Human pancreatic α-cell heterogeneity and trajectory inference analyses reveal *SMOC1* as a β-cell dedifferentiation gene

Randy B. Kang [1], Miguel Varela[1], Eunjin Oh [1], Jungeun Lee[1], Tuo Zhang [2], Esra Karakose [3], Andrew F. Stewart[3], Donald K. Scott [3], Debbie C. Thurmond [1], Adolfo Garcia-Ocana [1,4] ✉ & Geming Lu [1,4] ✉

β-cell dysfunction and dedifferentiation towards an α-cell-like phenotype are hallmarks of type 2 diabetes. However, the cell subtypes involved in β-to-α-cell transition are unknown. Using single-cell and single-nucleus RNA-seq, RNA velocity, PAGA/cell trajectory inference, and gene commonality, we interrogated α-β-cell fate switching in human islets. We found five α-cell subclusters with distinct transcriptomes. PAGA analysis showed bifurcating cell trajectories in non-diabetic while unidirectional cell trajectories from β-to-α-cells in type 2 diabetes islets suggesting dedifferentiation towards α-cells. Ten genes comprised the common signature genes in trajectories towards α-cells. Among these, the α-cell gene *SMOC1* was expressed in β-cells in type 2 diabetes. Enhanced SMOC1 expression in β-cells decreased insulin expression and secretion and increased β-cell dedifferentiation markers. Collectively, these studies reveal differences in α-β-cell trajectories in non-diabetes and type 2 diabetes human islets, identify signature genes for β-to-α-cell trajectories, and discover SMOC1 as an inducer of β-cell dysfunction and dedifferentiation.

Glucagon-producing pancreatic α-cells are recognized as important physiological regulators of life-threatening hypoglycemia by counteracting the effects of insulin on glucose homeostasis[1,2]. In diabetes, patients display postprandial hyperglucagonemia, which exacerbates hyperglycemia[3–6]. Despite these important observations, initial studies on human pancreatic islet cell heterogeneity have focused mainly on insulin-producing β-cells. Recently, however, more studies have emerged describing specific human islet α-cell subpopulations that participate in normal and dysregulated glucose homeostasis[7–10]. Indeed, recent evidence indicates that there is variation in glucagon content among human α-cells[7] and that α-cell functional heterogeneity is linked to α-cell maturation in type 2 diabetes (T2D)[8]. Furthermore, the subpopulation of glucagon-like peptide-1 (GLP-1) secreting α-cells is increased in human T2D islets[9]. Finally, elevated serum amino acids induce a subpopulation of α-cells to initiate pancreatic neuroendocrine tumor formation[10]. Collectively, these studies indicate that analyzing α-cell heterogeneity and the mechanisms controlling their identity can be of great importance in health and disease.

Single-cell RNA sequencing (scRNA-seq) has revolutionized the identification and analysis of different cell types within heterogeneous cell populations. By algorithmically clustering the data, it is possible to annotate distinct cell types, and with varying hyperparameters for granularity such as Louvain resolution, we can investigate cell subpopulations with distinct transcriptomes in an unbiased manner[11]. In the human islet, scRNA-seq has uncovered several α- and β-cell subtypes with different transcriptome profiles that can predict the maturity of these islet cell subtypes in basal and diabetic conditions[12–19]. However, how these human α- and β-cell subtypes can

[1]Department of Molecular and Cellular Endocrinology, Arthur Riggs Diabetes and Metabolism Research Institute, Beckman Research Institute of City of Hope, Duarte, CA 91010, USA. [2]Genomics Resources Core Facility, Weill Cornell Medicine, New York, NY 10065, USA. [3]Diabetes, Obesity and Metabolism Institute, and Division of Endocrinology, Diabetes and Bone Diseases, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [4]These authors contributed equally: Adolfo Garcia-Ocana, Geming Lu. ✉e-mail: agarciaocana@coh.org; gelu@coh.org

transcriptionally transition from one cell subtype to another remains understudied. Furthermore, T2D is characterized by a decrease in functional β-cells, in part due to dedifferentiation and conversion to other endocrine cells, including glucagon-producing α-like-cells[20,21], a concept that may in part support the hyperglucagonemia encountered in diabetes[3–6]. However, whether specific β-cell subpopulations can transition into α-cells, and whether there is a specific gene signature involved in this process, is unknown.

In the current study using scRNA-seq and snRNA-seq of human islets isolated from adult non-diabetic donors, we identified five *GCG*-expressing α-cell subtypes with different transcriptome profiles. We found that one of the α-cell subtypes, named "AB cells" is a multi-hormonal α-cell subpopulation that could potentially transition in a bifurcated manner into either mature α- or β-cells. However, trajectory analysis of scRNA-seq data from human islets isolated from T2D donors obtained from the Human Islet Research Network (HIRN)-Human Pancreas Analysis Program (HPAP) database showed uni-directional trajectories from β-cells to α-cells suggesting potential pressure on β-cells to become less differentiated or converted to α-like cells. Analysis of common genes on the trajectories from β- to α-cells in human islets isolated from adult non-diabetic donors identified S*MOC1, PLCE1, PAPPA2, ZNF331, ALDH1A1, SLC30A8, BTG2, TM4SF4, NR4A1* and *PCSK2* as signature genes. Among these, *SMOC1* (SPARC-related modular calcium-binding protein-1) which encodes for an extracellular glycoprotein of the SPARC (secreted protein, acidic and rich in cysteine)-related modular calcium-binding protein family, was recently identified as one of the top islet-derived genes encoding a secreted protein in obese mice[22–29]. However, the role of SMOC1 in islet cells is unknown. Here, we show that *SMOC1* is expressed mostly in α-cells in adult human non-diabetic islets and its expression negatively correlated with *INS* expression. Interestingly, SMOC1 mRNA and protein were detected in T2D β-cells. SMOC1 expression in non-diabetic human islets and EndoC-βH1 cells reduced *INS* expression, diminished glucose-stimulated insulin secretion (GSIS), decreased the expression of β-cell identity genes and enhanced the appearance of T2D α-cell-like and T2D β-cell-like gene features in these cells. Collectively, these studies identify α-cell subpopulations and analyze transcriptional transitions between β-cells and α-cells as a function of the pathophysiological context. We also identify *SMOC1* as a gene of previously unrecognized importance and relevance to β-cell dedifferentiation in T2D.

Results

scRNA-seq and snRNA-seq distinguish five α-cell subtypes in human islets

We have previously shown that integrated scRNA-seq and snRNA-seq analysis of human islets can distinguish three β-cell subtypes with different transcriptome profiles[17]. Combining the datasets from both RNA-seq platforms increases analytical power by providing additional information on cytoplasmic and nuclear transcriptomes. Using the same dataset and integrated reference, we aimed here to define α-cell subtypes in human islets (Fig. 1a, Supplementary Data 1). After initial sub-setting (Fig. 1b), we subclustered the α-cell cluster with a Louvain resolution of 0.8. We grouped resulting minor α-cell clusters containing fewer than ten cells into clusters with closest proximity, yielding five α-cell subclusters: α1, α2, α3, α4, and AB (an α-cell subpopulation with *INS* expression) (Fig. 1c, d). We employed this unbiased strategy instead of marker-based subcluster assignment since it considers both the entire transcriptome and the multiple PCA dimensions assigned to calculate clusters. The total number of cells analyzed per subcluster and transcriptomics platform in the three human islet preparations appear in Supplementary Fig. 1a. The α1, α2 and α3 subclusters comprised most α-cells (77–80%) and their proportions were similar between scRNA- and snRNA-seq datasets. In contrast, α4 and AB subclusters were different in proportion between

scRNA- and snRNA-seq datasets (Fig. 1e), with a larger proportion of α4 cells (20% vs. 8%) and fewer AB cells (3% vs. 12%) in scRNA-seq than in snRNA-seq, respectively. This suggests that snRNA-seq with pre-mRNA analysis can reveal more AB cells (α-cell with *INS* expression, see below) than scRNA-seq.

Gene expression and pathway enrichment analysis of human islet α-cell subtypes

GCG and *ALDH1A1* were highly expressed across the five α-cell sub-clusters in both scRNA- and snRNA-seq datasets. *TTR* and *CRYBA2* showed substantially higher expression levels in scRNA-seq than in snRNA-seq datasets, where their expression was marginally detectable or absent in some α-cell subtypes (Fig. 2a). This provides further support for our previous observation that the *GCG, CRYBA2, ALDH1A1, TTR* gene set optimally defines α-cells in scRNA-seq analysis of human islets but is suboptimal for snRNA-seq annotation[17]. Therefore, we next analyzed the expression of our recent α-cell gene set derived from snRNA-seq[17] and observed that *PTPRT, FAP, PDK4* and *LOXL4* showed greater relative expression in clusters in the snRNA-seq than in the scRNA-seq dataset (Fig. 2b). Interestingly, these four latter genes showed a progressively decreasing pattern from α1 to AB cells.

To better define gene sets that identify the different α-cell sub-clusters and their transcriptome differences, we performed differential gene expression (DEG) analysis for every cluster against all the remaining clusters (Fig. 2c, d, Supplementary Fig. 1b, Supplementary Data 2). The α1 subcluster displayed selectively higher expression of *NEAT1, ACTG1, PEAK1* and *ACTB*. The α2 subcluster favored the expression of *FAP, PCSK2, SLC30A8,* and *GLS*. The α3 subcluster most highly expressed *HSPA1A, HSPH1, DNAJB1,* and *PLCG2*. The α4 subcluster favored *PCSK1N, GAPDH, TTR,* and *SNHG29*. The AB cell subcluster displayed expression of pan-endocrine hormonal genes such as *INS, SST, PPY* and *IAPP* (Fig. 2c, Supplementary Fig. 1b, Supplementary Data 2). Interestingly, most of the differentially expressed genes in AB cells were defined by the snRNA-seq which employs intron-inclusive references. This suggests that a large portion of these genes were pre-mRNA, as might be expected of RNA collected from the nuclear compartment (Fig. 2d).

Bihormonal cells (glucagon-positive and insulin-positive cells) have been detected in the human pancreas and in isolated human islets by immunolabeling, reaching values between 0.3 and 4.5% of all insulin-positive cells in non-diabetic organ donors. This percentage increases up to 10.9% in human pancreas from insulin resistant organ donors[30–33]. Here, AB cells (*INS*+ and *GCG*+) were 2.2% (scRNA-seq) and 6.9% (snRNA-seq) of *INS*+ cells, percentages close to the ones reported for immunolabeled human islet bihormonal cells[31–33].

