-accessibility analysis between promoters and cis-regulatory elements to identify effector genes at T1D (183)

and T2D (182,183,262) risk loci. Here, we have extended this previous work to include the influence of developmental stage by performing a comparative analysis of human adult vs. fetal endocrine cells to uncover enrichment of diabetes genetic risk loci, permitting the assessment of regions of DNA harboring T1D and T2D-risk alleles that are accessible during development and may therefore affect the expression of developmental genes. T1D SNPs showed lower fGWAS enrichment in either adult or fetal endocrine tissue compared to T2D; this result is not surprising, given that most signals for T1D are believed to be related to immune cells (186,235). Still, we observed developmental enrichment of T1D signals in the fetal accessible chromatin, including loci that are known to have development-specific functions (252), highlighting the strength of our comparative analysis. For T2D signals, as a positive control, our analysis identified fetal enrichment of risk loci of known regulators of endocrine differentiation, such as *NEUROG3*. Furthermore, we also detected enrichment of monogenic diabetes genes, which are known to influence beta cell development and contribute to diabetes when disrupted. In addition, we also observed fetal enrichment in genes with no known function in endocrine development, such as *LRFN2*.

One intriguing observation from our analysis is the potential for further understanding context specificity of gene regulation at the *PROX1* locus. There are two independent signals at this locus: the first has been fine mapped to a single variant (rs340874), and the second has two SNPs (rs79687284 and rs17712208) in the credible set (263). An evaluation of these variants on transcriptional activity in both human HepG2 hepatocytes and EndoC-βH1 beta cell models using *in vitro* reporter assays demonstrated effects in both liver and beta cells for rs340874. At the second signal, however, only one of the two variants (rs17712208) influenced activity in beta cells and neither in HepG2 cells (263). Our data now raise the intriguing possibility that the rs79687284 variant could alter activity earlier in development, thus expanding the complexity of the regulatory impact at this locus. These analyses therefore provide a framework for the

identification of development-specific disease risk loci and a rich opportunity for further study of their function in islet biology.

Despite tremendous progress in recent decades in devising methods to generate beta-like cells *in vitro* from hPSCs, these protocols still suffer from the generation of unwanted cell types. A previous study aimed to classify hPSC-derived endocrine cells by referring to adult islet cells, but was constrained by the absence of EPs in adult tissue (59). By performing computational comparison of endogenous *in vivo* vs. *in vitro* endocrine development, we observed that the EP cell types generated *in vitro* are similar to those present *in vivo*. That said, the generation of mis-differentiated cell types, such as the stem cell-derived enterochromaffin (sc_EC) population, demonstrates that there remains significant room for improvement of the differentiation protocol with respect to purity and efficiency. Given the similarities between pancreas and intestinal endocrine development, the generation of the enterochromaffin population is likely due to the mis-expression of key genes that then tips the balance towards an enteroendocrine fate. To generate more pure and functionally mature beta cell populations, future work will focus on modulating the expression, ideally in a temporally constrained fashion, of key genes that are currently aberrantly expressed. Additionally, our data have important implications for the generation of *in vitro*-derived islet cells, as our *in vivo* developmental roadmap can now be used for the refined production of non-beta endocrine cells, including human alpha (88,89) and delta cells.

Lastly, the results of *FEV* gene ablation demonstrate how the generation of a detailed *in vivo* roadmap of endocrine differentiation can successfully be combined with *in vitro* genome editing techniques to functionally interrogate important regulators of human endocrine development. We validated that *FEV* not only marks beta cell progenitors in the developing fetal pancreas, but itself also plays a functional role in human endocrine differentiation *in vitro*. Further investigation of *FEV* through the use of TF binding studies will provide insight as to how it regulates endocrine differentiation or function. Applying this approach to other genes of

interest is a promising approach for understanding additional, uncharacterized regulators of human endocrine differentiation.

In summary, we provide here a comprehensive, single-cell, multi-omic roadmap of human fetal pancreatic endocrine development that represents a critical step towards generating bona fide beta cells *in vitro* for therapeutic use.

3.4 Materials and Methods

Data and code availability

Scripts used in this study are available at GitHub:
https://github.com/sneddonucsf/2021-Developmental-single-cell-multi-omic
Raw single-cell sequencing data of human fetal pancreas samples can be found on dbGaP (accession number phs002693.v1.p1). Raw and processed data of the FEV WT and KO bulk RNA-sequencing can be found on GEO (GSE196230). All data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request.

Experimental subject details

Informed consent was obtained for all human tissue collection, and protocols were approved by the Human Research Protection Program Committee at UCSF. Human fetal dorsal pancreas tissue was obtained from post-mortem fetuses at 8 to 20 weeks post conception (w) through two sources: University of Washington Birth Defects Research Laboratory (BDRL) and Advanced Bioscience Resources, Inc. (ABR). Identifiers were maintained at the source only, and the investigators received only de-identified specimens. After isolation, tissue was shipped overnight (O/N) on ice in RPMI medium. A portion of tissue was fixed in 4% paraformaldehyde O/N at 4 °C, washed three times with 1 x phosphate-buffered saline (PBS), and cryopreserved in 30% sucrose solution at 4 °C for O/N in preparation for embedding in optimal cutting

temperature (OCT) compound. Sections measuring 10 μm in thickness were cut using a cryostat and stored at -80°C for immunofluorescence staining or *in situ* hybridization, as described below.

Adult human islets were isolated from cadaveric donor tissue by the UCSF Islet Production Core with permission from the UCSF ethical committee. Consented cadaver donor pancreata were provided by the nationally recognized organization UNOS via local organ procurement agencies. The identifiers were maintained only at the source, and the investigators received de-identified specimens.

Processing of pancreas tissue scRNA-Seq and snATAC-Seq

To isolate cells for single-cell RNA-Sequencing, human fetal pancreas tissue was minced with scalpels and transferred to dissociation buffer containing Liberase TM and 0.1 mg/mL Dnase I for 30-55 minutes at 37°C on a Thermomixer at 1000 rpm. Enzyme was quenched with 1X HBSS containing 5mM EDTA and 10% FBS. The resulting cell suspension was filtered through a 30 μm strainer. All tissues were subject to removal of red blood cells (RBCs) using immunomagnetic negative selection with the EasySep RBC Depletion kit (STEMCELL Technologies, 18170). 12 w samples were further subjected to EasySep™ Human EpCAM Positive Selection Kit II (STEMCELL Technologies, 18356) to positively select for epithelial cells. Tissues at 19 w and 20 w were subjected to EasySep™ Human CD45 Depletion Kit II (STEMCELL Technologies, 18259) to remove CD45+ immune cells. Cell viability was measured for all samples using a MoxiFlow (Orflo) to confirm greater than 90% viability.

To isolate nuclei for single-nuclei ATAC-Sequencing, 12 w human fetal pancreatic tissue was placed in a dissociation buffer containing Liberase TM and 0.1 mg/mL Dnase I at 37 °C. Dissociated cells were filtered through a 30 μm strainer and further enriched for EpCAM+ epithelial cells using the EasySep™ Human EpCAM Positive Selection Kit II. Nuclei from

EpCAM+ cells were isolated following 10x Genomics protocol CG000169, Rev D. In brief, EpCAM+ cells were resuspended in PBS + 0.04% BSA and centrifuged at 1000 rpm and 4°C for 5 min. Chilled Lysis Buffer was added to the cell pellet, which was then incubated on ice for 3.5 min. After lysis, chilled Wash Buffer was added, cells were centrifuged at 1200 rpm, and isolated nuclei were suspended in 1X Nuclei Buffer. After isolation, nuclei were manually counted with a hemocytometer; quality was assessed under a 63x bright field microscope to ensure that the periphery of isolated nuclei appeared smooth.

Single-cell capture and sequencing

For scRNA-Seq of human fetal tissue, we utilized the Chromium Single Cell 3' Reagent Version 3.1 Kit (10x Genomics). For non-enriched human fetal samples, we loaded 25,000 cells each onto two lanes of the 10x chip, resulting in a total of 50,000 cells loaded per sample. For enriched human fetal samples, including EpCAM+ cells from two 12 w samples, EpCAM+ cells and CD45- cells from one 19 w sample, and CD45- cells from one 20 w sample, 25,000 cells from each enrichment condition were loaded onto a single lane of the 10x chip. Gel Bead-In EMulsions (GEMs) were generated and subjected to reverse transcription for RNA barcoding before cleanup and cDNA amplification. Libraries were then prepared with the Chromium Single Cell 3' Reagent Version 3.1 Kit according to the manufacturer's instructions. Each resulting library was sequenced on the Novaseq 6000 platform (Illumina) with the following parameters: Read 1 – 28 cycles, Index 1 i7 – 8 cycles, Index 2 i5 – 0 cycles, Read 2 – 91 cycles.

For snATAC-Seq of 12w human fetal tissue, we utilized Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit v1.1 (10x Genomics). 7,166 nuclei were loaded onto one lane of a 10x chip. Transposition, GEM generation and barcoding, cleanup and library construction were performed according to the manufacturer's protocol. The library was then sequenced on the Novaseq 6000 platform (Illumina) with the following parameters: Read1- 50 cycles, Index1 – 8 cycles, Index 2 – 16 cycles, Read 2 – 49 cycles.

Human fetal single-cell RNA-Sequencing analysis

To assemble the transcriptomic profiles of individual cells, we utilized CellRanger versions v3.0-4.0 with default settings to demultiplex, aligned reads to the human genome (GRCh38, supplied by 10x Genomics), and quantified unique molecular identifiers (UMIs). The resulting gene-barcode matrices were then analyzed and aggregated with the R package Seurat v3.1.2 (157). Each sample was subjected to filtering to exclude cells expressing fewer than 200 genes and genes expressed in fewer than three cells. Technical replicates (two 10x lanes of the same biological sample) were merged using the MergeSeurat() function. High-quality cells were retained by filtering on the number of expressed genes and mitochondrial content (**Supplemental file 3.1**). Each sample was normalized with NormalizeData(), and variable genes were identified with the FindVariableFeatures() function using 2,000 genes and the "vst" selection method. Integration anchors were found across all samples with the FindIntegrationAnchors() with 30 principal components and 2,000 genes. The samples were then integrated using the IntegrateData() function. The data were then scaled with ScaleData() function and principal component analysis (PCA) was performed, with 30 principal components selected based on the ElbowPlot(). Dimensionality reduction and initial clustering was performed with the FindNeighbors(), FindClusters() and RunUMAP() functions using 30 principal components and a resolution parameter of 2.0. The resulting Louvain clusters were then manually annotated into "Broad Groups" of known biological cell types using canonical marker genes. Cluster 17 from the initial clustering was removed, as it had no distinguishable marker genes and expressed low levels of features.

Cell clustering with CellFindR

To further sub-cluster the Broad Groups, we applied a novel clustering package, CellFindR (https://github.com/kevyu27/CellFindR) (122), to each Big Group individually. Each

Broad Group was subsetted, PCA and UMAP were recalculated, and the top level resolution was found with the res() function. Iterative sub-clustering was performed on each top level cluster with the sub_clustering() function. Clusters that were deemed non-biological (i.e. COL3A1+/hormone+ doublets) were manually removed from the endocrine data set, including a sub-cluster of *INS/CELA3A*-high cells that displayed a low number of features and counts, likely representing empty droplets containing ambient RNAs known to contaminate scRNA-Seq datasets of the pancreas[137].

Single cell differential gene expression analysis

Marker genes were identified with Seurat's FindAllMarkers() function and visualized with FeaturePlot(), VlnPlot() and DoHeatmap() functions. Pairwise volcano plots were created by utilizing the FindMarkers() function and plotting the results as a volcano plot from the EnhancedVolcano package (https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html). Transcription factors were identified through comparison to AnimalTFDB3.0 database (264).

Cell-cell communication analysis

To infer cell-cell communication within our human fetal pancreas dataset, we utilized the R package CellChat (https://github.com/sqjin/CellChat). For each analysis, we used normalized counts and cell-type specific labeling as input. We then followed the 'Inference and analysis of cell-cell communication using CellChat' vignette.

Pseudotemporal ordering and lineage analysis of scRNA-Seq data

For trajectory and pseudotime analyses, we utilized the R package Slingshot (https://github.com/kstreet13/slingshot) (206) and the Python package CellRank (https://github.com/theislab/cellrank) (205). Seurat-based UMAP dimensional reduction and

CellFindR clustering were used as input for the merged human fetal endocrine analysis. Lineage reconstruction was performed with the slingshot() function, with the cEP population (Cluster 3.1.0) designated as the beginning of pseudotime. Averaged pseudotime values for each cell were calculated with the slingAvgPseudotime() function, and the Seurat object was converted to an AnnData object with the Seurat Disk () functions SaveH5Seurat() and Convert().

The resulting file was then read into Scanpy (https://github.com/scverse/scanpy) (265). Neighbor graphs were calculated with ScVelo's (https://github.com/theislab/scvelo) pp.moments() function with n_pcs = 9. Next, the directed transition probability was calculated using the Slingshot pseudotime values with CellRank's PseudotimeKernel function and the transition matrix generated with compute_transition_matrix(), with a "soft" threshold scheme. The transition matrix was then projected onto UMAP reduction with compute.projection() and visualized with velocity_embedding_stream(). Coarse-grained transition probabilities and macrostates were calculated with GPCCA(), compute_schur() and compute_macrostates(), with n_states = 5 and n_cells = 30. Terminal states were calculated with set_terminal_states_from_macrostates(). Absorption probabilities were calculated with compute_absorption_probabilities() with preconditioner = "ilu" and tol = 1e^-10. Finally, absorption probabilities were visualized with cluster_fates() and plot_absorption_probabilities().

Pathway analysis

Pathway analysis and calculation of associated p values were performed using the ConsensusPathDB overrepresentation analysis for pathway based sets category (http://cpdb.molgen.mpg.de)**.**

Initial snATAC-Seq analysis

To assemble the chromatin profiles of individual cells, we utilized Cell Ranger ATAC v1.1 with default settings to demultiplex, align reads to the human genome (using the pre-built GRCh38 human genome supplied by 10x Genomics), and generate single-cell accessibility counts. The resulting files were then analyzed using the R package ArchR v0.9.5 (121). Arrow files were created from the fragment file from the CellRanger output with the createArrowFiles() function and an ArchR project created with the ArchRProject() function. Doublets were filtered out using the addDoubletScores() and filterDoublets() functions, resulting in a dataset comprising 6,010 nuclei. Dimensional reduction was calculated with the addIterativeLSI() function using the following settings: iterations = 2, resolution = 0.5, sampleCells = 2500, n.start = 10, varFeatures = 25000, dimsToUse = 1:30. Clustering was performed with addClusters() with method = Seurat and resolution = 0.1. UMAP was calculated with addUMAP() and clusters were annotated as “Broad Groups” based on marker gene expression (see above in Initial single-cell RNA-Sequencing analysis).

The endocrine cluster was then subsetted and dimensional reduction was calculated with the addIterativeLSI() function with the following settings: iterations = 2, resolution = 0.3, sampleCells = 1500, n.start = 10, varFeatures = 25000, dimsToUse = 2:10. UMAP was calculated with addUMAP() with default settings. Unconstrained integration with scRNA-Seq data were then performed using the final scRNA-Seq endocrine dataset (Fig. 2a) as input with the addIntegrationMatrix() function, transferring cluster labels and a pseudo-RNA-Seq profile. To maintain the robustness of the analysis, the endocrine progenitor (EP) clusters were collapsed into a single EP cluster, resulting in a final dataset comprising 5 clusters and 1,754 cells.

snATAC-Seq peak calling

To call peaks in the snATAC-Seq dataset, group coverages were added to the ArchR project with addGroupCoverages() and peaks called with Macs2 with the

addReproduciblePeakSet() function with a cutoff of 0.1. The peak set was then added to the project with addPeakMatrix(). Marker peaks were identified with the getMarkerFeatures() function using the PeakMatrix and peaks with a false discovery rate (FDR) less than or equal to 0.1 and log2FC of 0.5 or greater were visualized with markerHeatmap().

Motif analysis

To perform motif enrichment in regions of open chromatin, we first added motif annotations with the addMotifAnnotations() function, using “cisbp” as the motif set. To visualize motif deviation at single-cell resolution, we utilized ChromVAR (124) within the ArchR package. Background peaks were added with addBgdPeaks() and per-cell deviations calculated with addDeviationsMatrix(). Deviation scores were then visualized with the plotEmbedding() function.

To calculate “positive” transcription factors (TFs showing both motif enrichment and RNA expression), deviant motifs were first accessed with seGroupMotif(). The motif deviation scores and RNA expression profiles were then correlated with the correlateMatrices() function and filtered based on a threshold of correlation greater than 0.5, adjusted p-value less than 0.01 and a maximum inter-cluster difference in deviation z-score that is in the top quartile.

Pseudotemporal ordering analysis of snATAC-Seq data

To order cells in pseudotime, we first increased the clustering resolution, as ArchR does not allow for ordering of trajectories with fewer than 3 clusters. Clustering was performed with the addClusters() function with a resolution setting of 1.0. Alpha, Beta, Delta and Epsilon trajectories were then manually chosen by clusters; the cells along those trajectories were then ordered by pseudotime with the addTrajectory() function. The trajectories were then visualized with plotTrajectory().

To assess positive transcription factors across pseudotime, we accessed motif enrichment and RNA expression across pseudotime with getTrajectory(), using the motif matrix

and gene integration matrix, respectively. We then correlated these trajectories with cutoffs of correlation = 0.5 and variance quartile cutoff = 0.8 for both the motif matrix and gene integration matrix.

Enrichment in GWAS signals

Datasets for adult islet samples were accessed from GEO (GSE160472)and analyzed together with our snATAC-Seq data from the 12 w fetal pancreas sample. Combinatorial barcoding (CB) data were processed with the ENCODE ATAC-seq pipeline (v1.9.3) by aligning to the genome reference GRCh38. Cell barcoding information contained in the read names was added as CB tags in the bam files with a customized script. Only mapped reads with MAPQ score > 30 were retained. Cell type annotation of adult islet cells was applied from the metadata file provided on GEO. ArchR (v1.0.1) was used as a platform for the downstream analysis, including clustering, peak calling (MACS2, v2.2.7.1) and Wilcoxon testing for differential peaks.

To assess the overall enrichment of T1D or T2D signals in peaks called in each cell type of either fetal or adult pancreas, fGWAS (v0.3.6) was performed with a window size of 5000 bp on the BMI unadjusted summary statistics of T2D or summary statistics of T1D. To further prioritize the 380 loci of T2D and 157 loci of T1D in the context of differential peaks between adult and fetal beta cells (or between adult hormone+ and fetal EPs) (padj < 0.05 and abs(fold change) > 2), GoShifter (v0.3) was used to calculate the enrichment with 10000 permutation. Top 20 loci were shown in the forest plot with the number of overlapping SNPs and all SNPs in the credible sets for each loci shown on the right (**Figure 5F**, **Figure S5E-F**).

Analysis of scRNA-Seq data from *in vitro*-differentiated stem cell-derived cells

Processed counts and cell metadata were downloaded from GEO accession number GSE103412 for Stage 5 (EP stage) and Stage 6 (beta cell stage) cells (23). Counts and meta

data were read into R and Seurat objects were created with associated metadata. Data were normalized with NormalizeData() and variable features found with FindVariableFeatures(). Data were scaled and PCAs calculated with ScaleData() and RunPCA(), then data were clustered with FindNeighbors(), FindClusters() and data reduced with RunUMAP().

Endocrine cell types were classified with the R package scPred (248). First, normalized scRNA Seq data of human intestinal epithelium were downloaded from the gut cell atlas (https://www.gutcellatlas.org/#x) and analyzed with Seurat. The data were subsetted for enterochromaffin cells (both TAC1+ and NPW+ clusters), NEUROG3+ progenitors, and enteroendocrine cells based on the original authors' annotations. The cells were then subsetted to only include samples derived from fetal tissue. The resulting dataset was then normalized and scaled, and variable features and PCA were calculated. This intestinal dataset was merged with our fetal pancreatic endocrine cells using the Seurat function merge(). Next, the features used for classification were calculated using the merged dataset with the scPred function getFeatureSpace(). The classifier was trained with trainModel(), with "mda" as the model, then applied to the Stage 5 and Stage 6 datasets with ScPredict().

For *in vitro* and *in vivo* differential gene expression analysis, datasets were combined with the MergeSeurat() function and differential gene expression calculated as stated above.

Multiplexed *in situ* hybridization and immunofluorescence

Human fetal pancreas tissue was fixed in 4% paraformaldehyde (PFA) overnight (O/N) at 4 °C. hESC-derived clusters were fixed with 4% PFA for 20 min at room temperature (RT). Post-fixed tissues and clusters were washed three times with 1x phosphate-buffered saline (PBS), cryopreserved in 30% sucrose solution at 4 °C O/N, and embedded in optimal cutting temperature (OCT) compound. Sections measuring 10 μm in thickness were cut using a cryostat and stored at -80 °C for *in situ* hybridization and immunofluorescence, as described below.

After removal from -80 °C storage and incubation at RT for 30 minutes, cryosections were washed with 1 x PBS to remove OCT, and sequentially treated with hydrogen peroxide and proteinase III. Tissues were then hybridized with probe mixes for 2 hours at 40 °C. Inventoried or customized probes were purchased from Advanced Cell Diagnostics, Inc. Probes against *SUSD2* (42673), *PRPH* (410231-C2), *FEV* (471421-C3), *NEUROG3* (050798-C4), *LMX1B* (582661), *MEG3* (400821), *ARX* (486711-C2), *SFRP1* (428381-C4), *SFRP2* (476341-C3), *CCL21* (474371-C2), *GJA5* (471431-C2), *PLVAP* (437461), and *ACKR1* (515131) were used according to the manufacturer's instructions for the RNAScope multiplex fluorescent detection V2 kit (Advanced Cell Diagnostics, Inc, 323110). To validate probe specificity, negative control probe (DapB) and positive control probe (*POLR2A*/*PPIB*/*UBC*) were included in each experiment. Hybridization signals were amplified via sequential hybridization of amplifier AMP1, AMP2, and AMP3 and fluorophores Opal 570 (1:1500, PerkinElmer, FP1488001KT), Opal 650 (1:1500, PerkinElmer, FP1496001KT), Opal 690 (1:1500, PerkinElmer, FP1497001KT). Following signal amplification of the target probes, sections were either stained with DAPI and mounted for imaging or continued with standard immunofluorescence (IF) procedure. For IF, sections were incubated in 1 x blocking buffer (0.1% PBST containing 5% normal donkey serum) for 1 hr at RT then stained O/N at 4° C using the following primary antibodies diluted in blocking buffer: Chromogranin A (1:200, Abcam, ab15160), Glucagon (1:200, Cell Signaling, 2760S), Insulin (1:200, DAKO, A0564), Somatostatin (1:200, Santa Cruz Biotechnology, sc-7819), SMA (1:200, Abcam, ab21027), PECAM (1:100, Dako, M0823). The next day, sections were washed in 1X PBS three times and incubated with species-specific Alexa Fluor 488-, 555-, 594-, or 647- conjugated secondary antibodies (1:500) and DAPI in the blocking buffer for 1 hour at RT. Stained slides were mounted with ProLong Gold Antifade Mountant (ThermoFisher SCIENTIFIC, P36930) and stored at 4 °C prior to imaging.

To assess polyhormonality within human fetal pancreas samples, human fetal tissue was fixed, sectioned, washed, blocked and stained as described above with the following

antibodies: E-Cadherin (1:200, BD BioSciences, 610182), SST (1:200, Santa Cruz, sc-7189), C-peptide (1:200, Developmental Studies Hybridoma Bank (DSHB), GN-ID4), and Glucagon (1:200, Cell Signaling, 2760S).

Image acquisition

Optical sectioning images were acquired with a Leica confocal laser scanning SP8 microscope equipped with white light sources. Z-sections were captured for each imaging area with 10 steps x 1 mm thickness.

Genetic engineering to generate the FEV-KO hESC line

The HUES8 human embryonic stem cell (hESC) line was grown on Matrigel (Corning, 354230)-coated tissue culture plates in mTeSR1 (STEMCELL Technologies). Media was changed to mTeSR1 + 10 μM Rock inhibitor Y-27632 for 2 hours prior to nucleofection. Cells were dissociated into a single-cell suspension using TrypLE Express (Gibco, 11588846). A FEV-KO gRNA (5'-CTGATCAACATGTACCTGCC-3') was designed using Benchling software and purchased from Dharmacon. To carry out the nucleofection, 160 μM tracrRNA and 160 μM FEV-KO gRNA were mixed together to make the RNA-complex and incubated for 30 min. in a 37 °C cell culture incubator. Purified Cas9-NLS protein (QB3 UC Berkeley MacroLab) was added to the RNA-complex, gently mixed to make the RNP (ribonucleoprotein), and incubated at 37 °C. After 15 min., dissociated cells were resuspended in P3 buffer (Lonza, V4XP-3032). Cell suspension and RNP were mixed and inserted into the Lonza 4D-Nucelofector (Lonza, AAF-1002B) and nucleofected in the P3 buffer. Nucleofected cells were transferred to mTeSR1 supplemented with Rock inhibitor, then seeded onto Matrigel-coated T75 tissue culture flasks (ThermoFisher Scientific, 159910).

Validation of FEV-KO and -WT hESC lines

Cells were sorted with FACS and clonally plated onto Matrigel-coated 96 well plates and grown in mTeSR1 supplemented with Rock inhibitor Y-27632. Clonal colonies were hand-picked under a colony-picking microscope under sterile conditions and each colony was transferred into one well of a 96-well plate, then successively passed onto larger plate formats. To determine the efficiency of genomic editing of each colony, genomic DNA from each colony was harvested with QuickExtract DNA Extraction (Lucigen, QE09050) and then used for PCR amplification. The following forward and reverse primers targeting the FEV-KO editing site were used to produce a 491-bp amplicon: 5'-CCGTCTTCTCCTCCTTGTCACC3' and 5'-CTCGGCCACAGAGTACTCCAC-3'. PCR polymerase capable of handling GC-rich amplicons was used (PrimeSTAR GXL Premix, Clontech). The resulting DNA amplicon, along with a wildtype DNA amplicon, were sent to Quintara Biosciences for Sanger sequencing. The chromatographs of each sequencing run were used for TIDE (Tracking of Indels by Decomposition) analysis (https://tide.deskgen.com) and the cutting efficiency of hESCs nucleofected with FEV-KO gRNA was then determined. The FEV-KO hESC clonal line used here had a 1-bp deletion in one allele and a 1-bp insertion in the second allele, leading to a homozygous mutation in the *FEV* locus. The FEV-WT hESC clonal line had no detectable deletion in the *FEV* locus and was used as a control.

Human embryonic stem cell culture and differentiation to the beta cell lineage

FEV-KO and -WT HUES8 hESC lines were expanded and maintained in parallel in planar culture in mTeSR1 (STEMCELL Technologies) media. Once enough cells were obtained, the planar cells were then adapted to suspension culture and maintained as clusters in suspension in mTeSR1 in 500 mL spinner flasks (Corning, VWR) on a magnetic stir plate (Dura-Mag) within a 37 °C incubator at 5% CO2, 100% humidity, and a rotation rate of 60 rpm. Cells were routinely screened for mycoplasma contamination using the MycoProbe Mycoplasma Detection Kit (R&D Systems), according to the manufacturer's instructions.

