Revealing the Key Regulators of Cell Identity in the Human Adult Pancreas

Lotte Vanheer[1], Andrea Alex Schiavo[2,3], Matthias Van Haele[4], Tine Haesen[1], Adrian Janiszewski[1], Joel Chappell[1], Tania Roskams[4], Miriam Cnop[2,3], Vincent Pasque[1,5,*]

[1]Department of Development and Regeneration; KU Leuven - University of Leuven; Herestraat 49, B-3000 Leuven; Belgium.
[2]ULB Center for Diabetes Research; Medical Faculty, Université Libre de Bruxelles; Route de Lennik 808, B-1070 Brussels, Belgium.
[3]Division of Endocrinology; Erasmus Hospital, Université Libre de Bruxelles; Route de Lennik 808, B-1070 Brussels, Belgium.
[4]Department of Imaging and Pathology; Translational Cell and Tissue Research, KU Leuven and University Hospitals Leuven; Herestraat 49, B-3000 Leuven, Belgium.
[5]Lead Contact
*Correspondence: vincent.pasque@kuleuven.be

HIGHLIGHTS

- Reconstruction of gene regulatory networks for human adult pancreatic cell types
- An interactive resource to explore and visualize gene expression and regulatory states
- Predicting putative transcription factors driving pancreatic cell identity
- HEYL and JUND as candidate regulators of acinar and alpha cell identity, respectively

SUMMARY

Cellular identity during development is under the control of transcription factors that form gene regulatory networks. However, the transcription factors and gene regulatory networks underlying cellular identity in the human adult pancreas remain largely unexplored. Here, we integrate multiple single-cell RNA sequencing datasets of the human adult pancreas, totaling 7393 cells, and comprehensively reconstruct gene regulatory networks. We show that a network of 142 transcription factors forms distinct regulatory modules that characterize pancreatic cell types. We present evidence that our approach identifies key regulators of cell identity in the human adult pancreas. We predict that HEYL and JUND are active in acinar and alpha cells, respectively, and show that these proteins are present in the human adult pancreas as well as in human induced pluripotent stem cell-derived pancreatic cells. The comprehensive gene regulatory network atlas can be explored interactively online. We anticipate our analysis to be the starting point for a more sophisticated dissection of how transcription factors regulate cell identity in the human adult pancreas. Furthermore, given that transcription factors are major regulators of embryo development and are often

perturbed in diseases, a comprehensive understanding of how transcription factors work will be relevant in development and disease biology.

INTRODUCTION

A fundamental question in biology is how a single genome gives rise to the great diversity of cell types that make up organs and tissues. A key goal is to map all cell types of developing and mature organs such as the pancreas, an essential organ at the basis of multiple human disorders including diabetes and cancer (Kahn, Cooper and Del Prato, 2014; Kamisawa *et al.*, 2016; Han *et al.*, 2020). Single-cell RNA sequencing (scRNA-seq) provides a powerful tool to resolve cellular heterogeneity, identify cell types and capture high-resolution snapshots of gene expression in individual cells (Grün *et al.*, 2016). With the advent of single-cell transcriptomics, great progress has been made toward the creation of a reference cell atlas of the pancreas (Baron *et al.*, 2016; Segerstolpe *et al.*, 2016; Wang *et al.*, 2016; Xin *et al.*, 2016; Enge *et al.*, 2017; Lawlor *et al.*, 2017; Augsornworawat and Millman, 2020; Han *et al.*, 2020). Work from several groups provided cellular atlases of the pancreas during mouse development (Stanescu *et al.*, 2017; Byrnes *et al.*, 2018; Scavuzzo *et al.*, 2018), in adult mice (Muraro *et al.*, 2016) and in human fetal and adult pancreas (Liu *et al.*, 2014; Baron *et al.*, 2016; Muraro *et al.*, 2016; Segerstolpe *et al.*, 2016; Wang *et al.*, 2016; Enge *et al.*, 2017; Han *et al.*, 2020). Efforts have also been made to map cellular identity during pancreas development starting from human pluripotent stem cells (Rezania *et al.*, 2014; Hrvatin *et al.*, 2014; Zhu *et al.*, 2016; Han *et al.*, 2020; Hogrebe *et al.*, 2020; Peterson *et al.*, 2020). Taken together, these studies provide an opportunity to better understand the maintenance and establishment of cellular identity among different pancreatic cell types.

Work over the past decades indicated that cellular identity is established by combinations of transcription factors (TFs) that recognize and interact with cis-regulatory elements in the genome. These transcription factors, together with chromatin modifiers, give rise to gene expression programs. A small number of core TFs are thought to be sufficient for the establishment and maintenance of gene expression programs that define cellular identity during and after development (Ohno, 1979). Studies conducted in both mouse and human have successfully identified TFs that are pivotal for the acquisition and maintenance of pancreatic cell fates (Dassaye, Naidoo and Cerf, 2016). These include PDX1 (Zhou *et al.*, 2008; Shih *et al.*, 2015), MAFA (Nishimura, Takahashi and Yasuda, 2015), NGN3 (Gradwohl *et al.*, 2000), NKX2.2 (Sussel *et al.*, 1998), PAX4 (Sosa-Pineda *et al.*, 1997), NKX6.1, NEUROD1 (Mastracci *et al.*, 2013), ARX (Collombat *et al.*, 2003), MAFB (Artner *et al.*, 2006), RFX6 (Smith *et al.*, 2010), GATA4 (Ketola *et al.*, 2004), FOXA2 (Lee *et al.*, 2019) and SOX9 (Shroff *et al.*, 2014; Shih *et al.*, 2015). Conditional deletion of TFs such as FOXA2 and PDX1 in adult beta cells results in the loss of cellular identity and function (Sund *et al.*, 2001; Gao *et al.*, 2014). Genetic evidence for the role of these TFs in establishing human pancreatic cell identity is provided by the identification of TF loss-of-function mutations that cause pancreatic agenesis (Stoffers *et al.*, 1997; Sellick *et al.*, 2004; Allen *et al.*, 2012; Shaw-Smith *et al.*, 2014; De Franco *et al.*, 2019) or neonatal

or young-onset diabetes (Senée *et al.*, 2006; Solomon *et al.*, 2009; Rubio-Cabezas *et al.*, 2010, 2011; Smith *et al.*, 2010; Bonnefond *et al.*, 2013; Flanagan *et al.*, 2014). In addition, TF overexpression can reprogram somatic cells to adopt alternative identities (Takahashi and Yamanaka, 2006; Zhou *et al.*, 2008; Vierbuchen *et al.*, 2010; Lima *et al.*, 2016). For example, the induced expression of *Ngn3*, *Pdx1*, and *Mafa* was shown to reprogram mouse alpha cells into beta-like cells *in vivo* (Zhou *et al.*, 2008). However, how key TFs underlie the maintenance of cellular identity in the human pancreas remains incompletely understood.

Over the past decade, multiple approaches to reconstruct gene regulatory networks (GRNs) from bulk and single-cell omics data have been developed (Ghazanfar et al., 2016; Lim et al., 2016; Matsumoto et al., 2017; Fiers et al., 2018). In particular, it is now possible to combine single-cell transcriptomic data with cis-regulatory information to infer GRNs (Janky *et al.*, 2014; Aibar *et al.*, 2017; Van de Sande *et al.*, 2020). Because TFs recognize DNA motifs in the genome, one can measure if inferred target genes are expressed within single cells, and therefore quantify the activity of TFs. Such approaches have revealed the regulatory programs in distinct systems including the Drosophila brain (Davie et al , 2018), cancer (Wouters *et al.*, 2020), during early mouse embryonic development (Peng et al , 2019), in a mouse cell atlas (Suo et al., 2018) and a human cell atlas (Han *et al.*, 2020).

Analysis of GRNs in the human adult pancreas has identified distinct endocrine and exocrine regulatory states with multiple stable cell states for alpha, beta and ductal cells (Kumar and Vinod, 2019). Type 2 diabetes and body mass index (BMI) were shown not to impact GRN activity of alpha and beta cells (Kumar and Vinod, 2019). Previous data shows that type 2 diabetic (Faerch *et al.*, 2015; Dennis et al., 2019; Dybala and Hara, 2019; Zaharia et al., 2019) and non-diabetic human islet preparations vary greatly depending on age (Enge et al., 2017; Westacott et al., 2017) and BMI (Henquin, 2018) warranting the exploration of GRNs in larger cohorts. Hence, it remains unclear whether previous GRN findings can be extrapolated to a broader, highly heterogeneous non-diabetic and type 2 diabetes patient population. The development of integration methods provides an opportunity to analyze multiple scRNA-seq studies from multiple laboratories and patients (Butler *et al.*, 2018; Luecken *et al.*, 2020). Additional knowledge on how GRNs maintain cellular identity in the human adult pancreas may further the understanding of disease states as well as improve efforts to convert patient cells into functional, mature beta cells for diabetes treatment.

Here, we build an integrated human pancreas gene regulatory atlas. In this resource, we use single-cell transcriptomes of the human adult pancreas, taking advantage of integration strategies and computational tools to reconstruct GRNs. Our analysis identifies the GRN landscape and candidate regulators that are critical for cellular identity in the human adult pancreas.

RESULTS

Integrated Analysis of scRNA-seq Data Identifies 12 Human Adult Pancreatic Cell Types.

Integrating multiple human adult pancreas scRNA-seq datasets can further improve the power of scRNA-seq analyses to create a human adult pancreas cell atlas. We set out to analyse and integrate five publicly available datasets covering a total of 35 non-diabetic, 15 type 2 diabetic and 1 type 1 diabetic individuals using Seurat v3.0 CCA integration tools **(Figure 1A,** detailed donor information can be found in **Table S1)** (Segerstolpe *et al.*, 2016; Wang *et al.*, 2016; Xin *et al.*, 2016; Enge *et al.*, 2017; Lawlor *et al.*, 2017; Hafemeister and Satija, 2019; Stuart *et al.*, 2019). After filtering out low quality transcriptomes and data integration, uniform manifold approximation and projection for dimension reduction (UMAP) visualisation revealed that 7393 cells localize into distinct clusters (**Figure 1B**). Cells from each original dataset localize together suggesting that the location of cells on the UMAP is not driven by the dataset of origin **(Figure 1C)**.

We next sought to identify pancreatic cell types **(Figure 1D)**. Clustering analyses based on the expression of well-established cell type specific markers led to the identification of eight cell types in the human adult pancreas**:** beta, alpha, gamma, delta, acinar, ductal, stellate and endothelial cells **(Figure 1B, Table S2)**. UMAP visualization allowed for the segregation of endocrine, exocrine and other lineages (**Figure S1A**). Beta cells grouped together, away from other clusters and were marked by INS expression (**Figure 1E**). Other distinct clusters corresponded to alpha, gamma and delta cells based on global transcriptional similarity and elevated expression of GCG, PPY and SST, respectively, and other markers **(Figure 1E, Table S2)**. Using a similar approach, we reliably detected other, previously described, major pancreatic cell types (acinar, ductal, endothelial and stellate, **Figure 1B**). All cell types were detected in both non-diabetic and type 2 diabetic pancreases **(Figure 1F)**. An additional four rare cell populations, that cannot be robustly identified through clustering analyses, were identified manually by assessing the expression of GHRL (epsilon cells), TPS1AB (schwann cells), CD86 (mast cells) and SOX10 (Major histocompatibility complex (MHC) class 2 cells) (Wierup *et al.*, 2002; Segerstolpe *et al.*, 2016) **(Figure 1G)**. These rare cell types often cluster with other common cell types. Importantly, our annotation recapitulated previous annotations to a large extent **(Figure S1B-C)**. In summary, we reconstructed an integrated single-cell atlas of the human adult pancreas, and annotated 12 pancreatic cell types.

Reconstruction of Gene Regulatory Networks in the Human Adult Pancreas

Next, we set out to comprehensively reconstruct GRNs for all pancreatic cell types from single-cell transcriptomic data, applying single-cell regulatory network inference and clustering (pySCENIC) (Aibar *et al.*, 2017; Van de Sande *et al.*, 2020). PySCENIC links *cis*-regulatory sequence information together with single-cell transcriptomes in three sequential steps by 1) co-expression analysis, 2) target gene motif and ChIP-seq track enrichment analysis, and 3) regulon activity evaluation **(Figure 2A)**. Each regulon consists of a TF with its predicted target genes (co-expressed genes with an enriched TF motif), altogether forming a regulon. pySCENIC identified 142 regulons that characterize the GRNs of the human adult pancreas

(Figure 2B/C, Table S3). Multiple regulons identified here as active in the pancreas correspond to TF binding motifs enriched in accessible chromatin in the pancreas (assessed by ATAC-seq in FACS-purified pancreatic cells, (Arda *et al.*, 2018)), supporting the validity of the approach **(Figure S2A)**.