AB cells had (1) more than 500 UMI count, (2) more than 250 gene varieties, (3) more than 0.8 log10 genes per UMI (gene density), and (4) 20% or less mitochondrial genes. The median nFeatures_RNA for AB cells was 692, within the expected range for cells and not background/noise or potential doublets/aggregates as indicated by the analysis packages Seurat (nFeature_RNA > 200 for real cells) and Scanpy (n_genes_by_counts >100 for real cells). To eliminate doublets/aggregates, we used the DoubletFinder Algorithm pANN (proportion of artificial nearest neighbors), which provides a score as a decision factor to distinguish between doublets and singlets. The score of the AB cluster was 0.23, similar to the score of the rest of the islet cell types (Supplementary Fig. 1c). Furthermore, analysis of the available datasets from single nucleus multi-omics of human islets from non-diabetic islet donors performed by Millman's group[34] (Supplementary Fig. 2a–c), showed that the AB cluster (*GCG*+/*INS*+ cells) displayed enhanced chromatin accessibility for *INS* and *GCG* genes (Supplementary Fig. 2d). Finally, AB α cells expressed several exclusive genes compared with α and β cells (Supplementary Data 3). Collectively, these results strongly suggest that AB α cells are *GCG*+/*INS*+ cells and not low-quality cells or doublets/aggregates.
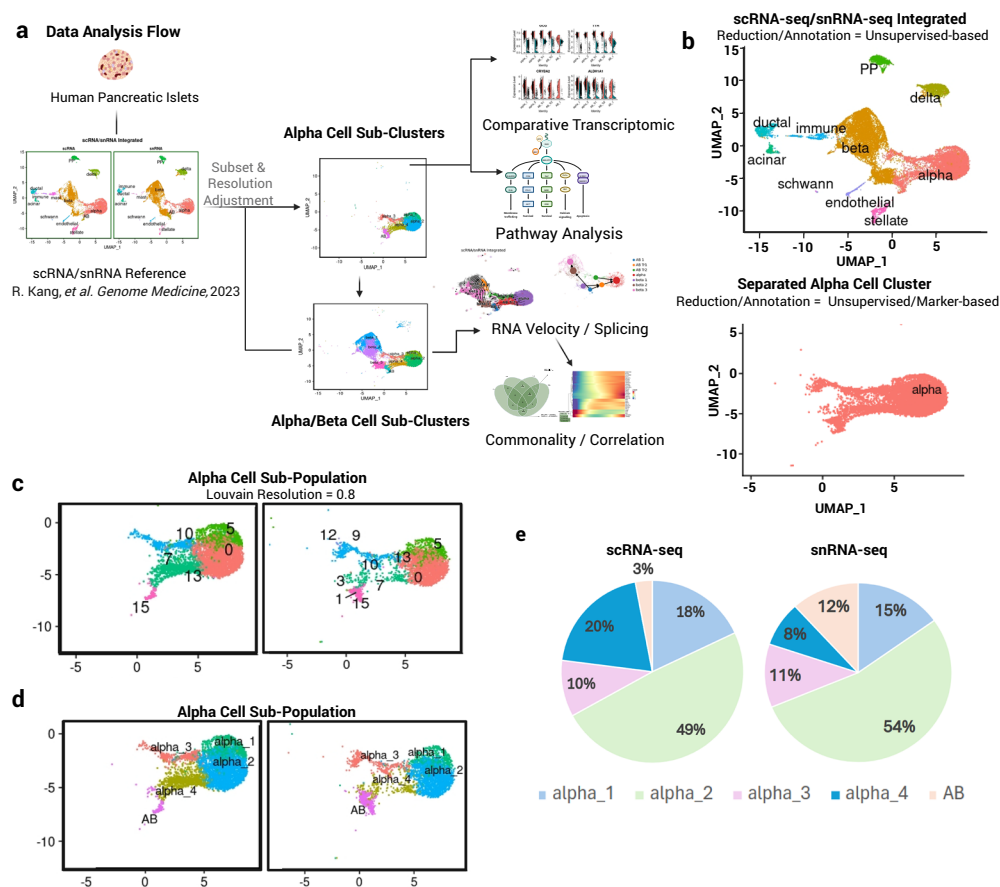
Fig. 1 | Experimental design, unsupervised clustering, and sub-clustering of human α cells by scRNA-seq and snRNA-seq of islets from adult human non-diabetic donors. a Human islet processing and data generation scheme. Created in BioRender. Garcia-Ocana, A. (2025) https://BioRender.com/nizfh1l. **b** Unsupervised clustering, cell type annotated UMAP and separated α-cell cluster (below). **c**. Pre-annotated α-cell subclusters by assigning Louvain resolution 0.8. **d**. Annotated α-cell clusters according to the UMAP location relative to neighboring β-cells[17] and gene expression. UMAP is split by the processing type – scRNA-seq (left) and snRNA-seq (right). **e** Proportion of the α-cell subtypes in scRNA-seq and snRNA-seq datasets.

While DEG analysis is useful to understand the differences among subclusters, their functional significance in a specific subcluster is uncertain. Therefore, we performed single-cell-level pathway enrichment analysis using the integrated scRNA- and snRNA-seq dataset[17], the Msigdb database[35], and the escape package[36]. We performed gene set enrichment analysis (GSEA) for all pathways in curated gene sets (C2) and ontology gene sets (C5) in pancreatic α-cells and screened for glucagon production and secretion, metabolism, differentiation, and translation-related pathways, then selected the pathways with qualifying *p*-value and FDR. As expected, we observed differential enrichment of pathways among the various α-cell subclusters (Fig. 2e). While α1 and α2 cell subclusters displayed higher enrichment for α-cell differentiation and glutamate/glutamine metabolism pathways, α3 cells showed enrichment for glucagon secretion (also present in α1 and α2 cells albeit at a lower level). The α1 and α4 subclusters demonstrated the highest enrichment for protein translation-related pathways, while the AB cluster displayed the highest enrichment for early pancreatic precursor cell gene expression, top β-cell markers, and β-cell transcription factors (Fig. 2e).

Non-targeted GSEA of each α-cell subcluster revealed the top 4 pathways for each subcluster according to the *p*-value and FDR (Supplementary Fig. 3). The α1 subcluster favored enrichment in genes involved in vesicle transport, cell adhesion, TCA cycle, and carbohydrate catabolic processes. The α2 subcluster showed enrichment in neuron and neurotransmitter-related pathways, seemingly suggesting a neuroendocrine-responsive cluster. Interestingly, it also showed high enrichment for the mitotic cell cycle process (Supplementary Fig. 3).

The α3 cell subcluster displayed enrichment in NAD metabolic process, inclusion body assembly, and acetylcholine response related pathways. In addition, it was the only cluster that showed enrichment in the trans-differentiation pathway, although the enriched cell population was small (~ 25% cells) (Supplementary Fig. 3). The α4 cell subcluster displayed high enrichment in ribosome-related pathways, suggesting enhanced protein translation capability, and in the polyamine transport pathway (Supplementary Fig. 3). Finally, the AB cell subcluster with *GCG, INS, SST, IAPP* and *PPY* gene expression and with a topographical location near to the β-cells, exhibited high enrichment for neurotransmitter/noradrenergic neuron differentiation, calcium ion regulation-related pathways such as postsynaptic cytosolic calcium ion concentration and intracellular calcium activated chloride channel activity (Supplementary Fig. 3). Cellular stress can promote non−β cells in the islet to exhibit β-cell–like properties[37,38]. To determine whether AB cells could represent α cells under stress leading to expression of other hormone genes, we performed GSEA of stress pathways in these cells. We found that stress pathways were reduced in the AB α-cell cluster compared with the other α-cell clusters. The stress module score was lower in AB α-cells compared with other α-cells and similar to the score for β1 and β3 cell clusters (Supplementary Fig. 4a-b).

RNA velocity and PAGA analysis of trajectories between α- and β-cells in human islets

We next analyzed RNA velocity using Scanpy and scVelo packages to interrogate potential transitions among α- and β-cell subclusters. The
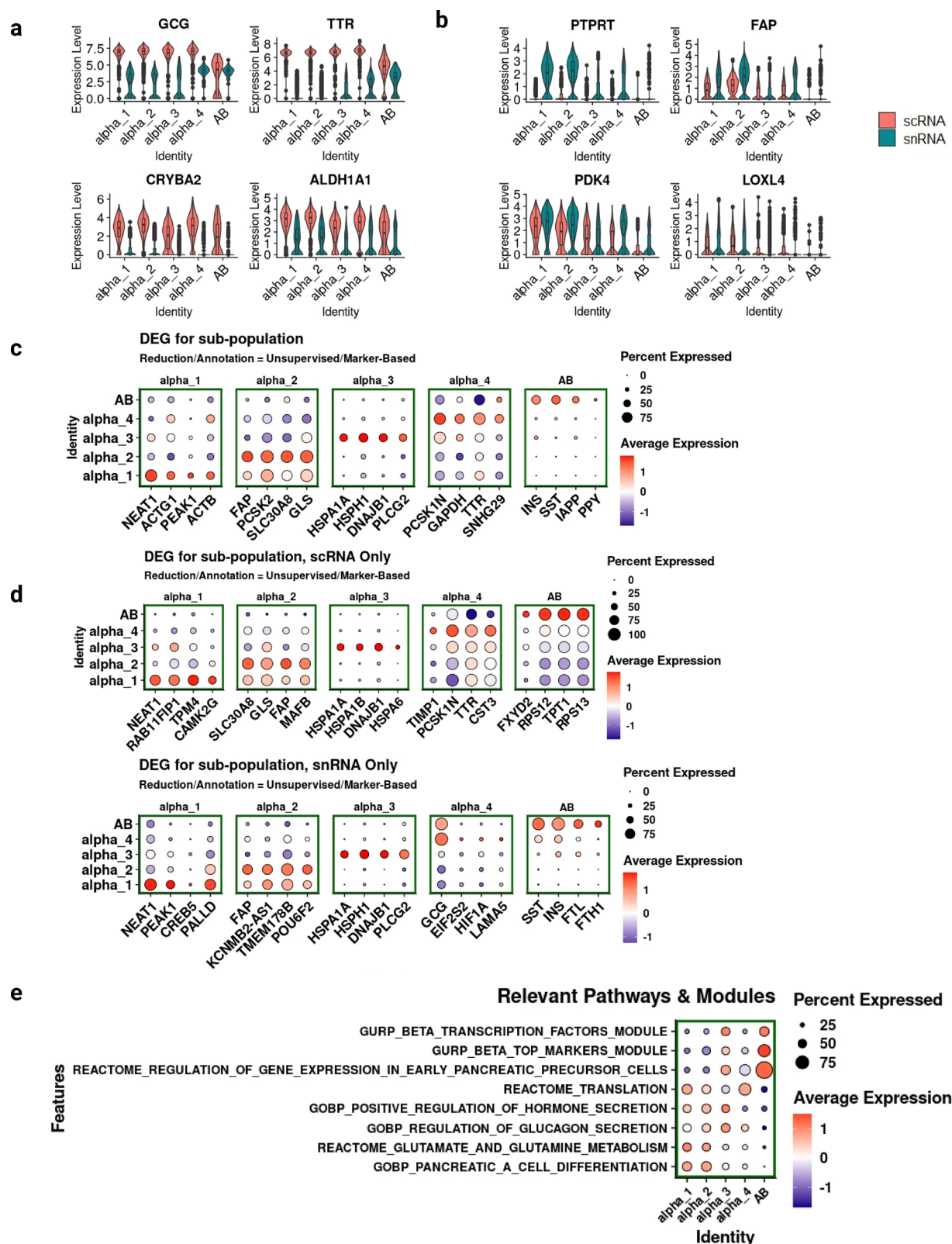
Fig. 2 | Gene expression, DEG, and pathway enrichment analysis of human α-cell subclusters by scRNA-seq and snRNA-seq of islets from adult human non-diabetic donors. a Gene expression of 4 canonical α-cell markers, split by processing types, in the different α-cell subclusters. **b** Gene expression of the previously identified[17] single nucleus α-cell markers in the different α-cell subclusters. **c** Dot-plot visualization of top 4 differentially expressed genes in each α-cell subcluster. **d** Differentially expressed genes for α-cell sub-populations in scRNA-seq data (top) and snRNA-seq data (bottom). **e** Pathway enrichment analysis for α-cell sub-populations. We searched pathways using keywords for pancreatic endocrine cells (α/β/δ/PP/ε), glucagon and hormone signaling/processing/secretion, metabolism, and cellular development (differentiation, precursor, dedifferentiation, development, senescence) and arrange the pathway order according to the enrichment patterns from α1 to AB cells.