Beta-like cells were generated as previously described (22). In brief, single hESCs were seeded into a spinner flask at a density of 1e6 cells/mL in mTeSR1 media containing 10 μM Rock inhibitor Y-27632 (STEMCELL Technologies) to allow formation of clusters. Differentiation was initiated 72 h later and was achieved in a step-wise fashion using the following growth factors and/or small molecules: Stage 1 (Day 1-3) medium : 500 mL MCDB 131 (Corning, 15-100-CV) + 0.22 g glucose (MilliporeSigma, G7528) + 1.23 g sodium bicarbonate (MilliporeSigma, S5761) + 10 g fatty-acid free bovine serum albumin (FAF-BSA) (Lampire Biological Laboratories, 7500812) + 10 μL Insulin-Transferrin-Selenium-Ethanolamine (ITS-X) (Invitrogen, 51500056) + 5 mL GlutaMAX (Invitrogen, 35050079) + 5 mL Penicillin-Streptomycin (P/S) solution (Corning, 30-002-CI). Stage 2 (Day 4-6) medium: 500 mL MCDB 131 + 0.22 g glucose + 0.615 g sodium bicarbonate + 10 g FAF-BSA + 10 μL ITS-X + 5 mL GlutaMAX + 0.022 g vitamin C (MilliporeSigma, A4544) + 5 mL P/S. Stage 3 (Day 7-8) / Stage 4 (Day 9-13) medium: 500 mL MCDB 131 + 0.22 g glucose + 0.615 g sodium bicarbonate + 10 g FAF-BSA + 2.5 mL ITS-X + 5 mL GlutaMAX + 0.022 g vitamin C + 5 mL P/S. Stage 5 (Day 14-20) medium:500 mL MCDB 131 + 1.8 g glucose + 0.877 g sodium bicarbonate + 10 g FAF-BSA + 2.5 mL ITS-X + 5 mL GlutaMAX + 0.022 mg vitamin C + 5 mL P/S + 5 mg heparin (MilliporeSigma, H3149). Stage 6 (Day 21-31) medium: 500 mL CMRL 1066 supplemented (Corning, 99-603-CV) + 10% Fetal Bovine Serum (FBS) (Corning, MT-35-011-CV) + 5 mL P/S.

Flow cytometric analysis of stem cell-derived cells

Stem cell-derived clusters at various stages of differentiation were washed in PBS (Corning, 21-040-CV) and dissociated with Accumax™ (Innovative Cell Technologies Inc, AM105) at 37 °C for the following times at each stage: 5 minutes (pluripotency), 5-7 minutes

(End Stage 1), 7-9 minutes (End Stage 3), 11-13 minutes (End Stage 5), 12-15 minutes (Stage 6). Cells were then fixed with 4% PFA for 10 minutes at RT and spun down at 1,200 rpm for 5 minutes and resuspended in PBS. Cells were filtered through a 37 μm cell strainer (Corning, 352235) on ice, washed with 1X Permeabilization Buffer (00-8333-56, Invitrogen™) and spun down at 1200 rpm for 5 minutes at 4° C. Cells were stained with flow antibodies, then diluted in CAS Blocking Buffer (Invitrogen, 8120) containing 0.2% Triton-X, 5% NDS, and 1% bovine serum albumin (BSA) O/N at 4° C. The following morning, cells were washed with Permeabilization Buffer, spun down at 1,500 rpm for 5 minutes at 4° C, resuspended in FACS buffer (PBS, 1% FBS and 2mM EDTA), and analyzed on an LSR-II flow cytometer (BD Biosciences). At the end of each differentiation stage, FEV-KO or -WT hESC-derived cells were subjected to flow cytometric analyses, as described above. Pluripotency was assessed by staining for OCT3/4 and NANOG. Efficiency of generating definitive endoderm at the end of Stage 1 (ES1) was determined by measuring the percentage of cells double positive for SOX17 and FOXA2. Efficiency of generating early pancreatic progenitors at the end of Stage 3 (ES3) was determined by measuring the percentage of cells positive for PDX1. Efficiency of generating endocrine progenitors at the end of Stage 5 (ES5) was determined by measuring the percentage of cells double positive for PDX1 and NKX6.1. Efficiency of generating beta-like cells at the end of Stage 6 (ES6) was determined by measuring the percentage of cells double positive for C-PEPTIDE and NKX6.1. Data were analyzed with FlowJo software (Tree Star Inc.).

Quantitative RT-PCR

hESCs were collected from various stages of directed differentiation and subjected to RNA extraction using the RNeasy Mini Kit (QIAGEN 74106). Reverse transcription was performed with the Clontech RT- PCR kit. RT-PCR was run on a 7900HT Fast Real-Time PCR instrument (Applied Biosystems) with Taqman probes for *FEV* (assay ID: Hs00232733_m1) and

GAPDH (assay ID: Hs02758991_g1) in triplicate. Expression of *FEV* was normalized to *GAPDH*.

Bulk-RNA sequencing

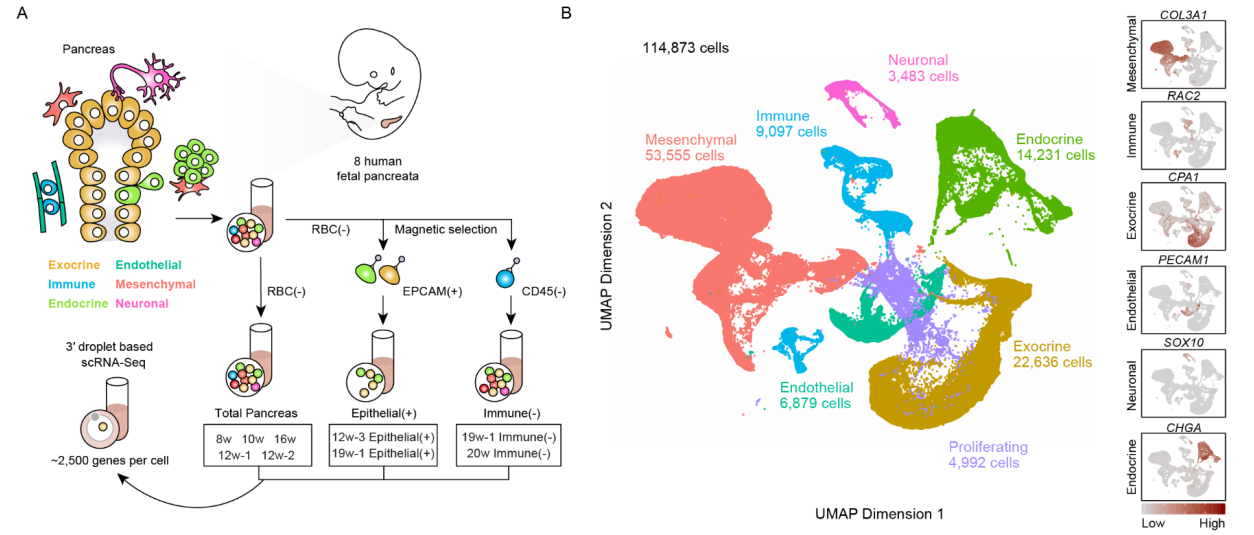
FEV-KO and -WT clusters were collected from four independent batches of stem cell differentiation at Stage 6 day 10 (S6D10). 2e6 cells were lysed in 350 ul RLT buffer (QIAGEN, 79216) and stored at -80 °C before RNA extraction. RNA was purified with RNeasy Mini Kit (QIAGEN, 74106). Samples with a RNA Integrity Number (RIN) greater than 9 were advanced to library construction using poly-A enrichment. Sequencing was conducted on a NovaSeq 6000 PE150 platform with the following parameters: Read 1 - 150 cycles, Index 1 i7 - 8 cycles, Index 2 i5 - 8 cycles, Read 2 - 150 cycles. The resulting files were mapped to the reference genome (GRCh38) with STAR (v2.6.1d) and counts were generated with FeatureCounts (v1.26.0-p3). Differentially expressed genes were calculated with the DESeq2 (v1.26.0) workflow (266).

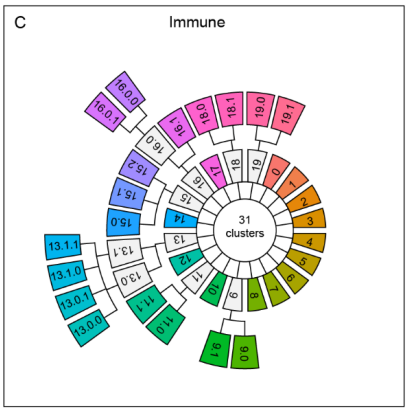
Quantification and statistical analysis

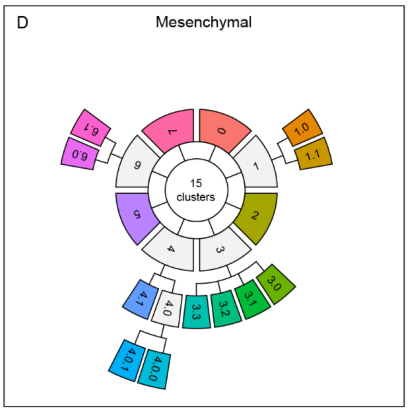
To analyze the population dynamics *in vivo* of each novel endocrine progenitor population over developmental time, nine samples of human fetal tissue from 8 w (n = 3), 12 w (n = 3), and 18 w (n = 3) were stained using multiplexed *in situ* hybridization and immunofluorescence. For each biological sample, images from five areas were taken at random and processed with the maximum intensity z-projection function with the ImageJ software package. Adjustments to brightness and contrast were applied equally across images in a series. The number of cells corresponding to each cell state was manually counted in each biological sample from the five image areas using the Image J plug-in Cell Counter. The proportion of each cell state present was then calculated using the sum of cells corresponding to all cell states as the denominator, and cells that scored positive for a given cell state as the numerator. Data were presented as Mean ± SEM (n = 3). Graphs were generated in GraphPad

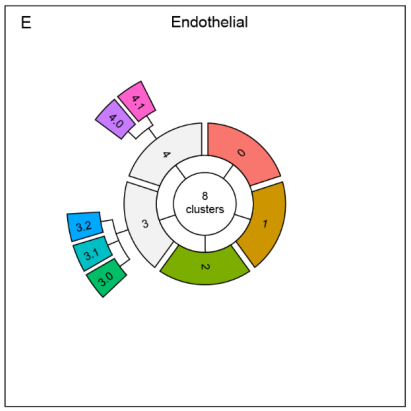
software (Prism 8). When assessing the proportional changes of FEV+NEUROG3+ progenitor cells over developmental time, an unpaired t-test was used to determine the statistical significance of the difference between the ratio of FEV+NEUROG3+ progenitor cells at 8 and 12 wpc, as well as the ratio difference between 12 and 18 w.

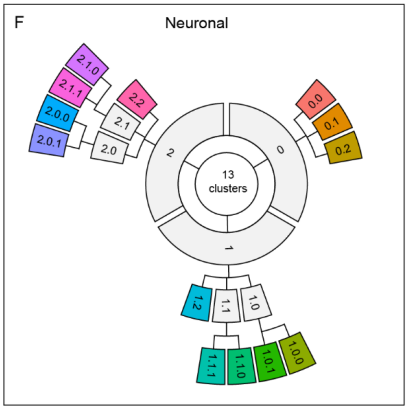
To assess differences between stem cell-derived cells from FEV-KO vs. -WT hESCs, data were quantified from flow cytometric analyses. At the early beta-like cell stage (Stage 6, day 4), the proportion of C-PEP+/NKX6-1+ double positive cells in FEV-KO or -WT cells was analyzed from three independent batches of differentiation. Statistical significance of difference between the two groups was determined using the paired t-test in GraphPad software (Prism 8). Data were presented as Mean ± SEM. To quantify the proportional changes of hormone-producing cells upon *FEV* knockout, immunofluorescence images of FEV-KO and -WT clusters from two independent batches of differentiation were manually counted using ImageJ software, and differences in cell proportion between WT and FEV-KO were assessed with an unpaired t-test in GraphPad software (Prism 8).

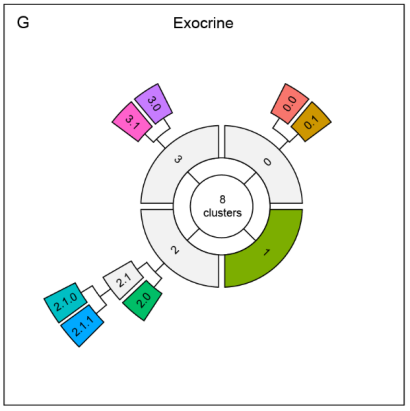

A
Pancreas
8 human
fetal pancreata
RBC(-)
Magnetic selection
RBC(-)
EPCAM(+)
CD45(-)
Exocrine
Endothelial
Immune
Mesenchymal
Endocrine
Neuronal
3' droplet based
scRNA-Seq
~2,500 genes per cell
Total Pancreas
8w 10w 16w
12w-1 12w-2
Epithelial(+)
12w-3 Epithelial(+)
19w-1 Epithelial(+)
Immune(-)
19w-1 Immune(-)
20w Immune(-)
B
114,873 cells
Neuronal
3,483 cells
Immune
9,097 cells
Mesenchymal
53,555 cells
Endocrine
14,231 cells
Exocrine
22,636 cells
Endothelial
6,879 cells
Proliferating
4,992 cells
UMAP Dimension 1
UMAP Dimension 2
COL3A1
RAC2
CPA1
PECAM1
SOX10
CHGA
Mesenchymal
Immune
Exocrine
Endothelial
Neuronal
Endocrine
Low
High


C
Immune
31
clusters


D
Mesenchymal
15
clusters


E
Endothelial
8
clusters


F
Neuronal
13
clusters


G
Exocrine
8
clusters

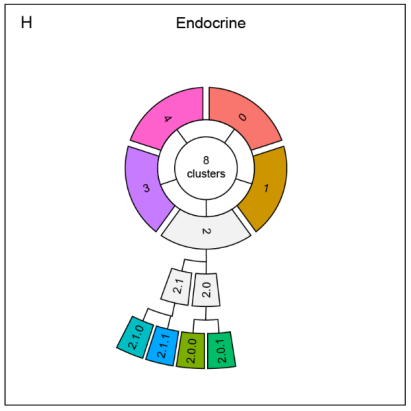

H
Endocrine
8
clusters

Figure 3.1: Large-scale single-cell RNA-Sequencing identifies striking cellular heterogeneity within the human fetal pancreas.

(**A**) Overview of experimental approach. Eight samples of human fetal pancreas tissue ranging from 8 to 20 weeks post-conception (w) were dissociated and subjected to red blood cell lysis ("RBC(-)") to deplete erythrocytes. Resulting single-cell suspensions were then used directly ("Total Pancreas"), or subjected to either magnetic bead-based enrichment for EpCAM+ cells ("Epithelial(+)"), or depletion of CD45+ cells ("Immune(-)"), followed by single-cell RNA-Sequencing (scRNA-Seq). The 19 w-1 Epithelial(+) and 19 w-1 Immune(-) cells were from the same tissue sample, and the 12w-1 Total Pancreas, 12w-2 Total Pancreas, and 12w-3 Epithelial(+) samples were three independent biological replicates. (**B**) UMAP visualization of the merged scRNA-Seq dataset from all nine conditions, derived from eight biological specimens at six developmental timepoints. Each cell is color-coded according to the Broad Group to which it belongs. Expression of marker genes *COL3A1, RAC2, CPA1, PECAM1, SOX10,* and *CHGA* are displayed in feature plots to the right. (**C**)-(**H**) Iterative clustering performed using the CellFindR algorithm revealed three layers of heterogeneity within each Broad Group, with Tier 1, 2, and 3 populations arranged in the inner, middle, and outer circles, respectively. Clustering hierarchies depict cellular populations within the Broad Groups of (**C**) immune; (**D**) mesenchymal; (**E**) endothelial; (**F**) neuronal; (**G**) exocrine; and (**H**) endocrine cells.

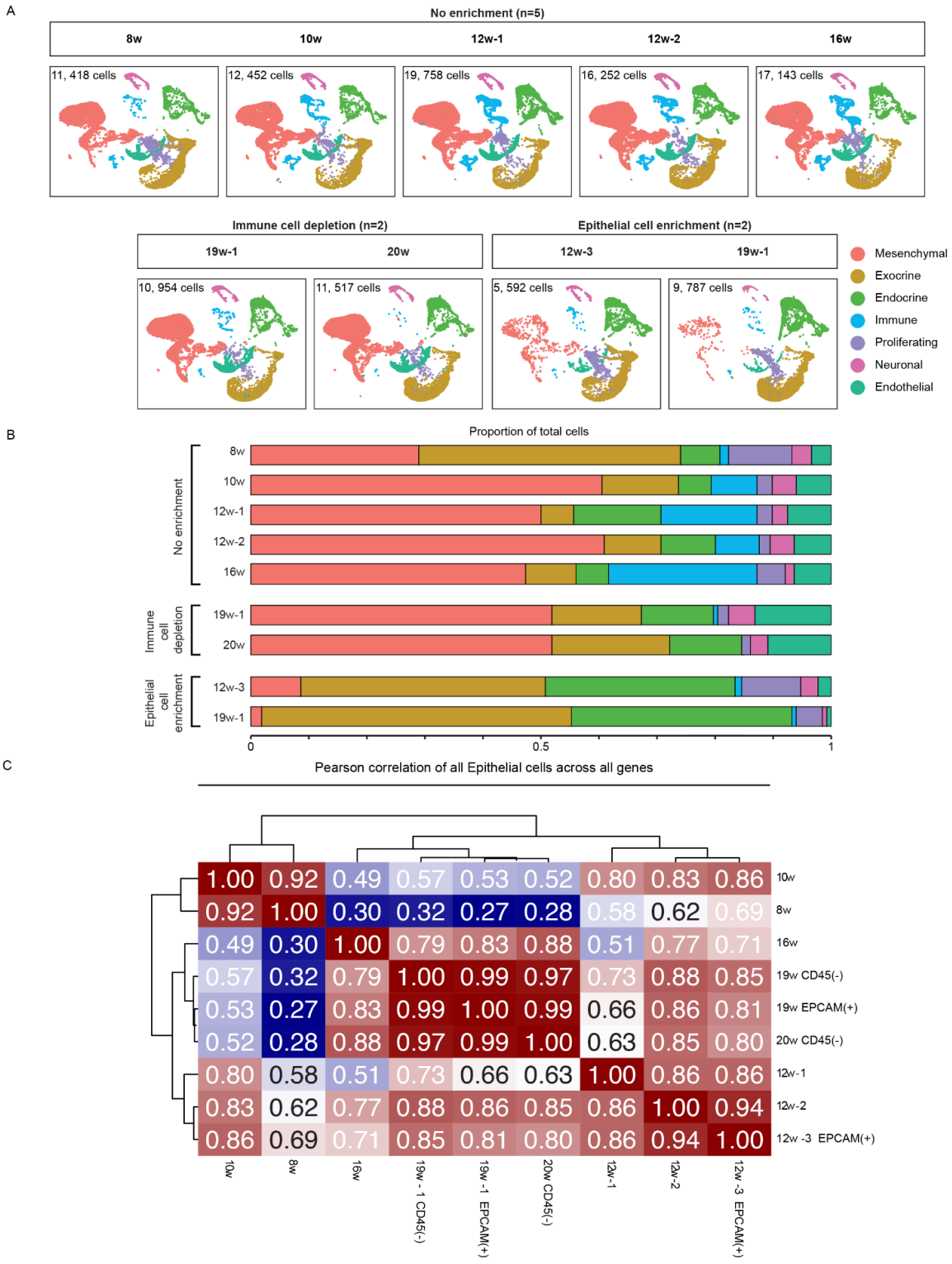

A
No enrichment (n=5)
8w
10w
12w-1
12w-2
16w
11, 418 cells
12, 452 cells
19, 758 cells
16, 252 cells
17, 143 cells
Immune cell depletion (n=2)
19w-1
20w
Epithelial cell enrichment (n=2)
12w-3
19w-1
10, 954 cells
11, 517 cells
5, 592 cells
9, 787 cells
Mesenchymal
Exocrine
Endocrine
Immune
Proliferating
Neuronal
Endothelial
B
Proportion of total cells
No enrichment
Immune cell depletion
Epithelial cell enrichment
0
0.5
1
C
Pearson correlation of all Epithelial cells across all genes
1.00 0.92 0.49 0.57 0.53 0.52 0.80 0.83 0.86
0.92 1.00 0.30 0.32 0.27 0.28 0.58 0.62 0.69
0.49 0.30 1.00 0.79 0.83 0.88 0.51 0.77 0.71
0.57 0.32 0.79 1.00 0.99 0.97 0.73 0.88 0.85
0.53 0.27 0.83 0.99 1.00 0.99 0.66 0.86 0.81
0.52 0.28 0.88 0.97 0.99 1.00 0.63 0.85 0.80
0.80 0.58 0.51 0.73 0.66 0.63 1.00 0.86 0.86
0.83 0.62 0.77 0.88 0.86 0.85 0.86 1.00 0.94
0.86 0.69 0.71 0.85 0.81 0.80 0.86 0.94 1.00
10w
8w
16w
19w CD45(-)
19w EPCAM(+)
20w CD45(-)
12w-1
12w-2
12w -3 EPCAM(+)

Figure 3.2: CellFindR clustering and proportional representation of Broad Groups in the human fetal pancreas, parsed by individual sample.

(**A**) Split UMAP visualization shows contribution of each sample to the overall merged dataset, with cells colored according to their Broad Group identity. The total number of cells per sample contributing to the merged data set is labeled on each UMAP. (**B**) Bar graph depicting the proportional representation of each Broad Group in each of the individual scRNA-Seq samples. (**C**) Hierarchical clustering of samples was performed based on gene expression of all epithelial cells. Heatmap depicting Pearson correlation between all Epithelial cell populations (Exocrine and Endocrine Broad Groups) across conditions.

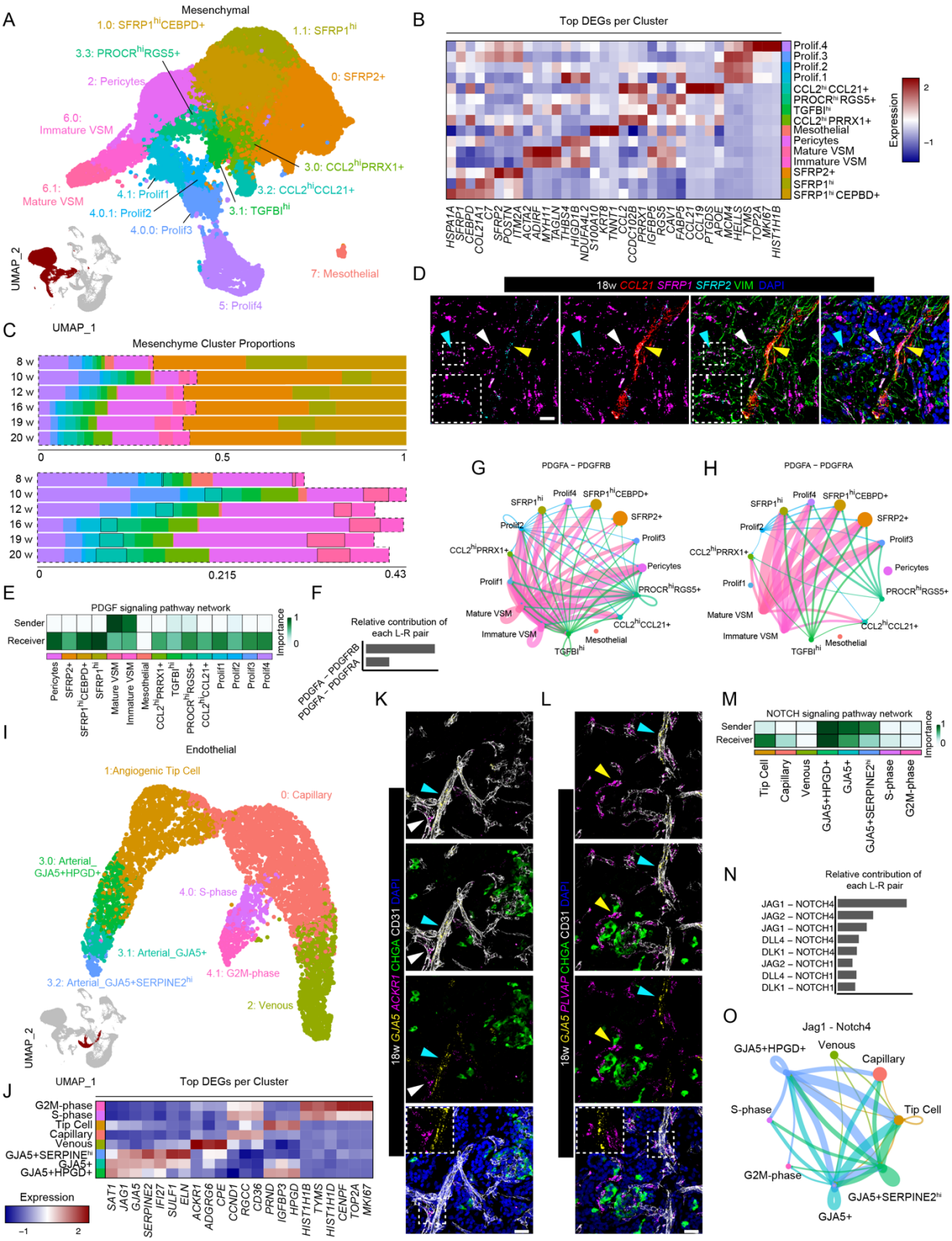

A
Mesenchymal
1.0: SFRP1hiCEBPD+
1.1: SFRP1hi
3.3: PROCRhiRGS5+
0: SFRP2+
2: Pericytes
6.0: Immature VSM
3.0: CCL2hiPRRX1+
6.1: Mature VSM
4.1: Prolif1
3.2: CCL2hiCCL21+
4.0.1: Prolif2
3.1: TGFBIhi
4.0.0: Prolif3
UMAP_2
7: Mesothelial
5: Prolif4
UMAP_1
B
Top DEGs per Cluster
Prolif.4
Prolif.3
Prolif.2
Prolif.1
CCL2hi CCL21+
PROCRhi RGS5+
TGFBIhi
CCL2hi PRRX1+
Mesothelial
Pericytes
Mature VSM
Immature VSM
SFRP2+
SFRP1hi
SFRP1hi CEPBD+
Expression
HSPA1A SFRP1 CEBPD COL21A1 C7 SFRP2 POSTN ITM2A ADIRF MYH11 TAGLN THBS4 HIGD1B NDUFA4L2 S100A10 KRT8 TNC CCL2 CCDC102B PRRX1 IGFBP5 RGS5 CAV1 FABP5 CCL21 CCL19 PTGDS APOE MCM4 HELLS TYMS TOP2A MKI67 HIST1H1B
C
Mesenchyme Cluster Proportions
8 w
10 w
12 w
16 w
19 w
20 w
0
0.5
1
0.215
0.43
D
18w CCL21 SFRP1 SFRP2 VIM DAPI
E
PDGF signaling pathway network
Sender
Receiver
Importance
Pericytes SFRP2+ SFRP1hiCEBPD+ SFRP1hi Mature VSM Immature VSM Mesothelial CCL2hiPRRX1+ TGFBIhi PROCRhiRGS5+ CCL2hiCCL21+ Prolif1 Prolif2 Prolif3 Prolif4
F
Relative contribution of each L-R pair
PDGFA – PDGFRB
PDGFA – PDGFRA
G
PDGFA – PDGFRB
H
PDGFA – PDGFRA
I
Endothelial
1:Angiogenic Tip Cell
0: Capillary
3.0: Arterial_ GJA5+HPGD+
4.0: S-phase
3.1: Arterial_GJA5+
4.1: G2M-phase
3.2: Arterial_GJA5+SERPINE2hi
2: Venous
UMAP_2
J
UMAP_1
Top DEGs per Cluster
G2M-phase
S-phase
Tip Cell
Capillary
Venous
GJA5+SERPINEhi
GJA5+
GJA5+HPGD+
Expression
SAT1 JAG1 GJA5 SERPINE2 IFI27 SULF1 ELN ACKR1 ADGRG6 CPE CCND1 RGCC CD36 PRND IGFBP3 HPGD HIST1H1B TYMS HIST1H1D CENPF TOP2A MKI67
K
18w GJA5 ACKR1 CHGA CD31 DAPI
L
18w GJA5 PLVAP CHGA CD31 DAPI
M
NOTCH signaling pathway network
Sender
Receiver
Importance
Tip Cell Capillary Venous GJA5+HPGD+ GJA5+ GJA5+SERPINE2hi S-phase G2M-phase
N
Relative contribution of each L-R pair
JAG1 – NOTCH4
JAG2 – NOTCH4
JAG1 – NOTCH1
DLL4 – NOTCH4
DLK1 – NOTCH4
JAG2 – NOTCH1
DLL4 – NOTCH1
DLK1 – NOTCH1
O
Jag1 - Notch4
GJA5+HPGD+
Venous
Capillary
S-phase
Tip Cell
G2M-phase
GJA5+SERPINE2hi
GJA5+

Figure 3.3: Identification of cell heterogeneity and cell-cell communication within the mesenchymal and endothelial lineages of the fetal human pancreas.