UMAP visualization based on the activity of 142 regulons in non-diabetic and type 2 diabetic pancreata revealed groups of cells that differ from one another based on their regulatory activity **(Figure 2B-C)**. In particular, there are distinct regulatory states for exocrine and endocrine pancreatic lineages, stellate and endothelial cells (**Figure 2B/C**). Endocrine cell types clustered together, indicating shared regulatory states, while exocrine cell types formed two distinct clusters. Stellate and endothelial cells differed most from other cell types in their regulatory states. These results are consistent with previous analyses (Baron, Veres, Samuel L. Wolock, *et al.*, 2016; Lawlor *et al.*, 2017; Kumar and Vinod, 2019) and are also in line with our findings above based on gene expression analysis (**Figure 1D**). As expected, regulons active in endocrine cell types include RFX6, PAX6 and NEUROD1 **(Figure 2D-G)**. These TFs have reported roles in endocrine cell fate commitment and maintenance of cell identity throughout adult life (Smith *et al.*, 2010; Hart *et al.*, 2013; Mastracci *et al.*, 2013). Using iRegulon for visualization, many of the NEUROD1 target genes identified here have been previously linked to beta cell survival and function, such as SNAP25, TSPAN2, ELAVL4, PLCXD3 and NRNX1 (Mosedale *et al.*, 2012; Hwang *et al.*, 2016; Daraio *et al.*, 2017; Juan-Mateu *et al.*, 2017; Aljaibeji *et al.*, 2019) **(Figure S2B)**. Interestingly, other NEUROD1 target genes are reported to be involved in cell fate specification during endocrine pancreas development such as PAX6, NKX2.2, INSM1 and HDAC6, suggesting an overlap in NEUROD1 target genes in adult life and embryonic development (Gierl *et al.*, 2006; Haumaitre *et al.*, 2008; Hart *et al.*, 2013; Churchill *et al.*, 2017). Clustering all cells based on the activity of all regulons identifies regulatory modules (**Figure 2D,** red squares). In the exocrine pancreas, one regulatory module, containing NR5A2, was shared between acinar and ductal cells, although with a tendency for increased regulon activity in ductal cells **(Figure 2D/H)**. Other exocrine regulons included ONECUT1, REST and HNF1B with reported roles in exocrine development (Nissim *et al.*, 2016; Kropp, Zhu and Gannon, 2019) and the adult exocrine pancreas (Quilichini *et al.*, 2019; Bray *et al.*, 2020) **(Figure 2D)**. In summary, this analysis confirms the expected separation of exocrine and endocrine cells with distinct gene regulatory programs, and identifies known and novel candidate regulators of pancreas cell states.

Several regulatory modules are shared between different cell types within the endocrine and exocrine pancreas. Additionally, each cell type is defined by cell-type specific regulatory modules **(Figure 1D)**. In the endocrine pancreas, alpha and beta cells shared endocrine regulons (MAFB, MEIS2), whereas we observed distinct activities for ARX and IRX2 regulons in alpha cells and RXRG and PDX1 in beta cells **(Figure 2D/H)**, expanding previous findings (Kumar and Vinod, 2019). Using iRegulon for visualization, PDX1 target genes include SLC6A17, PDIA6 and ABHD3, which have been reported to control insulin release (Thomas, Brown and Brown, 2014; Eletto *et al.*, 2016; Rorsman and Ashcroft, 2018) **(Figure S2C)**.

Interestingly, gamma and delta cells overlapped with alpha and beta cells, respectively, suggesting a shared regulatory state **(Figure 2C/D)**. This includes shared regulon activity for ARX in gamma and alpha and PDX1 in beta and delta cells **(Figure 2D/H)**, consistent with their reported expression in published scRNA-seq studies (Baron *et al.*, 2016; Segerstolpe *et al.*, 2016; Lawlor *et al.*, 2017). GATA4 and RBPJL, known acinar-specific TFs, were highly active in acinar cells (Masui *et al.*, 2010; Carrasco *et al.*, 2012) **(Figure 2H)**. Similarly, ductal cells were characterised by highly active SOX9 and POU2F3 regulons, in line with previous literature (Shroff *et al.*, 2014; Yamashita *et al.*, 2017) **(Figure 2H)**. In sum, this analysis confirms that alpha, beta, acinar and ductal cells are defined by the activity of distinct TF combinations that form gene regulatory modules.

In conclusion, the network approach recovers many of the expected regulators of pancreatic cellular identity allowing for the comprehensive characterisation of the gene regulatory state of all major human adult pancreatic cell types.

Prediction of Regulators of Endocrine Cell Identity in the Human Adult Pancreas

A comprehensive network analysis provides an opportunity to predict and identify critical regulators of cell identity. To identify regulons with highly cell type-specific activities within the human adult non-diabetic pancreas, we calculated regulon specificity scores (RSS) (a complete list of RSSs can be found in **Table S4)** (Suo *et al.*, 2018). The RSS utilises Jensen–Shannon divergence to measure the similarity between the probability distribution of the regulon's enrichment score and cell type annotation wherein outliers receive a higher RSS and are therefore considered cell type-specific (Suo *et al.*, 2018). It can therefore be used to rank the activity of TFs within specific cell types.

Among the top regulons identified in alpha cells, we recover well known regulators of alpha and endocrine cell fate such as ARX, IRX2, PAX6, MAFB, NEUROD1 and RFX6 **(Figure 3A/B)** (Collombat *et al.*, 2003; Artner *et al.*, 2006; Delporte *et al.*, 2008; Smith *et al.*, 2010; Dorrell *et al.*, 2011; Mastracci *et al.*, 2013). In addition, we identified JUND, EGR4, SREBF1 and STAT4, which have not yet been implicated in alpha cell identity. EGR1 (but not EGR4) has been shown to transcriptionally regulate glucagon expression (Leung-Theung-Long *et al.*, 2005) as well as the PDX1 promoter in beta cells (Eto, Kaur and Thomas, 2007). STAT4 and JUND have been described in pancreatic tissue in general and beta cells, respectively, but not in alpha cells (Yu and Kim, 2012; Good *et al.*, 2019) **(Figure 3B/C)**. These TFs respond to the JNK and EGFR signalling pathways and may have important physiological functions. Both JUND and the JUND/JNK signaling pathway have been implicated in pancreatic cancer (Shin *et al.*, 2009; Recio-Boiles *et al.*, 2016). Immunocytochemistry of the human adult pancreas confirmed the presence of nuclear JUND in islets **(Figure 3Di)**. We also detected nuclear JUND protein in a subset of human induced pluripotent stem cells (iPSCs) subjected to beta cell differentiation (**Figure 3E/F**). Surprisingly, we also detected JUND protein in ductal cells, despite lower JUND regulon activity in this cell type **(Figures 3C, 3Dii)**. Thus, protein

expression does not always predict regulatory activity. Nevertheless, these results show that JUND is present and active in a subset of pancreatic cell types in the human adult pancreas and human pluripotent stem cell derived islet cells. Altogether, this analysis predicts TFs active in human alpha cells, recovering known as well as new candidate TFs.

Among the top regulons identified in beta cells, we retrieved well-known as well as new candidate regulators of beta and endocrine cell identity. Known TFs include RXRG, PDX1, NEUROD1, PAX6 and RFX6 (Zhou *et al.*, 2008; Miyazaki *et al.*, 2010; Smith *et al.*, 2010; Mastracci and Sussel, 2012; Hart *et al.*, 2013) (**Figure 3G/H)**. In addition, we found that ZNF705D, ASCL2 and HOXD13 were highly ranked regulons (**Figure 3H/I).** HOXD13 and BHLHE41 have been shown to be present in the exocrine pancreas (Cantile *et al.*, 2009; Sato *et al.*, 2012). Interestingly, ASCL2 has been reported to interact with β-catenin of the Wnt pathway, the latter has an established role in endocrine fate specification during *in vitro* differentiation (Schuijers *et al.*, 2015; Sharon *et al.*, 2019; Vethe *et al.*, 2019). Many putative target genes of ASCL2 including PDX1, INS, ABCC8, FOXA1, KCNK16, FXYD2 are directly related to glucose sensing and beta cell identity, in line with the beta cell-specific regulatory activity of ASCL2 **(Figure S3A)** (Gao *et al.*, 2010; Arystarkhova *et al.*, 2013; Vierra *et al.*, 2015; Park, Lee and Park, 2016). FXYD2γa, a regulatory subunit of the Na^+-K^+-ATPase, is a transcript exclusively expressed in human beta cells (Flamez *et al.*, 2010). Immunohistochemistry of human adult pancreas sections showed that ASCL2 is expressed in INS^+ beta and islet cells **(Figure 3J)**. Surprisingly, ASCL2 was mainly localized to the cytoplasm **(Figure 3J)**, which is unexpected for TFs which tend to localize to the nucleus (Baranek, Sock and Wegner, 2005). Cytoplasmic localization of ASCL2 has been reported in the context of colon and breast cancer (Zhu *et al.*, 2012; Xu *et al.*, 2017). These results implicate additional TFs including ASCL2 in the regulation of beta cell identity. They also illustrate the value of network analyses to increase our understanding of the biology of the human pancreas.

In summary, GRN analysis and regulon ranking allowed us to pinpoint both known and novel candidate regulators of pancreatic endocrine cell identity, providing a resource for further investigation of their roles in cellular identity and function.

Prediction of Regulators of Exocrine Cell Identity in the Human Adult Pancreas

Similarly, the comprehensive network analysis provides an opportunity to predict and identify critical regulators of exocrine cell identity.

We also identify known and new TFs in acinar cells. Among the top acinar-specific regulons, we recovered well known regulators of acinar and exocrine cell identity such as PTF1A, RBPJL, GATA4 and NR5A2 (Ketola *et al.*, 2004; Masui *et al.*, 2010; Nissim *et al.*, 2016; Sakikubo *et al.*, 2018) **(Figure 4A/B)**. These findings are in line with a recent study that used single-nucleus RNA-seq on pancreatic acinar tissue (Tosti

et al., 2019). Furthermore, we identified MECOM, HEYL and TGIF1 as highly ranked regulons **(Figure 4B/C)**. Interestingly, aberrant *Mecom* expression has been linked to the induction of gastric genes in acinar cells, which disrupts acinar cell identity and increases susceptibility to malignancy (Hoang *et al.*, 2016). The loss of TGIF1 has been linked to pancreatic ductal adenocarcinoma progression making further exploration of these regulons interesting in the context of cancer biology (Weng *et al.*, 2019). Ectopic expression of Tgif2 (but not Tgif1) reprograms mouse liver cells towards a pancreas progenitor state (Cerdá-Esteban *et al.*, 2017). HEYL is a reported Notch signalling target gene in NGN3^{+} exocrine cells (Gomez *et al*, 2015). We confirmed nuclear expression of HEYL in human acinar and islet cells **(Figure 4Di/ii,** detailed donor information can be found in **Table S1)** by immunohistochemistry, in agreement with elevated HEYL regulon activity in acinar cells **(Figure 3C).**

Top ranked regulons in ductal cells included well known regulators of ductal and exocrine cell identity such as POU2F3, NR5A2 and HNF1B (Nissim *et al.*, 2016; Yamashita *et al.*, 2017; Quilichini *et al.*, 2019) **(Figure 3E/F)**. In addition, we observed CDX2 and PPARD as highly specific ductal regulons **(Figure 3F/G)**. Both PPARD and CDX2 have been reported to be involved in human pancreatic ductal carcinoma, warranting further functional studies in the healthy pancreatic context (Matsumoto *et al.*, 2004; Coleman *et al.*, 2013).

In summary, GRN analysis and regulon ranking allowed us to pinpoint both known and novel candidate regulators of pancreatic exocrine cell identity. Specifically, we identified HEYL as a candidate TF that might be important for acinar cell identity, warranting further investigation.

An Interactive Online Resource for Visualization of the Human Adult Pancreas Cell Network Atlas

To enable users to easily navigate the human pancreatic cell network atlas, we provide a loom file that allows for the visualisation and exploration of the data using the web-based portal SCope (Davie *et al.*, 2018) **(.loom file and tutorial available at http://scope.aertslab.org/#/PancreasAtlas/*/welcome and https://github.com/pasquelab/scPancreasAtlas).** Features such as cell type annotation as defined in this paper, gene expression and regulon activity can be explored on the regulon and gene expression based UMAP. This resource enables users to select and visualize up to three genes or regulons simultaneously and select subsets of cells for downstream analyses. For example, the expression of COVID-19 related genes can be interactively explored (Yang *et al.*, 2020). Target genes of a specific regulon can be downloaded to facilitate further exploration, for example in iRegulon or gene ontology analysis (Janky *et al.*, 2014). A list of predicted target genes of all 142 regulons can also be found in **Table S5.** Furthermore, a list of target genes can be manually defined to compute the activity of a custom regulon. This resource can be used to further study cell identity and gene regulation in the context of the pancreas, diabetes and cancer.

DISCUSSION

In this resource, we take advantage of integration strategies and new computational tools to reconstruct an integrated cell and GRN atlas of the human adult pancreas from single-cell transcriptome data. This approach provides a comprehensive analysis of the gene regulatory logic underlying cellular identity in the human adult pancreas in a broad range of individuals, limiting the influence of inter-donor variability. We recovered known regulators of pancreatic cell identity and uncovered novel candidate regulators of cell identity that can be further investigated for their roles in cellular identity and function. By validating regulon analyses and creating an easily accessible interactive online resource which allows for the exploration of the gene regulatory state of 7393 cells from 51 individuals, this approach extends beyond previous gene regulatory studies in the human adult pancreas (Augsornworawat and Millman, 2020).

The present analysis identified regulators of pancreatic development, function and survival that are known to be critical in humans because loss-of-gene function causes pancreatic agenesis or young onset diabetes. For example, PTF1A (Sellick *et al.*, 2004; Weedon *et al.*, 2014) and GATA4 (Shaw-Smith *et al.*, 2014), whose loss of function are linked to pancreatic agenesis and neonatal diabetes, were among the top acinar-specific regulons **(Figure 3)**. In addition, monogenic diabetes related genes PDX1 (Nicolino *et al.*, 2010), NEUROD1 (Rubio-Cabezas *et al.*, 2010), PAX6 (Solomon *et al.*, 2009), RFX6 (Smith *et al.*, 2010; Patel *et al.*, 2017) and GLIS3 (Senée *et al.*, 2006) were among the top beta cell-specific regulons **(Figure 3 and Table S4**).