α-cell subclusters in the current study were combined with the three β-cell subclusters that we previously reported[17]. Analysis of RNA velocity has been performed traditionally with scRNA-seq and not with snRNA-seq because snRNA-seq data contain transcripts with more intron-retention and, hence will have increased unspliced rates compared with scRNA-seq data where less intron-retention cytoplasmic mature transcripts are captured. However, we reasoned that integrating these two modalities in the same samples will allow us to capture both the early transcriptional dynamics (from snRNA-seq) and the downstream more mature expression profiles (from scRNA-seq), thereby offering a more complete temporal trajectory of cell state transitions, enhancing overall RNA velocity confidence[39]. Indeed, the RNA velocity confidence score, which is a quality metric for how well the predicted RNA velocity aligns with the splicing dynamic of each gene, was significantly improved using integrated scRNA-seq and snRNA-seq (mean 0.81; median 0.93, Mann-Whitney-Wilcoxon test, $p = 1.55 \times 10^{-135}$) compared with scRNA-seq alone (mean 0.71; median 0.80) (Supplementary Fig. 5a). We also observed a comparable ratio of spliced/unspliced RNA among the different cell subtypes in the integrated reference dataset, except for the α2 subcluster, which had the highest proportion of spliced mRNA (Fig. 3a). The β1 subcluster (most mature β-cell subcluster)[17], had the most dominant proportion of unspliced mRNA (Fig. 3a). We then visualized the streamlined velocity plot and observed a general bifurcating pattern from the AB/β3 cluster both towards the β1 and α2 subclusters (Fig. 3b). Using velocity length to characterize strength of transition or differentiation, both the β2 and the α4 clusters showed higher length than the other clusters, denoting vibrant splicing activity and satisfying overall velocity confidence across all the α- and β-cell subpopulations (Fig. 3c). Additionally, we measured the velocity of several representative α- and β-cell genes and found that the expression level and velocity did not always correlate with each other. *GCG* expression was highest in the transitional clusters (α3/α4), but RNA velocity was higher in α2 and part of the α1 subcluster. *INS* expression level was selectively higher in β-cells, while RNA velocity displayed slight induction in the α4 and AB subclusters. We also assessed the β-cell gene, *GLP1R*, and found that its expression pattern matched with RNA velocity. Similarly, we also assessed the α1 cell gene, *NEAT1* (see previous section), and found high expression in some β1 cells but very low RNA velocity (Fig. 3d). Together, these observations support the notion that cells in the β3 and AB subclusters may spontaneously transition to more mature β- and α-cell types. Indeed, PAGA trajectory analysis identified four main trajectories among α- and β-cell subpopulations: (1) AB to β2 to β1 (Trajectory A) which shows AB cells progressing towards the mature β1 subtype[17]; (2) β3 to α4 to α2 (Trajectory B), showing transition from the less mature β-cell subtype[17] to mature α-cells; (3) AB to α4 to α2 (Trajectory C), from multihormonal cell cluster to mature α-cells; and, (4) α3 to α2 (Trajectory D) from less differentiated α-cell cluster to a more differentiated α2-cell cluster (Fig. 3e). Cell trajectory analysis using only the scRNA-seq dataset is shown in Supplementary Fig. 5b-f. We observed a similar trajectory pattern as with the integrated scRNA-seq and snRNA-seq data (Fig. 3e) with a bifurcating trajectory pattern from the AB cluster both towards the α- and β-cell subclusters (Supplementary Fig. 5f). Additional trajectories in the scRNA-seq only analysis were (1) from less differentiated (β3) to more differentiated β-cells (β1), and (2) from α3 to AB cells that further reinforces the bifurcating trajectory potential of AB cells. Collectively, these results indicate that both integrated scRNA-seq and snRNA-seq or only scRNA-seq provide a similar bifurcating trajectory from AB to α- and β-cells in non-diabetic human islet cells in our dataset.

We subsequently analyzed the genes with high pseudo-temporal association for each trajectory in the integrated scRNA-seq and snRNA-seq data (Fig. 4a–d, Supplementary Data 4). First, we transformed the data set into the Monocle3 format and then assigned the root as the unbiased PAGA origin for pseudo-time analysis. We ranked the gene list according to Moran's I score for visualization and clustered the genes according to the patterns (Fig. 4a–d). For the transition from AB to β1 (Trajectory A), there were 2 main transitional patterns: increase or biphasic (increase then decrease) (Fig. 4a). The transcriptome in the increased gene pattern from AB to β1 included cellular ion regulation and signaling (*TRPM3, DPP6, CASR, KCNMA1*) and insulin release (*ABCC8*) genes. They also included neural development and synaptic function genes (*LSAMP, DLG, NRG1*). Genes with a biphasic pattern of expression related to ribosomal protein genes (RPLs, RPSs) and ferritin proteins (*FTL* and *FTH1*) (Fig. 4a). Transition from β3 to α2 (Trajectory B) showed increasing expression pattern of apoptosis and proliferation-involved genes (*KIAA1324, BTG2, MALAT1, NR4A2*), glucagon-related genes (*PCSK2, ARFGEF3*) and glucose homeostasis-related genes (*SLC30A8, PAPPA2*) (Fig. 4b). In contrast to Trajectory A, Trajectory B showed increasing patterns of expression of ribosomal protein genes (RPLs, RPSs), metabolism-associated transcriptomes (*SLC7A2, PDE10A, GLS*) and decreased expression of β-cell identity genes (*IAPP, INS*) (Fig. 4b). Trajectory C, the transition from AB to α2 showed decreasing pattern of expression β-cell identity genes (*IAPP, INS*) (Fig. 4c), while apoptosis and proliferation-involved genes (*TM4SF4, MALAT1, NR4A1, NR4A2, BTG2, KIAA1324*), ribosomal protein genes, and metabolism and cellular signaling genes (*GLS, SLC30A8, PDE10A*) were increased (Fig. 4c). Finally, Trajectory D, the α3-α2 transition displayed an increased expression pattern of cellular transport and metabolism (*SLC7A2, ALDH1A1*), glucose regulation (*ARFGEF3, SLC30A8, PAPPA2, PTPRN2*), and development/differentiation (*PLCE1, NR4A1, KIAA1324, TM4SF4*) genes. On the other hand, β-cell genes (*INS, IAPP, ZNF385D, TRPM3*) displayed reduced expression (Fig. 4d).

Gene commonality analysis in trajectory inference

To identify common genes among the three trajectories that might define the transition between β- and α-cell types (Trajectories B, C, and D), we performed gene commonality analysis (Fig. 4e). Common α-cell trajectory genes from β3 to α2, AB to α2 and α3 to α2 cells (trajectories B/C/D) were S*MOC1, PLCE1, PAPPA2, ZNF331, ALDH1A1, SLC30A8, BTG2, TM4SF4, NR4A1* and *PCSK2* (Fig. 4e). All these genes were upregulated in mature, more functional α1 and α2 cells while their levels were decreased in α3, α4, AB and the rest of β-cells except for *SLC30A8* (Fig. 4f), suggesting that alteration in the levels of *SMOC1, PLCE1, PAPPA2, ZNF331, ALDH1A1, BTG2, TM4SF4, NR4A1 and PCSK2* may drive the transition from AB cells to α-cells. This occurred in a similar way in both scRNA-seq and snRNA-seq datasets. Finally, the degree of expression of these genes positively correlated with the GOBP pathway of pancreatic α-cell differentiation (Fig. 4).

We next performed gene commonality analysis among the four trajectories (Trajectory A and inverse Trajectories B, C, D) to identify common genes that can define the transition from α-cells to β-cells (Supplementary Fig. 6a). Common β-cell trajectory genes were *ZNF385D, TRPM3, CASR, MEG3, HDAC9, INS, SST, RBP4, IAPP* and *SPP1*. All these genes, except *SST*, were upregulated in mature more functional β1 and β2 cells, while their levels were decreased in β3, AB and the rest of α-cells (Supplementary Fig. 6b), suggesting that alteration in the levels of these genes may drive the transition from AB cells to β-cells. Interestingly, the highest expression of *ZNF385D, TRPM3, CASR, MEG3 and HDAC9* occurred mostly in the snRNA-seq where pre-mRNA expression and processing helped to determine cellular trajectories towards transition to different states. Finally, the degree of expression of these genes positively correlated with GOBP pathway of pancreatic β-cell differentiation (Supplementary Fig. 6c).

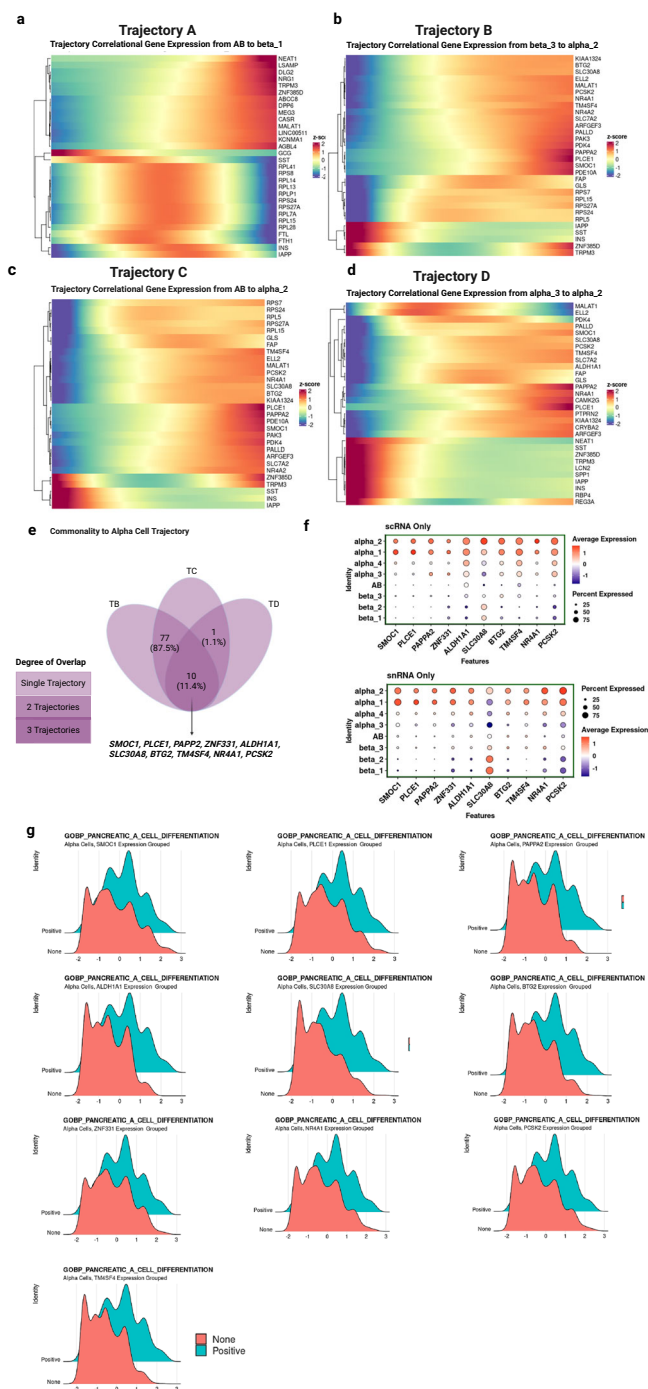## Trajectories between α- and β-cells in human islets from T2D donors