(**A**) UMAP visualization of the cell populations comprising the human fetal pancreatic mesenchyme. (**B**) Heatmap depicting the top differentially expressed genes per mesenchymal cell type (**C**) Bar graph depicting the proportional representation of endothelial cell types across developmental time. (**D**) 18 w human fetal pancreas tissue stained with an antibody against Vimentin (VIM; green) and *in situ* hybridization (ISH) probes against *SFRP1* (cyan), *SFRP2* (magenta), and *CCL21* (red) mRNA, counterstained with DAPI (blue) to detect nuclei. Aqua, White, and Yellow arrowheads mark presumptive SFRP1^{hi}/SFRP1^{hi}CEBPD+ cells (SFRP1+VIM+), SFRP2+ cells (SFRP2+VIM+) and CCL2^{hi}CCL21+ cells (CCL21+VIM+), respectively. Dashed line represents magnified region in inset. (**E**) Heatmap depicting predicted activity of the PDGF signaling pathway among the mesenchymal populations. (**F**) Comparison of the predicted relative contribution of both significant PDGF signaling ligand receptor pairs, PDGFA-PDGFRB and PDGFA-PDGFRA, across the mesenchymal compartment as a whole. (**G**), (**H**) Circle plots depicting signaling mediated by PDGFA-PDGFRB (**F**) and PDGFA-PDGFRA (**G**) ligand-receptor pairs between mesenchymal populations. Line thickness is proportional to signaling strength, and line colors represent which population is the predicted "Sender" of the signal; colors match populations in (**A**). (**I**) UMAP visualization of the populations detected within the human fetal pancreatic endothelial Broad Group. (**J**) Heatmap depicting the top differentially expressed genes per mesenchymal cell type. (**K**), (**L**) 18 w human fetal pancreas tissue stained with antibodies against pan-endothelial marker CD31 (gray) and pan-endocrine marker CHGA (green), along with ISH probes against (**K**) *GJA5* (yellow) and *ACKR1* (magenta) or (**L**) *GJA5* (yellow) and *PLVAP* (magenta) mRNA; nuclei are counterstained with DAPI (blue). Aqua, white, and yellow arrowheads mark presumptive GJA5+CD31+ Arterial cells, ACKR1+CD31+ Venous cells and PLVAP+CD31+ Capillary/Venous cells, respectively. Dashed insets represent magnified regions. (**M**) Heatmap depicting predicted activity of the NOTCH signaling pathway among the various endothelial populations. (**N**) Predicted relative contribution of each significant NOTCH signaling ligand-receptor pair within the endothelial compartment as a whole. (**O**) Circle plot depicting the communication between endothelial populations with respect to JAG1-NOTCH4 signaling. Line thickness is proportional to signaling strength, and line colors represent which population is the "Sender" of the signal; colors match populations in (**I**). Scale bars are 25 μm throughout.

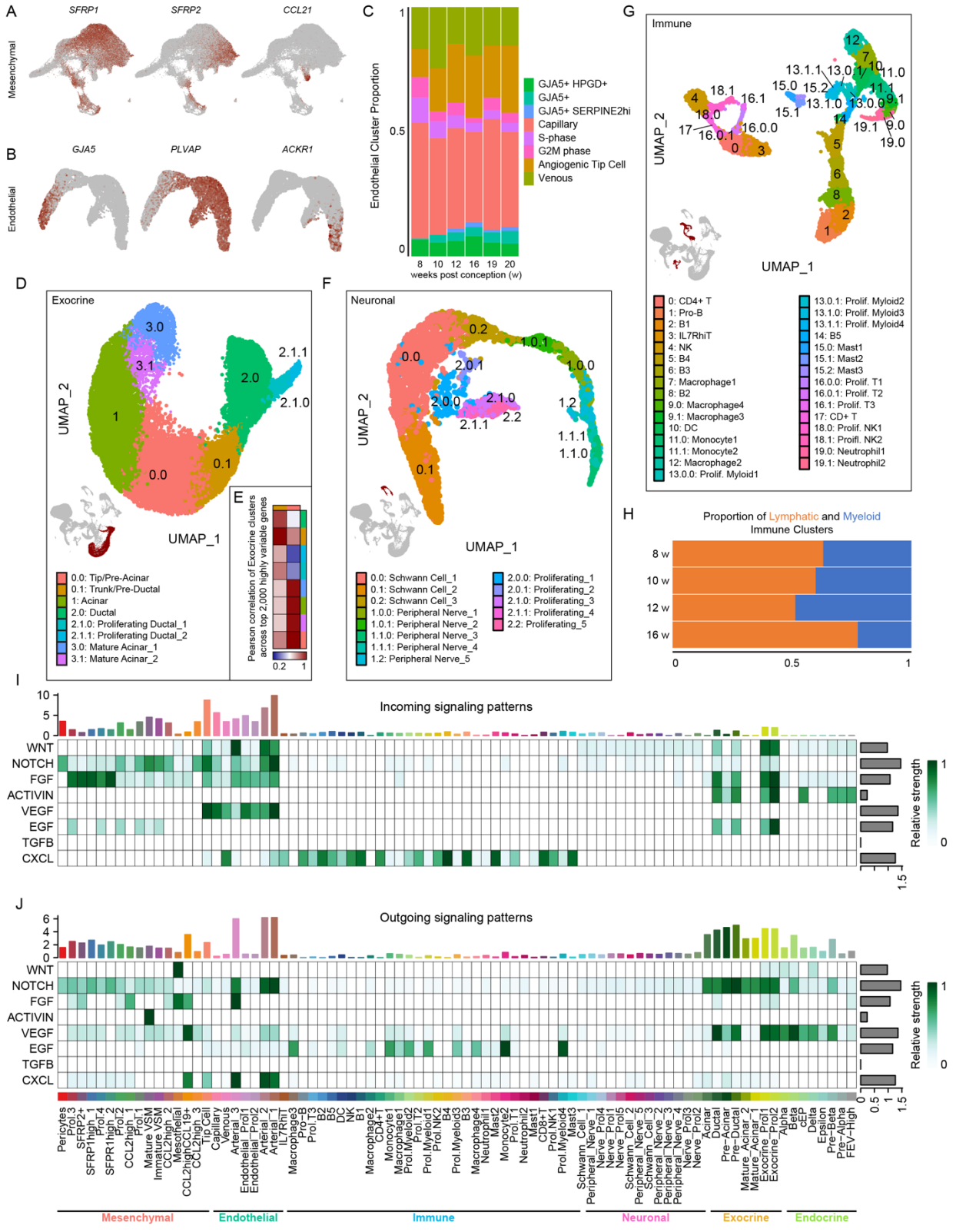

A
SFRP1
SFRP2
CCL21
Mesenchymal
B
GJA5
PLVAP
ACKR1
Endothelial
C
Endothelial Cluster Proportion
weeks post conception (w)
GJA5+ HPGD+
GJA5+
GJA5+ SERPINE2hi
Capillary
S-phase
G2M phase
Angiogenic Tip Cell
Venous
D
Exocrine
UMAP_1
UMAP_2
0.0: Tip/Pre-Acinar
0.1: Trunk/Pre-Ductal
1: Acinar
2.0: Ductal
2.1.0: Proliferating Ductal_1
2.1.1: Proliferating Ductal_2
3.0: Mature Acinar_1
3.1: Mature Acinar_2
E
Pearson correlation of Exocrine clusters across top 2,000 highly variable genes
F
Neuronal
0.0: Schwann Cell_1
0.1: Schwann Cell_2
0.2: Schwann Cell_3
1.0.0: Peripheral Nerve_1
1.0.1: Peripheral Nerve_2
1.1.0: Peripheral Nerve_3
1.1.1: Peripheral Nerve_4
1.2: Peripheral Nerve_5
2.0.0: Proliferating_1
2.0.1: Proliferating_2
2.1.0: Proliferating_3
2.1.1: Proliferating_4
2.2: Proliferating_5
G
Immune
0: CD4+ T
1: Pro-B
2: B1
3: IL7RhiT
4: NK
5: B4
6: B3
7: Macrophage1
8: B2
9.0: Macrophage4
9.1: Macrophage3
10: DC
11.0: Monocyte1
11.1: Monocyte2
12: Macrophage2
13.0.0: Prolif. Myloid1
13.0.1: Prolif. Myloid2
13.1.0: Prolif. Myloid3
13.1.1: Prolif. Myloid4
14: B5
15.0: Mast1
15.1: Mast2
15.2: Mast3
16.0.0: Prolif. T1
16.0.1: Prolif. T2
16.1: Prolif. T3
17: CD+ T
18.0: Prolif. NK1
18.1: Proifl. NK2
19.0: Neutrophil1
19.1: Neutrophil2
H
Proportion of Lymphatic and Myeloid Immune Clusters
8 w
10 w
12 w
16 w
I
Incoming signaling patterns
WNT
NOTCH
FGF
ACTIVIN
VEGF
EGF
TGFB
CXCL
Relative strength
J
Outgoing signaling patterns
Pericytes
Prol3
SFRP2+
SFRP1high_1
Prol4
SFRP1high_2
Prol2
CCL21high
Prol1
Mature_VSM
Immature_VSM
CCL2high_2
Mesothelial
CCL2highCCL19+
CCL2high_3
Tip_Cell
Capillary
Venous
Arterial_3
Endothelial_Prol1
Endothelial_Prol2
Arterial_2
Arterial_1
IL7Rhi
Macrophage3
Pro-B
Prol.T3
B2
B5
DC
NK
B1
Macrophage2
CD4+T
Monocyte1
Macrophage1
Prol.Myeloid2
Prol.T2
Prol.Myeloid1
Prol.NK2
B4
Prol.Myeloid3
B3
Macrophage4
Neutrophil1
Mast2
Monocyte2
Prol.T1
Neutrophil2
Mast1
CD8+T
Prol.NK1
Prol.Myeloid4
Mast3
Schwann_Cell_1
Peripheral_Nerve_1
Nerve_Prol4
Nerve_Prol1
Nerve_Prol5
Schwann_Cell_2
Schwann_Cell_3
Peripheral_Nerve_2
Peripheral_Nerve_3
Peripheral_Nerve_4
Nerve_Prol3
Nerve_Prol2
Acinar
Ductal
Pre-Acinar
Pre-Ductal
Mature_Acinar_2
Mature_Acinar_1
Exocrine_Prol1
Exocrine_Prol2
Alpha
Beta
cEP
Delta
Epsilon
Pre-Beta
Pre-Alpha
FEV-High
Mesenchymal
Endothelial
Immune
Neuronal
Exocrine
Endocrine

Figure 3.4: Novel populations discovered within the exocrine, neuronal, and immune compartments, and predicted intercellular communication between all Broad Group cell subtypes.

(**A**) Feature plots of cell-type specific markers of mesenchymal subclusters (**B**) Feature plots of cell-type specific markers of endothelial subclusters (**C**) Bar graph depicting the proportional representation of endothelial cell types across developmental time (**D**), (**F, G**) UMAP visualization of cell populations identified by CellFindR within the (**D**) exocrine, (**F**) neuronal, and (**G**) immune broad groups within the human fetal pancreas. (**E**) Heatmap depicting pearson correlation of Exocrine clusters across the top 2,000 highly variable genes. (**H**) Bar graph depicting the proportional representation of lymphatic and myeloid immune cell types across developmental time (**I**)-(**J**) Predicted paracrine signaling pathways that mediate cell-cell communication between different Broad Groups, as inferred by CellChat analysis. (**I**) Heatmap of incoming signaling patterns depicting the cell types predicted to receive significant paracrine signals. (**J**) Heatmap of outgoing signaling patterns depicting the cell types predicted to be the source of significant paracrine signals. Bar graphs at the top of (**I**) and (**J**) represent the aggregate signaling for each cluster across all signaling pathways; bar graphs on the right represent the aggregate signaling strength of each signaling pathway across all clusters.

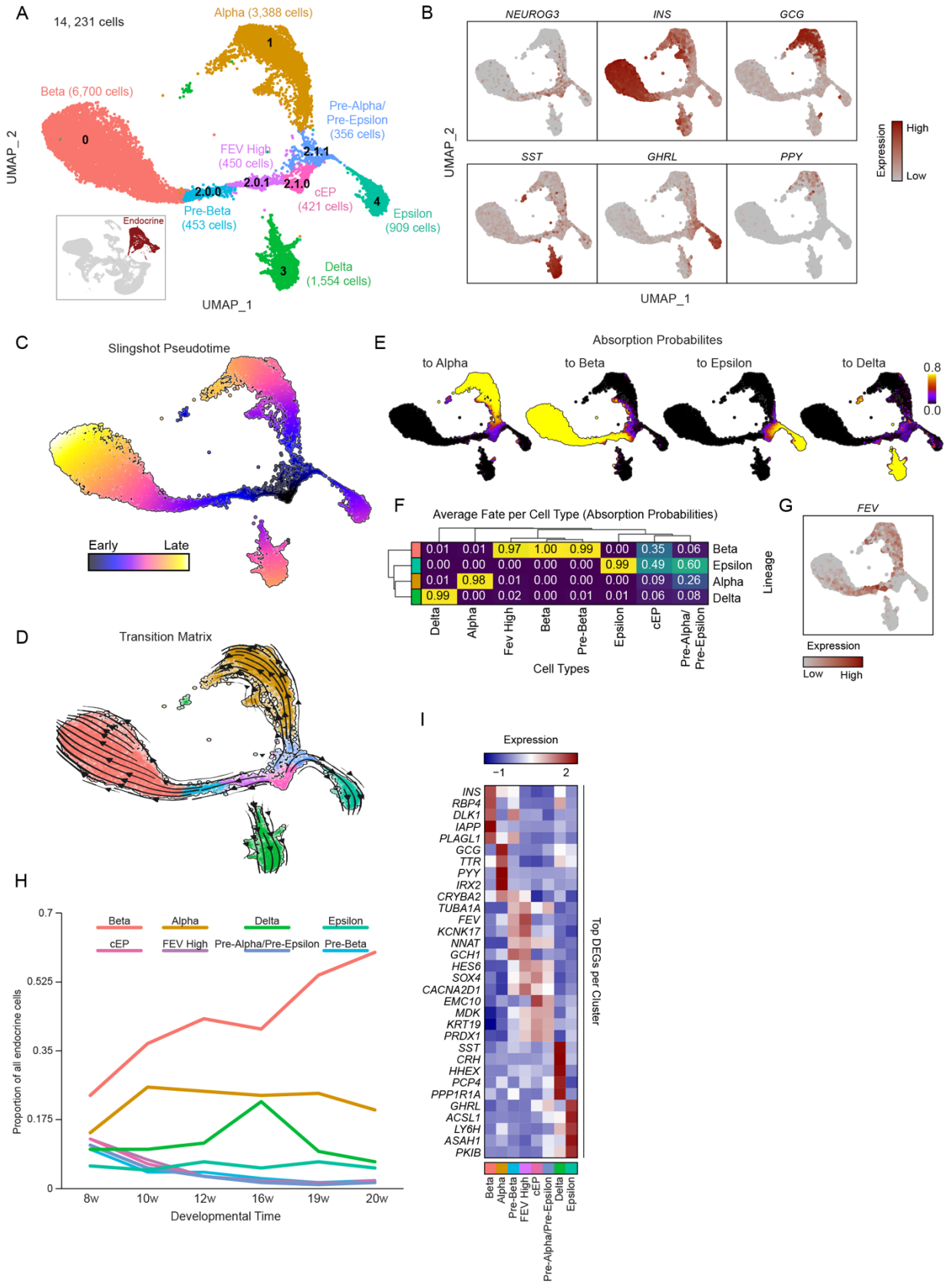

A
14, 231 cells
Alpha (3,388 cells)
Beta (6,700 cells)
Pre-Alpha/
Pre-Epsilon
(356 cells)
FEV High
(450 cells)
cEP
(421 cells)
Pre-Beta
(453 cells)
Epsilon
(909 cells)
Delta
(1,554 cells)
Endocrine
UMAP_2
UMAP_1
B
NEUROG3
INS
GCG
SST
GHRL
PPY
Expression
High
Low
UMAP_2
UMAP_1
C
Slingshot Pseudotime
Early
Late
D
Transition Matrix
E
Absorption Probabilites
to Alpha
to Beta
to Epsilon
to Delta
0.8
0.0
F
Average Fate per Cell Type (Absorption Probabilities)
Beta
Epsilon
Alpha
Delta
Lineage
Cell Types
G
FEV
Expression
Low
High
H
Beta
Alpha
Delta
Epsilon
cEP
FEV High
Pre-Alpha/Pre-Epsilon
Pre-Beta
Proportion of all endocrine cells
Developmental Time
I
Expression
Top DEGs per Cluster

Figure 3.5: Discovery of four novel putative progenitor populations and unique lineage dynamics in the developing human endocrine pancreas.

(**A**) UMAP visualization of sub-clustered endocrine populations identified in the merged dataset as shown in **Figure 3.1b** (inset). cEP = Common Endocrine Progenitor. (**B**) Feature plots show expression of known markers of endocrine cell types, including *NEUROG3* to mark endocrine progenitors, *INS* to mark beta cells, *GCG* to mark alpha cells, *SST* to mark delta cells, and *GHRL* to mark epsilon cells. Expression of the transcription factor *FEV* is also shown. (**C**) Pseudotime reconstruction of endocrine lineage trajectories assembled using Slingshot. (**D**) UMAP visualization of CellRank directed transition matrix. Arrows represent the directed transition matrix based on the KNN graph and pseudotime values. (**E**) UMAP visualization of the absorption probabilities, depicting the likelihood of each cell becoming each of the differentiated cell types listed. (**F**) Heatmap reflecting the average absorption probabilities depicted in (**E**) for each cluster. (**G**) Feature plot depicting expression of the transcription factor *FEV* in the endocrine cells (**H**) Line graph showing the representation of each endocrine cell population as a proportion of the total number of endocrine cells, across developmental time. (**I**) Heatmap depicting expression levels of the top 5 differentially-expressed genes per endocrine cluster.

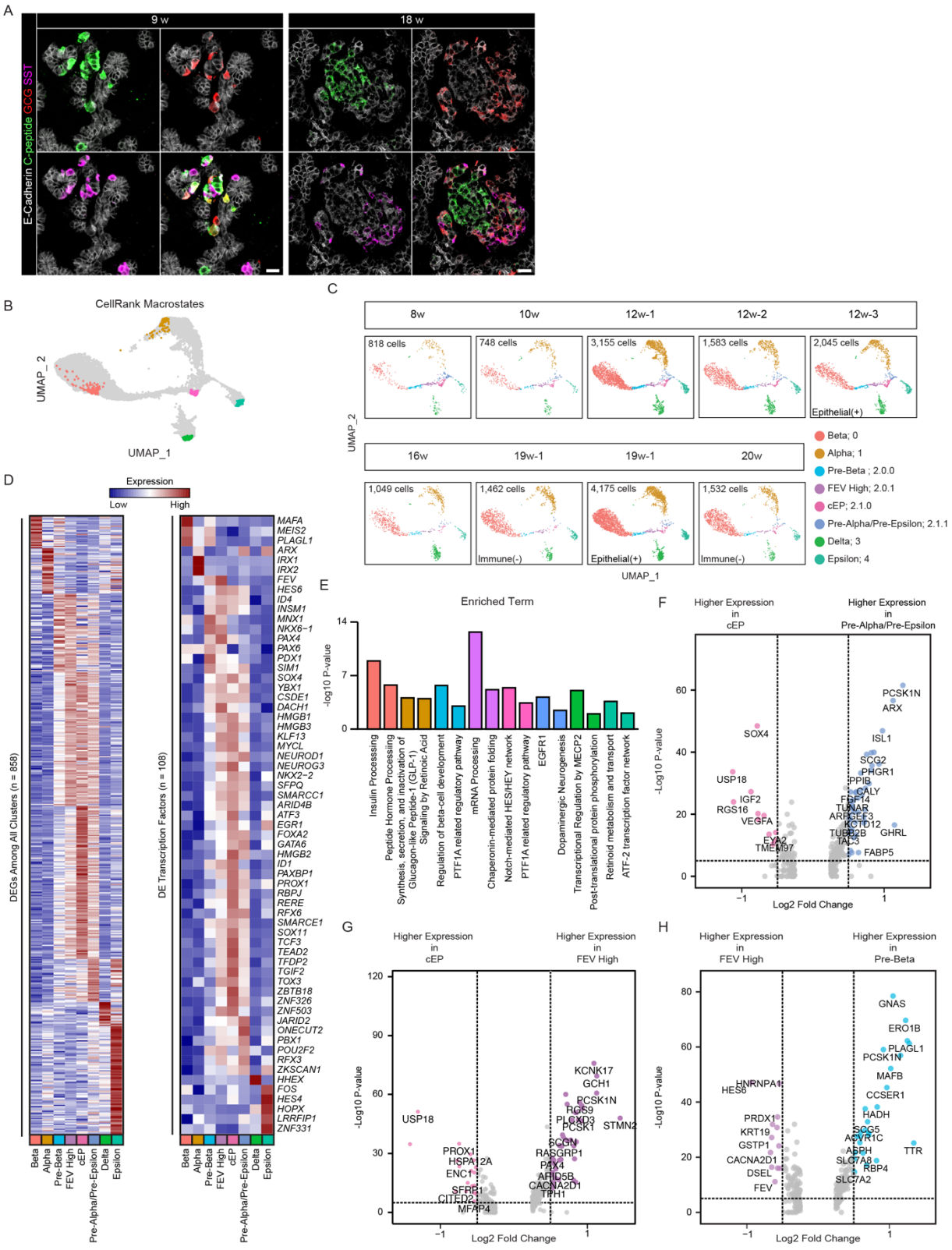

A
9 w
18 w
E-Cadherin C-peptide GCG SST
B
CellRank Macrostates
UMAP_2
UMAP_1
C
8w
10w
12w-1
12w-2
12w-3
818 cells
748 cells
3,155 cells
1,583 cells
2,045 cells
Epithelial(+)
16w
19w-1
19w-1
20w
1,049 cells
1,462 cells
4,175 cells
1,532 cells
Immune(-)
Epithelial(+)
Immune(-)
UMAP_2
UMAP_1
Beta; 0
Alpha; 1
Pre-Beta ; 2.0.0
FEV High; 2.0.1
cEP; 2.1.0
Pre-Alpha/Pre-Epsilon; 2.1.1
Delta; 3
Epsilon; 4
D
Expression
Low
High
DEGs Among All Clusters (n = 858)
DE Transcription Factors (n = 108)
MAFA
MEIS2
PLAGL1
ARX
IRX1
IRX2
FEV
HES6
ID4
INSM1
MNX1
NKX6−1
PAX4
PAX6
RDX1
SIM1
SOX4
YBX1
CSDE1
DACH1
HMGB1
HMGB3
KLF13
MYCL
NEUROD1
NEUROG3
NKX2−2
SFPQ
SMARCC1
ARID4B
ATF3
EGR1
FOXA2
GATA6
HMGB2
ID1
PAXBP1
PROX1
RBPJ
RERE
RFX6
SMARCE1
SOX11
TCF3
TEAD2
TFDP2
TGIF2
TOX3
ZBTB18
ZNF326
ZNF503
JARID2
ONECUT2
PBX1
POU2F2
RFX3
ZKSCAN1
HHEX
FOS
HES4
HOPX
LRRFIP1
ZNF331
Beta
Alpha
Pre-Beta
FEV High
cEP
Pre-Alpha/Pre-Epsilon
Delta
Epsilon
E
Enriched Term
-log10 P-value
Insulin Processing
Peptide Hormone Processing
Synthesis, secretion, and inactivation of Glucagon-like Peptide-1 (GLP-1)
Signaling by Retinoic Acid
Regulation of beta-cell development
PTF1A related regulatory pathway
mRNA Processing
Chaperonin-mediated protein folding
Notch-mediated HES/HEY network
PTF1A related regulatory pathway
EGFR1
Dopaminergic Neurogenesis
Transcriptional Regulation by MECP2
Post-translational protein phosphorylation
Retinoid metabolism and transport
ATF-2 transcription factor network
F
Higher Expression in cEP
Higher Expression in Pre-Alpha/Pre-Epsilon
-Log10 P-value
Log2 Fold Change
PCSK1N
ARX
SOX4
ISL1
SCG2
PHGR1
PPIB
CALY
FGF14
TUNAR
ARHGEF3
KCTD12
TUBB2B
GHRL
TAL1
FABP5
USP18
IGF2
RGS16
VEGFA
EYA2
TMEM97
G
Higher Expression in cEP
Higher Expression in FEV High
-Log10 P-value
Log2 Fold Change
KCNK17
GCH1
PCSK1N
RGS9
PLCXD3
PCSK1
STMN2
SCGN
RASGRP1
PAX4
ARID5B
CACNA2D1
TPH1
USP18
PROX1
HSPA12A
ENC1
SFRP1
CITED2
MFAP4
H
Higher Expression in FEV High
Higher Expression in Pre-Beta
-Log10 P-value
Log2 Fold Change
GNAS
ERO1B
PLAGL1
PCSK1N
MAFB
CCSER1
HADH
SCG5
ACVR1C
ABPH
SLC7A8
RBP4
SLC7A2
TTR
HES6
HNRNPA1
PRDX1
KRT19
GSTP1
CACNA2D1
DSEL
FEV

Figure 3.6: Transcriptomic and population analyses of human fetal endocrine cells.