Several top regulons in endocrine but not in exocrine cells are TFs involved in endoplasmic reticulum (ER) stress signalling (**Table S4**). CREB3 and CREB3L2 are non-canonical ER stress transducers that are induced in human islets and clonal beta cells upon exposure to the saturated fatty acid palmitate (Cnop *et al.*, 2014). Interestingly, SREBF1 and -2 undergo similar ER exit and proteolytic processing in the Golgi as these ER stress transducers, but they do so in response to changes in ER cholesterol content; both also have high regulon activity in alpha and beta cells. XBP1 is abundantly expressed in the exocrine and endocrine pancreas (Cnop *et al.*, 2017), but the XBP1 regulon has its highest specificity in beta cells. ATF3 and ATF4 are TFs that are activated upon eIF2α phosphorylation, an ER stress response pathway to which no less than 5 monogenic forms of diabetes belong (Eizirik, Pasquali and Cnop, 2020). Our data underscores the importance of these TFs for endocrine pancreatic cell identity.

Given that we predict novel regulators of cell identity in the human pancreas, it will be interesting to also expand this analysis to pancreas embryonic development. Our work may also be beneficial in guiding improvements of and better understanding the *in vitro* derivation of pancreatic cell types. For example, the emergence of SST-positive cells together with beta-like cells at the end of *in vitro* differentiation could be explained by the overlap in regulatory states between beta and delta cells (Baron *et al.*, 2016). GRN analyses are particularly interesting for *in vitro* derived beta cells since a better understanding of the

regulatory logic underlying control of beta cell fate may improve or facilitate future applications in regenerative medicine (Pagliuca *et al.*, 2014; Rezania *et al.*, 2014; Nostro *et al.*, 2015; Russ *et al.*, 2015; Baeyens *et al.*, 2018). Alternatively, many observed GRNs such as ASCL2, MECOM, PPARD, GATA6 and CDX2 are linked to pancreatic cancer making the additional exploration of GRNs interesting in the context of cancer biology (Matsumoto *et al.*, 2004; Zhu *et al.*, 2012; Coleman *et al.*, 2013; Hoang *et al.*, 2016; Xu *et al.*, 2017; Weng *et al.*, 2019; Brunton *et al.*, 2020). Finally, recent reports have stratified type 2 diabetes patients based on age at diagnosis, BMI, HbA1c and insulin secretion and sensitivity, and identified subtypes with different genetic predisposition, treatment response, disease progression and complication rates (Ahlqvist *et al.*, 2018). Hence, it would be interesting to assess differences in gene regulatory state and gene expression profiles of alpha and beta cells between different type 2 diabetic subgroups.

Limitations of the study

It is important to note that pySCENIC is a stochastic algorithm that does not produce precisely the same regulons for repeated applications, limiting reproducibility when comparing different datasets (Huynh-Thu *et al.*, 2010; Van de Sande *et al.*, 2020). To mitigate this uncertainty, we ran the full pySCENIC pipeline five times and only kept consistent regulons with the highest regulon activity. The performance of pySCENIC, and other GRN inference methods, suffers due to the large amount of drop-out events in scRNA-seq data warranting caution when interpreting results (Chen and Mar, 2018). This could explain the absence of well-established pancreas TFs such as MAFA (Olbrot *et al.*, 2002), MNX1 (Flanagan *et al.*, 2014), NEUROG3 (Krentz *et al.*, 2017), FOXA2 (Lee *et al.*, 2002) and NKX2-2 (Mastracci *et al.*, 2011) in this analysis. Nevertheless, in support of the validity of our findings, ATAC-seq, literature and immunohistochemistry of human pancreas sections corroborate several pySCENIC predictions. Chen and colleagues underline the importance of using large sample sizes to derive the most accurate network inference possible (Chen and Mar, 2018), highlighting the importance of dataset integration to increase the number of cells analysed. In the future, it will be interesting to extend our analyses to include many more cells and patients. In spite of current caveats, GRN analysis has enabled the capture of biological relevant information (Butte *et al.*, 2000).

One additional limitation of this study is the assumption that all TFs bind their binding motifs in the promoters of expressed genes. However, TF binding can be restricted to a subset of TF motifs in the genome due to influence of chromatin processes including the presence of nucleosomes as well as DNA methylation. Therefore, additional approaches such as single cell multi-omics that capture additional layers of genome regulation will be helpful to increase our understanding of gene regulation in the context of the human pancreas.

Taken together, our GRN atlas, containing 51 individuals, provides a valuable resource for future studies in the human pancreas development, donor variability, homeostasis and disease including type 2 diabetes

and pancreatic cancer. Finally, our results provide new insights into the activity of TFs and gene regulation in the human adult pancreas from a gene regulatory perspective.

ACKNOWLEDGEMENTS

We thank Stein Aerts, Kristofer Davie and the Stein Aerts lab for discussions and creating the permanent SCope link, Shengbao Suo for sharing the MATLAB script for calculating the regulon specificity score, the Flemish Supercomputer Center (VSC) and Leuven Stem Cell Institute. We thank San Kit To and Stein Aerts for feedback on the manuscript. This work was supported by Fonds Wetenschappelijk Onderzoek (FWO) Odysseus Return Grant G0F7716N to V.P., FWO grants G0C9320N and G0B4420N to VP), the KU Leuven Research Fund (BOFZAP starting grant StG/15/021BF to V.P., C1 grant C14/16/077 to V.P., and Project financing) and FWO SB Ph.D. Fellowship to L.V. (1S29419N). The M.C. lab is supported by the Fonds National de la Recherche Scientifique (FNRS), the Brussels Region Innoviris project DiaType, the Win2Wal program of the Walloon Region and the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115797 (INNODIA; this Joint Undertaking receives support from the Union's Horizon 2020 research and innovation program and "EFPIA" (European Federation of Pharmaceutical Industries Associations), "JDRF" (Juvenile Diabetes Research Foundation), and "The Leona M. and Harry B. Helmsley Charitable Trust").

AUTHOR CONTRIBUTIONS

Conceived project, VP, LV. Performed experiments, AAS, MVH. Analyses, AAS, MVH, TH, TR. scRNA-seq analyses, LV, AJ, JC. ATAC-seq analysis, LV. Wrote manuscript, LV, VP and MC with input from all authors. Supervision, VP.

DECLARATION OF INTERESTS

The authors declare no competing interests.

METHODS

Key Resources Table

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal mouse anti-human ASCL2	Cat# MAB4418, clone 7E2	Millipore, RRID:AB_10561764
Rabbit anti-human JUND	Cat# HPA063029	Atlas Antibodies, RRID:AB_2684925
Rabbit anti-human HEYL	Cat# HPA076960	Atlas Antibodies, RRID:AB_2732326
Guinea pig anti-human INS	Cat# IR002	Agilent, RRID:AB_2800361
Monoclonal mouse anti-human GCG	Cat# G2654	Sigma-Aldrich, RRID:AB_259852

Goat Anti-Guinea Pig IgG Alexa Fluor 568	Cat# A-11075	Thermo Fisher Scientific, RRID:AB_2534119
Donkey anti-Rabbit IgG Alexa Fluor 488	Cat# A-21206	Thermo Fisher Scientific, RRID:AB_2535792
Donkey Anti-Mouse IgG Alexa Fluor 647	Cat# A-31571	Thermo Fisher Scientific, RRID:AB_162542
Experimental Models: Cell Lines		
Control iPSC line HEL115.6	(Cosentino *et al.*, 2018)	
Biological Samples		
Human pancreas sections	Biobank of the University Hospital Leuven (Belgium)	
Software and Algorithms		
SRA toolkit	v2.9.4	
STAR	v2.5.3a	RRID:SCR_015899
Rsubread	v1.5.2	RRID:SCR_016945
ENCODE ATAC-seq pipeline		(Lee, 2016)
HOMER	v4.10.4	RRID:SCR_010881
Integrative Genomics Viewer	V2.8.0	RRID:SCR_011793
Seurat	v3.1.1	RRID:SCR_007322
seaborn	v0.9.0	RRID:SCR_018132
py SCENIC	v0.9.15	RRID:SCR_017247
Cytoscape	v3.7.1	RRID:SCR_003032
iRegulon (Cytoscape plugin)	v1.3	(Janky *et al.*, 2014)
MATLAB	R2019b	MathWorks
ggplot2	v3.1.1	RRID:SCR_014601
LoomPy	v2.0.17	RRID:SCR_016666
R	v3.6.1	RRID:SCR_001905
Python	3.6.9	RRID:SCR_008394
Deposited Data		
Reanalyzed human pancreatic scRNA-seq dataset	GEO	GEO: GSE86469
Reanalyzed human pancreatic scRNA-seq dataset	GEO	GEO: GSE81608

Reanalyzed human pancreatic scRNA-seq dataset	GEO	GEO: GSE83139
Reanalyzed human pancreatic scRNA-seq dataset	GEO	GEO: GSE81547
Reanalyzed human pancreatic scRNA-seq dataset	ArrayExpress	ArrayExpress: E-MTAB-5061
Reanalyzed human pancreatic bulk ATAC-seq dataset of FACS sorted beta, alpha, ductal and acinar cell populations	GEO	GEO: GSE79468
.loom file containing regulon and gene expression data	GEO	GEO: GSE156490

Resource availability

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Vincent Pasque (vincent.pasque@kuleuven.be).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

This study did not generate any new datasets. Regulon data, raw and integrated gene expression matrices and the .loom file are available in the Gene Expression Omnibus (GEO) repository under accession code GSE156490 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE156490]. A SCope tutorial is available at https://github.com/pasquelab/scPancreasAtlas. This study did not generate any new software; questions about data analysis should be directed to the Lead Contact, Vincent Pasque (vincent.pasque@kuleuven.be). The reviewer tokens for this GEO repository is cfsjciumfferjkd.

Method Details

Motif discovery of bulk ATAC-seq data

Paired-end raw reads for bulk ATAC-seq (see Key Resource Table) were downloaded from SRA using SRA toolkit (v2.9.4). Reads were aligned and further analyzed using the ENCODE ATAC-seq pipeline with default parameters using the ENCODE human reference genome GRCh38.15 (Lee, 2016). Bed files containing the global open chromatin landscape of Adult Alpha (Alpha_1; EA4 and Alpha_2; EA28), Beta (Beta_1; EA5 and Beta_2; EA29), Acinar (Acinar_1; EA7 and Acinar_2; EA27) and ductal (EA11) cells or cell type specific differentially accessible regions were used as input for motif discovery by HOMER (v4.10.4) using the 'findMotifsGenome.pl' with options using hg38 with size given (Heinz *et al.*, 2010). The

TFs whose motifs identified by HOMER correspond with TFs identified by pySCENIC are visualized in Figure S2A.

Analysis of publicly available scRNA-seq data

Raw reads for five publicly available scRNA-seq datasets (see Key Resource Table) were downloaded from SRA using SRA toolkit (v2.9.4). Afterwards, reads were aligned to the human reference genome GRCh38.95 using STAR (v2.5.3a) with default parameters followed by the conversion to the coordinate sorted BAM format. Next, the featureCounts command from the "Rsubread" (v1.5.2) package in R (v3.6.1) was used to assign mapped reads to genomic features. Low quality transcriptomes with a mitochondrial contamination greater than 5% and less than 200 expressed genes per cell were excluded from subsequent analyses. The resulting raw count matrix was batch corrected using the FindIntegrationAnchors and IntegrateData functions from the "Seurat" package (v3.1.1) after which subsequent analyses were carried out in the R package "Seurat" (v3.1.4). Gene expression was used to cluster all 7393 cells with UMAP, using Seurat's function RunUMAP.

Clusters for cell type annotation were defined using Seurat's shared nearest neighbour algorithm FindClusters function after which differential expression analysis was performed using Wilcoxon's rank sum test with a minimum cutoff of 0.25 average log fold change and min.pct of 0.25.

pySCENIC

GRNs were inferred using pySCENIC (python implementation of SCENIC, v0.9.15) in Python version 3.6.9 (Aibar *et al.*, 2017). Integrated read counts were used as input to run GENIE3 (Huynh-Thu *et al.*, 2010) which is part of arboreto (v0.1.5). GRNs were subsequently inferred using pySCENIC with the hg38_refseq-r80 motif database and default settings. To control for the stochasticity, which is inherent to pySCENIC, a consensus GRN was generated by merging results from five repeat pySCENIC runs. If regulons were identified in multiple pySCENIC runs, only the regulon with the highest AUC value was retained. Regulon activity represented by AUCell values was used to cluster all 7393 cells with UMAP, using Seurat's RunUMAP function.

All 142 regulons within non-diabetic cell types were visualized using the 'clustermap' function of the Python package "seaborn" (v0.9.0). The z-score for each regulon across all cells was calculated using the z-score parameter of the seaborn 'clustermap' function.

Extended analysis of the target genes of specific regulons was conducted in Cytoscape (v3.7.1) using the iRegulon application (v1.3). The list of target genes of a specific regulon was downloaded from the loom file through the SCope platform (https://github.com/pasquelab/scPancreasAtlas) (Davi*e et al.*, 2018).

Regulon ranking

To quantify the cell-type specificity of a regulon, we utilized an entropy-based strategy as described previously (Suo *et al.*, 2018) using the AUCell matrix as input in MATLAB R2019b. The top 10 most specific regulons were subsequently visualized using the R package ggplot2 (v3.1.1). The complete regulon ranking list is available in Table S2.

Beta cell differentiation and immunocytochemistry

Control iPSC line HEL115.6 (Cosentino *et al.*, 2018) was differentiated into beta cells using a previously published 7-step protocol (Cosentino *et al.*, 2018). At the end of the stage 4, cells were seeded into 24-well Aggrewell 400 microwell plates (Stem Cell Technologies) at a density of $0.9 \cdot 10^6$ cells per well after which differentiation was carried out as described previously (Cosentino *et al.*, 2018).

Stage 7 differentiated beta cells were washed twice with PBS containing 0.5 mM EDTA and incubated in Accumax (SIGMA #A7089) for 8 min at 37 °C after which 50% volume of KnockOut Serum Replacement (ThermoFisher #10828028) was added to stop the reaction. After centrifugation at 400 g for 5 min at room temperature, cells were resuspended in 1 mL HAM's F-10 medium, supplemented as indicated before (Demine *et al.*, 2020). 75,000 cells in 500 µL medium were seeded per square of a Nunc Lab-Tek II ICC chamber (ThermoFisher).