All the preceding studies were performed in adult human islets from non-diabetic donors. To evaluate whether α-β-cell plasticity might
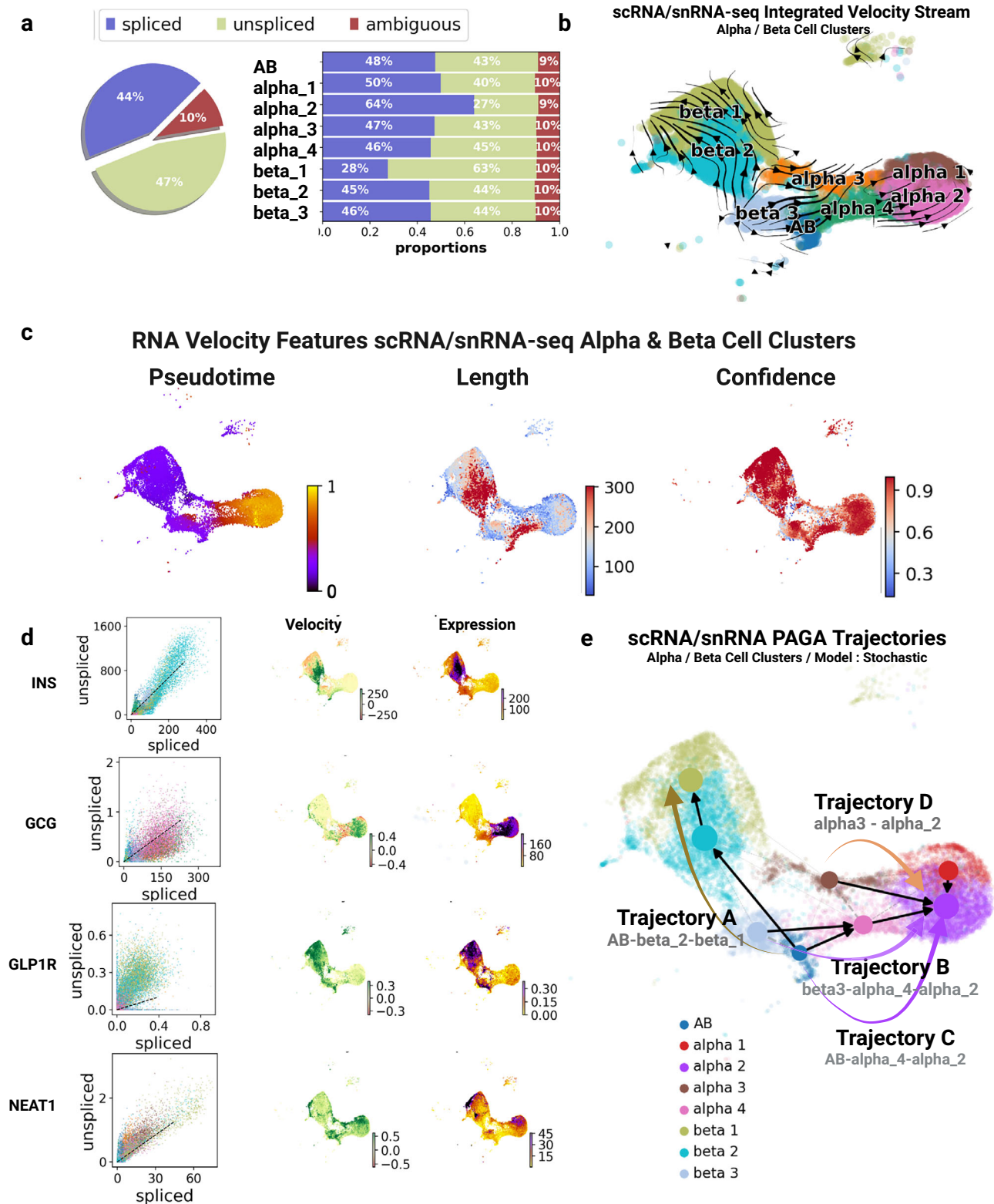
Fig. 3 | RNA velocity, splicing and trajectories in scRNA-seq and snRNA-seq data of islets from adult human non-diabetic donors. a Spliced and unspliced genes percentage in each α-and β-cell subclusters. **b** RNA velocity streamline among α- and β-cell subclusters with scVelo. Stochastic modeling was used. **c** Pseudotime, RNA velocity length and RNA velocity confidence in the α-and β-cell subclusters. **d** Single gene RNA velocity visualization with spliced/unspliced scatter plot (left), RNA velocity (mid), gene expression (right) for *INS, GCG, GLP1R* and *NEAT1*. **e** PAGA trajectory analysis between α-and β-cells based on stochastic modeled RNA velocity showing 4 main trajectories: Trajectory A – from AB to β1 (toward β-cells); Trajectory B- from β3 to α2 (toward α-cells); Trajectory C – from AB to α2 (toward α-cells); Trajectory D – from α3 to α2 (toward α-cells).

Fig. 4 | Trajectories between α-cells and β-cells and their associated gene expression pattern. a Expression of genes with selective high association with Trajectory A (from AB to β1). Genes were clustered according to the expression pattern. The visualization is done by Monocle3, by separating the relevant clusters and assigning the root according to the PAGA analysis. **b** Expression of genes with selective high association with Trajectory B (β3 to α2). **c** Expression of genes with selective high association with Trajectory C (AB to α2). **d** Expression of genes with selective high association with Trajectory D (α3 to α2). **e** Gene commonality analysis towards α-cell transition. Genes in the intersection of Trajectories B, C, and D revealed 10 common genes among the different trajectories. **f** Dot plots representing gene expression of the commonality genes between Trajectories B, C and D in scRNA-seq data (top) and snRNA-seq data (bottom). **g** Trajectories between α-cells and β-cells and their associated gene expression pattern. Enrichment ridge plot of GOBP's pancreatic α-cell differentiation pathway for the common genes among the Trajectories B, C and D. We grouped α-cells in two groups by expression level for each gene – negative (0% normalized expression), positive (>0%).

occur in T2D human islets, we analyzed the scRNA-seq data from human islets isolated from non-diabetic ($n = 13$) and T2D ($n = 13$) donors that are publicly available in the Human Pancreas Analysis Program (HPAP) database, consortia under the Human Islet Research Network (HIRN)[40]. We repeated identical data analysis strategies – projection to a scRNA-/snRNA-seq reference, extracting UMAP coordinates, cell type annotation and RNA velocity analysis. Analysis of total RNA splicing ratios within distinct α-β-cell clusters showed that the AB, α4 and β3 subclusters, the most immature α- and β-cells, exhibited the highest percentage in spliced mRNA in T2D islet cells (Fig. 5a). This suggests enhanced mRNA maturation in the most immature α- and β-cells and potentially enhanced mRNA variation by increased exon combination (alternative splicing) (Fig. 5a).

To compare non-diabetic and T2D datasets, we established a scale ranging from 0 to 600 for RNA velocity length measurements. This allowed us to discern significant differences between the two datasets. This revealed a distinct pattern in the T2D dataset, with a selective increase in velocity length observed within the α3, α4, AB and β3 subclusters (Fig. 5b-c). PAGA analysis showed that non-diabetic human α- and β-cells exhibited a relatively complex multi-directional pattern with trajectories from α- to β-cells and vice versa, implying mutual potential plasticity among α- and β-cell subclusters (Fig. 5d). This complex trajectory pattern partially resembled the pattern observed in our integrated scRNA-seq and snRNA-seq or only scRNA-seq analysis of human islet cells from non-diabetic donors (see sections above) with a distinct bifurcated transition pattern of AB cells into either mature α- or β-cells. This difference could be related to: (1) the smaller number of samples in our dataset compared with HPAP; (2) the wide range in age (4-55 years) and BMI (20.6-38.1) of the HPAP non-diabetic human islet donors compared with our samples (age 24–42, BMI 20-26.3) that can influence the transcriptional gene profile of islet cells; and, (3) the inclusion of 3 donors in the HPAP non-diabetic human islet samples with HbA1c in the range of 5.7–6.0 suggesting potential undiagnosed pre-diabetes that could affect the results.

Overall, and in contrast to non-diabetic human islets, PAGA analysis showed that human T2D α- and β-cells followed mostly β-cell to α-cell transition, moving from β1 to β3 and then to AB, from α4 to α2 cells or direct transition from β1 to α1 and then to α2 (Fig. 5d). This suggests potential pressure on β-cells to become less differentiated or converted to α-cells in subjects with T2D. Pseudotime plots with a pre-defined root by PAGA further supported distinct directions between non-diabetic and T2D human islet cell clusters in the HPAP data sets (Supplementary Fig. 7).