(**A**) Immunofluorescence staining of 9 and 18 w human fetal pancreas tissue for epithelial marker E-Cadherin and hormones GCG, C-peptide and SST. Scale bar, 20 μm. (**B**) UMAP visualization of macrostates identified by CellRank. (**C**) Split UMAP visualization shows the contribution of each individual sample to the merged endocrine scRNA-Seq dataset, with the number of cells per sample listed and each cell colored according to its endocrine cluster identity. (**D**) Heatmaps depicting expression levels of all differentially expressed genes among endocrine cells of the human fetal pancreas (left). All differentially expressed transcription factors are depicted in the heatmap on the right. (**E**) Pathway analysis of all genes with a log2-fold change (log2FC) in expression of at least 0.5 between a single endocrine cell population and all other endocrine cell populations of the human fetal pancreas. For each significant pathway, the clusters in which that pathway is active is colored to match the color in panel (**B**). (**F-H**) Volcano plots depicting pairwise comparisons of gene expression in Common Endocrine Progenitor (cEP) vs. Pre-Alpha/Pre-Epsilon cells (**F**), cEP vs. FEV High (**G**), and Pre-Beta vs. FEV High cells (**H**). Genes with a log2FC of at least 0.5 are highlighted.

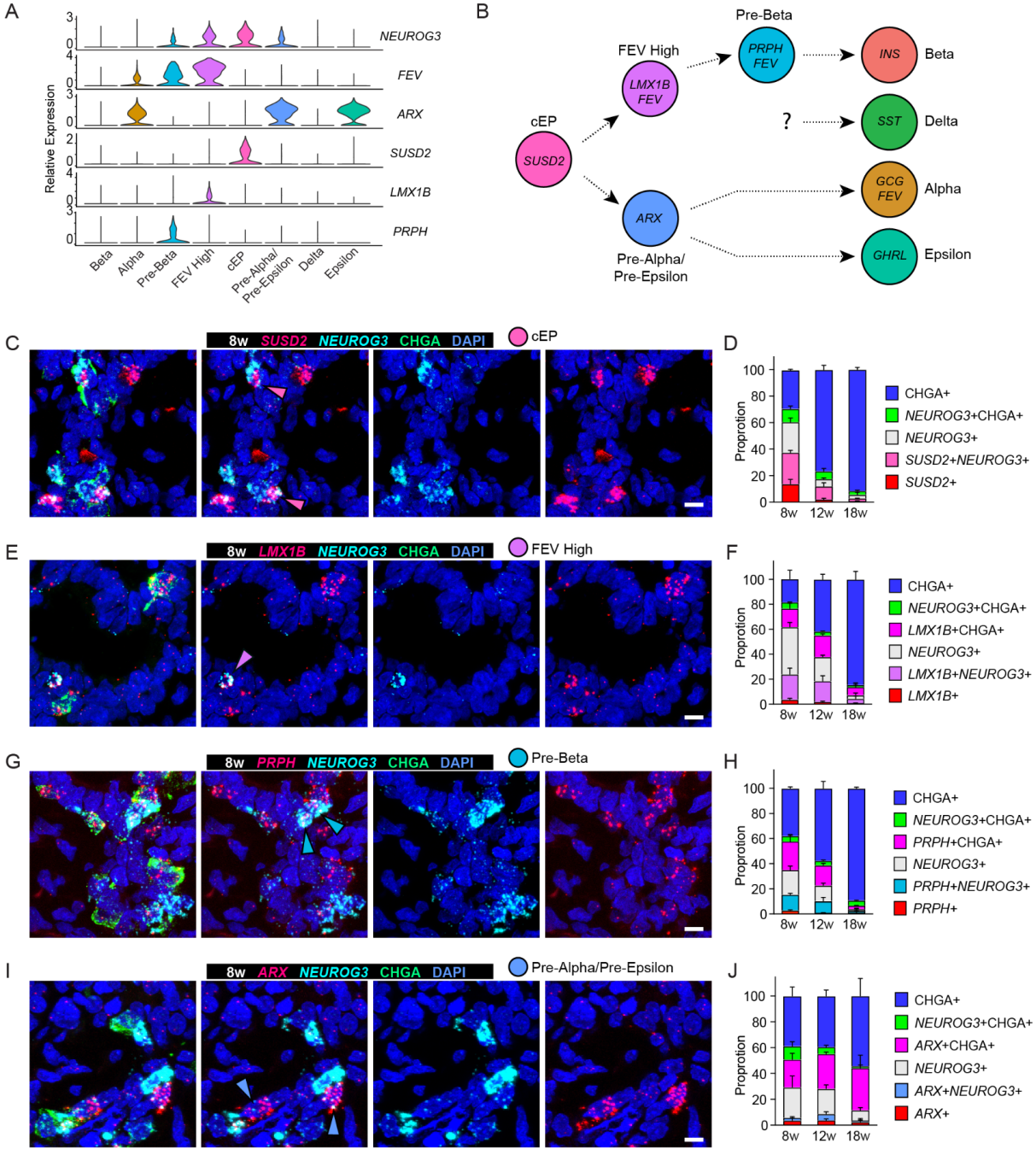

A
Relative Expression
NEUROG3
FEV
ARX
SUSD2
LMX1B
PRPH
Beta
Alpha
Pre-Beta
FEV High
cEP
Pre-Alpha/
Pre-Epsilon
Delta
Epsilon
B
cEP
SUSD2
FEV High
LMX1B
FEV
Pre-Beta
PRPH
FEV
INS
Beta
?
SST
Delta
GCG
FEV
Alpha
ARX
Pre-Alpha/
Pre-Epsilon
GHRL
Epsilon
C
8w SUSD2 NEUROG3 CHGA DAPI
cEP
D
Proprotion
CHGA+
NEUROG3+CHGA+
NEUROG3+
SUSD2+NEUROG3+
SUSD2+
8w
12w
18w
E
8w LMX1B NEUROG3 CHGA DAPI
FEV High
F
Proprotion
CHGA+
NEUROG3+CHGA+
LMX1B+CHGA+
NEUROG3+
LMX1B+NEUROG3+
LMX1B+
8w
12w
18w
G
8w PRPH NEUROG3 CHGA DAPI
Pre-Beta
H
Proprotion
CHGA+
NEUROG3+CHGA+
PRPH+CHGA+
NEUROG3+
PRPH+NEUROG3+
PRPH+
8w
12w
18w
I
8w ARX NEUROG3 CHGA DAPI
Pre-Alpha/Pre-Epsilon
J
Proprotion
CHGA+
NEUROG3+CHGA+
ARX+CHGA+
NEUROG3+
ARX+NEUROG3+
ARX+
8w
12w
18w

Figure 3.7: *In vivo* confirmation and quantification of novel endocrine progenitor cell populations.

(**A**) Violin plot depicting expression of endocrine progenitor (EP) marker genes. (**B**) Model showing predicted lineage relationships among developing human endocrine cells, along with genes that mark each population. The lineage trajectory of human delta cells remains unclear. (**C,E,G,I**) 8 w human fetal tissues were stained with an antibody against the differentiated endocrine cell marker CHGA (green), along with probes against *NEUROG3* (cyan) and putative EP cell markers (red) (**C**) *SUSD2*; (**E**) *LMX1B*; (**G**) *PRPH*; (**I**) *ARX* for detection by *in situ* hybridization (ISH). Nuclei were counterstained with DAPI (blue). Scale bars, 10 μm. Pink, Purple, Aqua, and Blue arrowheads mark presumptive cEP cells (SUSD2+NEUROG3+CHGA-), FEV High EPs (LMX1B+NEUROG3+CHGA+), Pre-Beta EPs (PRPH+NEUROG3+CHGA+), and Pre-Alpha/Pre-Epsilon EPs (ARX+NEUROG3+CHGA+), respectively. (**D,F,H,J**) Quantification of staining as performed in (**C,E,G,I**) extended across a time course of 8, 12, and 18 w. Y-axis represents the proportion of each type of EP as a fraction of all cells that stained positive for at least one of the markers. Data are presented as mean ± SEM (n=3 biological replicates per timepoint).

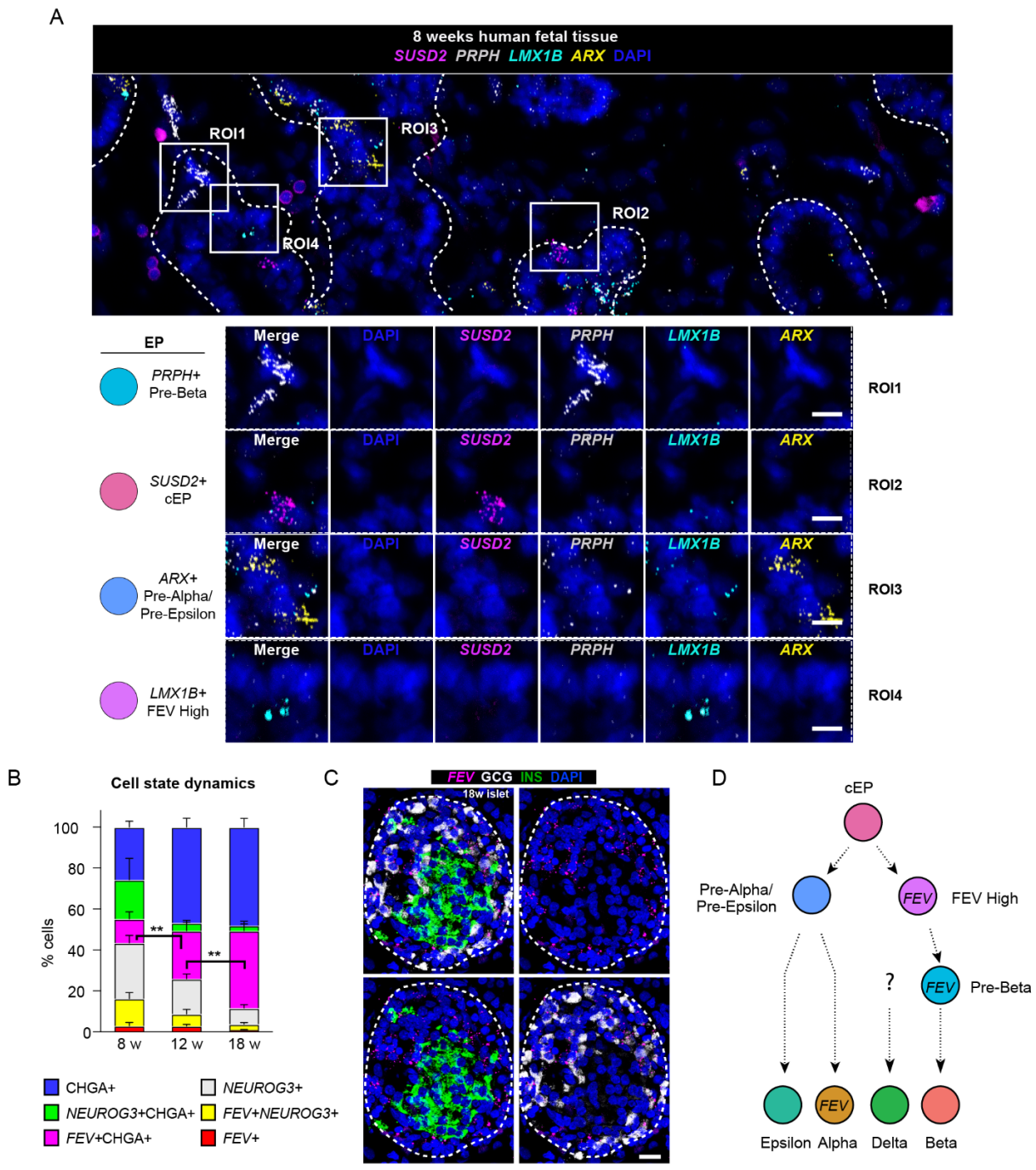

A
8 weeks human fetal tissue
SUSD2 PRPH LMX1B ARX DAPI
ROI1
ROI3
ROI2
ROI4
EP
PRPH+ Pre-Beta
SUSD2+ cEP
ARX+ Pre-Alpha/ Pre-Epsilon
LMX1B+ FEV High
Merge
DAPI
SUSD2
PRPH
LMX1B
ARX
ROI1
ROI2
ROI3
ROI4
B
Cell state dynamics
% cells
100
80
60
40
20
0
8 w
12 w
18 w
**
**
CHGA+
NEUROG3+
NEUROG3+CHGA+
FEV+NEUROG3+
FEV+CHGA+
FEV+
C
FEV GCG INS DAPI
18w islet
D
cEP
Pre-Alpha/ Pre-Epsilon
FEV
FEV High
?
FEV
Pre-Beta
Epsilon
Alpha
Delta
Beta

Figure 3.8: Multiplexed *in vivo* confirmation of novel putative endocrine progenitor populations in the developing human pancreas.

(**A**) Multiplexed *in situ* hybridization (ISH) and immunofluorescence (IF) staining of 8 w human fetal pancreas tissue with four probes against novel endocrine progenitor (EP) cell markers: *SUSD2* (red), *PRPH* (grey), *LMX1B* (cyan), and *ARX* (yellow); nuclei were counterstained with DAPI (blue). Regions of interest (ROIs) 1, 2, 3, and 4 highlight cells expressing single marker genes distinguishing Pre-Beta EP, cEP, Pre-Alpha/Pre-Epsilon EP, and FEV High EP cells, respectively. Dashed lines indicate epithelial border in the tissue. Scale bars, 10 μm. (**B**) Dynamics of cell states assessed by multiplexed ISH and IF staining for *FEV/NEUROG3*/CHGA, across developmental time. The relative proportion of FEV+CHGA+ cells significantly increases as development progresses. Significance was calculated using an unpaired t-test; **, P-value < 0.01. Data are presented as mean ± SEM (n=3 biological samples per timepoint). (**C**) Multiplexed ISH/IF staining for *FEV* (magenta), INS (green), and GCG (grey) on 18 w tissue. An islet is circled with a dashed line. Scale bar, 25 μm. (**D**) Diagram summarizing expression of *FEV* in EPs and GCG+ alpha cells.

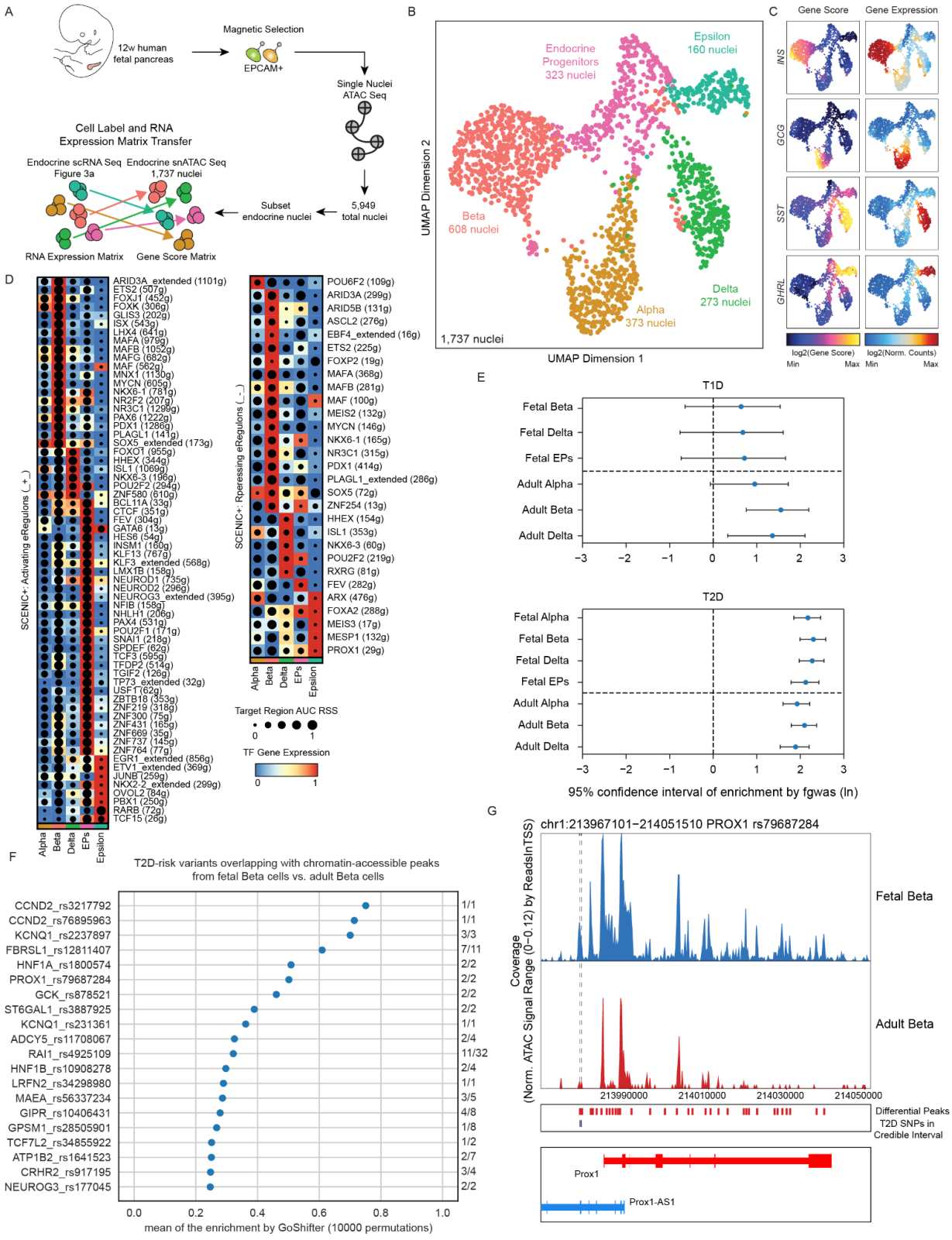

A
12w human fetal pancreas
Magnetic Selection
EPCAM+
Single Nuclei ATAC Seq
5,949 total nuclei
Subset endocrine nuclei
Cell Label and RNA Expression Matrix Transfer
Endocrine scRNA Seq Figure 3a
Endocrine snATAC Seq 1,737 nuclei
RNA Expression Matrix
Gene Score Matrix
B
Endocrine Progenitors 323 nuclei
Epsilon 160 nuclei
Beta 608 nuclei
Delta 273 nuclei
Alpha 373 nuclei
1,737 nuclei
UMAP Dimension 1
UMAP Dimension 2
C
Gene Score
Gene Expression
INS
GCG
SST
GHRL
log2(Gene Score) Min Max
log2(Norm. Counts) Min Max
D
SCENIC+: Activating eRegulons (_+_)
SCENIC+: Repressing eRegulons (_-_)
Alpha
Beta
Delta
EPs
Epsilon
Target Region AUC RSS
TF Gene Expression
E
T1D
Fetal Beta
Fetal Delta
Fetal EPs
Adult Alpha
Adult Beta
Adult Delta
T2D
Fetal Alpha
95% confidence interval of enrichment by fgwas (ln)
F
T2D-risk variants overlapping with chromatin-accessible peaks from fetal Beta cells vs. adult Beta cells
mean of the enrichment by GoShifter (10000 permutations)
G
chr1:213967101-214051510 PROX1 rs79687284
Fetal Beta
Adult Beta
Coverage (Norm. ATAC Signal Range (0-0.12) by ReadsInTSS)
Differential Peaks
T2D SNPs in Credible Interval
Prox1
Prox1-AS1

Figure 3.9: Single-nucleus ATAC-Sequencing reveals chromatin accessibility dynamics and diabetes genetic risk loci in the developing human endocrine pancreas.

(**A**) Schematic of workflow for single-nucleus ATAC-Sequencing (snATAC-Seq) performed on EpCAM+ enriched cells from 12 w human fetal pancreas. (**B**) UMAP visualization of snATAC-Seq data from all endocrine cells reveals populations of hormone-expressing cells (alpha, beta, delta, and epsilon cells) as well as endocrine progenitors (collapsed here into a single population; for endocrine progenitors broken down into 4 subpopulations see **Figure 3.10b**). (**C**) Feature plots showing ATAC gene scores (left) and corresponding RNA expression values (right) from integration of snATAC-Seq and scRNA-Seq data. (**D**) Heatmap depicting cell type specific SCENIC+ eRegulons. Heatmap colors depict TF expression, while dot size depicts TF target regions area-under-the-curve regulon specificity scores (RSS). (**E**) Forest plots depicting enrichment of T1D and T2D risk variants in fetal and adult endocrine cells as assessed with fGWAS. (**F**) T2D-risk loci enriched in differentially accessible peaks of fetal beta cells vs. adult beta cells as measured by GoShifter. Fractions listed to the right of the plot represent the number of SNPs in the genetic credible interval that are enriched in fetal vs. adult beta cells. (**G**) Track plots displaying accessibility of the *PROX1* locus in fetal (top) vs. adult (bottom) beta cells. Differential peaks and T2D risk variants in the genetic credible interval are highlighted. T2D risk variants overlapping with differential peaks are denoted by dashed lines.

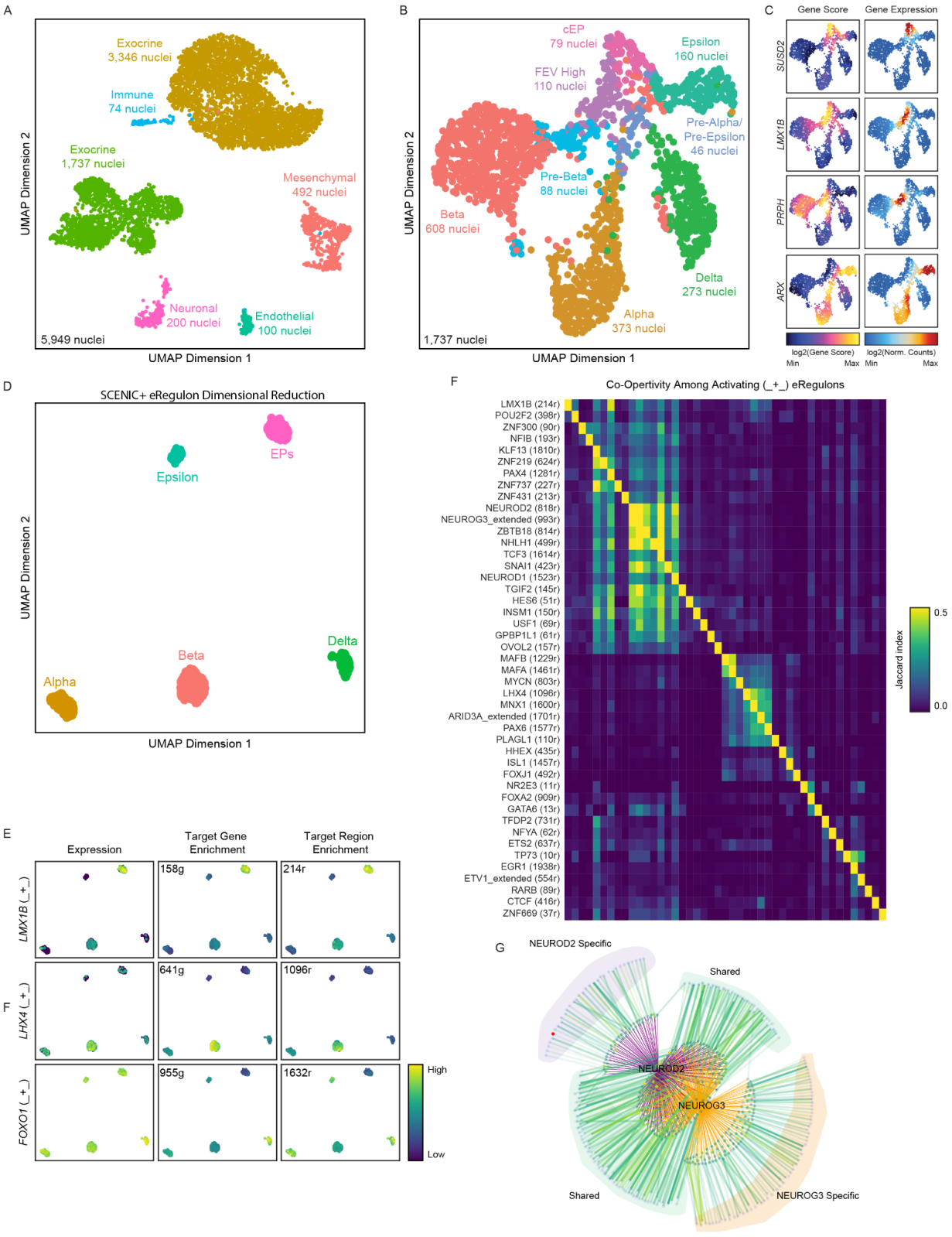
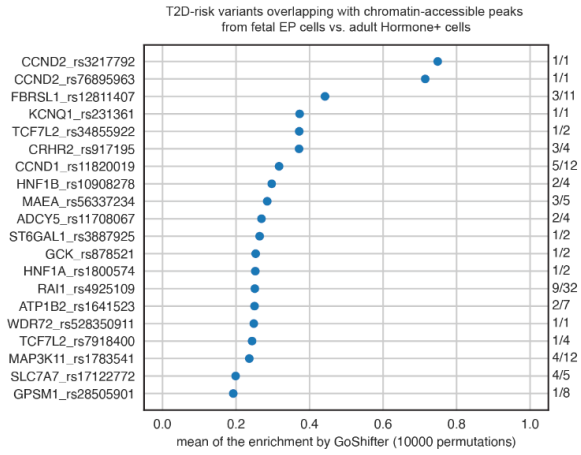

A
Exocrine
3,346 nuclei
Immune
74 nuclei
Exocrine
1,737 nuclei
Mesenchymal
492 nuclei
Neuronal
200 nuclei
Endothelial
100 nuclei
5,949 nuclei
UMAP Dimension 2
UMAP Dimension 1
B
cEP
79 nuclei
Epsilon
160 nuclei
FEV High
110 nuclei
Pre-Alpha/
Pre-Epsilon
46 nuclei
Pre-Beta
88 nuclei
Beta
608 nuclei
Delta
273 nuclei
Alpha
373 nuclei
1,737 nuclei
UMAP Dimension 2
UMAP Dimension 1
C
Gene Score
Gene Expression
SUSD2
LMX1B
PRPH
ARX
log2(Gene Score)
Min
Max
log2(Norm. Counts)
Min
Max
D
SCENIC+ eRegulon Dimensional Reduction
EPs
Epsilon
Delta
Beta
Alpha
UMAP Dimension 2
UMAP Dimension 1
F
Co-Opertivity Among Activating (_+_) eRegulons
LMX1B (214r)
POU2F2 (398r)
ZNF300 (90r)
NFIB (193r)
KLF13 (1810r)
ZNF219 (624r)
PAX4 (1281r)
ZNF737 (227r)
ZNF431 (213r)
NEUROD2 (818r)
NEUROG3_extended (993r)
ZBTB18 (814r)
NHLH1 (499r)
TCF3 (1614r)
SNAI1 (423r)
NEUROD1 (1523r)
TGIF2 (145r)
HES6 (51r)
INSM1 (150r)
USF1 (69r)
GPBP1L1 (61r)
OVOL2 (157r)
MAFB (1229r)
MAFA (1461r)
MYCN (803r)
LHX4 (1096r)
MNX1 (1600r)
ARID3A_extended (1701r)
PAX6 (1577r)
PLAGL1 (110r)
HHEX (435r)
ISL1 (1457r)
FOXJ1 (492r)
NR2E3 (11r)
FOXA2 (909r)
GATA6 (13r)
TFDP2 (731r)
NFYA (62r)
ETS2 (637r)
TP73 (10r)
EGR1 (1938r)
ETV1_extended (554r)
RARB (89r)
CTCF (416r)
ZNF669 (37r)
Jaccard index
0.5
0.0
E
Expression
Target Gene Enrichment
Target Region Enrichment
LMX1B (_+_)
158g
214r
F
LHX4 (_+_)
641g
1096r
FOXO1 (_+_)
955g
1632r
High
Low
G
NEUROD2 Specific
Shared
NEUROD2
NEUROG3
Shared
NEUROG3 Specific

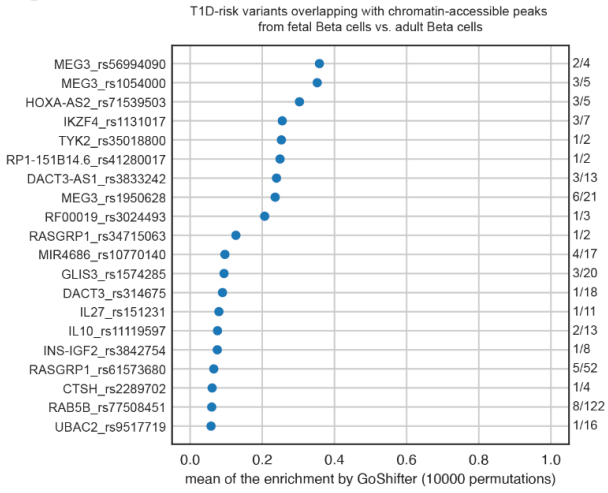
Figure 3.10: Gene regulatory networks in the human fetal endocrine pancreas.