Immunohistochemistry analyses were carried out largely as described previously (Demine *et al.*, 2020), using primary antibodies against the following proteins: INS (Dako (Agilent), IR002, Ready to use solution), GCG (Sigma-Aldrich, G2654, 1/1000) and JUND (Atlas Antibodies, HPA063029, 1/100) and the following secondary antibodies (ThermoFisher, A-11075, A-21206, A-31571, 1/500). Pictures were taken using an Axiovert fluorescence microscope (Zeiss).

Immunohistochemistry

Briefly, 4 µm tissue sections were retrieved from 4% formalin-fixed, paraffin-embedded tissue blocks of normal pancreas tissue (68 year old male, 59 year old male and 64 year old female) from the biobank of the University Hospital in Leuven (Belgium). Two pathologists (M.V.H. and T.R.) separately evaluated all histological sections. This study was approved by the ethical committee of the University Hospital in Leuven (Belgium) (S32980).

Immunohistochemistry analyses were carried out largely as described previously (Ceulemans *et al.*, 2017), using primary antibodies against the following proteins: ASCL2 (Merck, MAB4418, clone 7E2, 1/5000), JUND (Atlas Antibodies, HPA063029, 1/50), HEYL (Atlas Antibodies, HPA076960, 1/200) and INS (Agilent, IR002, 1/100). Pictures were taken using a Leica DMLB (Leica Microsystems).

SCope

The integrated single-cell RNA-seq data and pySCENIC results can be explored interactively in SCope (Davie *et al.*, 2018). Loompy (v2.0.17) (Linnarsson Lab., 2015) was used to create the loom files which were uploaded to SCope. The embedding of the regulon and integrated gene expression based UMAP clustering, as seen in this article, were added to the loom file.

REFERENCES

Ahlqvist, E. *et al.* (2018) 'Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables', *The Lancet Diabetes and Endocrinology*. Lancet Publishing Group, 6(5), pp. 361–369. doi: 10.1016/S2213-8587(18)30051-2.

Aibar, S. *et al.* (2017) 'SCENIC: single-cell regulatory network inference and clustering', *Nature Methods*. Nature Publishing Group, 14(11), pp. 1083–1086. doi: 10.1038/nmeth.4463.

Aljaibeji, H. *et al.* (2019) 'Reduced Expression of PLCXD3 Associates With Disruption of Glucose Sensing and Insulin Signaling in Pancreatic β-Cells', *Frontiers in Endocrinology*. Frontiers Media S.A., 10, p. 735. doi: 10.3389/fendo.2019.00735.

Allen, H. L. *et al.* (2012) 'GATA6 haploinsufficiency causes pancreatic agenesis in humans', *Nature Genetics*, 44(1), pp. 20–22. doi: 10.1038/ng.1035.

Arda, H. E. *et al.* (2018) 'A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas', *Cell Systems*, 7(3), pp. 310-322.e4. doi: 10.1016/j.cels.2018.07.007.

Artner, I. *et al.* (2006) 'An activator of the glucagon gene expressed in developing islet α- and β-cells', *Diabetes*, 55(2), pp. 297–304. doi: 10.2337/diabetes.55.02.06.db05-0946.

Arystarkhova, E. *et al.* (2013) 'Hyperplasia of pancreatic beta cells and improved glucose tolerance in mice deficient in the FXYD2 subunit of Na,K-ATPase', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 288(10), pp. 7077–7085. doi: 10.1074/jbc.M112.401190.

Augsornworawat, P. and Millman, J. R. (2020) 'Single-cell RNA sequencing for engineering and studying human islets', *Current Opinion in Biomedical Engineering*. Elsevier BV, 16, pp. 27–33. doi: 10.1016/j.cobme.2020.06.003.

Baeyens, L. *et al.* (2018) '(Re)generating Human Beta Cells: Status, Pitfalls, and Perspectives', *Physiological Reviews*. American Physiological Society, 98(3), pp. 1143–1167. doi:

10.1152/physrev.00034.2016.

Baranek, C., Sock, E. and Wegner, M. (2005) 'The POU protein Oct-6 is a nucleocytoplasmic shuttling protein.', *Nucleic acids research*, 33(19), pp. 6277–86. doi: 10.1093/nar/gki947.

Baron, M., Veres, A., Wolock, Samuel L, *et al.* (2016) 'A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure.', *Cell systems*. NIH Public Access, 3(4), pp. 346-360.e4. doi: 10.1016/j.cels.2016.08.011.

Bonnefond, A. *et al.* (2013) 'Transcription factor gene MNX1 is a novel cause of permanent neonatal diabetes in a consanguineous family', *Diabetes and Metabolism*, 39(3), pp. 276–280. doi: 10.1016/j.diabet.2013.02.007.

Bray, J. K. *et al.* (2020) 'Loss of RE-1 silencing transcription factor accelerates exocrine damage from pancreatic injury', *Cell Death and Disease*. Springer Nature, 11(2), pp. 1–14. doi: 10.1038/s41419-020-2269-7.

Brunton, H. *et al.* (2020) 'HNF4A and GATA6 Loss Reveals Therapeutically Actionable Subtypes in Pancreatic Cancer', *Cell Reports*. Elsevier B.V., 31(6). doi: 10.1016/j.celrep.2020.107625.

Butler, A. *et al.* (2018) 'Integrating single-cell transcriptomic data across different conditions, technologies, and species', *Nature Biotechnology*. Nature Publishing Group, 36(5), pp. 411–420. doi: 10.1038/nbt.4096.

Butte, A. J. *et al.* (2000) 'Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks', *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), pp. 12182–12186. doi: 10.1073/pnas.220392197.

Byrnes, L. E. *et al.* (2018) 'Lineage dynamics of murine pancreatic development at single-cell resolution', *Nature Communications*. Nature Publishing Group, 9(1), p. 3922. doi: 10.1038/s41467-018-06176-3.

Cantile, M. *et al.* (2009) 'HOX D13 expression across 79 tumor tissue types', *International Journal of Cancer*, 125(7), pp. 1532–1541. doi: 10.1002/ijc.24438.

Carrasco, M. *et al.* (2012) 'GATA4 and GATA6 control mouse pancreas organogenesis', *Journal of Clinical Investigation*. American Society for Clinical Investigation, 122(10), pp. 3504–3515. doi: 10.1172/JCI63240.

Cerdá-Esteban, N. *et al.* (2017) 'Stepwise reprogramming of liver cells to a pancreas progenitor state by

the transcriptional regulator Tgif2', *Nature Communications*. Nature Publishing Group, 8(1), pp. 1–13. doi: 10.1038/ncomms14127.

Ceulemans, A. *et al.* (2017) 'RNA-sequencing-based comparative analysis of human hepatic progenitor cells and their niche from alcoholic steatohepatitis livers', *Cell Death and Disease*. Nature Publishing Group, 8(11), p. e3164. doi: 10.1038/cddis.2017.543.

Chen, S. and Mar, J. C. (2018) 'Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data', *BMC Bioinformatics*. BioMed Central, 19(1), p. 232. doi: 10.1186/s12859-018-2217-z.

Churchill, A. J. *et al.* (2017) 'Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development', *eLife*. eLife Sciences Publications Ltd, 6. doi: 10.7554/eLife.20010.

Cnop, M. *et al.* (2014) 'RNA sequencing identifies dysregulation of the human pancreatic islet transcriptome by the saturated fatty acid palmitate', *Diabetes*. American Diabetes Association Inc., 63(6), pp. 1978–1993. doi: 10.2337/db13-1383.

Cnop, M. *et al.* (2017) 'Endoplasmic reticulum stress and eIF2α phosphorylation: The Achilles heel of pancreatic β cells.', *Molecular metabolism*. Elsevier, 6(9), pp. 1024–1039. doi: 10.1016/j.molmet.2017.06.001.

Coleman, J. D. *et al.* (2013) 'Role of Peroxisome Proliferator-Activated Receptor β/δ and B-Cell Lymphoma-6 in Regulation of Genes Involved in Metastasis and Migration in Pancreatic Cancer Cells.', *PPAR research*, 2013, p. 121956. doi: 10.1155/2013/121956.

Collombat, P. *et al.* (2003) 'Opposing actions of Arx and Pax4 in endocrine pancreas development', *Genes and Development*, 17(20), pp. 2591–2603. doi: 10.1101/gad.269003.

Cosentino, C. *et al.* (2018) 'Pancreatic -cell tRNA hypomethylation and fragmentation link TRMT10A deficiency with diabetes', *Nucleic Acids Research*. Oxford University Press, 46(19), pp. 10302–10318. doi: 10.1093/nar/gky839.

Daraio, T. *et al.* (2017) 'SNAP-25b-deficiency increases insulin secretion and changes spatiotemporal profile of Ca 2+ oscillations in β cell networks', *Scientific Reports*. Nature Publishing Group, 7(1). doi: 10.1038/s41598-017-08082-y.

Dassaye, R., Naidoo, S. and Cerf, M. E. (2016) 'Transcription factor regulation of pancreatic organogenesis, differentiation and maturation', *Islets*. Taylor and Francis Inc., pp. 13–34. doi: 10.1080/19382014.2015.1075687.

Davie, K. *et al.* (2018) 'A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain', *Cell*, 174(4), pp. 982-998.e20. doi: 10.1016/j.cell.2018.05.057.

Delporte, F. M. *et al.* (2008) 'Expression of zebrafish pax6b in pancreas is regulated by two enhancers containing highly conserved cis-elements bound by PDX1, PBX and PREP factors', *BMC Developmental Biology*, 8(1), p. 53. doi: 10.1186/1471-213X-8-53.

Demine, S. *et al.* (2020) 'Pro-inflammatory cytokines induce cell death, inflammatory responses, and endoplasmic reticulum stress in human iPSC-derived beta cells', *Stem cell research & therapy*. NLM (Medline), 11(1), p. 7. doi: 10.1186/s13287-019-1523-3.

Dennis, J. M. *et al.* (2019) 'Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data', *The Lancet Diabetes and Endocrinology*. Lancet Publishing Group, 7(6), pp. 442–451. doi: 10.1016/S2213-8587(19)30087-7.

Dorrell, C. *et al.* (2011) 'Transcriptomes of the major human pancreatic cell types', *Diabetologia*. Springer Verlag, 54(11), pp. 2832–2844. doi: 10.1007/s00125-011-2283-5.

Dybala, M. P. and Hara, M. (2019) 'Heterogeneity of the human pancreatic islet', *Diabetes*. American Diabetes Association Inc., 68(6), pp. 1230–1239. doi: 10.2337/db19-0072.

Eizirik, D. L., Pasquali, L. and Cnop, M. (2020) 'Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure', *Nature Reviews Endocrinology*. Nature Publishing Group, pp. 1–14. doi: 10.1038/s41574-020-0355-7.

Eletto, Daniela *et al.* (2016) 'PDIA6 regulates insulin secretion by selectively inhibiting the RIDD activity of IRE1', *FASEB Journal*. FASEB, 30(2), pp. 653–665. doi: 10.1096/fj.15-275883.

Enge, M. *et al.* (2017) 'Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns.', *Cell*. NIH Public Access, 171(2), pp. 321-330.e14. doi: 10.1016/j.cell.2017.09.004.

Eto, K., Kaur, V. and Thomas, M. K. (2007) 'Regulation of pancreas duodenum homeobox-1 expression by early growth response-1', *Journal of Biological Chemistry*, 282(9), pp. 5973–5983. doi: 10.1074/jbc.M607288200.

Faerch, K., Hulman, A. and P.J. Solomon, T. (2015) 'Heterogeneity of Pre-diabetes and Type 2 Diabetes: Implications for Prediction, Prevention and Treatment Responsiveness', *Current Diabetes Reviews*. Bentham Science Publishers Ltd., 12(1), pp. 30–41. doi: 10.2174/1573399811666150416122903.

Fiers, M. W. E. J. *et al.* (2018) 'Mapping gene regulatory networks from single-cell omics data', *Briefings in Functional Genomics*. Oxford University Press, 17(4), pp. 246–254. doi: 10.1093/bfgp/elx046.

Flamez, D. *et al.* (2010) 'A genomic-based approach identifies FXYD domain containing ion transport regulator 2 (FXYD2)γa as a pancreatic beta cell-specific biomarker', *Diabetologia*, 53(7), pp. 1372–1383. doi: 10.1007/s00125-010-1714-z.

Flanagan, S. E. *et al.* (2014) 'Analysis of transcription factors key for mouse pancreatic development establishes NKX2-2 and MNX1 mutations as causes of neonatal diabetes in man', *Cell Metabolism*, 19(1), pp. 146–154. doi: 10.1016/j.cmet.2013.11.021.

De Franco, E. *et al.* (2019) 'A Specific CNOT1 Mutation Results in a Novel Syndrome of Pancreatic Agenesis and Holoprosencephaly through Impaired Pancreatic and Neurological Development', *American Journal of Human Genetics*. Cell Press, 104(5), pp. 985–989. doi: 10.1016/j.ajhg.2019.03.018.

Gao, N. *et al.* (2010) 'Foxa1 and Foxa2 maintain the metabolic and secretory features of the mature β-cell', *Molecular Endocrinology*, 24(8), pp. 1594–1604. doi: 10.1210/me.2009-0513.

Gao, T. *et al.* (2014) 'Pdx1 maintains β cell identity and function by repressing an α cell program', *Cell Metabolism*. NIH Public Access, 19(2), pp. 259–271. doi: 10.1016/j.cmet.2013.12.002.