β-to-α cell trajectory genes and T2D

The decrease in β-cell numbers in T2D has been attributed in part to β-cell dedifferentiation[20,21]. In the HPAP-HIRN database, we found that β-cells in T2D islets display decreased *INS* expression and α-cells display increased *GCG* expression (Fig. 6a-b). We then explored whether any of the 10 shared genes involved in β-to-α cell trajectories (see preceding sections) might be upregulated in β-cells in T2D islets, perhaps directing their phenotype towards α-cells. Among these genes, *ZNF311*, *PAPPA2*, *TM4SF4* and *PLCE1* showed no noticeable expression in β-cells in T2D islets (Supplementary Fig. 8a). Of the remaining genes, *NR4A1, PCSK2, SLC30A8, and BTG2* were expressed in both non-diabetic α- and β-cells; their expression was enhanced in T2D β-cells, but they showed a positive correlation with *INS* expression in T2D β-cells (Supplementary Fig. 8a, b). *ALDH1A1* expression occurred in both α- and β-cells although its expression was enhanced in T2D β-cells and showed a negative correlation with *INS* expression (Supplementary Fig. 8a, b). All the genes showed a positive correlation between their expression and *GCG* expression in α-cells (Supplementary Fig. 8c). Finally, expression of the α-cell gene *SMOC1* appeared in β-cells (β1) of T2D islets and negatively correlated with
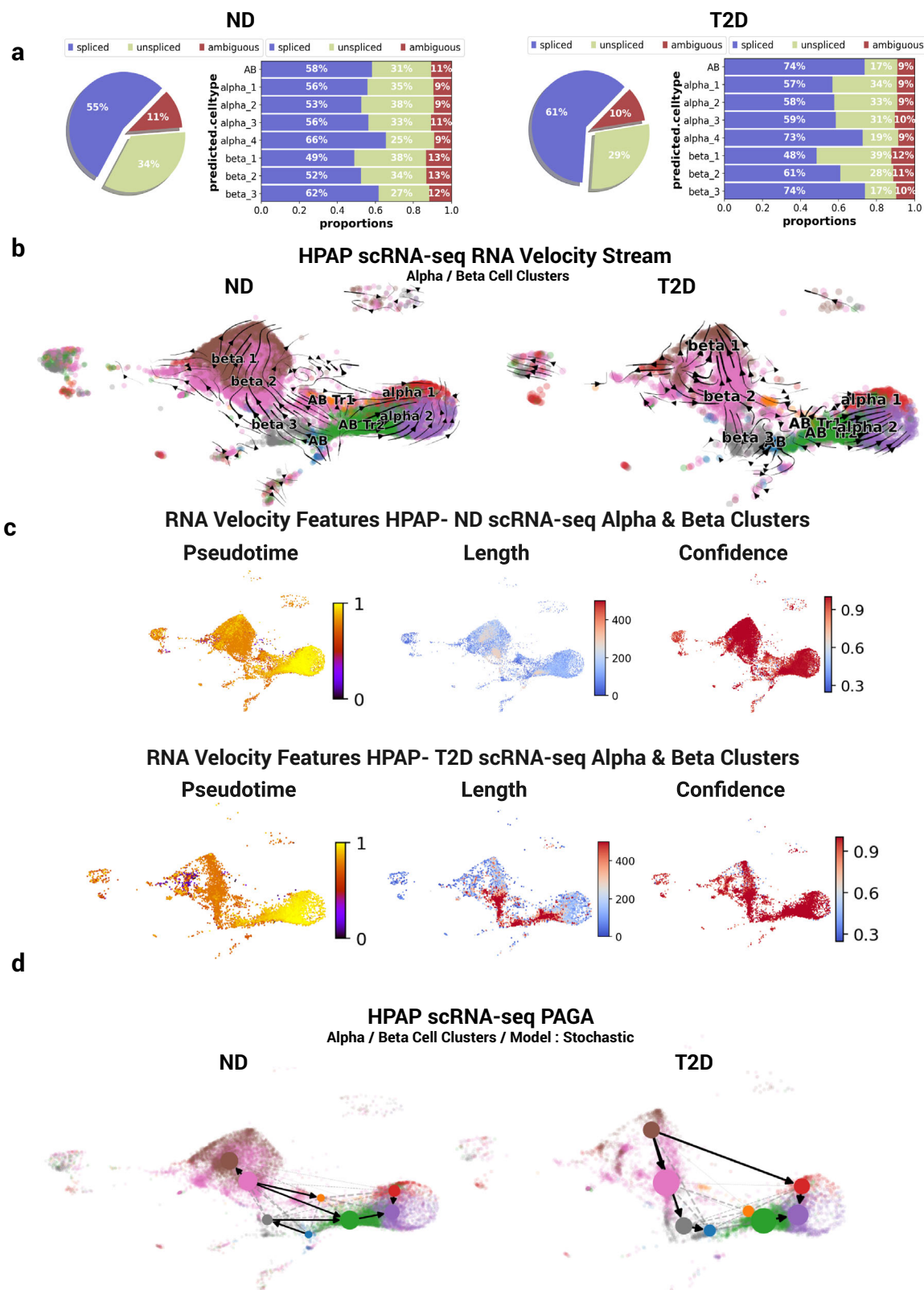
Fig. 5 | RNA velocity, splicing and trajectories in scRNA-seq data of islets from adult human non-diabetic and T2D donors in HPAP-HIRN. a Spliced and unspliced genes percentage in α-and β-subclusters in non-diabetic (ND) (left) and in T2D (right) human islet data from HPAP−HIRN. **b** Pseudotime, RNA velocity length and RNA velocity confidence in the α-and β-cell subclusters of human islets from non-diabetic (ND) donors in HPAP-HIRN dataset. **c** Pseudotime, RNA velocity length and RNA velocity confidence in the α-and β-cell subclusters of human islets from T2D donors in the HPAP-HIRN dataset. **d** PAGA trajectory with stochastic modeling for non-diabetic (left) and T2D islet data (right).

C-PEPTIDE/SMOC1/DAPI

GCG/SMOC1/DAPI
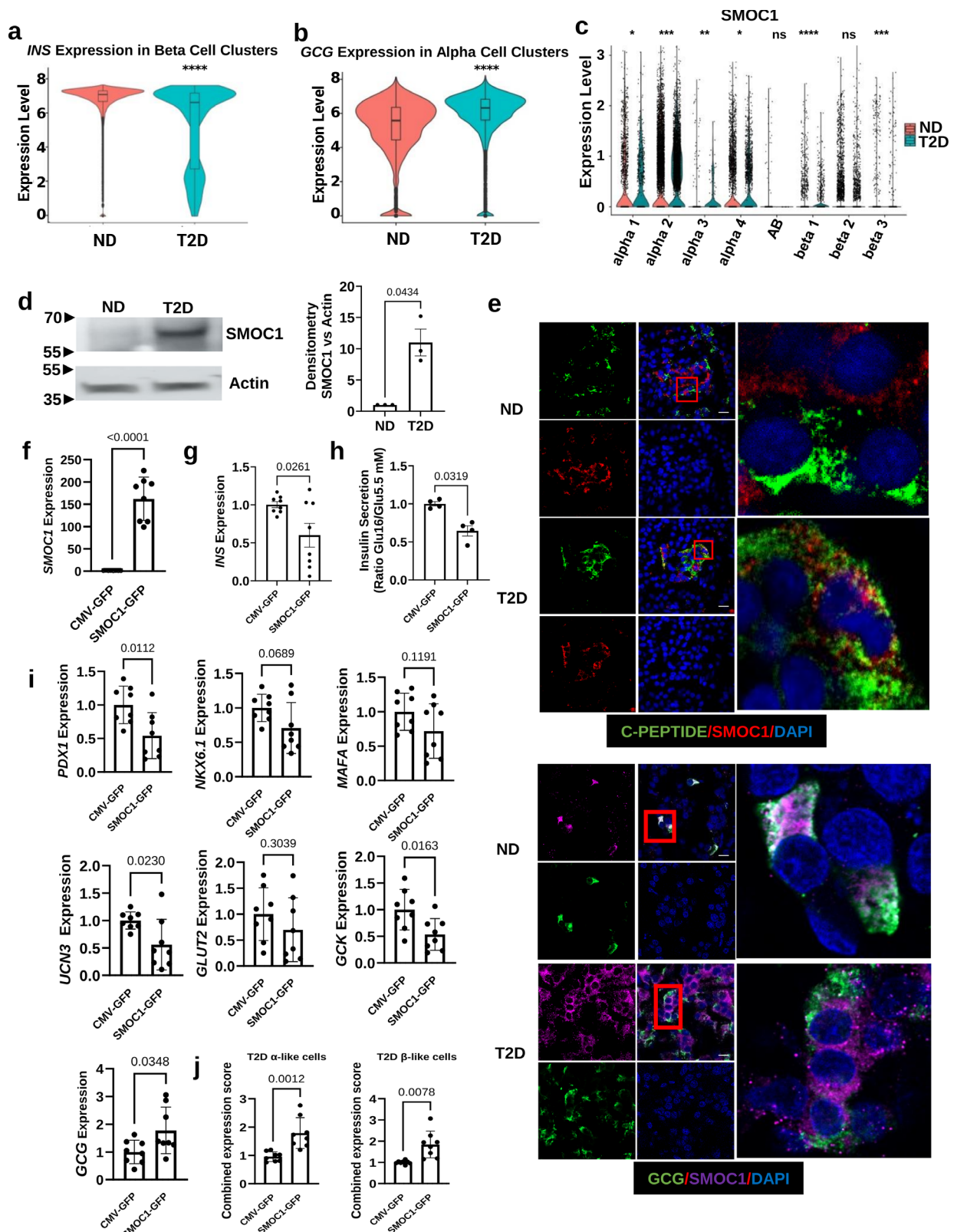
INS expression in non-diabetic β-cells while positively correlated with *GCG* expression in α-cells (Fig. 6c, Supplementary Fig. 8b, c). Collectively, these observations suggest that SMOC1 may play a functional role in normal α-cell biology and may contribute to β-cell dedifferentiation in T2D. We therefore focused on SMOC1 for the remainder of the study.

SMOC1 is a T2D gene in β-cells

We next investigated SMOC1 protein levels in islets of three human non-diabetic and three T2D donors (Supplementary Data 1). SMOC1 abundance was significantly enhanced in islets from human T2D vs. non-diabetic donors as assessed by western blot (Fig. 6d). SMOC1, C-peptide and glucagon immunolabeling of pancreas sections from