(**A**) UMAP displaying snATAC-Seq data for all fetal pancreatic cells. (**B**) UMAP displaying snATAC-Seq data for all fetal endocrine cells annotated via unconstrained integration with the scRNA-Seq dataset from **Figure 3.5**. In contrast to **Figure 3.9b**, here the endocrine progenitor (EP) subpopulations are labeled individually rather than pooled. (**C**) ATAC gene scores (left) and corresponding RNA expression values (right) of EP markers as defined in **Figure 3.7a**. (**D**) UMAP displaying SCENIC+ eRegulon dimensional reduction of fetal endocrine cells. (**E**) UMAPs displaying TF expression (left column), TF target gene enrichment (middle column) and TF target region enrichment (right column) in the SCENIC+ endocrine dataset. (**F**) Heatmap depicting co-operativity among the target regions of eRegulon TFs, clustered based on the Jaccard Index (overlap between target regions of the given eRegulon TF). (**G**) Network diagram depicting eRegulons for *NEUROG3* and *NEUROD2*. Target regions (diamond shape) are colored by their log2 FC enrichment in the EP cell type, and target genes are represented by circles.

A


T2D-risk variants overlapping with chromatin-accessible peaks
from fetal EP cells vs. adult Hormone+ cells
CCND2_rs3217792 1/1
CCND2_rs76895963 1/1
FBRSL1_rs12811407 3/11
KCNQ1_rs231361 1/1
TCF7L2_rs34855922 1/2
CRHR2_rs917195 3/4
CCND1_rs11820019 5/12
HNF1B_rs10908278 2/4
MAEA_rs56337234 3/5
ADCY5_rs11708067 2/4
ST6GAL1_rs3887925 1/2
GCK_rs878521 1/2
HNF1A_rs1800574 1/2
RAI1_rs4925109 9/32
ATP1B2_rs1641523 2/7
WDR72_rs528350911 1/1
TCF7L2_rs7918400 1/4
MAP3K11_rs1783541 4/12
SLC7A7_rs17122772 4/5
GPSM1_rs28505901 1/8
0.0 0.2 0.4 0.6 0.8 1.0
mean of the enrichment by GoShifter (10000 permutations)

B


T1D-risk variants overlapping with chromatin-accessible peaks
from fetal Beta cells vs. adult Beta cells
MEG3_rs56994090 2/4
MEG3_rs1054000 3/5
HOXA-AS2_rs71539503 3/5
IKZF4_rs1131017 3/7
TYK2_rs35018800 1/2
RP1-151B14.6_rs41280017 1/2
DACT3-AS1_rs3833242 3/13
MEG3_rs1950628 6/21
RF00019_rs3024493 1/3
RASGRP1_rs34715063 1/2
MIR4686_rs10770140 4/17
GLIS3_rs1574285 3/20
DACT3_rs314675 1/18
IL27_rs151231 1/11
IL10_rs11119597 2/13
INS-IGF2_rs3842754 1/8
RASGRP1_rs61573680 5/52
CTSH_rs2289702 1/4
RAB5B_rs77508451 8/122
UBAC2_rs9517719 1/16
0.0 0.2 0.4 0.6 0.8 1.0
mean of the enrichment by GoShifter (10000 permutations)

C

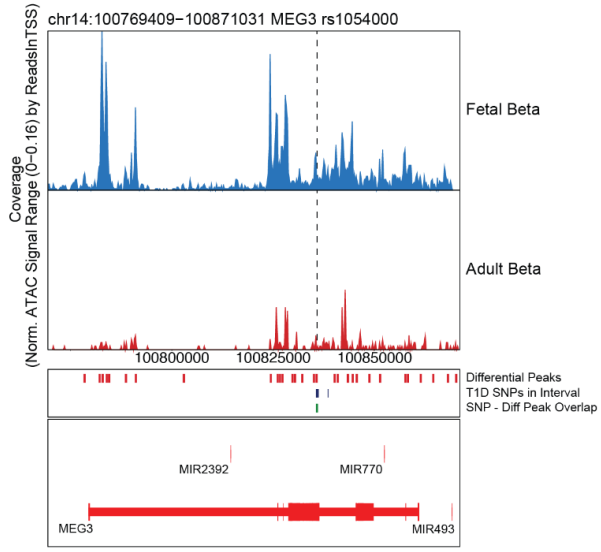

chr14:100769409−100871031 MEG3 rs1054000
Coverage
(Norm. ATAC Signal Range (0−0.16) by ReadsInTSS)
Fetal Beta
Adult Beta
100800000 100825000 100850000
Differential Peaks
T1D SNPs in Interval
SNP - Diff Peak Overlap
MIR2392
MIR770
MEG3
MIR493

Figure 3.11: T1D and T2D GWAS risk loci in the human fetal and adult pancreas.

(**A**) T2D-risk loci enriched in differentially accessible peaks of fetal EP vs. adult hormone+ endocrine cells, as identified by GoShifter. Fractions listed to the right of the plot represent the number of SNPs in the genetic credible interval that are enriched in fetal EP cells vs. adult hormone+ cells. (**B**) T1D-risk loci enriched in differentially accessible peaks of fetal beta cells vs. adult beta cells, as determined by GoShifter. Fractions listed to the right of the plot represent the number of SNPs in the genetic credible interval that are enriched in fetal vs. adult beta cells. (**C**) Track plot displaying accessibility of the *MEG3* locus in fetal (top) vs. adult (bottom) beta cells. Differential peaks and T1D risk variants in the genetic credible interval are highlighted. T1D risk variant overlapping with differential peaks is denoted by dashed line.

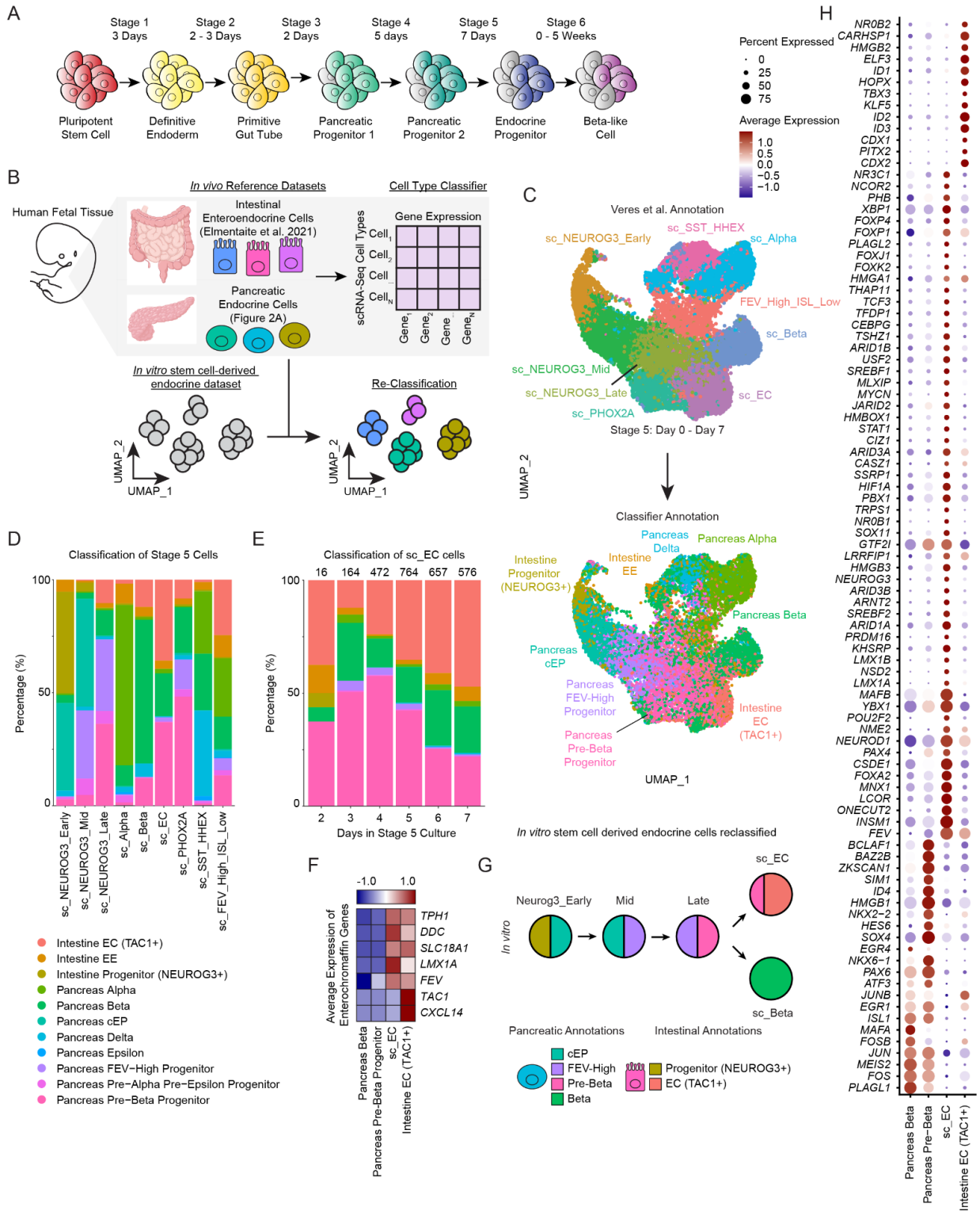

A
Stage 1
3 Days
Stage 2
2 - 3 Days
Stage 3
2 Days
Stage 4
5 days
Stage 5
7 Days
Stage 6
0 - 5 Weeks
Pluripotent Stem Cell
Definitive Endoderm
Primitive Gut Tube
Pancreatic Progenitor 1
Pancreatic Progenitor 2
Endocrine Progenitor
Beta-like Cell
B
In vivo Reference Datasets
Cell Type Classifier
Human Fetal Tissue
Intestinal Enteroendocrine Cells (Elmentaite et al. 2021)
Pancreatic Endocrine Cells (Figure 2A)
Gene Expression
scRNA-Seq Cell Types
In vitro stem cell-derived endocrine dataset
Re-Classification
UMAP_1
UMAP_2
C
Veres et al. Annotation
sc_NEUROG3_Early
sc_SST_HHEX
sc_Alpha
FEV_High_ISL_Low
sc_Beta
sc_NEUROG3_Mid
sc_NEUROG3_Late
sc_EC
sc_PHOX2A
Stage 5: Day 0 - Day 7
Classifier Annotation
Pancreas Delta
Pancreas Alpha
Intestine Progenitor (NEUROG3+)
Intestine EE
Pancreas Beta
Pancreas cEP
Pancreas FEV-High Progenitor
Intestine EC (TAC1+)
Pancreas Pre-Beta Progenitor
UMAP_1
UMAP_2
In vitro stem cell derived endocrine cells reclassified
D
Classification of Stage 5 Cells
Percentage (%)
sc_NEUROG3_Early
sc_NEUROG3_Mid
sc_NEUROG3_Late
sc_Alpha
sc_Beta
sc_EC
sc_PHOX2A
sc_SST_HHEX
sc_FEV_High_ISL_Low
Intestine EC (TAC1+)
Intestine EE
Intestine Progenitor (NEUROG3+)
Pancreas Alpha
Pancreas Beta
Pancreas cEP
Pancreas Delta
Pancreas Epsilon
Pancreas FEV−High Progenitor
Pancreas Pre−Alpha Pre−Epsilon Progenitor
Pancreas Pre−Beta Progenitor
E
Classification of sc_EC cells
16
164
472
764
657
576
Percentage (%)
Days in Stage 5 Culture
F
Average Expression of Enterochromaffin Genes
TPH1
DDC
SLC18A1
LMX1A
FEV
TAC1
CXCL14
Pancreas Beta
Pancreas Pre-Beta Progenitor
sc_EC
Intestine EC (TAC1+)
G
In vitro
Neurog3_Early
Mid
Late
sc_EC
sc_Beta
Pancreatic Annotations
Intestinal Annotations
cEP
FEV-High
Pre-Beta
Beta
Progenitor (NEUROG3+)
EC (TAC1+)
H
Percent Expressed
0
25
50
75
Average Expression
1.0
0.5
0.0
−0.5
−1.0
NR0B2
CARHSP1
HMGB2
ELF3
ID1
HOPX
TBX3
KLF5
ID2
ID3
CDX1
PITX2
CDX2
NR3C1
NCOR2
PHB
XBP1
FOXP4
FOXP1
PLAGL2
FOXJ1
FOXK2
HMGA1
THAP11
TCF3
TFDP1
CEBPG
TSHZ1
ARID1B
USF2
SREBF1
MLXIP
MYCN
JARID2
HMBOX1
STAT1
CIZ1
ARID3A
CASZ1
SSRP1
HIF1A
PBX1
TRPS1
NR0B1
SOX11
GTF2I
LRRFIP1
HMGB3
NEUROG3
ARID3B
ARNT2
SREBF2
ARID1A
PRDM16
KHSRP
LMX1B
NSD2
LMX1A
MAFB
YBX1
POU2F2
NME2
NEUROD1
PAX4
CSDE1
FOXA2
MNX1
LCOR
ONECUT2
INSM1
FEV
BCLAF1
BAZ2B
ZKSCAN1
SIM1
ID4
HMGB1
NKX2−2
HES6
SOX4
EGR4
NKX6−1
PAX6
ATF3
JUNB
EGR1
ISL1
MAFA
FOSB
JUN
MEIS2
FOS
PLAGL1
Pancreas Beta
Pancreas Pre−Beta
sc_EC
Intestine EC (TAC1+)

Figure 3.12: Transcriptional comparison of *in vitro* stem cell-derived endocrine cells with their endogenous *in vivo* counterparts.

(**A**) Cartoon representing the stages of beta cell differentiation *in vitro* from human pluripotent stem cells (SCs). (**B**) Cartoon depicting the workflow for classifying *in vitro* derived endocrine cells (**C**) Classification of *in vitro* SC-derived endocrine cells generated during Stage 5 of the differentiation protocol (23), using our human fetal pancreatic endocrine dataset and a human fetal intestinal enteroendocrine dataset (247) as a reference. UMAPs show annotation by Veres *et al.* (left) versus by cell-based classifier scPred (right), according to similarity to endogenous human fetal pancreatic endocrine and/or intestinal enteroendocrine cells. EC cell, enterochromaffin-like cell. EE cell, enteroendocrine cell. (**D**) Proportions of each cluster of SC-derived endocrine cells from the Veres *et al.* dataset that are annotated as intestinal or pancreatic cell types according to the cell type classifier. (**E**) Classification of the *in vitro* stem cell-derived enterochromaffin-like cell (sc_EC) cluster from the Veres *et al.* dataset, broken down according the number of days they have been cultured the number of days they have been cultured in Stage 5 media. (**F**) Heatmap showing the average gene expression of EC-associated genes among the pancreatic Pre-Beta and Beta cells, intestinal *TAC1*+ EC cells, and *in vitro* sc_EC cells. (**G**) Model representing the major cell type classifications along the sc_Beta and sc_EC lineages. (**H**) Dot plot depicting expression of all transcription factors differentially expressed between *in vitro*-derived sc_EC cells and endogenous pancreatic Beta, pancreatic Pre-Beta, and intestinal ECs.

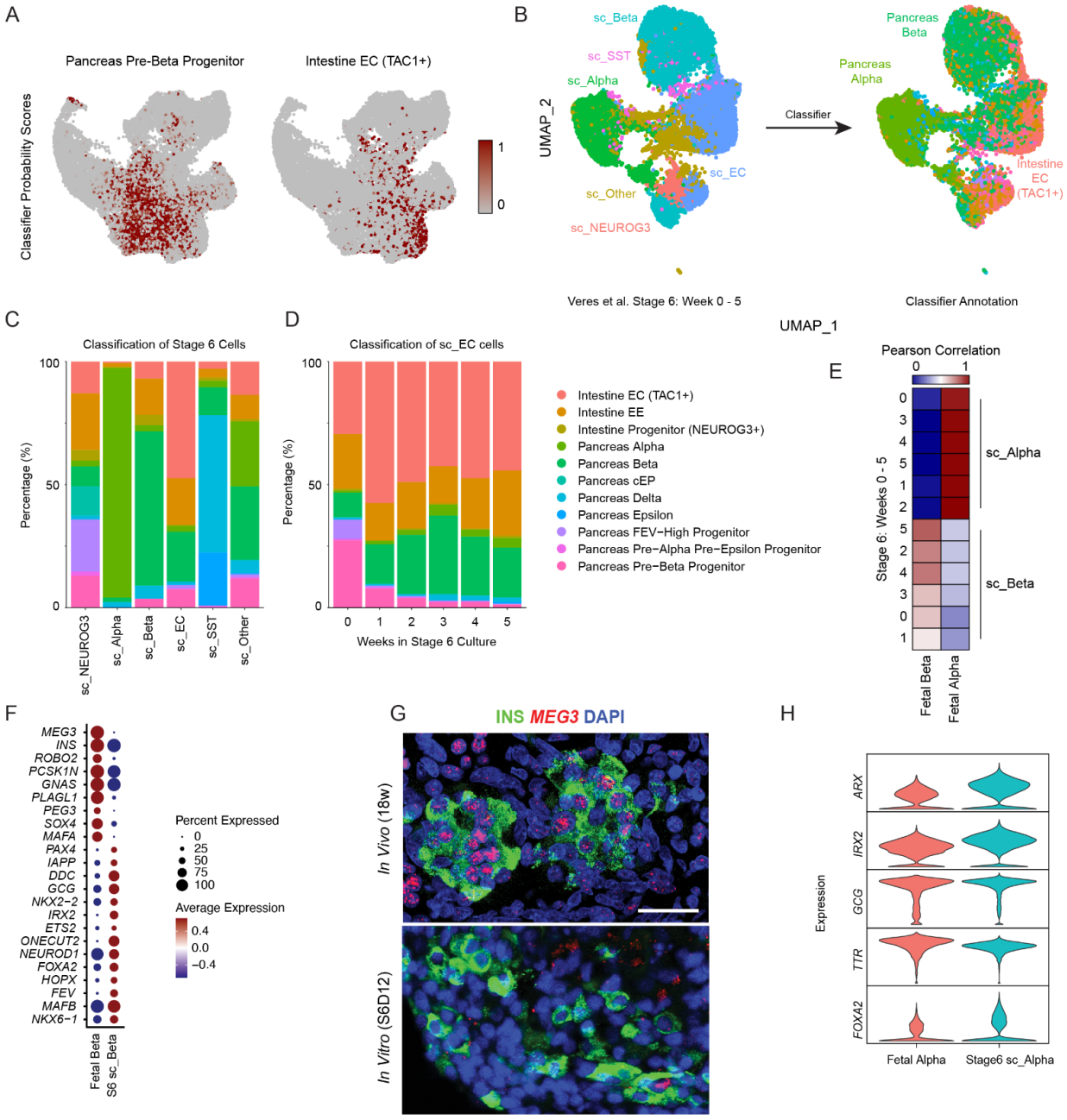

A
Pancreas Pre-Beta Progenitor
Intestine EC (TAC1+)
Classifier Probability Scores
1
0
B
sc_Beta
sc_SST
sc_Alpha
sc_EC
sc_Other
sc_NEUROG3
UMAP_2
Classifier
Pancreas Beta
Pancreas Alpha
Intestine EC (TAC1+)
Veres et al. Stage 6: Week 0 - 5
Classifier Annotation
C
Classification of Stage 6 Cells
Percentage (%)
100
50
0
sc_NEUROG3
sc_Alpha
sc_Beta
sc_EC
sc_SST
sc_Other
D
Classification of sc_EC cells
Weeks in Stage 6 Culture
UMAP_1
Intestine EC (TAC1+)
Intestine EE
Intestine Progenitor (NEUROG3+)
Pancreas Alpha
Pancreas Beta
Pancreas cEP
Pancreas Delta
Pancreas Epsilon
Pancreas FEV−High Progenitor
Pancreas Pre−Alpha Pre−Epsilon Progenitor
Pancreas Pre−Beta Progenitor
E
Pearson Correlation
Stage 6: Weeks 0 - 5
sc_Alpha
sc_Beta
Fetal Beta
Fetal Alpha
F
MEG3
INS
ROBO2
PCSK1N
GNAS
PLAGL1
PEG3
SOX4
MAFA
PAX4
IAPP
DDC
GCG
NKX2−2
IRX2
ETS2
ONECUT2
NEUROD1
FOXA2
HOPX
FEV
MAFB
NKX6−1
Fetal Beta
S6 sc_Beta
Percent Expressed
Average Expression
0.4
0.0
−0.4
G
INS MEG3 DAPI
In Vivo (18w)
In Vitro (S6D12)
H
ARX
IRX2
GCG
TTR
FOXA2
Expression
Fetal Alpha
Stage6 sc_Alpha

Figure 3.13: Classification of endocrine cells generated *in vitro*.

(**A**) Feature plots of *in vitro* stem cell (SC)-derived Stage 5 endocrine cells (**Figure 3.12c**). Classifier probability score indicates the likelihood that each cell is classified as a pancreatic Pre-Beta progenitor (left) or TAC1+ intestinal enterochromaffin (EC) cell (right). (**B**) UMAP of cells generated at the final (beta cell) Stage 6 of the stem cell differentiation protocol, with annotation by Veres *et al.* (left) or by cell-based classifier scPred (right). (**C**) Proportions of each SC-derived cluster from Stage 6 of the Veres *et al.* dataset that are annotated as either pancreatic or intestinal cells by the cell type classifier. (**D**) Proportions of the *in vitro* sc_EC cluster from the Stage 6 dataset that are annotated by the cell type classifier as pancreatic or intestinal cells, broken down by the number of weeks (0 to 5) in culture with Stage 6 media. (**E**) Heatmap depicting the Pearson correlation between cells generated at Stage 6 *in vitro* and fetal alpha or beta clusters, based on expression of genes that are highly variable in both fetal and *in vitro* datasets. Both fetal beta and fetal alpha cells are labeled according to the number of weeks that they were cultured in Stage 6 media. (**F**) Dot plot showing selected genes differentially expressed between fetal beta cells and Stage 6 SC-derived beta-like cells. (**G**) Dual *in situ* hybridization/immunofluorescence staining of 18 w human fetal pancreas tissue (top) and stem cell-derived clusters (Stage 6, day 12) (bottom). Cells were stained to detect *MEG3* mRNA and INS protein, with DAPI staining nuclei in blue. Scale bar, 25 μm. (**H**) Violin plots comparing the expression of known regulators of alpha cell fate in endogenous fetal alpha cells vs. in SC-derived alpha-like cells generated *in vitro* during Stage 6 of the differentiation.