Ghazanfar, S. *et al.* (2016) 'Integrated single cell data analysis reveals cell specific networks and novel coactivation markers', *BMC Systems Biology*. BioMed Central Ltd., 10(S5), p. 127. doi: 10.1186/s12918-016-0370-4.

Gierl, M. S. *et al.* (2006) 'The Zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic β cells and intestinal endocrine cells', *Genes and Development*, 20(17), pp. 2465–2478. doi: 10.1101/gad.381806.

Gomez, D. L. *et al.* (2015) 'Neurogenin 3 Expressing Cells in the Human Exocrine Pancreas Have the Capacity for Endocrine Cell Fate', *PLOS ONE*. Edited by Z. Rakonczay. Public Library of Science, 10(8), p. e0133862. doi: 10.1371/journal.pone.0133862.

Good, A. L. *et al.* (2019) 'JUND regulates pancreatic β cell survival during metabolic stress', *Molecular Metabolism*. Elsevier GmbH, 25, pp. 95–106. doi: 10.1016/j.molmet.2019.04.007.

Gradwohl, G. *et al.* (2000) 'neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas', *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), pp. 1607–1611. doi: 10.1073/pnas.97.4.1607.

Grün, D. *et al.* (2016) 'De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data', *Cell Stem Cell*. Cell Press, 19(2), pp. 266–277. doi: 10.1016/j.stem.2016.05.010.

Hafemeister, C. and Satija, R. (2019) 'Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression', *bioRxiv*. Cold Spring Harbor Laboratory, p. 576827. doi: 10.1101/576827.

Han, X. *et al.* (2020) 'Construction of a human cell landscape at single-cell level', *Nature*. Nature Publishing Group, pp. 1–9. doi: 10.1038/s41586-020-2157-4.

Hart, A. W. *et al.* (2013) 'The Developmental Regulator Pax6 Is Essential for Maintenance of Islet Cell Function in the Adult Mouse Pancreas', *PLoS ONE*. Edited by R. Zhou. Public Library of Science, 8(1), p. e54173. doi: 10.1371/journal.pone.0054173.

Haumaitre, C., Lenoir, O. and Scharfmann, R. (2008) 'Histone deacetylase inhibitors modify pancreatic cell fate determination and amplify endocrine progenitors.', *Molecular and cellular biology*. American Society for Microbiology Journals, 28(20), pp. 6373–83. doi: 10.1128/MCB.00413-08.

Heinz, S. *et al.* (2010) 'Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities', *Molecular Cell*, 38(4), pp. 576–589. doi: 10.1016/j.molcel.2010.05.004.

Henquin, J. C. (2018) 'Influence of organ donor attributes and preparation characteristics on the dynamics of insulin secretion in isolated human islets', *Physiological Reports*. American Physiological Society, 6(5), pp. 1–13. doi: 10.14814/phy2.13646.

Hoang, C. Q. *et al.* (2016) 'Transcriptional Maintenance of Pancreatic Acinar Identity, Differentiation, and Homeostasis by PTF1A', *Molecular and Cellular Biology*. American Society for Microbiology, 36(24), pp. 3033–3047. doi: 10.1128/mcb.00358-16.

Hogrebe, N. J. *et al.* (2020) 'Targeting the cytoskeleton to direct pancreatic differentiation of human pluripotent stem cells', *Nature Biotechnology*. Nature Research, 38(4), pp. 460–470. doi: 10.1038/s41587-020-0430-6.

Huynh-Thu, V. A. *et al.* (2010) 'Inferring Regulatory Networks from Expression Data Using Tree-Based Methods', *PLoS ONE*. Edited by M. Isalan, 5(9), p. e12776. doi: 10.1371/journal.pone.0012776.

Hwang, I. H. *et al.* (2016) 'Tetraspanin-2 promotes glucotoxic apoptosis by regulating the JNK/β-catenin signaling pathway in human pancreatic β cells', *FASEB Journal*. FASEB, 30(9), pp. 3107–3116. doi: 10.1096/fj.201600240RR.

Janky, R. *et al.* (2014) 'iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections', *PLoS Computational Biology*. Public Library of Science, 10(7). doi: 10.1371/journal.pcbi.1003731.

Juan-Mateu, J. *et al.* (2017) 'Neuron-enriched RNA-binding proteins regulate pancreatic beta cell function and survival', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 292(8), pp. 3466–3480. doi: 10.1074/jbc.M116.748335.

Kahn, S. E., Cooper, M. E. and Del Prato, S. (2014) 'Pathophysiology and treatment of type 2 diabetes: Perspectives on the past, present, and future', *The Lancet*. Lancet Publishing Group, pp. 1068–1083. doi: 10.1016/S0140-6736(13)62154-6.

Kamisawa, T. *et al.* (2016) 'Pancreatic cancer', *The Lancet*. Lancet Publishing Group, pp. 73–85. doi: 10.1016/S0140-6736(16)00141-0.

Ketola, I. *et al.* (2004) 'Transcription factor GATA-6 is expressed in the endocrine and GATA-4 in the exocrine pancreas', *Molecular and Cellular Endocrinology*. Elsevier Ireland Ltd, 226(1–2), pp. 51–57. doi: 10.1016/j.mce.2004.06.007.

Khoo, C. *et al.* (2012) 'Research resource: The Pdx1 cistrome of pancreatic islets', *Molecular Endocrinology*. The Endocrine Society, 26(3), pp. 521–533. doi: 10.1210/me.2011-1231.

Krentz, N. A. J. *et al.* (2017) 'Phosphorylation of NEUROG3 Links Endocrine Differentiation to the Cell Cycle in Pancreatic Progenitors', *Developmental Cell*. Cell Press, 41(2), pp. 129-142.e6. doi: 10.1016/j.devcel.2017.02.006.

Kropp, P. A., Zhu, X. and Gannon, M. (2019) 'Regulation of the Pancreatic Exocrine Differentiation Program and Morphogenesis by Onecut 1/Hnf6', *CMGH*. Elsevier Inc, 7(4), pp. 841–856. doi: 10.1016/j.jcmgh.2019.02.004.

Kumar, S. and Vinod, P. K. (2019) 'Single-cell transcriptomic analysis of pancreatic islets in health and type 2 diabetes', *International Journal of Advances in Engineering Sciences and Applied Mathematics*. Springer Science and Business Media LLC, 11(2), pp. 105–118. doi: 10.1007/s12572-018-0239-4.

Lawlor, N. *et al.* (2017) 'Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes.', *Genome research*. Cold Spring Harbor Laboratory Press, 27(2), pp. 208–222. doi: 10.1101/gr.212720.116.

Lee, C. S. *et al.* (2002) 'Foxa2 controls Pdx1 gene expression in pancreatic β-cells in vivo', *Diabetes*. American Diabetes Association Inc., 51(8), pp. 2546–2551. doi: 10.2337/diabetes.51.8.2546.

Lee, J. (2016) *kundajelab/atac_dnase_pipelines: ATAC-seq and DNase-seq processing pipeline*. Available at: https://github.com/kundajelab/atac_dnase_pipelines (Accessed: 3 May 2019).

Lee, K. *et al.* (2019) 'FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation', *Cell Reports*. Elsevier B.V., 28(2), pp. 382-393.e7. doi: 10.1016/j.celrep.2019.06.034.

Leung-Theung-Long, S. *et al.* (2005) 'Essential interaction of Egr-1 at an islet-specific response element for basal and gastrin-dependent glucagon gene transactivation in pancreatic α-cells', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 280(9), pp. 7976–7984. doi: 10.1074/jbc.M407485200.

Lim, C. Y. *et al.* (2016) 'BTR: training asynchronous Boolean models using single-cell expression data', *BMC Bioinformatics*. BioMed Central Ltd., 17(1), p. 355. doi: 10.1186/s12859-016-1235-y.

Lima, M. J. *et al.* (2016) 'Generation of functional beta-like cells from human exocrine pancreas', *PLoS ONE*. Public Library of Science, 11(5). doi: 10.1371/journal.pone.0156204.

Liu, J. *et al.* (2014) 'A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine

methylation.', *Nature chemical biology*. NIH Public Access, 10(2), pp. 93–5. doi: 10.1038/nchembio.1432.

Luecken, M. D. *et al.* (2020) 'Benchmarking atlas-level data integration in single-cell genomics', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.05.22.111161. doi: 10.1101/2020.05.22.111161.

Mastracci, T. L. *et al.* (2011) 'Nkx2.2 and Arx genetically interact to regulate pancreatic endocrine cell development and endocrine hormone expression', *Developmental Biology*. Academic Press Inc., 359(1), pp. 1–11. doi: 10.1016/j.ydbio.2011.08.001.

Mastracci, T. L. *et al.* (2013) 'Regulation of Neurod1 Contributes to the Lineage Potential of Neurogenin3+ Endocrine Precursor Cells in the Pancreas', *PLoS Genetics*, 9(2). doi: 10.1371/journal.pgen.1003278.

Mastracci, T. L. and Sussel, L. (2012) 'The endocrine pancreas: insights into development, differentiation, and diabetes', *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(5), pp. 609–628. doi: 10.1002/wdev.44.

Masui, T. *et al.* (2010) 'Replacement of Rbpj With Rbpjl in the PTF1 Complex Controls the Final Maturation of Pancreatic Acinar Cells', *Gastroenterology*. W.B. Saunders, 139(1), pp. 270–280. doi: 10.1053/j.gastro.2010.04.003.

Matsumoto, H. *et al.* (2017) 'SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation', *Bioinformatics*. Oxford University Press, 33(15), pp. 2314–2321. doi: 10.1093/bioinformatics/btx194.

Matsumoto, K. *et al.* (2004) 'Cdx2 expression in pancreatic tumors: Relationship with prognosis of invasive ductal carcinomas', *Oncology Reports*. Spandidos Publications, 12(6), pp. 1239–1243. doi: 10.3892/or.12.6.1239.

Miyazaki, S. *et al.* (2010) 'Nuclear hormone Retinoid X Receptor (RXR) negatively regulates the glucose-stimulated insulin secretion of pancreatic β-cells', *Diabetes*, 59(11), pp. 2854–2861. doi: 10.2337/db09-1897.

Mosedale, M. *et al.* (2012) 'Neurexin-1α contributes to insulin-containing secretory granule docking', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 287(9), pp. 6350–6361. doi: 10.1074/jbc.M111.299081.

Muraro, M. J. *et al.* (2016) 'A Single-Cell Transcriptome Atlas of the Human Pancreas', *Cell Systems*. Cell

Press, 3(4), pp. 385-394.e3. doi: 10.1016/j.cels.2016.09.002.

Nicolino, M. *et al.* (2010) 'A novel hypomorphic PDX1 mutation responsible for permanent neonatal diabetes with subclinical exocrine deficiency', *Diabetes*. American Diabetes Association Inc., 59(3), pp. 733–740. doi: 10.2337/db09-1284.

Nishimura, W., Takahashi, S. and Yasuda, K. (2015) 'MafA is critical for maintenance of the mature beta cell phenotype in mice', *Diabetologia*. Springer Verlag, 58(3), pp. 566–574. doi: 10.1007/s00125-014-3464-9.

Nissim, S. *et al.* (2016) 'Iterative use of nuclear receptor Nr5a2 regulates multiple stages of liver and pancreas development', *Developmental Biology*. Academic Press Inc., 418(1), pp. 108–123. doi: 10.1016/j.ydbio.2016.07.019.

Nostro, M. C. *et al.* (2015) 'Efficient Generation of NKX6-1+ Pancreatic Progenitors from Multiple Human Pluripotent Stem Cell Lines', *Stem Cell Reports*, 4(4), pp. 591–604. doi: 10.1016/j.stemcr.2015.02.017.

Ohno, S. (1979) 'The Number of Genes in the Mammalian Genome and the Need for Master Regulatory Genes', in. Springer, Berlin, Heidelberg, pp. 17–21. doi: 10.1007/978-3-642-81261-3_4.

Olbrot, M. *et al.* (2002) 'Identification of β-cell-specific insulin gene transcription factor RIPE3b1 as mammalian MafA', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 99(10), pp. 6737–6742. doi: 10.1073/pnas.102168499.

Pagliuca, F. W. *et al.* (2014) 'Generation of functional human pancreatic β cells in vitro.', *Cell*. NIH Public Access, 159(2), pp. 428–39. doi: 10.1016/j.cell.2014.09.040.

Park, J. S., Lee, H. J. and Park, C. H. (2016) 'A novel mutation of ABCC8 gene in a patient with diazoxide-unresponsive congenital hyperinsulinism', *Korean Journal of Pediatrics*. Korean Pediatric Society, 59, pp. S116–S120. doi: 10.3345/kjp.2016.59.11.S116.

Parker, S. C. J. *et al.* (2013) 'Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(44), pp. 17921–17926. doi: 10.1073/pnas.1317023110.

Patel, K. A. *et al.* (2017) 'Heterozygous RFX6 protein truncating variants are associated with MODY with reduced penetrance', *Nature Communications*. Nature Publishing Group, 8(1). doi: 10.1038/s41467-017-

00895-9.

Peng, G. *et al.* (2019) 'Molecular architecture of lineage allocation and tissue organization in early mouse embryo', *Nature*. Nature Publishing Group, pp. 528–532. doi: 10.1038/s41586-019-1469-8.

Peterson, Q. P. *et al.* (2020) 'A method for the generation of human stem cell-derived alpha cells', *Nature Communications*. Nature Publishing Group, 11(1), p. 2241. doi: 10.1038/s41467-020-16049-3.

Quilichini, E. *et al.* (2019) 'Pancreatic Ductal Deletion of Hnf1b Disrupts Exocrine Homeostasis, Leads to Pancreatitis, and Facilitates Tumorigenesis', *CMGH*. Elsevier Inc, 8(3), pp. 487–511. doi: 10.1016/j.jcmgh.2019.06.005.