Fig. 6 | SMOC1 expression in human islets from non-diabetic (ND) and T2D donors and the effect in insulin secretion and β-cell identity genes. Expression of **a** *INS* and **b** *GCG* in α-and β-cell clusters of adult human ND (red) and T2D islets (green) from the HPAP-HIRN dataset. In **a** ND = 17103 and T2D = 4873 biologically independent cells were analyzed. ND and T2D groups were separated by split.by parameter in Seurat's VlnPlot function, and the statistical test done by the default two-sided Wilcox rank-sum (Mann-Whitney U) method with ggpubr package. For the ND group, the distribution was: min. = 0.00, 25th percentile = 7.55, median = 8.17, 75th percentile = 8.46, and max. = 8.86; and for the T2D group: min. = 0.00, 25th percentile = 7.55, median = 8.02, 75th percentile = 8.39, and max. = 8.82, from normalized expression level stored in RNA assay slot. In **b** ND = 7606 and T2D = 5661 biologically independent cells were analyzed. For the ND group, the distribution was: min. = 0.00, 25th percentile = 5.37, median = 6.02, 75th percentile = 6.57, and max. = 8.10; For the T2D group: min. = 0.00, 25th percentile = 5.68, median = 6.39, 75th percentile = 6.86, and max. = 8.24 from normalized expression level stored in RNA assay slot. For **a** and **b**, the *p*-value annotation legend assigned by default asterisk annotation in ggpubr package refers to: ns:0.05 < *p* < 1.0; *:0.01 < *p* < 0.05; **:0.001<*p* < 0.01; ***:0.0001 < p < 0.001; ****:*p* < 0.0001. **c** Expression of *SMOC1* in α-and β-cell subclusters of adult human non-diabetic (red) and T2D islets (green). ns:0.05 < *p* < 1.0; *:0.01 < *p* < 0.05; **:0.001 < *p* < 0.01; ***:0.0001 < *p* < 0.001; ****:*p* < 0.0001. **d** Representative western blot of human islet protein extracts from adult human non-diabetic and T2D donors and quantitation of *n* = 3 different donors per condition. **e** Representative images of pancreas sections from adult human ND (*n* = 3) and T2D (*n* = 3) donors immunolabeled for DAPI (blue) and SMOC1 (red) and C-peptide (green), or SMOC1 (magenta) and glucagon (green). Enhanced image of the inset (box). **f** Quantitation of *SMOC1* mRNA expression in adenovirus-CMV-GFP or adenovirus-CMV-SMOC1-GFP transduced human islet cells from adult human ND donors by qPCR (*n* = 8 different human islet donors). **g.** *INS* mRNA expression in these adenovirus-CMV-GFP or adenovirus-CMV-SMOC1-GFP transduced-cells by qPCR (*n* = 8 different human islet donors). **h** Insulin secretion index (ratio of insulin secretion at 16 mM and 5.5 mM corrected per insulin content) in adenovirus-CMV-GFP or adenovirus-CMV-SMOC1-GFP transduced-cells (*n* = 4 different human islet donors). **i** *PDX-1, NKX6.1, MAFA, UCN3, GCK, GLUT-2*, and *GCG* mRNA expression in adenovirus-CMV-GFP or adenovirus-CMV-SMOC1-GFP transduced-human islet cells from adult human non-diabetic donors by qPCR (*n* = 8 different human islet donors). **j** T2D α-cell and β-cell scores based on the expression of T2D genes in adenovirus-CMV-GFP or adenovirus-CMV-SMOC1-GFP transduced-human islet cells from adult human non-diabetic donors by qPCR and calculated as indicated in Methods. (*n*= 8 different human islet donors). In all the panels, bars indicate mean+ SEM; *p < 0.05 and **p < 0.01 or as indicated in the figure, using two-tailed Student's *t*-test.

non-diabetic and T2D tissue donors revealed SMOC1 expression in glucagon-positive cells but not in C-peptide-positive cells in non-diabetic pancreas tissue (Fig. 6e). In contrast, SMOC1 expression was easily detectable in C-peptide- and glucagon-positive cells in T2D islets (Fig. 6e). To address the role of enhanced expression of SMOC1 in human islets, we transduced dispersed human islet cells from non-diabetic donors with adenovirus-hSMOC1. Enhanced expression of SMOC1 (Fig. 6f) significantly decreased *INS* gene expression (Fig. 6g) and GSIS (Fig. 6h) and reduced the expression of β-cell transcription factor genes (*PDX-1, MAFA, NKX6.1*), glucose-sensing genes (*SLC2A2, GCK*), and the β-cell maturation gene *UCN3*, while increasing the expression of *GCG* (Fig. 6i). Collectively, these results suggest that enhanced expression of SMOC1 in T2D islet cells might exert detrimental effects on β-cell function and identity.

To address whether SMOC1 upregulation might have an impact in the expression of shared genes in the trajectory from β- to α-cells, we measured the levels of these genes in non-diabetic human islets transduced with adenovirus-hSMOC1. As shown in Supplementary Fig. 9a, SMOC1 upregulation positively influenced the expression of common genes in the β- to α-cell trajectory. We also analyzed the impact of SMOC1 upregulation on gene signatures associated with T2D in α-cells (*IRX2, ZNHIT1, ZFPL1, PAX6, and DRAP1*) and β-cells (*RFX6, RFX7, FOXO1, PPARα, PPARγ, RB1, POU5F1, NANOG, and MYCL*), all previously associated with metabolic inflexibility, stress response and endocrine progenitors/stemness[33]. As shown in Fig. 6j and Supplementary Fig. 9b, c, SMOC1 upregulation in non-diabetic human islet cells positively enhanced the expression of T2D α-cell and T2D β-cell signature genes. These findings support the notion that *SMOC1* may be a bona fide marker as well as a driver gene in human T2D islet cells.

To analyze the direct effect of SMOC1 in human β-cells, we expressed SMOC1 in EndoC-βH1 cells (Fig. 7a). This resulted in decreased *INS* expression (Fig. 7b), diminished GSIS (Fig. 7c), reduced expression of *PDX-1, MAFA, GK, SLC2A2 and GLP1R*, while increasing expression of *FOXO1* (Fig. 7d). In addition, SMOC1 enhanced the expression of some common genes in the trajectory from β- to α-cells and α-cell gene markers such as *MAFB, TTR* and *CRYBA2* (Fig. 7e and Supplementary Figs. 10 and 11), suggesting that SMOC1 might be a potential participant in β- to α-cell conversion. However, the expression of additional characteristic genes that define α-cells such as *GCG, ARX, SCG2, FAP, PDK4* and *LOXL4* was not altered indicating that SMOC1 expression in β-cells is not enough for the conversion of β-cells into bihormonal or α-cells (Supplementary Fig. 11).

Increased FOXO1 expression and activity in T2D contributes to the reduction of insulin secretion while enhancing the pathogenesis of the disease[41,42]. To determine whether FOXO1 regulates SMOC1 expression, we analyzed the effect of FOXO1 upregulation in EndoC-βH1 cells. Enhanced FOXO1 expression led to increased SMOC1 protein expression and secretion suggesting a feedback loop regulating both proteins (Fig. 7f, g).

SMOC1 is a secreted calcium-binding protein with circulating levels in non-diabetic individuals in the range of 50 to 500 ng/ml[28]. Importantly, circulating levels in T2D subjects are significantly increased compared to the levels in non-diabetic subjects reaching up to 800 ng/ml (Fig. 8a). We next wondered whether over-abundance of SMOC1 in the islet environment might influence GSIS and *INS* expression in a paracrine or autocrine manner. To address this possibility, we treated adult human islets from non-diabetic donors and EndoC-βH1 cells with doses of recombinant human SMOC1 in the range of the ones present in the circulation of non-diabetic and T2D individuals. SMOC1 treatment reduced GSIS in both EndoC-βH1 and human islet cell cultures (Fig. 8b, c), suggesting that excessive secretion of this protein in the islet milieu leads to β-cell dysfunction. In addition, exogenously added SMOC1 significantly reduced *INS* expression in EndoC-βH1 cells (Fig. 8c). To address whether the decrease in GSIS by SMOC1 could be related to its calcium-binding properties, we measured the effect of exogenously added SMOC1 on Ca²⁺ flux in EndoC-βH1 cells. SMOC1 induced a substantial decrease in glucose-induced Ca^{2+} flux (Fig. 8d). In addition, SMOC1 significantly reduced glucose- and exendin-4-enhanced cAMP intracellular levels in EndoC-βH1 cells (Fig. 8e). Finally, exogenous SMOC1 significantly decreased the number of mature insulin granules in human islet β-cells (Fig. 8f). Taken together, these results indicate that SMOC1 induces a significant detrimental effect in the function of the β-cell that may contribute to T2D pathogenesis (Fig. 8g).

Discussion

Single cell and single nucleus RNA sequencing can help to identify specific α-cell and β-cell subtypes involved in the development and progression of T2D[2,8,12–19]. It can also reveal specific gene sets that might participate in the β-cell dedifferentiation process that occurs in T2D providing additional therapeutic targets for diabetes treatment. Conversely, deciphering specific genes and pathways in α-cell subpopulations that can enable their reprogramming into insulin-producing cells for β-cell replacement therapies can facilitate the identification of new targets for diabetes therapy[1–6,8,43–47].
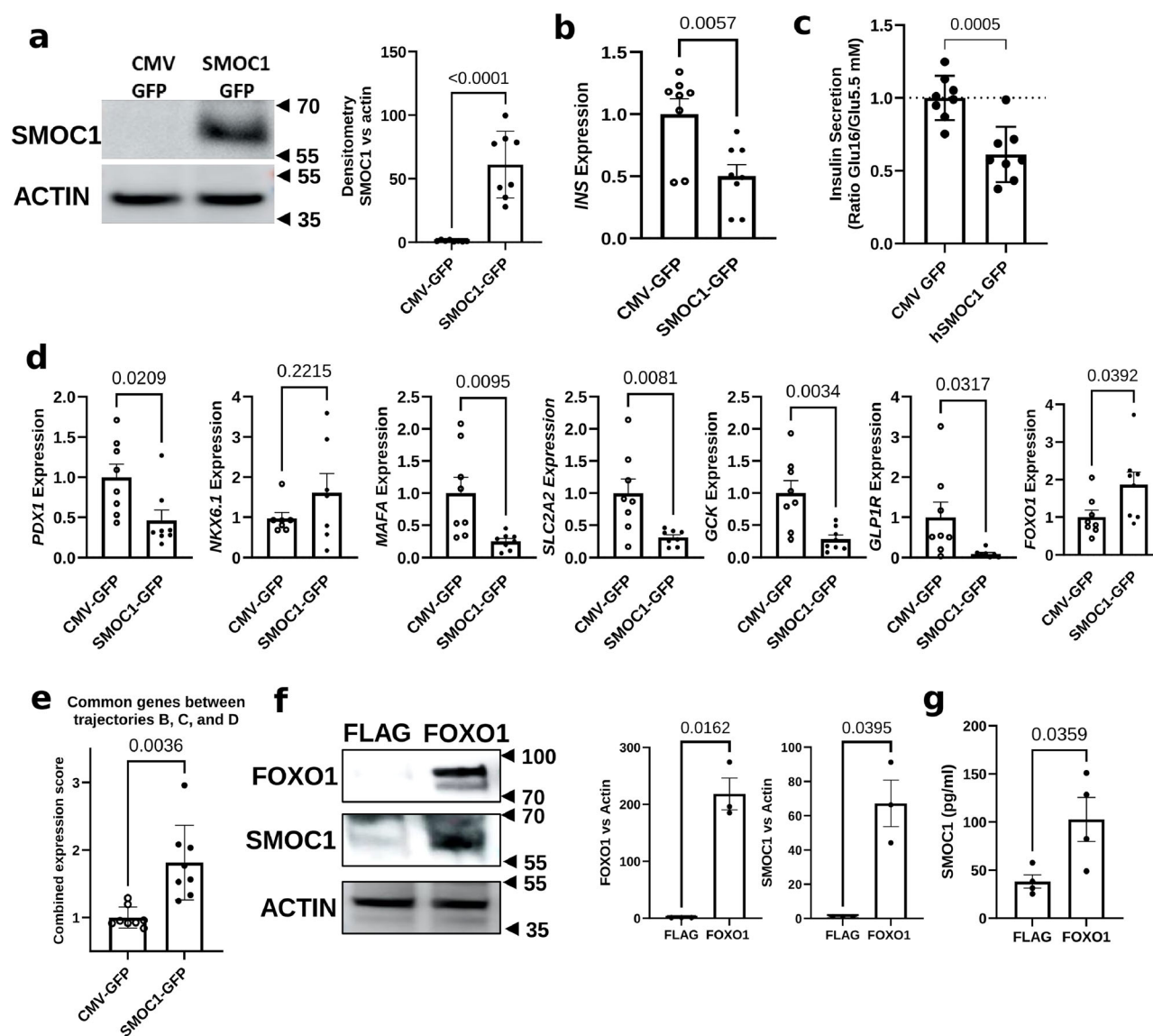
Fig. 7 | **Expression of SMOC1 in EndoC-βH1 cells leads to β-cell dysfunction and de-differentiation. a** Transduction of EndoC-βH1 cells with adenovirus-CMV-SMOC1-GFP to induce the expression of SMOC1 in these cells and quantitation of $n = 8$ experiments. **b** *INS* expression by qPCR in EndoC-βH1 cells expressing SMOC1 ($n = 8$). **c** Insulin secretion index (ratio of insulin secretion at 20 mM and 2 mM corrected by insulin content) in EndoC-βH1 cells expressing SMOC1 ($n = 8$). **d** *PDX-1, NKX.6.1, MAFA, SLC2A2 (GLUT-2), GCK, GLP-1R* and *FOXO1* expression by qPCR in EndoC-βH1 cells expressing SMOC1 ($n = 8$). **e** Combined expression score calculated as indicated in Methods for the common genes in the trajectories from β-cells to α-cells in EndoC-βH1 cells expressing SMOC1 ($n = 8$). **f** Expression of FOXO1 and SMOC1 in EndoC-βH1 cells transfected with plasmid-CMV-FOXO1 ($n = 3$ experiments). **g** SMOC1 levels in the medium from EndoC-βH1 cells transfected with plasmid-CMV-FOXO1 measured by ELISA ($n = 4$ experiments). Bars indicate means ± SEM;*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ and ****$p < 0.0001$ or as indicated in the figure, using two-tailed Student's t-test.