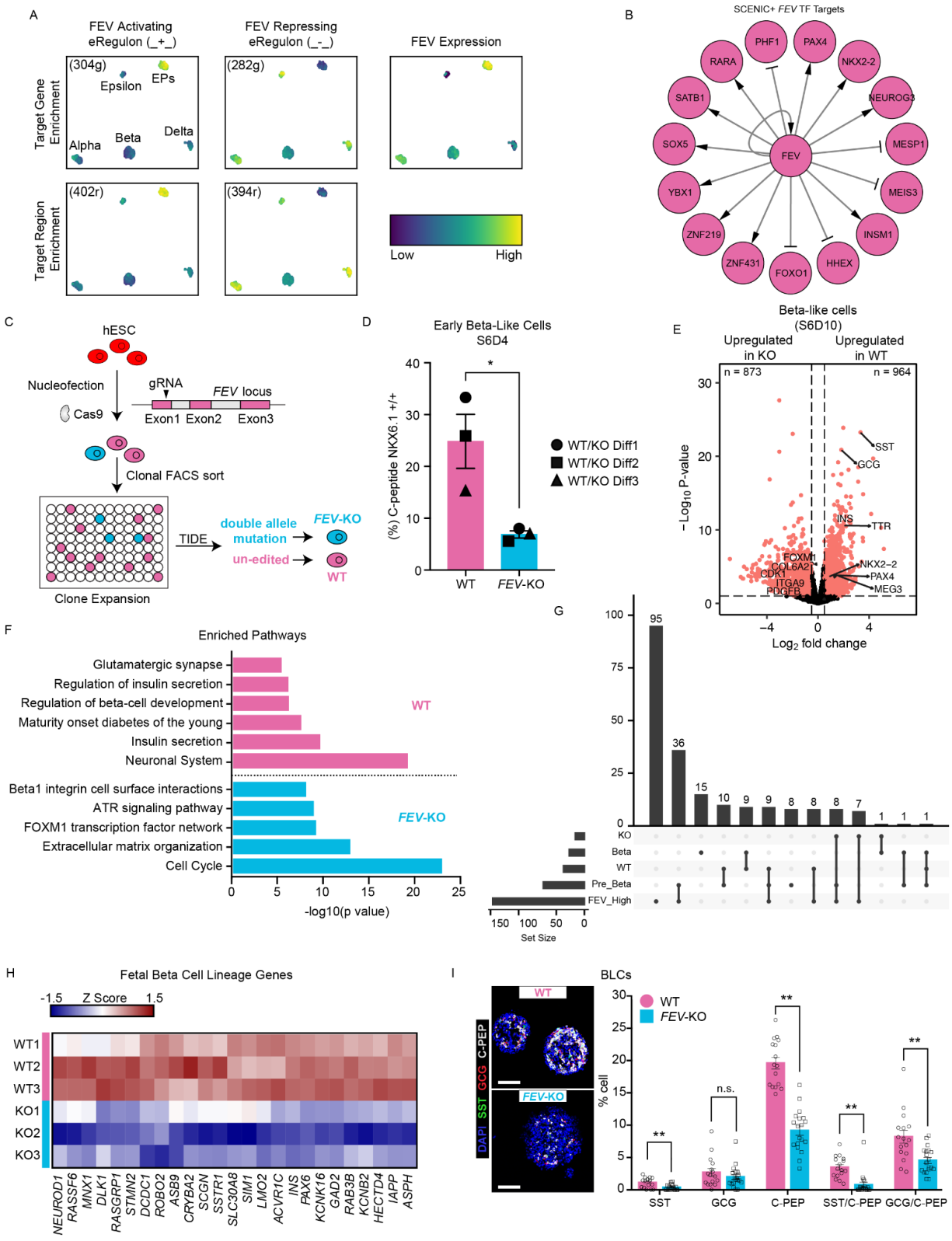

A
FEV Activating eRegulon (_+_)
FEV Repressing eRegulon (_-_)
FEV Expression
Target Gene Enrichment
Target Region Enrichment
(304g)
Epsilon
EPs
Alpha
Beta
Delta
(282g)
(402r)
(394r)
Low
High
B
SCENIC+ FEV TF Targets
PHF1
PAX4
RARA
NKX2-2
SATB1
NEUROG3
SOX5
FEV
MESP1
YBX1
MEIS3
ZNF219
INSM1
ZNF431
FOXO1
HHEX
C
hESC
Nucleofection
gRNA
FEV locus
Cas9
Exon1 Exon2 Exon3
Clonal FACS sort
TIDE
double allele mutation
FEV-KO
un-edited
WT
Clone Expansion
D
Early Beta-Like Cells S6D4
(%) C-peptide NKX6.1 +/+
WT
FEV-KO
WT/KO Diff1
WT/KO Diff2
WT/KO Diff3
E
Beta-like cells (S6D10)
Upregulated in KO
Upregulated in WT
n = 873
n = 964
−Log10 P-value
Log2 fold change
SST
GCG
INS
TTR
FOXM1
COL6A2
CDK1
ITGA9
PDGFB
NKX2−2
PAX4
MEG3
F
Enriched Pathways
Glutamatergic synapse
Regulation of insulin secretion
Regulation of beta-cell development
Maturity onset diabetes of the young
Insulin secretion
Neuronal System
Beta1 integrin cell surface interactions
ATR signaling pathway
FOXM1 transcription factor network
Extracellular matrix organization
Cell Cycle
WT
FEV-KO
-log10(p value)
G
KO
Beta
WT
Pre_Beta
FEV_High
Set Size
H
Fetal Beta Cell Lineage Genes
Z Score
WT1
WT2
WT3
KO1
KO2
KO3
NEUROD1
RASSF6
MNX1
DLK1
RASGRP1
STMN2
DCDC1
ROBO2
ASB9
CRYBA2
SCGN
SSTR1
SLC30A8
SIM1
LMO2
ACVR1C
INS
PAX6
KCNK16
GAD2
RAB3B
KCNB2
HECTD4
IAPP
ASPH
I
BLCs
WT
FEV-KO
DAPI SST GCG C-PEP
% cell
n.s.
SST
GCG
C-PEP
SST/C-PEP
GCG/C-PEP

Figure 3.14: Loss of *FEV* diminishes differentiation of human stem cells to beta-like cells *in vitro*.

(**A**) UMAPs displaying *FEV* expression (left column), *FEV* target gene enrichment (middle column) and *FEV* target region enrichment (right column) in the SCENIC+ endocrine dataset. Cluster labels are as annotated in **Figure 3.10d**. (**B**) Network diagram of *FEV* targets in the endocrine GRN. (**C**) Schematic for the generation of a *FEV* knockout (KO) human embryonic stem cell (hESC) line using CRISPR-Cas9 mediated gene editing. A WT (non-edited) line was used as a control. (**D**) Quantification of flow cytometry data from three independent paired differentiations of FEV-KO vs. WT control hESCs towards the early beta-like cell (BLC) stage (Stage 6, day 4). Efficiency of generating beta cells was quantified as percent cells staining positive for both C-PEPTIDE and NKX6.1. Data are presented as mean ± SEM (n = 3 independent batches of paired differentiations). *, p-value < 0.05, paired t-test. (**E**) Volcano plot depicting the genes differentially expressed between the FEV-KO vs. WT BLCs at Stage 6, day 10 of the directed differentiation, as assessed by bulk RNA-Sequencing. Red dots depict genes with a log2FC of at least 0.5. (**F**) Pathway analysis of the genes differentially expressed between FEV-KO vs. WT cells at Stage 6, day 10. Pathways deemed enriched in WT cells are shown in pink, and those enriched in FEV-KO cells are shown in blue. (**G**) Up-set plot showing intersection of differentially-expressed genes (DEGs) (defined as log2 FC at least 0.5) specific to any one of the fetal beta lineage cell populations (*i.e.,* FEV High, Pre-Beta, or Beta), along with DEGs identified in a comparison of WT vs. FEV-KO BLCs. (**H**) Heatmap depicting the expression levels of representative fetal beta lineage genes in paired differentiations of FEV-KO and WT cells. (**I**) Representative staining of WT and FEV-KO BLCs (Stage 6, day 12) for SST (green), GCG (red), C-PEP (gray), and DAPI (blue) (left panels). Scale bars: 100 μm. Right panel: Quantification of aggregate immunofluorescence staining data across two batches of differentiation (N=1,770 WT cells, N=2,300 FEV-KO cells). Each dot represents the cell ratio quantified per cluster. Graphs are represented as mean ± SEM; n.s., not significant; ** p-value < 0.01; unpaired t-test.

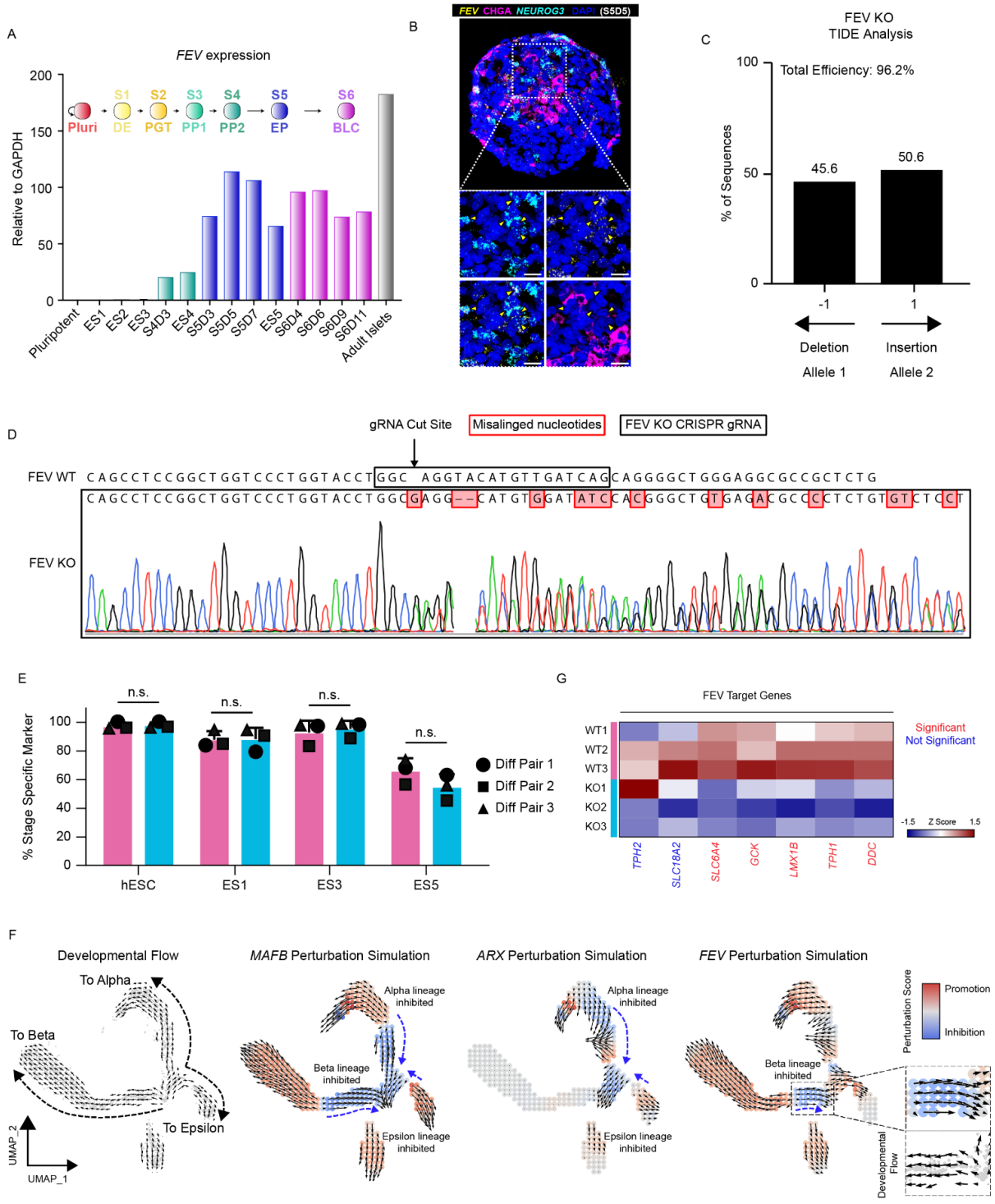

A
FEV expression
Relative to GAPDH
Pluri
S1 DE
S2 PGT
S3 PP1
S4 PP2
S5 EP
S6 BLC
Pluripotent
ES1
ES2
ES3
S4D3
ES4
S5D3
S5D5
S5D7
ES5
S6D4
S6D6
S6D9
S6D11
Adult Islets
B
FEV CHGA NEUROG3 DAPI (S5D5)
C
FEV KO
TIDE Analysis
Total Efficiency: 96.2%
% of Sequences
45.6
50.6
Deletion
Allele 1
Insertion
Allele 2
D
gRNA Cut Site
Misalinged nucleotides
FEV KO CRISPR gRNA
FEV WT
FEV KO
E
% Stage Specific Marker
n.s.
Diff Pair 1
Diff Pair 2
Diff Pair 3
hESC
ES1
ES3
ES5
G
FEV Target Genes
WT1
WT2
WT3
KO1
KO2
KO3
Significant
Not Significant
Z Score
TPH2
SLC18A2
SLC6A4
GCK
LMX1B
TPH1
DDC
F
Developmental Flow
To Alpha
To Beta
To Epsilon
UMAP_2
UMAP_1
MAFB Perturbation Simulation
Alpha lineage inhibited
Beta lineage inhibited
Epsilon lineage inhibited
ARX Perturbation Simulation
FEV Perturbation Simulation
Perturbation Score
Promotion
Inhibition
Developmental Flow

Figure 3.15: *FEV* marks endocrine progenitor cells *in vitro*, and its loss results in impaired endocrine development.

(**A**) Beta-like cells (BLCs) were generated from pluripotent (Pluri) human embryonic stem cells (hESCs) using a six-stage differentiation protocol. *FEV* mRNA expression was measured throughout the differentiation by qRT-PCR Taqman analysis. Adult human islets are included as control. (**B**) Representative dual *in situ* hybridization (ISH) and immunofluorescence (IF) staining of a cluster of hESC-derived endocrine stage cells (Stage 5, day 5) to detect *FEV* mRNA (yellow) and *NEUROG3* mRNA (cyan), along with CHGA protein (magenta). Nuclei were counterstained with DAPI (blue). Scale bar: 25 μm. Yellow arrowheads indicate FEV+/NEUROG3+ double-positive putative endocrine progenitors. (**C**) Tracking of Indels by DEcompositon (TIDE) analysis reveals efficiency of indels (insertions or deletions) in both alleles of the FEV locus in the FEV-KO line. (**D**) Sanger sequencing of the region of the *FEV* locus flanking the CRISPR gRNA cut site (outlined in black) for both FEV-KO and WT hESCs. KO sequence was aligned to WT, and misaligned nucleotides are outlined in red. (**E**) Quantification of flow cytometry data from three independent paired differentiations of FEV-KO vs. un-edited WT control hESCs at various differentiation stages towards the BLC stage. Data are presented as mean ± SEM. n.s. not significant, paired t-test. hESC represents cells in a pluripotent stage, ES1 (end of Stage 1) represents cells in a definitive endoderm stage, ES3 (end of Stage 3) represents cells in an early pancreatic progenitor stage and ES5 (end of Stage 5) represents cells after endocrine progenitor specification. (**F**) UMAP embeddings of CellOracle *in silico* transcription factor perturbation results. (**G**) Heatmap showing gene expression levels of known targets of *FEV* in FEV-KO and WT cells at the BLC stage (Stage 6, day 10). Target genes with a log2FC of at least 0.5 are deemed significantly differentially expressed (red).

Chapter 4: Conclusions and Future Directions

4.1 Overview of findings

In this work, we have constructed a roadmap of human and mouse pancreatic development by generating multi-omic, single-cell transcriptomic and epigenetic data of known and novel cell types across developmental time. In Chapter 2, we detailed our analysis of murine pancreatic endocrine and mesenchymal development, using snATAC-Seq data to construct gene regulatory networks and identify critical transcription factors and downstream target genes that potentially govern cell fate decisions in these populations. Together, our work will serve the field as a foundation for improved understanding of pancreatic development and of how to better recapitulate it *in vitro*. In Chapter 3, we detailed our analysis on human fetal pancreas development, identifying novel cellular heterogeneity and transcriptional control of cell fate decisions. By performing *in silico* lineage reconstruction analysis, we have expanded the field's understanding of how an endocrine progenitor cell differentiates into a hormone-expressing endocrine cell and identified candidate regulators of this process by constructing the gene regulatory networks operant in each fetal endocrine cell type. Using epigenomic data, we identified type I and type II diabetes genome wide association study (GWAS) single nucleotide polymorphisms (SNPs) that are enriched in developing endocrine cells compared to their adult counterparts, potentially identifying GWAS SNPs that exert developmental-specific effects. Furthermore, we have compared endogenous human cells undergoing endocrinogenesis *in vivo* to stem cell-derived cells undergoing directed differentiation *in vitro*, thereby classifying *in vitro* endocrine cells according to their *in vivo* counterparts and identifying discrepancies in gene expression between the two. Lastly, we have identified a role for the transcription factor *FEV* in human beta cell differentiation.

4.2 Cellular heterogeneity and transcriptional control within the non-endocrine compartments of the developing human and murine pancreas

Although this work largely focuses on the developing human and murine endocrine cells of the pancreas, the importance of non-endocrine cell types cannot be ignored. Using scRNA-Seq, we identified significant heterogeneity in the non-endocrine compartments of the human fetal pancreas. While some of these cell types had previously been described during pancreatic development (i.e., arterial and venous endothelial cells), other compartments had little to no previous knowledge of cellular heterogeneity and transcriptional regulation among these cell types. This included the mesenchymal cells of the pancreas, which have been described as consisting of only a few cell types (i.e., activated/inactivated stellate cells, vascular-associated mesenchymal cells) by the time of adulthood (267), but are described here as consisting of at least 15 distinct cell types. It is still not understood when or how these diverse mesenchymal cell types are resolved into the apparently more generic stellate cells by adulthood. This would be difficult to study in human fetal development, as fetal pancreas tissue past 20 w is difficult to obtain. One alternative approach would be to use less conventional animal models such a pig, which is thought to develop more similarly to human fetal pancreas and to follow more similar gestational timing to humans than mice (268,269). One could then track the emergence and disappearance of mesenchymal cell types through the use of scRNA-Seq, sampling cells across developmental time. Once an atlas of mesenchymal cells is identified, Cre-mediated lineage tracing could then be used to determine which mesenchymal cell type(s) ultimately differentiates or resolves to the adult pancreatic stellate cell. Similarly, we also detected additional neuronal, immune, and exocrine populations during human development that also warrant further study. Understanding the cellular changes across human pancreas development in these non-endocrine cell types may inform the field as to how these populations may drive endocrine differentiation and development through direct or indirect ways. Isolated human fetal

mesenchymal cells have already been shown to increase glucose-mediated insulin secretion of stem cell-derived beta cells *in vitro* (270), although this effect was not narrowed down to any specific mesenchymal cell type. Our atlas of non-endocrine cell types in the fetal pancreas could be used to discover ways to isolate different cell types, i.e. using expressed surface markers, and could be then be co-cultured with stem cell-derived beta cells to assess any changes in functional maturation. Our single-cell datasets could also be used to identify and test potential ligand-receptor pairs, as done with our CellChat analysis.

In the murine pancreas, we integrated scRNA-Seq data previously published by the Sneddon lab with novel snATAC-Seq data to generate GRNs of the developing murine pancreatic mesenchyme. This analysis elucidated novel candidate transcription factor regulators of mesenchymal cell fate and potential downstream target chromatin and target genes of these transcription factors. Future directions could involve confirming predicted TF binding targets of some of these factors using TF-chromatin precipitation techniques such as ChIP-Seq or CUT & RUN. Although ChIP-Seq requires large cellular input, CUT & RUN has been shown to work with as few as <10,000 cells (271), increasing the feasibility of TF binding confirmation. Candidates could first be screened by the availability of high-quality antibodies for ChIP-Seq/CUT & RUN, and then by previously described importance for mesenchymal cell maintenance in other tissues and contexts. Knowledge of critical factors could aid in the generation of pluripotent stem cell-derived mesenchymal cells for analyzing heterotypic cell interactions between mesenchymal and endocrine cells of both the human and murine pancreas.

Although no formal comparison of human and mouse murine pancreas was performed in either chapter of this study, some similarities and differences between the two can be gleaned. In the mesenchymal compartment, there are *Sfrp2/SFRP2*+ clusters in both human and mouse developing pancreas, as well as vascular smooth muscle cells. Mesothelial cells were much more prominent in the mouse pancreas, likely due to differences in the anatomical structure and

dissection process between the two. In the future, comparison of mouse and human mesenchymal GRNs can help identify any species-specific gene regulation occurring during mesenchymal differentiation and maturation.

4.3 Cellular heterogeneity and lineage reconstruction in the human fetal endocrine pancreas

Identification of all endocrine progenitor populations is crucial for *in silico* lineage reconstruction analyses in the developing human pancreas. Previous work by Yu *et. al.* identified four novel EP populations (EP1-EP4) in the developing human fetal pancreas using scRNA-Seq (177). Through computational lineage reconstruction, they postulate that these EP populations represent sequential progenitor states that have different differentiation potential, where EPs 1-3 follow sequentially in the lineage. EP3 can then differentiate either to EP4 or epsilon cells, with the EP4 cluster giving rise to alpha, beta, or delta cells. In our work, we also identified four EP populations with seemingly different differentiation potencies towards the hormone+ endocrine cells. While the lineage reconstruction in Yu *et. al.* is very reminiscent of murine endocrine development, our lineage construction analysis predicts that human endocrine differentiation is distinct from mouse. These differences in predictions might be due to differences in lineage reconstruction algorithms used between Yu *et. al.* and our group; Yu *et. al.* utilized a Force Directed Layout (FDL) to draw a graph representation of their endocrine scRNA-Seq data, and then inferred endocrine lineage with Slingshot (206), which has been shown to be very accurate compared to other lineage reconstruction methods in benchmarking tests (272). Our study also utilized Slingshot to calculate pseudotime values, which were then fed into CellRank (205) to calculate a transition matrix and absorption probabilities (i.e., likelihood of one cell state, or cluster, to transition to another). Of note, Yu *et. al.* utilized mSMRT-Seq, a well-based scRNA-Seq method that allows for deeper sequencing and gene capture to calculate a gene co-expression network (GCN) of the single-cell data for variable

feature selection, which was then applied to a 10x Genomics dataset of fetal endocrine cells for data processing and lineage reconstruction. Which set of methods, and therefore lineage inference, are more robust than the other likely cannot be determined with computational tools alone and will need to include some form of *in vivo* lineage tracing. *In vivo* lineage tracing methods have been utilized on some human tissue types, inferring cellular differentiation and lineage through somatic DNA mutations (273), lentiviral-mediated expression of cellular barcodes and CRISPR (274). These methods, however, require some sort of cellular division to carry the lineage information from EP to hormone expressing cell type; as cell division is not thought to occur in endocrine differentiation, utilizing these techniques to determine fetal endocrine development may be difficult.

Of note, our clustering did not predict a pre-delta progenitor population and dimensional reduction of our data did not "connect" the delta cluster to any of the EP populations, making lineage reconstruction impossible for the delta population in our dataset. Yu *et. al.* predicted that the EP4 population serves as a common progenitor to alpha, beta, and delta cells. As stated above, *in vivo* lineage tracing would likely be needed to prove this prediction. Given the importance of delta cell signaling in glucose homeostasis (275), unraveling the lineage differentiation towards a delta cell fate is crucial for our understanding for understanding delta cell development and potential *in vitro* delta cell directed differentiation protocols.

Numerous studies have reported the heterogeneity in hormone-expressing pancreatic cells, particularly in the beta cell. These include beta cells that display functional heterogeneity, such as so-called hub cells (276), "extreme" beta cells (277), and virgin beta cells (278). Other studies have reported beta cells that are transcriptionally heterogenous using scRNA-Seq, either during homeostasis or in a disease state (57,277). Notably, a recent publication investigating homeostatic beta cell heterogeneity identified through scRNA-Seq found that very few of the features that drive this heterogeneity are shared across multiple scRNA-Seq datasets, calling into question the validity of these findings (259). This result highlights the

importance of verifying *in silico* findings *in vivo* with protein labeling or *in situ* hybridization. Our dataset, as well as the scRNA-Seq dataset from Yu *et. al.* did not contain any heterogeneity in the hormone-expressing populations, possibly because they do not exist during fetal development.

4.4 Transcriptional control and epigenomics of endocrine development

Gene expression is largely governed by transcription factors that control the expression of downstream target genes by binding to regulatory genomic regions. These regions, commonly called *cis* regulatory elements (CREs), serve as enhancers or silencers, activating or repressing gene expression, respectively. This is usually achieved through the binding of TFs to these sites through the recognition of TF DNA binding motifs and downstream recruitment of other cofactors to either enhance or repress the expression of their associated gene (279). Therefore, understanding the general chromatin landscape of particular cell types and which transcription factors bind to these regulatory regions is critical for gaining a deeper understanding of cellular processes such as cell fate decisions and functional maturation. In our work, we utilized snATAC-Seq to understand the epigenomic landscape of endocrinogenesis in both mouse and human development. Through computational tools, we combined both chromatin accessibility and gene expression to construct GRNs of both human and mouse endocrine cells, allowing us to infer important regulators of endocrine differentiation. When comparing human and mouse GRNs, we can see shared TFs between the two, identifying conserved features between the two species. For instance, beta cells share known TFs *Mnx1/MNX1, Nkx6-1/NKX6-1, Pax6/PAX6*, and *Pdx1/PDX1.* Because these GRN analyses were not performed with the same computational methods, identifying TFs that are present in one species and not the other is difficult, as the difference being due to true biology and not computational methods cannot be definitely answered. While powerful, these analyses are still

computational predictions and therefore need to be validated through the use of TF binding assays such as ChIP-Seq or CUT & RUN.

4.5 Comparison of *in vivo* and *in vitro* human endocrinogenesis

Current stem cell-derived beta cell differentiation protocols suffer from a lack of efficiency in generating pure populations of beta cells, as well as functional immaturity of the beta cells produced. These hurdles might be due to missteps in cell fate decisions and precocious gene expression. Therefore, understanding *in vivo* human endocrinogenesis is essential for the improvement of *in vitro* stem cell differentiation protocols. A previous study identified a population of cells within *in vitro* beta like-cell clusters that resemble enterochromaffin cells found within the intestine (dubbed sc-EC cells) (23). These cells comprised a significant proportion of the total endocrine cell population and express known markers of enterochromaffin cells such as *TPH1, DDC, SLC18A1, LMX1A, ADRA2A, TAC1, CXCL14*, and *FEV*. Interestingly, another recent study postulates that these sc-EC cells are not an erroneous cell type, but rather pre-beta cell progenitors that are present during human *in vivo* endocrine development (280). To address this discrepancy in the literature, we compared stage 5 (endocrine progenitor) and stage 6 (beta-like cell) scRNA-Seq datasets against our fetal endocrine pancreas dataset, as well as a recently published scRNA-Seq dataset of enteroendocrine cells from the fetal intestine (247). We found that the sc-EC cells do in fact more closely resemble intestinal EC cells and not any pre-beta endocrine progenitors, and therefore are likely an off-target, non-pancreatic population that arises during *in vitro* differentiation. This is not surprising, given that the transcriptional programs that govern endocrine development in both the intestine and pancreas are very similar (249); sc-EC cells likely arise from small changes in gene expression that push the endocrine progenitors to an intestinal vs. pancreas cell fate. By comparing the transcriptional profiles of *in vivo* and *in vitro* endocrine differentiation, we hope to have identified candidate genes and transcription factors

whose expression can be modulated to not only increase the efficiency of the generation of beta-like cells, but also improve their functionality. Indeed, with the advent of multi-omic single-cell technologies, this approach is already being applied (155,280).

4.6 Role of *FEV* in endocrine development

The transcription factor *Fev*, also known as *Pet1* in mice, has been described as a master regulator of serotonergic neuron differentiation and function (143,281). *Fev* has also been investigated in the context of murine pancreas function, where adult knockout mice were found to have an impairment of glucose clearance due to a decrease in insulin content in beta cells (144). Indeed, ChIP-Seq studies of *Fev* in a beta cell related cell line revealed that *Fev* binds directly to the *Insulin* promoter and regulates its expression. Importantly, global loss of Fev did not lead to any defects in differentiated hormone+ cell number, indicating that while Fev does regulate insulin expression, it is not a critical regulator of beta cell differentiation (144). This result is consistent with our GRN analysis of the developing endocrine cell of the murine pancreas, where we predict that *Fev* binds to genes related to synaptic vesicle formation and pancreatic secretion.

Our work in human endocrine development points to potential differences between human and mouse Fev. While *Fev* marks progenitor populations in both human and mouse endocrine cells, the differentiation capacity of *Fev*-expressing progenitors is different between the two species. Furthermore, while *Fev* is seemingly dispensable for endocrine specification in the mouse pancreas, we observed a significant reduction in the number of CPEP+/NKX6.1+ cells in our FEV knockout hESC line. The predicted targets of *FEV* in our human GRN included important endocrine-related genes such as *PAX4, NEUROG3, NKX2-2, INSM1*, and *MAFA/MAFB*, indicating that *FEV* might play a more critical role in cell allocation in human development. Confirmation of *Fev/FEV* GRN targets with ChIP-Seq or CUT & RUN would greatly aid in elucidating the phenotypic effect of *Fev/FEV* KO in the murine and human

pancreas, and further highlights the need to study both mouse and human pancreatic development.