Recio-Boiles, A. *et al.* (2016) 'JNK pathway inhibition selectively primes pancreatic cancer stem cells to TRAIL-induced apoptosis without affecting the physiology of normal tissue resident stem cells', *Oncotarget*. Impact Journals LLC, 7(9), pp. 9890–9906. doi: 10.18632/oncotarget.7066.

Rezania, A. *et al.* (2014) 'Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells', *Nature Biotechnology*. Nature Publishing Group, 32(11), pp. 1121–1133. doi: 10.1038/nbt.3033.

Rorsman, P. and Ashcroft, F. M. (2018) 'Pancreatic β-cell electrical activity and insulin secretion: Of mice and men', *Physiological Reviews*. American Physiological Society, 98(1), pp. 117–214. doi: 10.1152/physrev.00008.2017.

Rubio-Cabezas, O. *et al.* (2010) 'Homozygous mutations in NEUROD1 are responsible for a novel syndrome of permanent neonatal diabetes and neurological abnormalities', *Diabetes*. American Diabetes Association Inc., 59(9), pp. 2326–2331. doi: 10.2337/db10-0011.

Rubio-Cabezas, O. *et al.* (2011) 'Permanent neonatal diabetes and enteric anendocrinosis associated with biallelic mutations in NEUROG3', *Diabetes*, 60(4), pp. 1349–1353. doi: 10.2337/db10-1008.

Russ, H. A. *et al.* (2015) 'Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro.', *The EMBO journal*. European Molecular Biology Organization, 34(13), pp. 1759–72. doi: 10.15252/embj.201591058.

Sakikubo, M. *et al.* (2018) 'Ptf1a inactivation in adult pancreatic acinar cells causes apoptosis through activation of the endoplasmic reticulum stress pathway', *Scientific Reports*. Nature Publishing Group, 8(1).

doi: 10.1038/s41598-018-34093-4.

Van de Sande, B. *et al.* (2020) 'A scalable SCENIC workflow for single-cell gene regulatory network analysis', *Nature Protocols*. Nature Research, 15(7), pp. 2247–2276. doi: 10.1038/s41596-020-0336-2.

Sato, F. *et al.* (2012) 'The basic helix-loop-helix transcription factor DEC2 inhibits TGF-β-induced tumor progression in human pancreatic cancer BxPC-3 cells', *International Journal of Molecular Medicine*, 30(3), pp. 495–501. doi: 10.3892/ijmm.2012.1037.

Scavuzzo, M. A. *et al.* (2018) 'Endocrine lineage biases arise in temporally distinct endocrine progenitors during pancreatic morphogenesis', *Nature Communications*. Nature Publishing Group, 9(1), p. 3356. doi: 10.1038/s41467-018-05740-1.

Schuijers, J. *et al.* (2015) 'Ascl2 acts as an R-spondin/wnt-responsive switch to control stemness in intestinal crypts', *Cell Stem Cell*. Cell Press, 16(2), pp. 158–170. doi: 10.1016/j.stem.2014.12.006.

Segerstolpe, Å. *et al.* (2016) 'Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes', *Cell Metabolism*. Cell Press, 24(4), pp. 593–607. doi: 10.1016/J.CMET.2016.08.020.

Sellick, G. S. *et al.* (2004) 'Mutations in PTF1A cause pancreatic and cerebellar agenesis', *Nature Genetics*, 36(12), pp. 1301–1305. doi: 10.1038/ng1475.

Senée, V. *et al.* (2006) 'Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism', *Nature Genetics*, 38(6), pp. 682–687. doi: 10.1038/ng1802.

Sharon, N. *et al.* (2019) 'Wnt Signaling Separates the Progenitor and Endocrine Compartments during Pancreas Development', *Cell Reports*. Elsevier B.V., 27(8), pp. 2281-2291.e5. doi: 10.1016/j.celrep.2019.04.083.

Shaw-Smith, C. *et al.* (2014) 'GATA4 mutations are a cause of neonatal and childhood-onset diabetes', *Diabetes*. American Diabetes Association Inc., 63(8), pp. 2888–2894. doi: 10.2337/db14-0061.

Shih, H. P. *et al.* (2015) 'A Gene Regulatory Network Cooperatively Controlled by Pdx1 and Sox9 Governs Lineage Allocation of Foregut Progenitor Cells', *Cell Reports*, 13(2), pp. 326–336. doi: 10.1016/j.celrep.2015.08.082.

Shin, S. *et al.* (2009) 'Activator protein-1 has an essential role in pancreatic cancer cells and is regulated

by a novel Akt-mediated mechanism', *Molecular Cancer Research*. NIH Public Access, 7(5), pp. 745–754. doi: 10.1158/1541-7786.MCR-08-0462.

Shroff, S. *et al.* (2014) 'SOX9: A useful marker for pancreatic ductal lineage of pancreatic neoplasms', *Human Pathology*, 45(3), pp. 456–463. doi: 10.1016/j.humpath.2013.10.008.

Sinisa Hrvatin, Charles W. O'Donnell, Francis Deng, Jeffrey R. Millman, Felicia Walton Pagliuca, Philip DiIorio, Alireza Rezania, David K. Gifford, and D. A. M. (2014) 'Differentiated human stem cells resemble fetal, not adult, β cells', *PNAS . 111 (8) 3038-3043; https://doi.org/10.1073/pnas.1400709111*. National Academy of Sciences. doi: 10.1073/pnas.96.18.10152.

Smith, S. B. *et al.* (2010) 'Rfx6 directs islet formation and insulin production in mice and humans', *Nature*, 463(7282), pp. 775–780. doi: 10.1038/nature08748.

Solomon, B. D. *et al.* (2009) 'Compound heterozygosity for mutations in PAX6 in a patient with complex brain anomaly, neonatal diabetes mellitus, and microophthalmia', *American Journal of Medical Genetics, Part A*, 149(11), pp. 2543–2546. doi: 10.1002/ajmg.a.33081.

Sosa-Pineda, B. *et al.* (1997) 'The Pax4 gene is essential for differentiation of insulin-producing β cells in the mammalian pancreas', *Nature*, 386(6623), pp. 399–402. doi: 10.1038/386399a0.

Stanescu, D. E. *et al.* (2017) 'Single cell transcriptomic profiling of mouse pancreatic progenitors', *Physiological Genomics*. American Physiological Society, 49(2), pp. 105–114. doi: 10.1152/physiolgenomics.00114.2016.

Stoffers, D. A. *et al.* (1997) 'Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence', *Nature Genetics*, 15(1), pp. 106–110. doi: 10.1038/ng0197-106.

Stuart, T. *et al.* (2019) 'Comprehensive Integration of Single-Cell Data', *Cell*. Cell Press, 177(7), pp. 1888-1902.e21. doi: 10.1016/j.cell.2019.05.031.

Sund, N. J. *et al.* (2001) 'Tissue-specific deletion of Foxa2 in pancreatic β cells results in hyperinsulinemic hypoglycemia', *Genes and Development*. Cold Spring Harbor Laboratory Press, 15(13), pp. 1706–1715. doi: 10.1101/gad.901601.

Suo, S. *et al.* (2018) 'Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas', *Cell Reports*. Cell Press, 25(6), pp. 1436-1445.e3. doi: 10.1016/J.CELREP.2018.10.045.

Sussel, L. *et al.* (1998) 'Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic β cells', *Development*, 125(12), pp. 2213–2221.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.

Thomas, G., Brown, A. L. and Brown, J. M. (2014) 'In vivo metabolite profiling as a means to identify uncharacterized lipase function: Recent success stories within the alpha beta hydrolase domain (ABHD) enzyme family', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. Elsevier B.V., pp. 1097–1101. doi: 10.1016/j.bbalip.2014.01.004.

Tosti, L. *et al.* (2019) 'Single nucleus RNA sequencing maps acinar cell states in a human pancreas cell atlas', *bioRxiv*. Cold Spring Harbor Laboratory, p. 733964. doi: 10.1101/733964.

Vethe, H. *et al.* (2019) 'The Effect of Wnt Pathway Modulators on Human iPSC-Derived Pancreatic Beta Cell Maturation', *Frontiers in Endocrinology*. Frontiers Media S.A., 10(MAY), p. 293. doi: 10.3389/fendo.2019.00293.

Vierbuchen, T. *et al.* (2010) 'Direct conversion of fibroblasts to functional neurons by defined factors', *Nature*. Nature Publishing Group, 463(7284), pp. 1035–1041. doi: 10.1038/nature08797.

Vierra, N. C. *et al.* (2015) 'Type 2 diabetes-associated K+ channel TALK-1 modulates β-cell electrical excitability, second-phase insulin secretion, and glucose homeostasis', *Diabetes*. American Diabetes Association Inc., 64(11), pp. 3818–3828. doi: 10.2337/db15-0280.

Wang, Y. J. *et al.* (2016) 'Single-Cell Transcriptomics of the Human Endocrine Pancreas.', *Diabetes*. American Diabetes Association, 65(10), pp. 3028–38. doi: 10.2337/db16-0405.

Weedon, M. N. *et al.* (2014) 'Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis.', *Nature genetics*. Europe PMC Funders, 46(1), pp. 61–64. doi: 10.1038/ng.2826.

Weng, C. C. *et al.* (2019) 'Loss of the transcriptional repressor TGIF1 results in enhanced Kras-driven development of pancreatic cancer', *Molecular Cancer*. BioMed Central Ltd., 18(1). doi: 10.1186/s12943-019-1023-1.

Westacott, M. J. *et al.* (2017) 'Age-dependent decline in the coordinated [Ca2+] and insulin secretory

dynamics in human pancreatic islets', *Diabetes*. American Diabetes Association Inc., 66(9), pp. 2436–2445. doi: 10.2337/db17-0137.

Wierup, N. *et al.* (2002) 'The ghrelin cell: A novel developmentally regulated islet cell in the human pancreas', *Regulatory Peptides*, 107(1–3), pp. 63–69. doi: 10.1016/S0167-0115(02)00067-8.

Wouters, J. *et al.* (2020) 'Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma', *Nature Cell Biology*. Nature Research, 22(8), pp. 986–998. doi: 10.1038/s41556-020-0547-3.

Xin, Y. *et al.* (2016) 'RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes', *Cell Metabolism*. Cell Press, 24(4), pp. 608–615. doi: 10.1016/J.CMET.2016.08.018.

Xu, H. *et al.* (2017) 'Elevated ASCL2 expression in breast cancer is associated with the poor prognosis of patients', *American Journal of Cancer Research*. E-Century Publishing Corporation, 7(4), pp. 955–961.

Yamashita, J. *et al.* (2017) 'Skn-1a/Pou2f3 functions as a master regulator to generate Trpm5-expressing chemosensory cells in mice', *PLoS ONE*. Public Library of Science, 12(12). doi: 10.1371/journal.pone.0189340.

Yang, L. *et al.* (2020) 'A Human Pluripotent Stem Cell-based Platform to Study SARS-CoV-2 Tropism and Model Virus Infection in Human Cells and Organoids', *Cell Stem Cell*. Cell Press, 27(1), pp. 125-136.e7. doi: 10.1016/j.stem.2020.06.015.

Yu, J. H. and Kim, H. (2012) 'Role of janus kinase/signal transducers and activators of transcription in the pathogenesis of pancreatitis and pancreatic cancer', *Gut and Liver*, pp. 417–422. doi: 10.5009/gnl.2012.6.4.417.

Zaharia, O. P. *et al.* (2019) 'Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study', *The Lancet Diabetes and Endocrinology*. Lancet Publishing Group, 7(9), pp. 684–694. doi: 10.1016/S2213-8587(19)30187-1.

Zhou, Q. *et al.* (2008) 'In vivo reprogramming of adult pancreatic exocrine cells to β-cells', *Nature*, 455(7213), pp. 627–632. doi: 10.1038/nature07314.

Zhu, R. *et al.* (2012) 'Ascl2 knockdown results in tumor growth arrest by mirna-302b-related inhibition of colon cancer progenitor cells', *PLoS ONE*. Public Library of Science, 7(2). doi:

10.1371/journal.pone.0032170.

Zhu, Z. *et al.* (2016) ‘Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes’, *Cell Stem Cell*. Cell Press, 18(6), pp. 755–768. doi: 10.1016/J.STEM.2016.03.015.