We have recently reported that three human β-cell subpopulations, with their corresponding distinct transcriptome profiles, can be identified using integrated scRNA-seq and snRNA-seq analysis of human islets[17]. Extending this earlier dataset from non-diabetic adult human islets, we have now identified five different *GCG*-expressing α-cell subclusters - α1, α2, α3, α4, and AB - displaying different transcriptome profiles. α1, α2, α3 and α4 clusters are enriched in genes specific for α-cell function and translation pathways. α2 cells are enriched in mitotic/cell cycle pathways suggesting that these might be the cycling α-cells recently described[33,48,49]. The *GCG*-expressing AB subcluster is a multihormonal gene expression cluster enriched in genes related to pancreatic progenitor and β-cell pathways. The identification of *GCG⁺INS⁺* islet cells in scRNA-seq studies has been controversial due to the potential presence of ambient RNA contamination and cell doublets/aggregates in the samples[50]. However,

here we used a stringent hyperparameter threshold in SoupX and employed the Doubltfinder package to algorithmically eliminate ambient RNA contamination or cell doublets/aggregates in our dataset, increasing the reliability of our findings. Indeed, DoubletFinder pANN score in the AB cell cluster was similar to the score of the rest of islet cell types. In addition, stress pathways were reduced, and the stress module score was similar in the AB cell cluster compared with the other α and β cell clusters suggesting that AB cells are not α-cells under stress that could lead to expression of other hormone genes. We also found that the percentage of AB cells with respect to *INS⁺* cells resembled the ratios of bihormonal cells detected by insulin and glucagon immunolabeling of human pancreas or isolated islets previously reported[31–33]. Furthermore, single nucleus multi-omics analysis of human islets from non-diabetic donors performed by Millman's group[34] showed that the AB cluster (*GCG⁺INS⁺* cells) display enhanced

chromatin accessibility for *INS* and *GCG* genes. Finally, AB cells expressed several exclusive genes compared with α and β cells. Taken together, these results strongly suggest that AB cells are *INS*+ and GCG+ cells and not low-quality cells or doublets/aggregates. Using single cell transcriptomics and pseudotime analysis, Saikia et al. identified four human α-cell subclusters with a varying degree of *GCG* expression, two of them displaying mature gene cell markers, one with β-cell like features (*INS* and *GCG* expression) and one intermediate[38]. Recently, a transcriptional cross species map of pancreatic islet cells from mouse, pig and humans has annotated four α-cell states in which 50% were mature α-cells and almost 20% had an immature or precursor-like profile[51]. Our findings in human islets resemble these previous studies with similar proportions of mature (α1, α2, α3 cells) and precursor-like profile cells (α4, AB cells).

RNA velocity and PAGA analysis of integrated scRNA-seq and snRNA-seq or only scRNA-seq datasets identify AB cells as the root of a bifurcated trajectory towards both α- and β-cells in isolated human islets from non-diabetic donors. Trajectory analysis of scRNA-seq data from human islets isolated from adult non-diabetic donors in the HPAP dataset partly resembles this trajectory and a more complex multi-directional trajectory pattern is observed, perhaps, reflecting a wider range in age, BMI and HbA1c in the HPAP samples compared with our samples (40, https://hpap.pmacs.upenn.edu/). Importantly, trajectory analysis of human islets from T2D donors from the HPAP database shows clear unidirectional trajectories from β-cells to α-cells, suggesting potential pressure on β-cells to become less differentiated or converted to α-like cells.

Gene commonality analysis of the trajectories from AB to α-cells (trajectories B, C and D) identifies *ZNF331, PLCE1, PAPPA2, TM4SF4, NR4A1, BTG2, SLC30A8, PCSK2, SMOC1*, and *ALDH1A1* as exclusive common genes in the three trajectories. *ZNF331* encodes for zinc finger protein 331 and has been identified as a putative tumor suppressor silenced or downregulated in gastrointestinal cancers[52], but its function in islet cells remains unknown. Phospholipase C-ε−1 (PLCE1) links Epac2 activation to the potentiation of GSIS in mouse islets[53] and its expression is decreased in β-cells while maintained in α-cells in T2D islets[54]. In our studies, we find that *PLCE1* expression increases in α1 and α2 (mature α-cells) and decreases in mature β1 cells in T2D islets highlighting this phospholipase as a potential candidate for β-cell dysfunction in T2D. Pappalysin-2 (PAPPA2), an enzyme that cleaves and inactivates insulin-growth factor binding proteins (IGFBPs), is upregulated in α-cells during gestation[55]. In the current study, we find that *PAPPA2* is expressed in non-diabetic α-cells and is upregulated in T2D α-cells where PAPPA2 upregulation could limit the activity and potential effects of IGFBP1 on α- to β-cell transdifferentiation[56]. Thus, PAPPA2 could suppress the effect of IGFBPs favoring the α-cell phenotype. Interestingly, a prospective human study has shown that high IGFBP1 levels reduce the risk of developing T2D by more than 85%[57] perhaps reflecting PAPPA2 upregulation in T2D. Transmembrane 4 L Six Family Member 4 (TM4SF4) is a tetraspanin protein localized in the membrane of human α-cells, but not β-cells[58]. Loss-of-function studies have revealed that tm4sf4 loss inhibits α- and β-cell specification in zebrafish[59]. Here, we observe that *TM4SF4* is upregulated in α-cells in T2D islets suggesting further specification of endocrine cells into α-cells. However, its function in adult α-cells is unknown. The orphan nuclear receptor NR4A1 is important for β-cell mitochondrial function and insulin secretion[60]. In the current study, we find that *NR4A1* is expressed in α-cells and β-cells in non-diabetic islets, and it is upregulated in β-cells of T2D subjects, perhaps as a compensatory mechanism for insulin resistance in T2D. This result is in contrast with a previous report indicating that *NR4A1* expression assessed by qPCR is not altered in whole islets from T2D subjects[61]. The different methods employed for gene expression analysis and the use of whole islets (qPCR) vs. single cells (scRNA-seq) could explain the difference in the results. B-cell translocation gene 2 (BTG2) is upregulated by high

glucose and induces apoptosis in β-cells[62]. However, its role in α-cells is unknown. Here we observe that *BTG2* expression is enhanced in β-cells of T2D islets perhaps suggesting involvement in the β-cell death process that occurs in T2D[63]. *SLC30A8* encodes the Solute Carrier Family 30 (Zinc Transporter), Member 8, an islet zinc transporter restricted to α- and β-cells which is responsible for zinc accumulation into secretory granules[64]. Importantly, loss-of-function mutations in *SLC30A8* protect against T2D, suggesting SLC30A8 inhibition as a therapeutic strategy in T2D prevention[65]. In our studies, we observe that *SLC30A8* expression is present in both non-diabetic mature α- and β-cells and increases in T2D β-cells. *PCSK2* encodes the pre-pro-glucagon processing enzyme. In the current studies, *PCSK2* expression is higher in α-cells than β-cells, as expected, and is increased in T2D islet cells. Elevated PCSK2 protein is present in T2D β-cells but it does not contribute to impaired proinsulin processing[66]. Therefore, *PCSK2* appears to be required for transition from β-cells to α-cells but its role in human β-cells is not completely known. *ALDH1A1* is mostly expressed in α-cells as shown in a previous report[58] and in the current study. Although the ALDH1A3 isoform has been implicated in β-cell dedifferentiation in T2D[21,67,68], the involvement of the α-cell ALDH1A1 isoform in β-cell dedifferentiation in T2D has not been previously reported. We observe that *ALDH1A1* expression increases in all the β-cell subtypes in T2D and negatively correlates with *INS* expression. These results clearly position ALDH1A1 as a mediator of - or participant in - β- to α-cell conversion and reinforce the idea that this family of enzymes may serve as therapeutic targets for diabetes treatment[68]. Collectively, this gene set merits further exploration in α- and β-cells in human and animal models of T2D-associated islet dysfunction.

One additional common gene in the trajectory from β- to α-cells is *SMOC1*. Gene expression network analysis identified *SMOC1* as a gene encoding a secreted protein that could potentially affect islet function in obese T2D mouse islets[29]. However, whether this is the case remains unexplored. In our dataset, *SMOC1* appears as a α-cell-exclusive gene in non-diabetic adult human islets. Importantly, SMOC1 protein and mRNA expression are present in β-cells in T2D islets suggesting a potential role in β-cell dedifferentiation. Indeed, forced expression of SMOC1 in human β-cells, or exogenously added SMOC1 protein to non-diabetic adult human islets, results in detrimental effects on *INS* expression, GSIS and β-cell identity. SMOC1 is expressed in the early mouse embryo in the endodermal basement membrane and is essential for ocular and limb development in both humans and mice[27]. *SMOC1* mRNA is expressed in mesenchymal as well as epithelial cells deriving from all three germ layers during mouse development, findings that underscore SMOC1's multi-functional roles in mouse embryogenesis[69]. However, the role of SMOC1 in pancreas development and islet endocrine cell specification is unknown and warrants further studies.