References

1. Tokarz VL, MacDonald PE, Klip A. The cell biology of systemic insulin function. J Cell Biol. 2018 Jul 2;217(7):2273–89.
2. Tomic D, Shaw JE, Magliano DJ. The burden and risks of emerging complications of diabetes mellitus. Nat Rev Endocrinol. 2022 Sep;18(9):525–39.
3. Yoon J-W, Jun H-S. Autoimmune destruction of pancreatic beta cells. Am J Ther. 2005 Dec;12(6):580–91.
4. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers. 2015 Jul 23;1:15019.
5. Akturk HK, Rompicherla S, Rioles N, Desimone M, Weinstock RS, Haw SJ, et al. Factors associated with improved A1C among adults with type 1 diabetes in the united states. Clin Diabetes. 2022 Oct 17;41(1):76–80.
6. Kontoangelos K, Raptis A, Lambadiari V, Economou M, Tsiori S, Katsi V, et al. Burnout related to diabetes mellitus: A critical analysis. Clin Pract Epidemiol Ment Health. 2022 Oct 21;18:e174501792209010.
7. Brown SA, Kovatchev BP, Raghinaru D, Lum JW, Buckingham BA, Kudva YC, et al. Six-Month Randomized, Multicenter Trial of Closed-Loop Control in Type 1 Diabetes. N Engl J Med. 2019 Oct 31;381(18):1707–17.
8. Ware J, Hovorka R. Closed-loop insulin delivery: update on the state of the field and emerging technologies. Expert Rev Med Devices. 2022 Nov 4;19(11):859–75.
9. Cartwright A, Wallymahmed M, Macfarlane IA, Wallymahmed A, Williams G, Gill GV. The outcome of brittle type 1 diabetes--a 20 year study. QJM. 2011 Jul;104(7):575–9.
10. Kelly WD, Lillehei RC, Merkel FK, Idezuki Y, Goetz FC. Allotransplantation of the

pancreas and duodenum along with the kidney in diabetic nephropathy. Surgery. 1967 Jun;61(6):827–37.

11. Linetsky E, Bottino R, Lehmann R, Alejandro R, Inverardi L, Ricordi C. Improved human islet isolation using a new enzyme blend, liberase. Diabetes. 1997 Jul;46(7):1120–3.

12. Lakey JR, Warnock GL, Shapiro AM, Korbutt GS, Ao Z, Kneteman NM, et al. Intraductal collagenase delivery into the human pancreas using syringe loading or controlled perfusion. Cell Transplant. 1999;8(3):285–92.

13. Shapiro AM, Lakey JR, Ryan EA, Korbutt GS, Toth E, Warnock GL, et al. Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. N Engl J Med. 2000 Jul 27;343(4):230–8.

14. Biarnés M, Montolio M, Nacher V, Raurell M, Soler J, Montanya E. Beta-cell death and mass in syngeneically transplanted islets exposed to short- and long-term hyperglycemia. Diabetes. 2002 Jan;51(1):66–72.

15. Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature. 2008 Oct 2;455(7213):627–32.

16. Banga A, Akinci E, Greder LV, Dutton JR, Slack JMW. In vivo reprogramming of Sox9+ cells in the liver to insulin-secreting ducts. Proc Natl Acad Sci USA. 2012 Sep 18;109(38):15336–41.

17. Chen Y-J, Finkbeiner SR, Weinblatt D, Emmett MJ, Tameire F, Yousefi M, et al. De novo formation of insulin-producing “neo-β cell islets” from intestinal crypts. Cell Rep. 2014 Mar 27;6(6):1046–58.

18. Moede T, Leibiger IB, Berggren P-O. Alpha cell regulation of beta cell function.

Diabetologia. 2020 Oct;63(10):2064–75.

19. Furuyama K, Chera S, van Gurp L, Oropeza D, Ghila L, Damond N, et al. Diabetes relief in mice by glucose-sensing insulin-secreting human α-cells. Nature. 2019 Mar;567(7746):43–8.

20. Anson DS. The use of retroviral vectors for gene therapy-what are the risks? A review of retroviral pathogenesis and its relevance to retroviral vector-mediated gene delivery. Genet Vaccines Ther. 2004 Aug 13;2(1):9.

21. Nair GG, Tzanakakis ES, Hebrok M. Emerging routes to the generation of functional β-cells for diabetes mellitus cell therapy. Nat Rev Endocrinol. 2020 Sep;16(9):506–18.

22. Pagliuca FW, Millman JR, Gürtler M, Segel M, Van Dervort A, Ryu JH, et al. Generation of functional human pancreatic β cells in vitro. Cell. 2014 Oct 9;159(2):428–39.

23. Veres A, Faust AL, Bushnell HL, Engquist EN, Kenty JH-R, Harb G, et al. Charting cellular identity during human in vitro β-cell differentiation. Nature. 2019 May 8;569(7756):368–73.

24. Rezania A, Bruin JE, Arora P, Rubin A, Batushansky I, Asadi A, et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat Biotechnol. 2014 Nov;32(11):1121–33.

25. Velazco-Cruz L, Goedegebuure MM, Millman JR. Advances Toward Engineering Functionally Mature Human Pluripotent Stem Cell-Derived β Cells. Front Bioeng Biotechnol. 2020 Jul 9;8:786.

26. Kim SK, Hebrok M, Melton DA. Notochord to endoderm signaling is required for pancreas development. Development. 1997 Nov;124(21):4243–52.

27. Hebrok M, Kim SK, Melton DA. Notochord repression of endodermal Sonic hedgehog

permits pancreas development. Genes Dev. 1998 Jun 1;12(11):1705–13.

28. Deutsch G, Jung J, Zheng M, Lóra J, Zaret KS. A bipotential precursor population for pancreas and liver within the embryonic endoderm. Development. 2001 Mar;128(6):871–81.

29. Gu G, Dubauskaite J, Melton DA. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development. 2002 May;129(10):2447–57.

30. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, et al. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development. 1996 Mar;122(3):983–95.

31. Jonsson J, Carlsson L, Edlund T, Edlund H. Insulin-promoter-factor 1 is required for pancreas development in mice. Nature. 1994 Oct 13;371(6498):606–9.

32. Fishman MP, Melton DA. Pancreatic lineage analysis using a retroviral vector in embryonic mice demonstrates a common progenitor for endocrine and exocrine cells. Int J Dev Biol. 2002 Mar;46(2):201–7.

33. Herrera PL. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. Development. 2000 Jun;127(11):2317–22.

34. Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA. A multipotent progenitor domain guides pancreatic organogenesis. Dev Cell. 2007 Jul;13(1):103–14.

35. Villasenor A, Chong DC, Cleaver O. Biphasic Ngn3 expression in the developing pancreas. Dev Dyn. 2008 Nov;237(11):3270–9.

36. Kesavan G, Sand FW, Greiner TU, Johansson JK, Kobberup S, Wu X, et al. Cdc42-mediated tubulogenesis controls cell specification. Cell. 2009 Nov 13;139(4):791–801.

37. Villasenor A, Chong DC, Henkemeyer M, Cleaver O. Epithelial dynamics of pancreatic branching morphogenesis. Development. 2010 Dec;137(24):4295–305.

38. Seymour PA, Freude KK, Tran MN, Mayes EE, Jensen J, Kist R, et al. SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc Natl Acad Sci USA. 2007 Feb 6;104(6):1865–70.

39. Schaffer AE, Freude KK, Nelson SB, Sander M. Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. Dev Cell. 2010 Jun 15;18(6):1022–9.

40. Solar M, Cardalda C, Houbracken I, Martín M, Maestro MA, De Medts N, et al. Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. Dev Cell. 2009 Dec;17(6):849–60.

41. Pictet RL, Clark WR, Williams RH, Rutter WJ. An ultrastructural analysis of the developing embryonic pancreas. Dev Biol. 1972 Dec;29(4):436–67.

42. Desai BM, Oliver-Krasinski J, De Leon DD, Farzad C, Hong N, Leach SD, et al. Preexisting pancreatic acinar cells contribute to acinar cell, but not islet beta cell, regeneration. J Clin Invest. 2007 Apr;117(4):971–7.

43. Wollny D, Zhao S, Everlien I, Lun X, Brunken J, Brüne D, et al. Single-Cell Analysis Uncovers Clonal Acinar Cell Heterogeneity in the Adult Pancreas. Dev Cell. 2016 Nov 7;39(3):289–301.

44. Shih HP, Kopp JL, Sandhu M, Dubois CL, Seymour PA, Grapin-Botton A, et al. A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. Development. 2012 Jul;139(14):2488–99.

45. Esni F, Ghosh B, Biankin AV, Lin JW, Albert MA, Yu X, et al. Notch inhibits Ptf1

function and acinar cell differentiation in developing mouse and zebrafish pancreas. Development. 2004 Sep;131(17):4213–24.

46. Ahnfelt-Rønne J, Hald J, Bødker A, Yassin H, Serup P, Hecksher-Sørensen J. Preservation of proliferating pancreatic progenitor cells by Delta-Notch signaling in the embryonic chicken pancreas. BMC Dev Biol. 2007 Jun 7;7:63.

47. Gradwohl G, Dierich A, LeMeur M, Guillemot F. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc Natl Acad Sci USA. 2000 Feb 15;97(4):1607–11.

48. Schwitzgebel VM, Scheel DW, Conners JR, Kalamaras J, Lee JE, Anderson DJ, et al. Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. Development. 2000 Aug;127(16):3533–42.

49. Heller RS, Jenny M, Collombat P, Mansouri A, Tomasetto C, Madsen OD, et al. Genetic determinants of pancreatic epsilon-cell development. Dev Biol. 2005 Oct 1;286(1):217–24.

50. Suissa Y, Magenheim J, Stolovich-Rain M, Hija A, Collombat P, Mansouri A, et al. Gastrin: a distinct fate of neurogenin3 positive progenitor cells in the embryonic pancreas. PLoS One. 2013 Aug 5;8(8):e70397.

51. Desgraz R, Herrera PL. Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. Development. 2009 Nov;136(21):3567–74.

52. Johansson KA, Dursun U, Jordan N, Gu G, Beermann F, Gradwohl G, et al. Temporal control of neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. Dev Cell. 2007 Mar;12(3):457–65.

53. Scavuzzo MA, Hill MC, Chmielowiec J, Yang D, Teaw J, Sheng K, et al. Endocrine

lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis. Nat Commun. 2018 Aug 22;9(1):3356.

54. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May 21;161(5):1202–14.

55. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015 May 21;161(5):1187–201.

56. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8:14049.

57. Szlachcic WJ, Ziojla N, Kizewska DK, Kempa M, Borowiak M. Endocrine Pancreas Development and Dysfunction Through the Lens of Single-Cell RNA-Sequencing. Front Cell Dev Biol. 2021 Apr 29;9:629212.

58. Byrnes LE, Wong DM, Subramaniam M, Meyer NP, Gilchrist CL, Knox SM, et al. Lineage dynamics of murine pancreatic development at single-cell resolution. Nat Commun. 2018 Sep 25;9(1):3922.

59. Krentz NAJ, Lee MYY, Xu EE, Sproul SLJ, Maslova A, Sasaki S, et al. Single-Cell Transcriptome Profiling of Mouse and hESC-Derived Pancreatic Progenitors. Stem Cell Rep. 2018 Dec 11;11(6):1551–64.

60. Bastidas-Ponce A, Tritschler S, Dony L, Scheibner K, Tarquis-Medina M, Salinno C, et al. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. Development. 2019 Jun 17;146(12).

61. van Gurp L, Muraro MJ, Dielen T, Seneby L, Dharmadhikari G, Gradwohl G, et al. A transcriptomic roadmap to α- and β-cell differentiation in the embryonic pancreas. Development. 2019 Jun 24;146(12).

62. Yu X-X, Qiu W-L, Yang L, Zhang Y, He M-Y, Li L-C, et al. Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. EMBO J. 2019 Apr 15;38(8).

63. Krueger KC, Deneris ES. Serotonergic transcription of human FEV reveals direct GATA factor interactions and fate of Pet-1-deficient serotonin neuron precursors. J Neurosci. 2008 Nov 26;28(48):12748–58.

64. Kubo A, Stull R, Takeuchi M, Bonham K, Gouon-Evans V, Sho M, et al. Pdx1 and Ngn3 overexpression enhances pancreatic differentiation of mouse ES cell-derived endoderm population. PLoS One. 2011 Sep 13;6(9):e24058.

65. Oropeza D, Horb M. Transient expression of Ngn3 in Xenopus endoderm promotes early and ectopic development of pancreatic beta and delta cells. Genesis. 2012 Mar;50(3):271–85.

66. Rukstalis JM, Habener JF. Neurogenin3: a master regulator of pancreatic islet differentiation and regeneration. Islets. 2009 Dec;1(3):177–84.

67. Huang HP, Liu M, El-Hodiri HM, Chu K, Jamrich M, Tsai MJ. Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. Mol Cell Biol. 2000 May;20(9):3292–307.

68. Smith SB, Gasa R, Watada H, Wang J, Griffen SC, German MS. Neurogenin3 and hepatic nuclear factor 1 cooperate in activating pancreatic expression of Pax4. J Biol Chem. 2003 Oct 3;278(40):38254–9.

69. Smith SB, Watada H, German MS. Neurogenin3 activates the islet differentiation program while repressing its own expression. Mol Endocrinol. 2004 Jan;18(1):142–9.

70. Ait-Lounis A, Baas D, Barras E, Benadiba C, Charollais A, Nlend Nlend R, et al. Novel function of the ciliogenic transcription factor RFX3 in development of the endocrine pancreas. Diabetes. 2007 Apr;56(4):950–9.

71. Smith SB, Qu H-Q, Taleb N, Kishimoto NY, Scheel DW, Lu Y, et al. Rfx6 directs islet formation and insulin production in mice and humans. Nature. 2010 Feb 11;463(7282):775–80.

72. Cheng C, Lu J, Cao X, Yang F-Y, Liu J-Y, Song L-N, et al. Identification of Rfx6 target genes involved in pancreas development and insulin translation by ChIP-seq. Biochem Biophys Res Commun. 2019 Jan 8;508(2):556–62.

73. Sosa-Pineda B, Chowdhury K, Torres M, Oliver G, Gruss P. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. Nature. 1997 Mar 27;386(6623):399–402.

74. Ashery-Padan R, Zhou X, Marquardt T, Herrera P, Toube L, Berry A, et al. Conditional inactivation of Pax6 in the pancreas causes early onset of diabetes. Dev Biol. 2004 May 15;269(2):479–88.

75. Jennings RE, Berry AA, Kirkwood-Wilson R, Roberts NA, Hearn T, Salisbury RJ, et al. Development of the human pancreas from foregut to endocrine commitment. Diabetes. 2013 Oct;62(10):3514–22.

76. Jennings RE, Berry AA, Strutt JP, Gerrard DT, Hanley NA. Human pancreas development. Development. 2015 Sep 15;142(18):3126–37.

77. Lyttle BM, Li J, Krishnamurthy M, Fellows F, Wheeler MB, Goodyer CG, et al.

Transcription factor expression in the developing human fetal endocrine pancreas. Diabetologia. 2008 Jul;51(7):1169–80.

78. Sarkar SA, Kobberup S, Wong R, Lopez AD, Quayum N, Still T, et al. Global gene expression profiling and histochemical analysis of the developing human fetal pancreas. Diabetologia. 2008 Feb;51(2):285–97.

79. Polak M, Bouchareb-Banaei L, Scharfmann R, Czernichow P. Early pattern of differentiation in the human pancreas. Diabetes. 2000 Feb;49(2):225–32.

80. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren P-O, Caicedo A. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci USA. 2006 Feb 14;103(7):2334–9.

81. Assady S, Maor G, Amit M, Itskovitz-Eldor J, Skorecki KL, Tzukerman M. Insulin production by human embryonic stem cells. Diabetes. 2001 Aug;50(8):1691–7.

82. Kroon E, Martinson LA, Kadoya K, Bang AG, Kelly OG, Eliazer S, et al. Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat Biotechnol. 2008 Apr;26(4):443–52.

83. D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, et al. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat Biotechnol. 2006 Nov;24(11):1392–401.

84. Rezania A, Bruin JE, Riedel MJ, Mojibian M, Asadi A, Xu J, et al. Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice. Diabetes. 2012 Aug;61(8):2016–29.

85. Nair GG, Liu JS, Russ HA, Tran S, Saxton MS, Chen R, et al. Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived β cells. Nat

Cell Biol. 2019 Feb 1;21(2):263–74.

86. Balboa D, Barsby T, Lithovius V, Saarimäki-Vire J, Omar-Hmeadi M, Dyachok O, et al. Functional, metabolic and transcriptional maturation of human pancreatic islets derived from stem cells. Nat Biotechnol. 2022 Jul;40(7):1042–55.

87. Hogrebe NJ, Maxwell KG, Augsornworawat P, Millman JR. Generation of insulin-producing pancreatic β cells from multiple human stem cell lines. Nat Protoc. 2021 Sep;16(9):4109–43.

88. Peterson QP, Veres A, Chen L, Slama MQ, Kenty JHR, Hassoun S, et al. A method for the generation of human stem cell-derived alpha cells. Nat Commun. 2020 May 7;11(1):2241.

89. Rezania A, Riedel MJ, Wideman RD, Karanu F, Ao Z, Warnock GL, et al. Production of functional glucagon-secreting α-cells from human embryonic stem cells. Diabetes. 2011 Jan;60(1):239–47.

90. Henquin J-C, Ishiyama N, Nenquin M, Ravier MA, Jonas J-C. Signals and pools underlying biphasic insulin secretion. Diabetes. 2002 Feb;51 Suppl 1:S60-7.

91. Zhu H, Wang G, Nguyen-Ngoc K-V, Kim D, Miller M, Goss G, et al. Understanding cell fate acquisition in stem-cell-derived pancreatic islets using single-cell multiome-inferred regulomes. Dev Cell. 2023 May 8;58(9):727–743.e11.

92. Guz Y, Montminy MR, Stein R, Leonard J, Gamer LW, Wright CV, et al. Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. Development. 1995 Jan;121(1):11–8.

93. Ohlsson H, Karlsson K, Edlund T. IPF1, a homeodomain-containing transactivator of the

insulin gene. EMBO J. 1993 Nov;12(11):4251–9.

94. van der Meulen T, Huising MO. Role of transcription factors in the transdifferentiation of pancreatic islet cells. J Mol Endocrinol. 2015 Apr;54(2):R103-17.

95. Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev. 1998 Jun 15;12(12):1763–8.

96. Yang Y-P, Thorel F, Boyer DF, Herrera PL, Wright CVE. Context-specific α- to-β-cell reprogramming by forced Pdx1 expression. Genes Dev. 2011 Aug 15;25(16):1680–5.

97. Collombat P, Mansouri A, Hecksher-Sorensen J, Serup P, Krull J, Gradwohl G, et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev. 2003 Oct 15;17(20):2591–603.

98. Collombat P, Hecksher-Sørensen J, Broccoli V, Krull J, Ponte I, Mundiger T, et al. The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the alpha- and beta-cell lineages in the mouse endocrine pancreas. Development. 2005 Jul;132(13):2969–80.

99. Prado CL, Pugh-Bernard AE, Elghazi L, Sosa-Pineda B, Sussel L. Ghrelin cells replace insulin-producing beta cells in two mouse models of pancreas development. Proc Natl Acad Sci USA. 2004 Mar 2;101(9):2924–9.

100. Sander M, Sussel L, Conners J, Scheel D, Kalamaras J, Dela Cruz F, et al. Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. Development. 2000 Dec;127(24):5533–40.

101. Sussel L, Kalamaras J, Hartigan-O'Connor DJ, Meneses JJ, Pedersen RA, Rubenstein JL, et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to

arrested differentiation of pancreatic beta cells. Development. 1998 Jun;125(12):2213–21.

102. Schaffer AE, Taylor BL, Benthuysen JR, Liu J, Thorel F, Yuan W, et al. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. PLoS Genet. 2013 Jan 31;9(1):e1003274.

103. Golosow N, Grobstein C. Epitheliomesenchymal interaction in pancreatic morphogenesis. Dev Biol. 1962 Apr;4:242–55.

104. Bhushan A, Itoh N, Kato S, Thiery JP, Czernichow P, Bellusci S, et al. Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. Development. 2001 Dec;128(24):5109–17.

105. Landsman L, Nijagal A, Whitchurch TJ, Vanderlaan RL, Zimmer WE, Mackenzie TC, et al. Pancreatic mesenchyme regulates epithelial organogenesis throughout development. PLoS Biol. 2011 Sep 6;9(9):e1001143.

106. Cozzitorto C, Mueller L, Ruzittu S, Mah N, Willnow D, Darrigrand J-F, et al. A specialized niche in the pancreatic microenvironment promotes endocrine differentiation. Dev Cell. 2020 Oct 26;55(2):150–162.e6.

107. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013 Dec;10(12):1213–8.

108. Duvall E, Benitez CM, Tellez K, Enge M, Pauerstein PT, Li L, et al. Single-cell transcriptome and accessible chromatin dynamics during endocrine pancreas development. Proc Natl Acad Sci USA. 2022 Jun 28;119(26):e2201267119.

109. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015 Jul

23;523(7561):486–90.

110. Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science. 2015 May 22;348(6237):910–4.

111. Ranzoni AM, Tangherloni A, Berest I, Riva SG, Myers B, Strzelecka PM, et al. Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. Cell Stem Cell. 2021 Mar 4;28(3):472–487.e7.

112. Duong TE, Wu Y, Sos BC, Dong W, Limaye S, Rivier LH, et al. A single-cell regulatory map of postnatal lung alveologenesis in humans and mice. Cell Genomics. 2022 Mar 9;2(3).

113. Lyu P, Hoang T, Santiago CP, Thomas ED, Timms AE, Appel H, et al. Gene regulatory networks controlling temporal patterning, neurogenesis, and cell-fate specification in mammalian retina. Cell Rep. 2021 Nov 16;37(7):109994.

114. Ziffra RS, Kim CN, Ross JM, Wilfert A, Turner TN, Haeussler M, et al. Single-cell epigenomics reveals mechanisms of human cortical development. Nature. 2021 Oct 6;598(7879):205–13.

115. Trevino AE, Müller F, Andersen J, Sundaram L, Kathiria A, Shcherbina A, et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell. 2021 Sep 16;184(19):5053–5069.e23.

116. Scott MM, Wylie CJ, Lerch JK, Murphy R, Lobur K, Herlitze S, et al. A genetic approach to access serotonin neurons for in vivo and in vitro studies. Proc Natl Acad Sci USA. 2005 Nov 8;102(45):16472–7.

117. Wyler SC, Spencer WC, Green NH, Rood BD, Crawford L, Craige C, et al. Pet-1

Switches Transcriptional Targets Postnatally to Regulate Maturation of Serotonin Neuron Excitability. J Neurosci. 2016 Feb 3;36(5):1758–74.

118. Sugiyama T, Benitez CM, Ghodasara A, Liu L, McLean GW, Lee J, et al. Reconstituting pancreas development from purified progenitor cells reveals genes essential for islet differentiation. Proc Natl Acad Sci USA. 2013 Jul 30;110(31):12691–6.

119. Mellitzer G, Martín M, Sidhoum-Jenny M, Orvain C, Barths J, Seymour PA, et al. Pancreatic islet progenitor cells in neurogenin 3-yellow fluorescent protein knock-add-on mice. Mol Endocrinol. 2004 Nov;18(11):2765–76.

120. White P, May CL, Lamounier RN, Brestelli JE, Kaestner KH. Defining pancreatic endocrine precursors and their descendants. Diabetes. 2008 Mar;57(3):654–68.

121. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet. 2021 Mar;53(3):403–11.

122. Yu KS, Frumm SM, Park JS, Lee K, Wong DM, Byrnes L, et al. Development of the mouse and human cochlea at single cell resolution. BioRxiv. 2019 Aug 20;

123. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008 Sep 17;9(9):R137.

124. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods. 2017 Oct;14(10):975–8.

125. Xu EE, Krentz NAJ, Tan S, Chow SZ, Tang M, Nian C, et al. SOX4 cooperates with neurogenin 3 to regulate endocrine pancreas formation in mouse models. Diabetologia. 2015 May;58(5):1013–23.

126. Kim SK, Selleri L, Lee JS, Zhang AY, Gu X, Jacobs Y, et al. Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus. Nat Genet. 2002 Apr;30(4):430–5.

127. Anderson KR, White P, Kaestner KH, Sussel L. Identification of known and novel pancreas genes expressed downstream of Nkx2.2 during development. BMC Dev Biol. 2009 Dec 10;9:65.

128. Suriben R, Kaihara KA, Paolino M, Reichelt M, Kummerfeld SK, Modrusan Z, et al. β-Cell Insulin Secretion Requires the Ubiquitin Ligase COP1. Cell. 2015 Dec 3;163(6):1457–67.

129. Schreiber V, Mercier R, Jiménez S, Ye T, García-Sánchez E, Klein A, et al. Extensive NEUROG3 occupancy in the human pancreatic endocrine gene regulatory network. Mol Metab. 2021 Nov;53:101313.

130. Zhu Z, Li QV, Lee K, Rosen BP, González F, Soh C-L, et al. Genome editing of lineage determinants in human pluripotent stem cells reveals mechanisms of pancreatic development and diabetes. Cell Stem Cell. 2016 Jun 2;18(6):755–68.

131. Miyatsuka T, Li Z, German MS. Chronology of islet differentiation revealed by temporal cell labeling. Diabetes. 2009 Aug;58(8):1863–8.

132. Lynn FC, Smith SB, Wilson ME, Yang KY, Nekrep N, German MS. Sox9 coordinates a transcriptional network in pancreatic progenitor cells. Proc Natl Acad Sci USA. 2007 Jun 19;104(25):10500–5.

133. Cebola I, Rodríguez-Seguí SA, Cho CH-H, Bessa J, Rovira M, Luengo M, et al. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. Nat Cell Biol. 2015 May;17(5):615–26.

134. Breslin MB, Wang H-W, Pierce A, Aucoin R, Lan MS. Neurogenin 3 recruits CBP co-activator to facilitate histone H3/H4 acetylation in the target gene INSM1. FEBS Lett. 2007 Mar 6;581(5):949–54.

135. Jia S, Ivanov A, Blasevic D, Müller T, Purfürst B, Sun W, et al. Insm1 cooperates with Neurod1 and Foxa2 to maintain mature pancreatic β-cell function. EMBO J. 2015 May 12;34(10):1417–33.

136. Apelqvist A, Li H, Sommer L, Beatus P, Anderson DJ, Honjo T, et al. Notch signalling controls pancreatic cell differentiation. Nature. 1999 Aug 26;400(6747):877–81.

137. Elhanani O, Salame TM, Sobel J, Leshkowitz D, Povodovski L, Vaknin I, et al. REST Inhibits Direct Reprogramming of Pancreatic Exocrine to Endocrine Cells by Preventing PDX1-Mediated Activation of Endocrine Genes. Cell Rep. 2020 May 5;31(5):107591.

138. Rovira M, Atla G, Maestro MA, Grau V, García-Hurtado J, Maqueda M, et al. REST is a major negative regulator of endocrine differentiation during pancreas organogenesis. Genes Dev. 2021 Sep 1;35(17–18):1229–42.

139. Martin D, Kim Y-H, Sever D, Mao C-A, Haefliger J-A, Grapin-Botton A. REST represses a subset of the pancreatic endocrine differentiation program. Dev Biol. 2015 Sep 15;405(2):316–27.

140. Takiishi T, Gysemans C, Bouillon R, Mathieu C. Vitamin D and diabetes. Endocrinol Metab Clin North Am. 2010 Jun;39(2):419–46, table of contents.