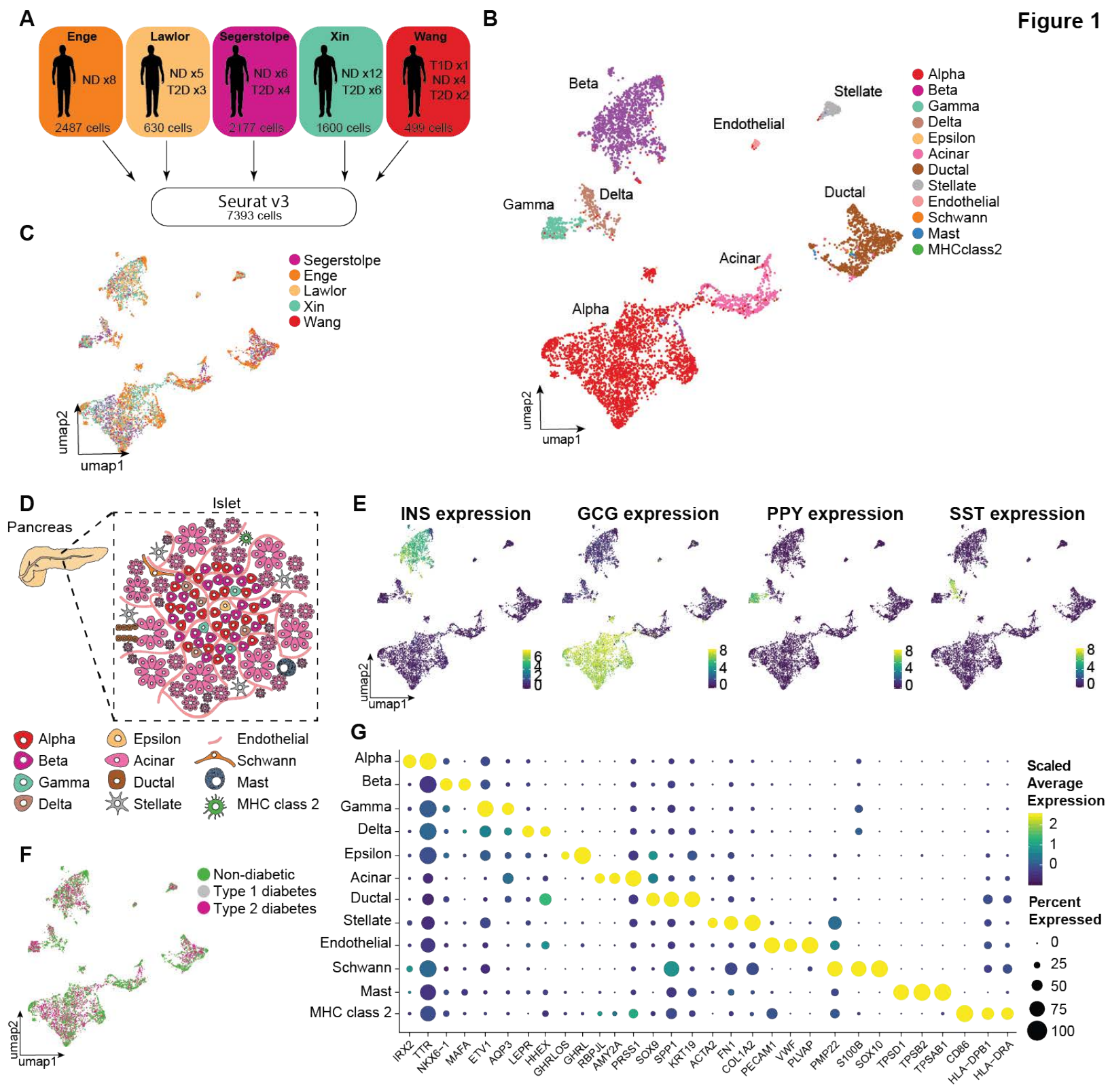

Figure 1. Integrated Analysis of scRNA-seq Identifies 12 Pancreatic Cell Types.
(A) Schematic of workflow used in this paper. Cells from five public datasets were processed uniformly from raw SRA files and integrated, resulting in one dataset of 7393 cells of 51 individuals. **(B)** Integrated gene expression based UMAP of 7393 single cells annotated by cell type. **(C)** Gene expression based UMAP of 7393 single cells annotated by dataset of origin. **(D)** Schematic overview of diverse cell types in the human adult pancreas. Created with BioRender.com **(E)** Gene expression based UMAP of 7393 single cells colored by non-integrated INS, GCG, PPY and SST expression.**(F)** Integrated gene expression based UMAP of 7393 single cells annotated by disease status. **(G)** Bubble plot showing the expression of various known marker genes across all annotated cell types. The bubble size is proportional to the percentage of cells that express a specific marker gene with the colour scale representing the average non-integrated scaled gene expression within the specific cell population.
MHC: major histocompatibility complex

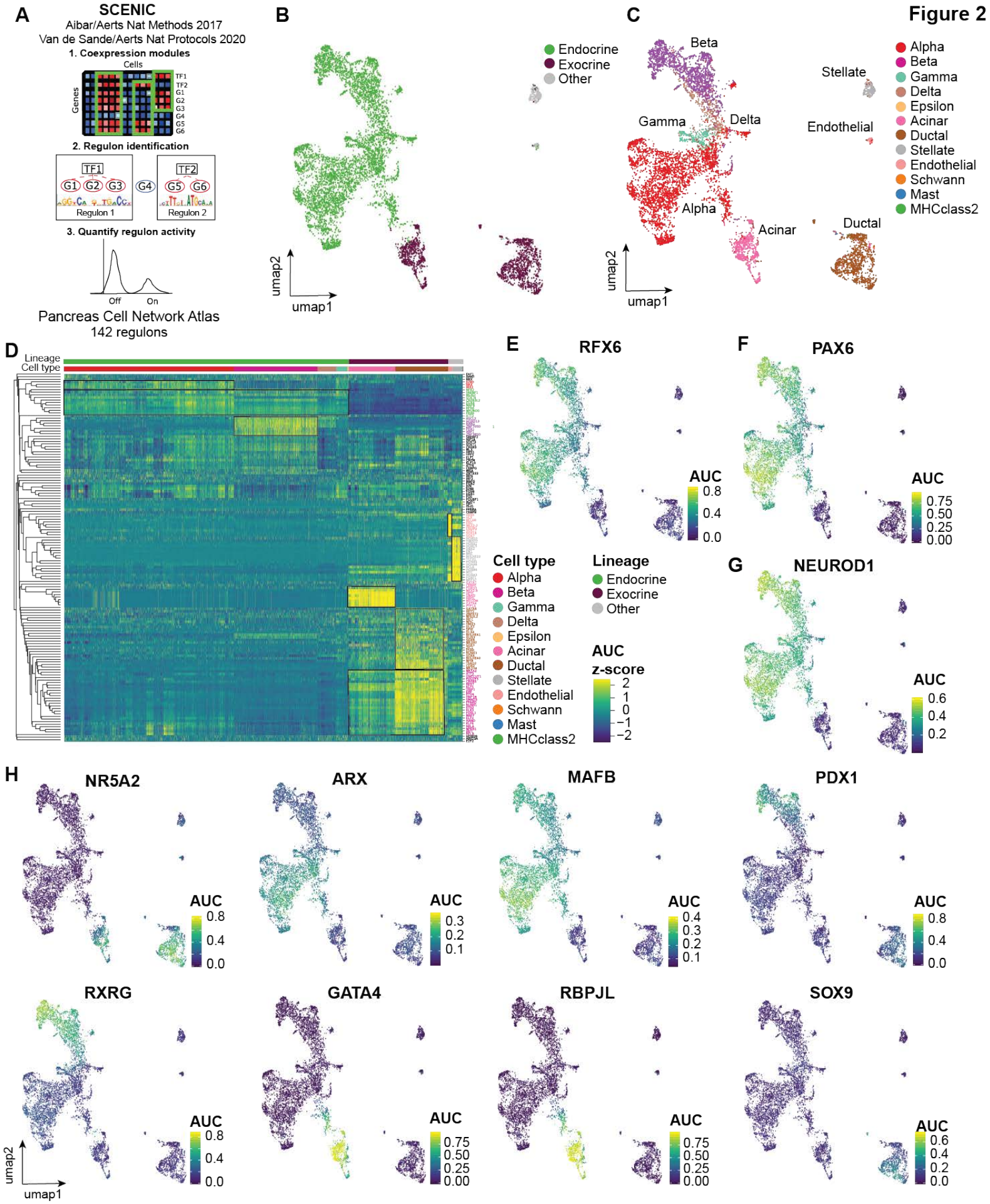

Figure 2. Reconstruction of Gene Regulatory Networks in the Human Adult Pancreas

(A) Schematic overview of the pySCENIC workflow used in this study. **(B-C)** Regulon activity based UMAP of 7393 single cells colored by pancreatic lineage (B) and cell type annotation (C). Endocrine cells include alpha, beta, epsilon, gamma and delta cells. Exocrine cells include acinar and ductal cells. Other cells represent schwann, MHC class 2, mast, stellate and endothelial cells. **(D)** Heatmap based on unsupervised clustering of all 142 regulons (rows) for 4795 non-diabetic cells (columns). Color scaling is based on a z-score calculated based on the activity of each regulon. **(E-G)** Regulon activity based UMAP colored by the regulon activity of RFX6 (E), PAX6 (F), NEUROD1 (G). **(H)** Regulon activity based UMAP colored by the regulon activity of NR5A2, ARX, MAFB, PDX1, RXRG, GATA4, RBPJL and SOX9.

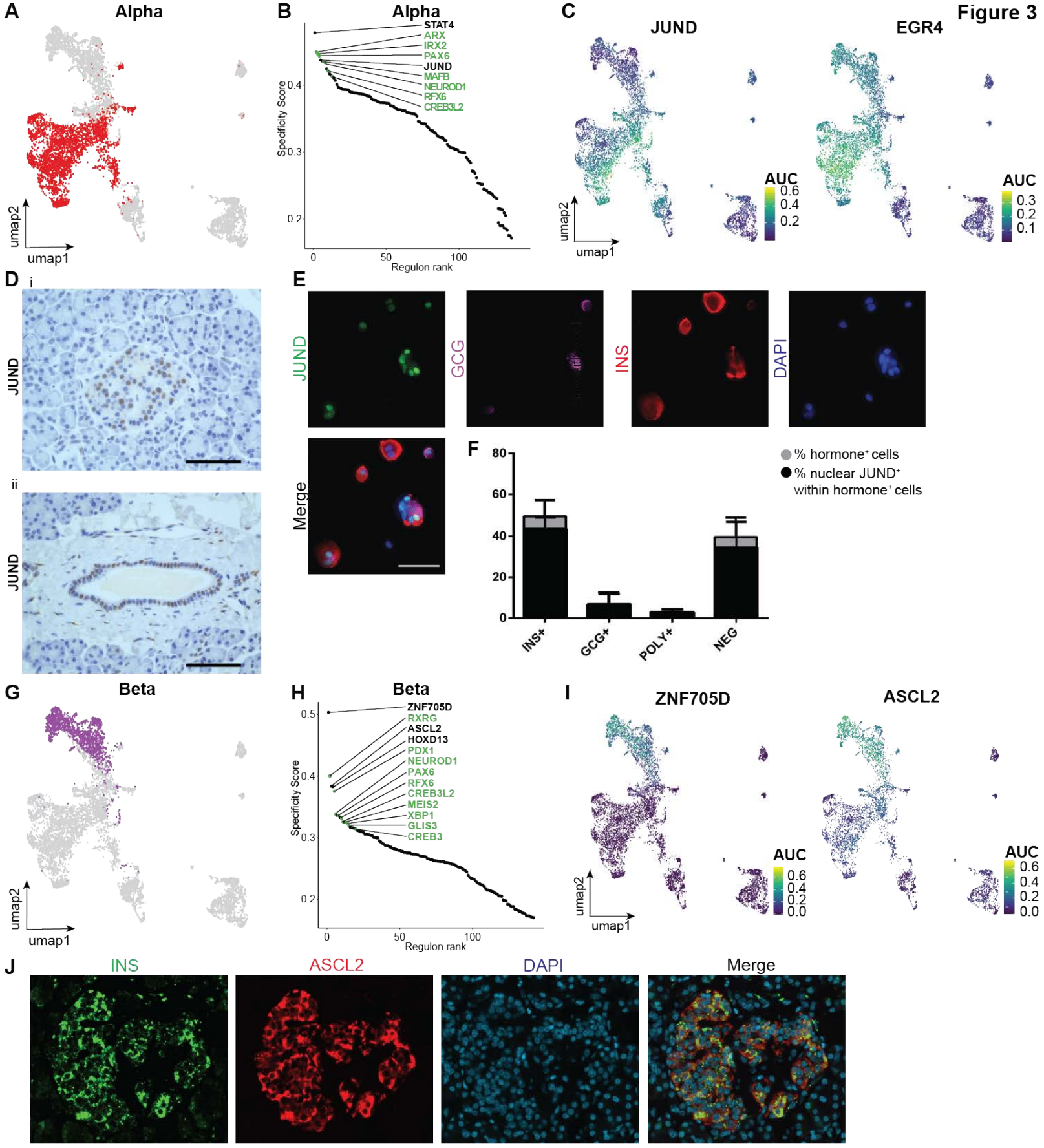

Figure 3. Prediction of Regulators of Endocrine Cell Identity in the Human Adult Pancreas
(A) Regulon activity based UMAP of 7393 single cells with the alpha cell population highlighted. **(B)** Top 10 ranked regulons in non-diabetic alpha cells based on regulon specificity score (RSS). Known cell type specific regulons are colored green. **(C)** Regulon activity based UMAP colored by the regulon activity of JUND and EGR4 showing the cell type specificity of regulons. **(D)** Immunohistochemical evaluation of JUND in healthy human adult pancreata showing nuclear JUND expression in alpha, beta (i) and ductal (ii) cells. Representative images are shown (n = 3 patients), Scale bar: 50 μm. **(E)** Immunofluorescence evaluation of JUND, INS, GCG and SST at stage 7 of human iPSC beta cell differentiation. Box indicates a GCG+ cell with nuclear JUND expression. Representative images examined for JUND (green), INS (red), GCG (purple) and DAPI (blue, nuclei counterstaining) are shown (n=3 beta cell differentiations). Scale bars: 50 μm. **(F)** The percentage of cells with nuclear JUND expression within the INS+, GCG+, polyhormonal and INS+/GCG+ cell populations. Results are shown as the normalized mean±s.d. (n=3 beta cell differentiations). **(G)** Regulon activity based UMAP of 7393 single cells with the beta cell population highlighted. **(H)** Top 10 ranked regulons in non-diabetic beta cells based on regulon specificity score (RSS). Known cell type specific regulons are colored green. **(I)** Regulon activity based UMAP colored by the regulon activity of ZNF705D and ASCL2 showing the cell type specificity of regulons. **(J)** Immunohistochemical evaluation of ASCL2 in healthy human adult pancreata showing cytoplasmic ASCL2 expression in mainly beta cells. Representative images examined for INS (green), ASCL2 (red) and DAPI (blue, nuclei counterstaining) are shown (n = 3 patients), Scale bar: 50 μm.