141. Chen C, Luo Y, Su Y, Teng L. The vitamin D receptor (VDR) protects pancreatic beta cells against Forkhead box class O1 (FOXO1)-induced mitochondrial dysfunction and cell apoptosis. Biomed Pharmacother. 2019 Sep;117:109170.

142. Morró M, Vilà L, Franckhauser S, Mallol C, Elias G, Ferré T, et al. Vitamin D Receptor

Overexpression in β-Cells Ameliorates Diabetes in Mice. Diabetes. 2020 May;69(5):927–39.

143. Fyodorov D, Nelson T, Deneris E. Pet-1, a novel ETS domain factor that can activate neuronal nAchR gene transcription. J Neurobiol. 1998 Feb 5;34(2):151–63.

144. Ohta Y, Kosaka Y, Kishimoto N, Wang J, Smith SB, Honig G, et al. Convergence of the insulin and serotonin programs in the pancreatic β-cell. Diabetes. 2011 Dec;60(12):3208–16.

145. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 Feb 17;12(1):1088.

146. Mutsaers SE. Mesothelial cells: their structure, function and role in serosal repair. Respirology. 2002 Sep;7(3):171–91.

147. Que J, Wilm B, Hasegawa H, Wang F, Bader D, Hogan BLM. Mesothelium contributes to vascular smooth muscle and mesenchyme during lung development. Proc Natl Acad Sci USA. 2008 Oct 28;105(43):16626–30.

148. Ariza L, Cañete A, Rojas A, Muñoz-Chápuli R, Carmona R. Role of the Wilms' tumor suppressor gene Wt1 in pancreatic development. Dev Dyn. 2018 Jul;247(7):924–33.

149. Domcke S, Hill AJ, Daza RM, Cao J, O'Day DR, Pliner HA, et al. A human cell atlas of fetal chromatin accessibility. Science. 2020 Nov 13;

150. Chung C-Y, Ma Z, Dravis C, Preissl S, Poirion O, Luna G, et al. Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships. Cell Rep. 2019 Oct 8;29(2):495–510.e6.

151. Thomas ED, Timms AE, Giles S, Harkins-Perry S, Lyu P, Hoang T, et al. Cell-specific

cis-regulatory elements and mechanisms of non-coding genetic disease in human retina and retinal organoids. Dev Cell. 2022 Mar 28;57(6):820–836.e6.

152. Simsek S, Zhou T, Robinson CL, Tsai S-Y, Crespo M, Amin S, et al. Modeling Cystic Fibrosis Using Pluripotent Stem Cell-Derived Human Pancreatic Ductal Epithelial Cells. Stem Cells Transl Med. 2016 May;5(5):572–9.

153. Delaspre F, Massumi M, Salido M, Soria B, Ravassard P, Savatier P, et al. Directed pancreatic acinar differentiation of mouse embryonic stem cells via embryonic signalling molecules and exocrine transcription factors. PLoS One. 2013 Jan 17;8(1):e54243.

154. Wilson ME, Scheel D, German MS. Gene expression cascades in pancreatic development. Mech Dev. 2003 Jan;120(1):65–80.

155. Augsornworawat P, Marquez E, Maestas MM, Ishahak M, Gale SE, Schmidt MD, et al. Multiomic profiling defines cell fate plasticity of in vitro-derived islets. BioRxiv. 2022 Feb 27;

156. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 Jul;9(7):671–5.

157. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888–1902.e21.

158. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun. 2020 Aug 26;11(1):4267.

159. Roopra A. MAGIC: A tool for predicting transcription factors and cofactors driving gene sets using ENCODE data. PLoS Comput Biol. 2020 Apr 6;16(4):e1007800.

160. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and

regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019 Nov;157:107843.

161. Melton D. The promise of stem cell-derived islet replacement therapy. Diabetologia. 2021 May;64(5):1030–6.

162. Migliorini A, Nostro MC, Sneddon JB. Human pluripotent stem cell-derived insulin-producing cells: A regenerative medicine perspective. Cell Metab. 2021 Apr 6;33(4):721–31.

163. Russ HA, Parent AV, Ringler JJ, Hennings TG, Nair GG, Shveygert M, et al. Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J. 2015 Jul 2;34(13):1759–72.

164. Nostro MC, Sarangi F, Yang C, Holland A, Elefanty AG, Stanley EG, et al. Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines. Stem Cell Rep. 2015 Apr 14;4(4):591–604.

165. Dolenšek J, Rupnik MS, Stožer A. Structural similarities and differences between the human and the mouse pancreas. Islets. 2015;7(1):e1024405.

166. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst. 2016 Oct 26;3(4):346–360.e4.

167. MacDonald MJ, Longacre MJ, Stoker SW, Kendrick M, Thonpho A, Brown LJ, et al. Differences between human and rodent pancreatic islets: low pyruvate carboxylase, atp citrate lyase, and pyruvate carboxylation and high glucose-stimulated acetoacetate in human pancreatic islets. J Biol Chem. 2011 May 27;286(21):18383–96.

168. Shih HP, Wang A, Sander M. Pancreas organogenesis: from lineage determination to morphogenesis. Annu Rev Cell Dev Biol. 2013 Jul 31;29:81–105.

169. Salisbury RJ, Blaylock J, Berry AA, Jennings RE, De Krijger R, Piper Hanley K, et al. The window period of NEUROGENIN3 during human gestation. Islets. 2014;6(3):e954436.

170. Cleaver O, Dor Y. Vascular instruction of pancreas development. Development. 2012 Aug;139(16):2833–43.

171. Lammert E, Cleaver O, Melton D. Induction of pancreatic differentiation by signals from blood vessels. Science. 2001 Oct 19;294(5542):564–7.

172. Borden P, Houtz J, Leach SD, Kuruvilla R. Sympathetic innervation during development is necessary for pancreatic islet architecture and functional maturation. Cell Rep. 2013 Jul 25;4(2):287–301.

173. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. 2016 Oct 26;3(4):385–394.e3.

174. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, Sun X, et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab. 2016 Oct 11;24(4):593–607.

175. Xin Y, Dominguez Gutierrez G, Okamoto H, Kim J, Lee A-H, Adler C, et al. Pseudotime Ordering of Single Human β-Cells Reveals States of Insulin Production and Unfolded Protein Response. Diabetes. 2018 Sep;67(9):1783–94.

176. Petersen MBK, Azad A, Ingvorsen C, Hess K, Hansson M, Grapin-Botton A, et al. Single-Cell Gene Expression Analysis of a Human ESC Model of Pancreatic Endocrine Development Reveals Different Paths to β-Cell Differentiation. Stem Cell Rep. 2017 Oct

10;9(4):1246–61.

177. Yu X-X, Qiu W-L, Yang L, Wang Y-C, He M-Y, Wang D, et al. Sequential progenitor states mark the generation of pancreatic endocrine lineages in mice and humans. Cell Res. 2021 Aug;31(8):886–903.

178. Gonçalves CA, Larsen M, Jung S, Stratmann J, Nakamura A, Leuschner M, et al. A 3D system to model human pancreas development and its reference single-cell transcriptome atlas identify signaling pathways required for progenitor expansion. Nat Commun. 2021 May 25;12(1):3144.

179. Ramond C, Beydag-Tasöz BS, Azad A, van de Bunt M, Petersen MBK, Beer NL, et al. Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. Development. 2018 Aug 15;145(16).

180. Wesolowska-Andersen A, Jensen RR, Alcántara MP, Beer NL, Duff C, Nylander V, et al. Analysis of differentiation protocols defines a common pancreatic progenitor molecular signature and guides refinement of endocrine differentiation. Stem Cell Rep. 2020 Jan 14;14(1):138–53.

181. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, et al. Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. Nat Genet. 2021 Apr 1;53(4):455–66.

182. Rai V, Quang DX, Erdos MR, Cusanovich DA, Daza RM, Narisu N, et al. Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of rare cells reveals cell-specific type 2 diabetes regulatory signatures. Mol Metab. 2020 Feb;32:109–21.

183. Chiou J, Geusz RJ, Okino M-L, Han JY, Miller M, Melton R, et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature. 2021 Jun;594(7863):398–

402.

184. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018 Nov;50(11):1505–13.

185. Pociot F. Type 1 diabetes genome-wide association studies: not to be lost in translation. Clin Transl Immunology. 2017 Dec 1;6(12):e162.

186. Robertson CC, Inshaw JRJ, Onengut-Gumuscu S, Chen W-M, Santa Cruz DF, Yang H, et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. Nat Genet. 2021 Jul;53(7):962–71.

187. Rust K, Byrnes LE, Yu KS, Park JS, Sneddon JB, Tward AD, et al. A single-cell atlas and lineage analysis of the adult Drosophila ovary. Nat Commun. 2020 Nov 6;11(1):5628.

188. Attali M, Stetsyuk V, Basmaciogullari A, Aiello V, Zanta-Boussif MA, Duvillie B, et al. Control of beta-cell differentiation by the pancreatic mesenchyme. Diabetes. 2007 May;56(5):1248–58.

189. Wagner B, Ricono JM, Gorin Y, Block K, Arar M, Riley D, et al. Mitogenic signaling via platelet-derived growth factor beta in metanephric mesenchymal cells. J Am Soc Nephrol. 2007 Nov;18(11):2903–11.

190. Karlsson L, Lindahl P, Heath JK, Betsholtz C. Abnormal gastrointestinal development in PDGF-A and PDGFR-(alpha) deficient mice implicates a novel mesenchymal structure with putative instructive properties in villus morphogenesis. Development. 2000 Aug;127(16):3457–66.

191. Lammert E, Cleaver O, Melton D. Role of endothelial cells in early pancreas and liver development. Mech Dev. 2003 Jan;120(1):59–64.

192. Azizoglu DB, Chong DC, Villasenor A, Magenheim J, Barry DM, Lee S, et al. Vascular development in the vertebrate pancreas. Dev Biol. 2016 Dec 1;420(1):67–78.

193. Roost MS, van Iperen L, de Melo Bernardo A, Mummery CL, Carlotti F, de Koning EJ, et al. Lymphangiogenesis and angiogenesis during human fetal pancreas development. Vasc Cell. 2014 Nov 1;6:22.

194. Henderson JR, Moss MC. A morphometric study of the endocrine and exocrine capillaries of the pancreas. Q J Exp Physiol. 1985 Jul;70(3):347–56.

195. Zanone MM, Favaro E, Camussi G. From endothelial to beta cells: insights into pancreatic islet microendothelium. Curr Diabetes Rev. 2008 Feb;4(1):1–9.

196. Buschmann I, Pries A, Styp-Rekowska B, Hillmeister P, Loufrani L, Henrion D, et al. Pulsatile shear and Gja5 modulate arterial identity and remodeling events during flow-driven arteriogenesis. Development. 2010 Jul;137(13):2187–96.

197. Guo L, Zhang H, Hou Y, Wei T, Liu J. Plasmalemma vesicle-associated protein: A crucial component of vascular homeostasis. Exp Ther Med. 2016 Sep;12(3):1639–44.

198. Akil A, Gutiérrez-García AK, Guenter R, Rose JB, Beck AW, Chen H, et al. Notch signaling in vascular endothelial cells, angiogenesis, and tumor progression: an update and prospective. Front Cell Dev Biol. 2021 Feb 16;9:642352.

199. Benedito R, Roca C, Sörensen I, Adams S, Gossler A, Fruttiger M, et al. The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. Cell. 2009 Jun 12;137(6):1124–35.

200. Hyde K, Reid CJ, Tebbutt SJ, Weide L, Hollingsworth MA, Harris A. The cystic fibrosis transmembrane conductance regulator as a marker of human pancreatic duct development. Gastroenterology. 1997 Sep;113(3):914–9.

201. Ye F, Duvillié B, Scharfmann R. Fibroblast growth factors 7 and 10 are expressed in the human embryonic pancreatic mesenchyme and promote the proliferation of embryonic pancreatic epithelial cells. Diabetologia. 2005 Feb 3;48(2):277–81.

202. Bocian-Sobkowska J, Zabel M, Wozniak W, Surdyk-Zasada J. Polyhormonal aspect of the endocrine cells of the human fetal pancreas. Histochem Cell Biol. 1999 Aug;112(2):147–53.

203. Riedel MJ, Asadi A, Wang R, Ao Z, Warnock GL, Kieffer TJ. Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas. Diabetologia. 2012 Feb;55(2):372–81.

204. Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity-current challenges and future perspectives. Mol Syst Biol. 2021 Aug;17(8):e10282.

205. Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, et al. CellRank for directed single-cell fate mapping. Nat Methods. 2022 Feb;19(2):159–70.

206. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018 Jun 19;19(1):477.

207. Da Silva Xavier G. The cells of the islets of langerhans. J Clin Med. 2018 Mar 12;7(3).

208. Aiken J, Buscaglia G, Bates EA, Moore JK. The α-Tubulin gene TUBA1A in Brain Development: A Key Ingredient in the Neuronal Isotype Blend. J Dev Biol. 2017 Sep 19;5(3).

209. Kanno N, Fujiwara K, Yoshida S, Kato T, Kato Y. Dynamic Changes in the Localization of Neuronatin-Positive Cells during Neurogenesis in the Embryonic Rat Brain. Cells Tissues Organs (Print). 2019 Dec 5;207(3–4):127–37.

210. Ajiro M, Jia R, Yang Y, Zhu J, Zheng Z-M. A genome landscape of SRSF3-regulated splicing events and gene expression in human osteosarcoma U2OS cells. Nucleic Acids Res. 2016 Feb 29;44(4):1854–70.

211. Zhou KI, Shi H, Lyu R, Wylder AC, Matuszek Ż, Pan JN, et al. Regulation of Co-transcriptional Pre-mRNA Splicing by m6A through the Low-Complexity Protein hnRNPG. Mol Cell. 2019 Oct 3;76(1):70–81.e9.

212. Wang H, Brun T, Kataoka K, Sharma AJ, Wollheim CB. MAFA controls genes implicated in insulin biosynthesis and secretion. Diabetologia. 2007 Feb;50(2):348–58.

213. Pound LD, Sarkar SA, Benninger RKP, Wang Y, Suwanichkul A, Shadoan MK, et al. Deletion of the mouse Slc30a8 gene encoding zinc transporter-8 results in impaired insulin secretion. Biochem J. 2009 Jul 15;421(3):371–6.

214. Taylor AJ, Chen Y-C, Verchere CB. Beta-cell prohormone convertase 1/3 plays a larger role in glucose homeostasis than beta-cell prohormone convertase 2 in mice. FASEB J. 2020 Apr;34(S1):1–1.

215. Ramzy A, Asadi A, Kieffer TJ. Revisiting Proinsulin Processing: Evidence That Human β-Cells Process Proinsulin With Prohormone Convertase (PC) 1/3 but Not PC2. Diabetes. 2020 Jul;69(7):1451–62.

216. Mondal MS, Yamaguchi H, Date Y, Toshinai K, Kawagoe T, Tsuruta T, et al. Neuropeptide W is present in antral G cells of rat, mouse, and human stomach. J Endocrinol. 2006 Jan;188(1):49–57.

217. Childs GV, Rougeau D, Unabia G. Corticotropin-releasing hormone and epidermal growth factor: mitogens for anterior pituitary corticotropes. Endocrinology. 1995 Apr;136(4):1595–602.

218. Conrad E, Stein R, Hunter CS. Revealing transcription factors during human pancreatic β cell development. Trends Endocrinol Metab. 2014 Aug;25(8):407–14.

219. Churchill AJ, Gutiérrez GD, Singer RA, Lorberbaum DS, Fischer KA, Sussel L. Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. Elife. 2017 Jan 10;6.

220. Gasa R, Mrejen C, Lynn FC, Skewes-Cox P, Sanchez L, Yang KY, et al. Induction of pancreatic islet cell differentiation by the neurogenin-neuroD cascade. Differentiation. 2008 Apr;76(4):381–91.

221. Murtaugh LC, Stanger BZ, Kwan KM, Melton DA. Notch signaling controls multiple steps of pancreatic differentiation. Proc Natl Acad Sci USA. 2003 Dec 9;100(25):14920–5.

222. Gao T, McKenna B, Li C, Reichert M, Nguyen J, Singh T, et al. Pdx1 maintains β cell identity and function by repressing an α cell program. Cell Metab. 2014 Feb 4;19(2):259–71.

223. Pan FC, Brissova M, Powers AC, Pfaff S, Wright CVE. Inactivating the permanent neonatal diabetes gene Mnx1 switches insulin-producing β-cells to a δ-like fate and reveals a facultative proliferative capacity in aged β-cells. Development. 2015 Nov 1;142(21):3637–48.

224. Swisa A, Avrahami D, Eden N, Zhang J, Feleke E, Dahan T, et al. PAX6 maintains β cell identity by repressing genes of alternative islet cell types. J Clin Invest. 2017 Jan 3;127(1):230–43.

225. Jacobson D, Shyng S-L. Ion channels of the islets in type 2 diabetes. J Mol Biol. 2019 Aug 29;

226. Liu H, Yang H, Zhu D, Sui X, Li J, Liang Z, et al. Systematically labeling developmental stage-specific genes for the study of pancreatic β-cell differentiation from human embryonic stem cells. Cell Res. 2014 Oct;24(10):1181–200.

227. Ramond C, Glaser N, Berthault C, Ameri J, Kirkegaard JS, Hansson M, et al. Reconstructing human pancreatic differentiation by mapping specific cell populations during development. Elife. 2017 Jul 21;6.

228. Itoh M, Takizawa Y, Hanai S, Okazaki S, Miyata R, Inoue T, et al. Partial loss of pancreas endocrine and exocrine cells of human ARX-null mutation: consideration of pancreas differentiation. Differentiation. 2010 Oct;80(2–3):118–22.

229. Camunas-Soler J, Dai X-Q, Hang Y, Bautista A, Lyon J, Suzuki K, et al. Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. Cell Metab. 2020 May 5;31(5):1017–1031.e4.

230. Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol. 2018 Jan;36(1):70–80.

231. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell. 2018 May 31;173(6):1535–1548.e16.

232. Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods. 2023 Jul 13;

233. De Franco E. From Biology to Genes and Back Again: Gene Discovery for Monogenic Forms of Beta-Cell Dysfunction in Diabetes. J Mol Biol. 2020 Mar 6;432(5):1535–50.

234. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet. 2022 May 12;54(5):560–72.

235. Kim S-S, Hudgins AD, Yang J, Zhu Y, Tu Z, Rosenfeld MG, et al. A comprehensive integrated post-GWAS analysis of Type 1 diabetes reveals enhancer-based immune dysregulation. PLoS One. 2021 Sep 16;16(9):e0257265.

236. Dimas AS, Lagou V, Barker A, Knowles JW, Mägi R, Hivert M-F, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes. 2014 Jun;63(6):2158–71.

237. Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. Elife. 2018 Feb 7;7.

238. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet. 2014 Apr 3;94(4):559–73.

239. He LM, Sartori DJ, Teta M, Opare-Addo LM, Rankin MM, Long SY, et al. Cyclin D2 protein stability is regulated in pancreatic beta-cells. Mol Endocrinol. 2009 Nov;23(11):1865–75.

240. Travers ME, Mackay DJG, Dekker Nitert M, Morris AP, Lindgren CM, Berry A, et al. Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. Diabetes. 2013 Mar;62(3):987–92.

241. Cho J-H, Kim J-W, Shin J-A, Shin J, Yoon K-H. β-cell mass in people with type 2 diabetes. J Diabetes Investig. 2011 Jan 24;2(1):6–17.

242. Li Y, Kim R, Cho YS, Song WS, Kim D, Kim K, et al. Lrfn2-Mutant Mice Display Suppressed Synaptic Plasticity and Inhibitory Synapse Development and Abnormal Social Communication and Startle Response. J Neurosci. 2018 Jun 27;38(26):5872–87.

243. Wesolowska-Andersen A, Zhuo Yu G, Nylander V, Abaitua F, Thurner M, Torres JM, et al. Deep learning models predict regulatory variants in pancreatic islets and refine type 2 diabetes association signals. Elife. 2020 Jan 27;9.

244. Katsura KA, Horst JA, Chandra D, Le TQ, Nakano Y, Zhang Y, et al. WDR72 models of structure and function: a stage-specific regulator of enamel mineralization. Matrix Biol. 2014 Sep;38:48–58.

245. Chandra V, Ibrahim H, Halliez C, Prasad RB, Vecchio F, Dwivedi OP, et al. The type 1 diabetes gene TYK2 regulates β-cell development and its responses to interferon-α. Nat Commun. 2022 Oct 26;13(1):6363.

246. Millman JR, Xie C, Van Dervort A, Gürtler M, Pagliuca FW, Melton DA. Generation of stem cell-derived β-cells from patients with type 1 diabetes. Nat Commun. 2016 May 10;7:11463.

247. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021 Sep 8;597(7875):250–5.

248. Alquicira-Hernandez J, Sathe A, Ji HP, Nguyen Q, Powell JE. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. Genome Biol. 2019 Dec 12;20(1):264.

249. Lavergne A, Tarifeño-Saldivia E, Pirson J, Reuter A-S, Flasse L, Manfroid I, et al. Pancreatic and intestinal endocrine cells in zebrafish share common transcriptomic

signatures and regulatory programmes. BMC Biol. 2020 Aug 31;18(1):109.

250. You L, Wang N, Yin D, Wang L, Jin F, Zhu Y, et al. Downregulation of long noncoding RNA meg3 affects insulin synthesis and secretion in mouse pancreatic beta cells. J Cell Physiol. 2016 Apr;231(4):852–62.

251. Wang N, Zhu Y, Xie M, Wang L, Jin F, Li Y, et al. Long noncoding RNA meg3 regulates mafa expression in mouse beta cells by inactivating rad21, smc3 or sin3α. Cell Physiol Biochem. 2018 Mar 6;45(5):2031–43.

252. Augsornworawat P, Maxwell KG, Velazco-Cruz L, Millman JR. Single-Cell Transcriptome Profiling Reveals β Cell Maturation in Stem Cell-Derived Islets after Transplantation. Cell Rep. 2020 Aug 25;32(8):108067.

253. Hrvatin S, O'Donnell CW, Deng F, Millman JR, Pagliuca FW, DiIorio P, et al. Differentiated human stem cells resemble fetal, not adult, β cells. Proc Natl Acad Sci USA. 2014 Feb 25;111(8):3038–43.

254. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature. 2023 Feb 8;614(7949):742–51.

255. Gage BK, Asadi A, Baker RK, Webber TD, Wang R, Itoh M, et al. The role of ARX in human pancreatic endocrine specification. PLoS One. 2015 Dec 3;10(12):e0144100.

256. Mastracci TL, Wilcox CL, Arnes L, Panea C, Golden JA, May CL, et al. Nkx2.2 and Arx genetically interact to regulate pancreatic endocrine cell development and endocrine hormone expression. Dev Biol. 2011 Nov 1;359(1):1–11.

257. Russell R, Carnese PP, Hennings TG, Walker EM, Russ HA, Liu JS, et al. Loss of the transcription factor MAFB limits β-cell derivation from human PSCs. Nat Commun. 2020

Jun 2;11(1):2742.

258. Artner I, Le Lay J, Hang Y, Elghazi L, Schisler JC, Henderson E, et al. MafB: an activator of the glucagon gene expressed in developing islet alpha- and beta-cells. Diabetes. 2006 Feb;55(2):297–304.

259. Mawla AM, Huising MO. Navigating the depths and avoiding the shallows of pancreatic islet cell transcriptomes. Diabetes. 2019 Jul;68(7):1380–93.

260. Blodgett DM, Redick SD, Harlan DM. Surprising heterogeneity of pancreatic islet cell subsets. Cell Syst. 2016 Oct 26;3(4):330–2.

261. Lin L, Zhang Y, Qian W, Liu Y, Zhang Y, Lin F, et al. Single-cell transcriptome lineage tracing of human pancreatic development identifies distinct developmental trajectories of alpha and beta cells. BioRxiv. 2021 Jan 15;

262. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Huang S, et al. Single cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. BioRxiv. 2019 Jul 9;

263. Mahajan A, Spracklen CN, Zhang W, Ng MC, Petty LE, Kitajima H, et al. Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. medRxiv. 2020 Sep 23;

264. Hu H, Miao Y-R, Jia L-H, Yu Q-Y, Zhang Q, Guo A-Y. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. 2019 Jan 8;47(D1):D33–8.

265. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15.

266. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for

RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

267. Tosti L, Hang Y, Debnath O, Tiesmeyer S, Trefzer T, Steiger K, et al. Single-Nucleus and In Situ RNA-Sequencing Reveal Cell Topographies in the Human Pancreas. Gastroenterology. 2021 Mar;160(4):1330–1344.e11.

268. McKnight KD, Wang P, Kim SK. Deconstructing pancreas development to reconstruct human islets from pluripotent stem cells. Cell Stem Cell. 2010 Apr 2;6(4):300–8.

269. Pan FC, Brissova M. Pancreas development in humans. Curr Opin Endocrinol Diabetes Obes. 2014 Apr;21(2):77–82.

270. Chmielowiec J, Szlachcic WJ, Yang D, Scavuzzo MA, Wamble K, Sarrion-Perdigones A, et al. Human pancreatic microenvironment promotes β-cell differentiation via non-canonical WNT5A/JNK and BMP signaling. Nat Commun. 2022 Apr 12;13(1):1952.

271. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife. 2017 Jan 16;6.

272. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019 May;37(5):547–54.

273. Spencer Chapman M, Ranzoni AM, Myers B, Williams N, Coorens THH, Mitchell E, et al. Lineage tracing of human development through somatic mutations. Nature. 2021 Jul;595(7865):85–90.

274. Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat Rev Genet. 2020 Jul;21(7):410–27.

275. Arrojo E Drigo R, Jacob S, García-Prieto CF, Zheng X, Fukuda M, Nhu HTT, et al. Structural basis for delta cell paracrine regulation in pancreatic islets. Nat Commun. 2019 Aug 16;10(1):3700.

276. Johnston NR, Mitchell RK, Haythorne E, Pessoa MP, Semplici F, Ferrer J, et al. Beta Cell Hubs Dictate Pancreatic Islet Responses to Glucose. Cell Metab. 2016 Sep 13;24(3):389–401.

277. Farack L, Golan M, Egozi A, Dezorella N, Bahar Halpern K, Ben-Moshe S, et al. Transcriptional heterogeneity of beta cells in the intact pancreas. Dev Cell. 2019 Jan 7;48(1):115–125.e4.

278. van der Meulen T, Mawla AM, DiGruccio MR, Adams MW, Nies V, Dólleman S, et al. Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets. Cell Metab. 2017 Apr 4;25(4):911–926.e6.

279. Wittkopp PJ, Kalay G. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2011 Dec 6;13(1):59–69.

280. Zhu H, Wang G, Nguyen-Ngoc K-V, Kim D, Miller M, Goss G, et al. Improving stem cell-derived pancreatic islets using single-cell multiome-inferred regulomes. BioRxiv. 2022 Sep 26;

281. Hendricks TJ, Fyodorov DV, Wegman LJ, Lelutiu NB, Pehek EA, Yamamoto B, et al. Pet-1 ETS gene plays a critical role in 5-HT neuron development and is required for normal anxiety-like and aggressive behavior. Neuron. 2003 Jan 23;37(2):233–47.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Sean de la O
F5708E43CFF24ED...
Author Signature

8/25/2023
Date