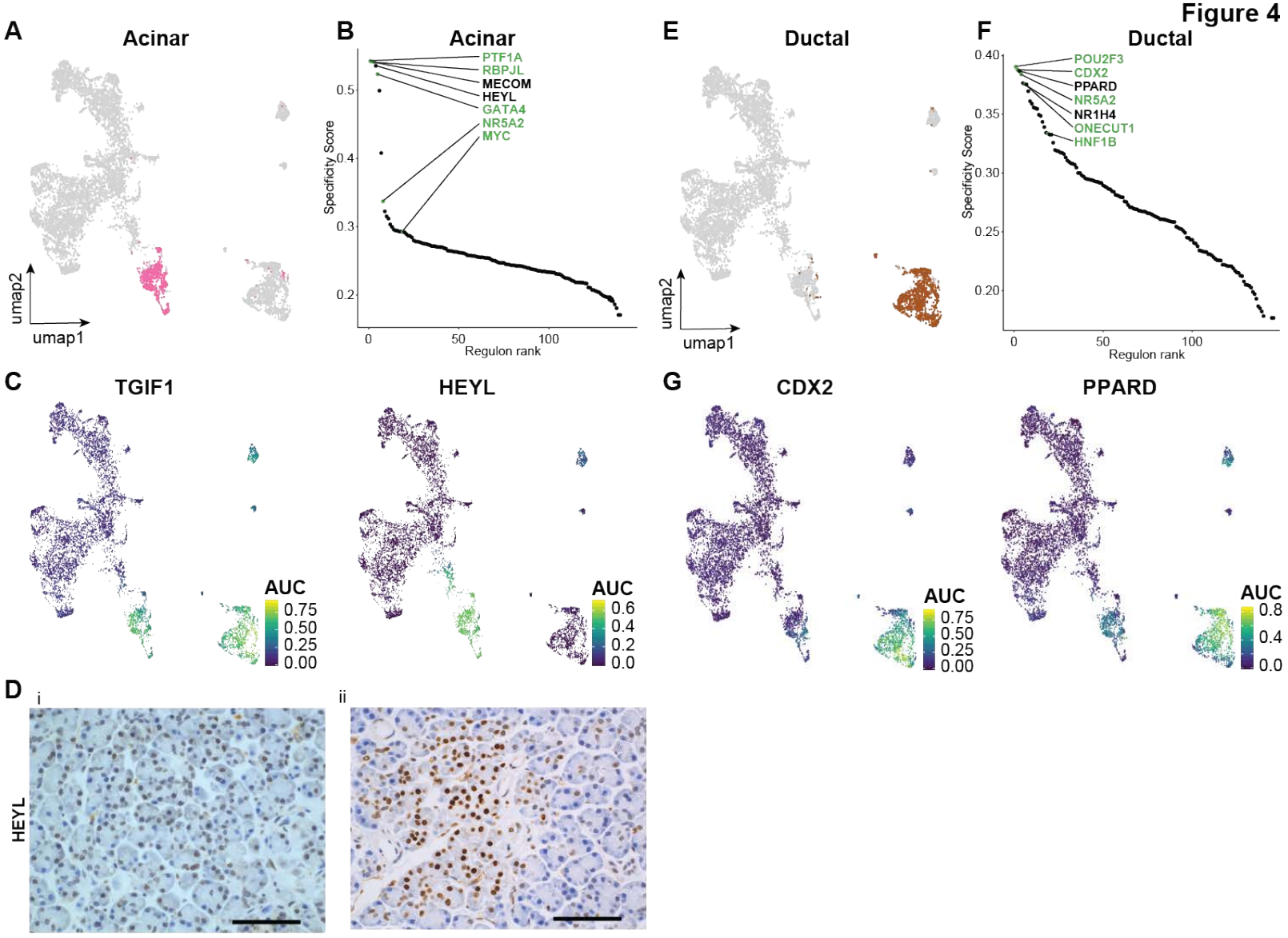

Figure 4. Prediction of Regulators of Exocrine Cell Identity in the Human Adult Pancreas
(A) Regulon activity based UMAP of 7393 single cells with the acinar cell population highlighted. **(B)** Top 10 ranked regulons in non-diabetic acinar cells based on regulon specificity score (RSS). Known cell type specific regulons are colored green. **(C)** Regulon activity based UMAP colored by the regulon activity of TGIF1 and HEYL showing the cell type specificity of regulons. **(D)** Immunohistochemical evaluation of HEYL in healthy human adult pancreata showing nuclear HEYL expression in acinar (i) and islet (ii) cells. Representative images are shown (n = 3 patients), Scale bar: 50 μm. **(E)** Regulon activity based UMAP of 7393 single cells with the ductal cell population highlighted. **(F)** Top 10 ranked regulons in non-diabetic beta cells based on regulon specificity score (RSS). Known cell type specific regulons are colored green. **(G)** Regulon activity based UMAP colored by the regulon activity of CDX2 and PPARD showing the cell type specificity of regulons.

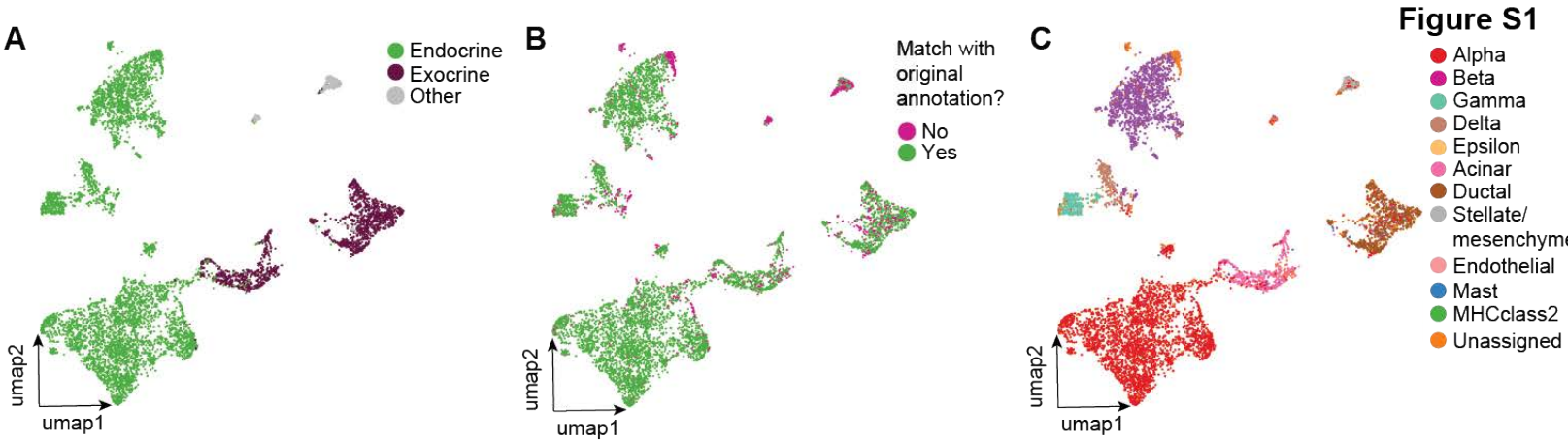

Figure S1. Integrated Analysis of scRNA-seq Identifies 12 Pancreatic Cell Types
(A) Gene expression based UMAP of 7393 single cells colored by pancreatic lineage. Endocrine cells include alpha, beta, epsilon, gamma and delta cells. Exocrine cells include acinar and ductal cells. Other cells represent schwann, MHC class 2, mast, stellate and endothelial cells. **(B)** Gene expression based UMAP of 7393 single cells showing the large degree of overlap of novel and original cell type annotation. Cells were annotated as matching when the novel cell type annotation corresponded with the original annotation from each paper. **(C)** Gene expression based UMAP of 7393 single cells annotated by the original cell type annotation defined in the individual papers.

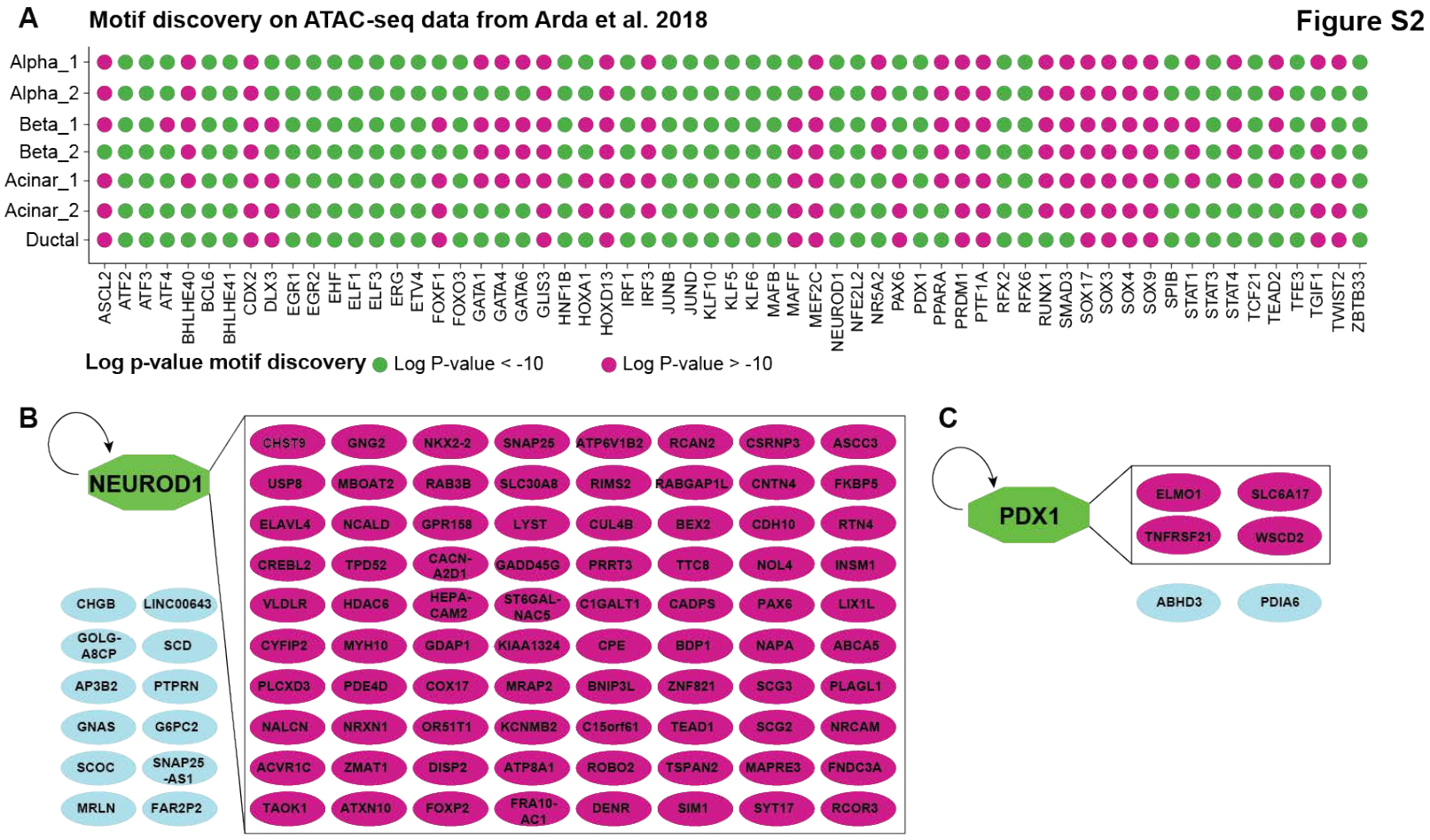

Figure S2. Reconstruction of Gene Regulatory Networks in the Human Pancreas Predicts a Small Set of Essential Regulators for Each Cell Type
(A) Bubble plot summarizing the enrichment of pySCENIC related transcription factor motifs. Bulk ATAC-seq data of FACS sorted alpha, beta, ductal and acinar cells from (Arda et al., 2018) was downloaded and analysed after which motif discovery by HOMER based motif discovery was performed on the global open chromatin landscape. Significantly enriched motifs, with logarithmic p-values below -10, are represented in green. TF motifs with logarithmic p-values above -10 were considered not significant and are represented in purple. **(B)** Extended analysis of putative NEUROD1 target genes. NEUROD1 is colored green. Target genes are color coded (purple/blue) based on whether they are part or not of the network as defined by iRegulon. **(C)** Extended analysis of putative PDX1 target genes. PDX1 is colored green. Target genes are color coded (purple/blue) based on whether they are part or not of the network as defined by iRegulon.

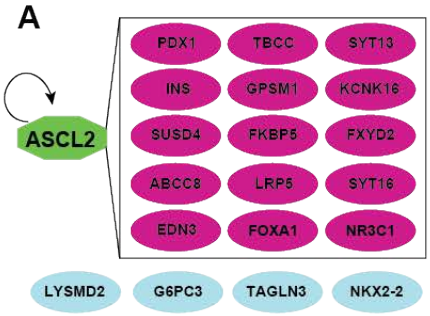

Figure S3. ASCL2 Target Genes can be Linked to Beta Cell Functioning
(A) Extended analysis of the target genes of the ASCL2 regulon by iRegulon. ASCL2 is colored green. Target genes are color coded (purple/blue) based on whether they are part of a network.

SUPPLEMENTARY TABLES AND FILES

Table S1. **Overview of donor information and number of cell types identified for each donor.**
This table is related to Figure 1, 2, 3 and 4.

Table S2. Overview of marker genes used to annotate cell populations.
This table is related to Figure 1.

Table S3. Overview of all 142 regulons.
This table is related to Figure 2, 3 and 4.

Table S4. Overview of regulon ranking.
List of regulon specificity scores for all 142 regulons of non-diabetic alpha, beta, acinar and ductal cells.
This table is related to Figure 3 and 4.

Table S5. List of putative target genes for each regulon.
List of putative target genes for each regulon. This table is related to Figure 2, 3 and 4.