Ca^{2+} participates in the assembly, stability, and storage of the insulin hexamer in the insulin granules. Ca^{2+} influx is responsible for the fusion of insulin-containing vesicles to the plasma membrane in β-cells to release insulin into the bloodstream[70,71]. SMOC1 impairs transmembrane Ca^{2+} influx in keratinocytes by decreasing the amplitude of Ca^{2+} peak response[25]. Here we observe that exogenously added SMOC1, at doses similar to the ones in the circulation of non-diabetic and T2D subjects, remarkably impairs Ca^{2+} flux, reduces insulin granule maturation and reduces GSIS. Thus, it seems likely that SMOC1 secreted from α- or β-cells in T2D islets could impair Ca^{2+} entry into β-cells decreasing insulin release. In addition, SMOC1 reduces cAMP formation induced by the GLP1R agonist, exendin-4. This might reflect the decrease in *GLP1R* expression observed in cells with forced expression of SMOC1 and could have potential therapeutic implications for the treatment of T2D subjects already under GLP1R agonist therapy. Alternative mechanisms involved in SMOC1-induced β-cell dysfunction could also account for the observed effects, including a potential undiscovered receptor for SMOC1 or the interaction of
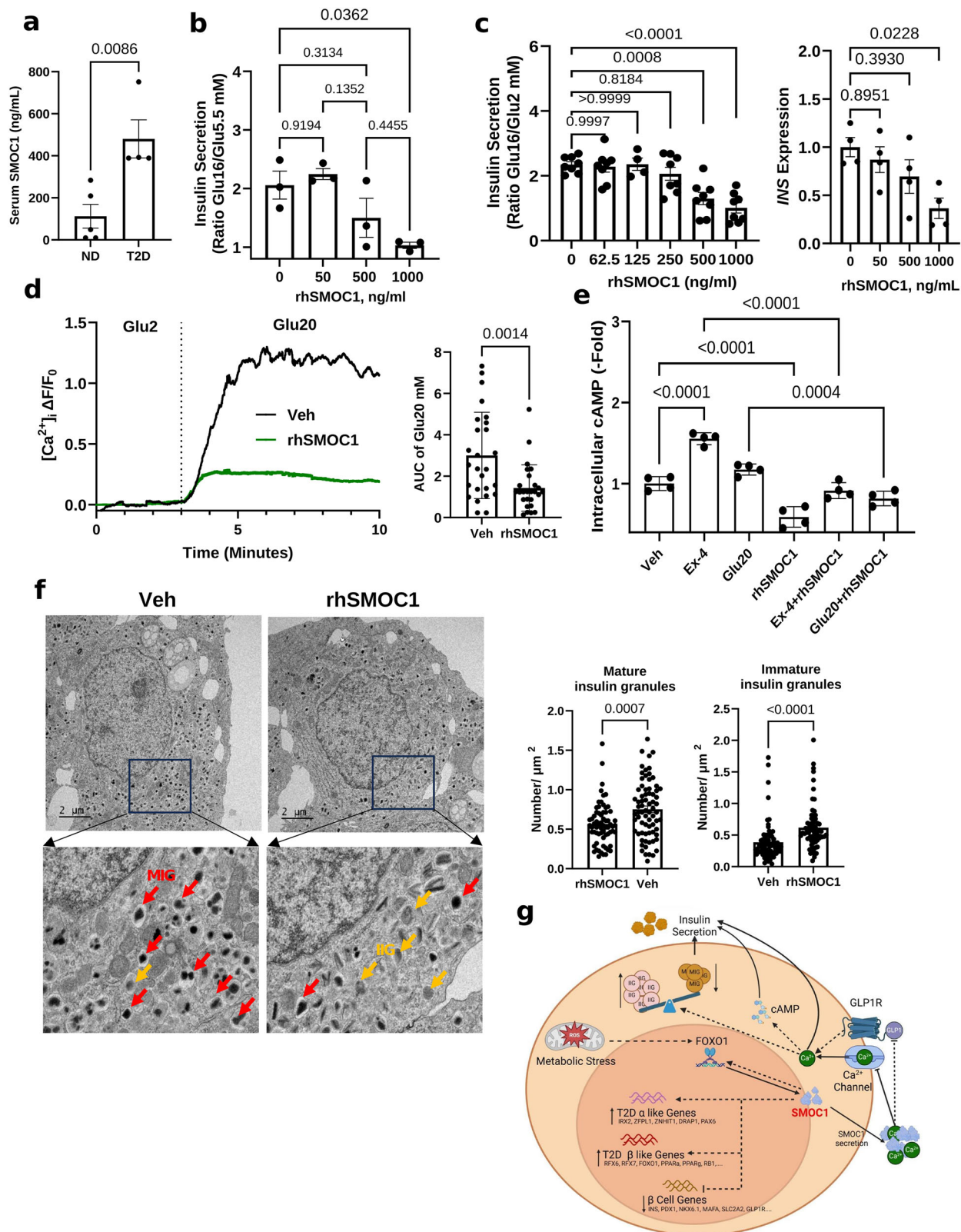
SMOC1 with proteins altering their receptor signaling[23,24]. Indeed, SMOC1 modulates the signaling of BMP-2, an inducer of β-cell dysfunction that is upregulated in T2D[23,72]. Therefore, impairing SMOC1 action in β-cells may be of relevance for therapeutic approaches in T2D. On the other hand, SMOC1 suppresses hepatic glucose production and administration of SMOC1-FC increases insulin sensitivity in preclinical models of T2D[28], suggesting directionally opposite effects on insulin secretion and insulin action. Unraveling these possibilities and others will be the focus of future studies. In the interim, however, these studies highlight the power of common gene analysis of multiple trajectory inferences to identify T2D genes such as *SMOC1* that could be of therapeutic value for treating the disease.

Fig. 8 | SMOC1 protein reduces GSIS, *INS* expression, Ca²⁺ flux, intracellular cAMP formation and insulin granule maturation. a Circulating levels of SMOC1 in non-diabetic (*n* = 5) and T2D subjects (*n* = 4). **b** Effect of rhSMOC1 on insulin secretion in human islets from adult non-diabetic donors (ratio of insulin secretion at 16 mM and 5.5 mM per insulin content) (*n* = 3 different human islet preparations). **c** Insulin secretion (ratio of insulin secretion at 20 mM and 2 mM corrected by insulin content) (left) and *INS* expression by qPCR (right) in EndoC-βH1 cells (*n* = 8 experiments) treated with different doses of rhSMOC1 protein. **d** Measurement of calcium flux in EndoC-βH1 cells transduced with adenovirus-GCaMP5, treated with 1000 ng/ml rhSMOC1 for 48 h and exposed to 2 and 20 mM glucose in a perfusion system (*n* = 3, representative tracking from a single experiment is shown). Quantitation of area under the curve (AUC) of 30-40 cells in these experiments (right). **e** Intracellular cAMP levels in EndoC-βH1 cells treated with 1000 ng/ml rhSMOC1

and exposed to 2 and 20 mM glucose or 10 nM exendin-4 for 1 h (*n* = 4 independent replicates). **f** Electron microscopy of human islet cells treated with 1000 ng/ml rhSMOC1 for 48 h. Representative images focus on human β-cells showing mature insulin granules (MIG) and immature insulin granules (IIG); mature granules contain cores of condensed insulin with angular sides and large surrounding halos. Immature granules are situated mainly near the trans-Golgi network and have no halos. Quantitation of the number of MIG and IIG in *n* = 3 independent biological replicates per condition (right). **g** Schematic representation of the effects of SMOC1 in human β-cells. Created in BioRender. Garcia-Ocana, A. (2025) https://BioRender.com/nizfh1l. In all the panels, bars indicate mean+ SEM, **p* < 0.05, ***p* < 0.01, ****p* < 0.001 and *****p* < 0.0001 or as indicated in the figure, using two-tailed Student's t-test.

In summary, these studies identify five α-cell subtypes and their potential transcriptional status. They also provide evidence of the potential transcriptional transition among α- and β-cell subtypes depending on the pathophysiological context. Finally, they highlight the utility of scRNA-seq and snRNA-seq transcriptomic platforms to find T2D-related genes such as *SMOC1* which could be a therapeutic target for treating diabetes.

Methods

Human islet samples

Adult human pancreata and pancreatic islets from brain-dead donors were provided by Prodo Laboratories (Aliso Viejo, CA) or the City of Hope Human Islet Isolation Center according to the standard procedure and used for the studies described here (Supplementary Data 1)[17,73]. Briefly, islets were harvested from pancreata from deceased organ donors without any identifying information and with informed consent properly and legally secured, and Western Institutional Review Board (WIRB) approval. Transcriptomics analysis of some of these samples has been published[17]. In addition, we mined the raw FASTQ data from the Human Pancreas Analysis Program (HPAP) database, consortia under the Human Islet Research Network (HIRN) and performed analysis of 13 non-diabetic islet and 13 T2D islet scRNA-sequencing datasets through SFTP (secure file transfer protocol) (details below)[40].

Human islet cells and nuclei processing

Human islets from three different cadaveric donors (3000 IEQs/donor) were used in these studies[17]. Briefly, islet cells were dissociated using pre-warmed Accutase (cat# 25−058-CL, Corning) and half of the cells were resuspended in binding buffer (cat# 130−090-101, Miltenyi Biotec) with dead cell removal beads and applied onto the dead cell removal column (cat # 130−042-401, Miltenyi Biotec), which was attached to the MACS separator. Then the cell concentration was measured with the Countess-3 Automated Cell Counter (Thermo-Fisher). The other half of the cells was homogenized with a pestle, and their nuclei were isolated with the Minute™ single nucleus isolation kit for tissue/cells (Cat# SN-047, Invent Biotechnologies, INC) and the nuclei concentration measured with the Countess-3 Automated Cell Counter. After this, nuclei samples were processed in an identical way as to the cell samples.

Single-cell and single-nucleus RNA sequencing, alignment, and matrix generation

Description of the methods used to obtain cells and nuclei samples for these studies has been already published[17]. Briefly, cells and nuclei were prepared according to the 10X Genomics Single Cell 3′ V3.1 Reagent Kit protocol, processed with 10X Genomic Chromium Controller for partitioning and barcoding, followed by the cDNA library generation. Samples were sequenced by NovaSeq 6000 System (Illumina) at the Weill Cornell Medicine, Genomics and Epigenomics Core.

FASTQ files were aligned with Cell Ranger V.6.1.1 with Single Cell 3′ V3 chemistry on the 10X Cloud's pipeline. In the analysis, we included the intronic reads only in the snRNA-seq data with GRCh38-2020-A library. After the 10X h5 format files were generated, data were analyzed on the R platform with Seurat package V.4.3.0[74].

Quality control, integration, and projection

Quality control parameters have been recently described for these studies[17]. Briefly, ambient mRNA adjustment was performed using SoupX (20% contamination estimation)[75]. Cells with less than a 500 gene count, less than 250 gene varieties, less than 0.8 log10 genes per UMI, and a greater than 20% mitochondrial gene ratio were filtered out. Doublets were algorithmically removed with the Doubltfinder package[76]. After data quality control, scRNA-seq and snRNA-seq data were integrated using Seurat's SCTransform function without allocating method parameters, and cell type identity was assigned according to the normalized gene expression level, referencing the canonical pancreatic cell type genes[17,73,77].

Data processing with α-cell cluster

We previously integrated three donor-matched datasets (single-cell RNA and single-nucleus RNA sequencing)[17] and the identified α cell cluster (*GCC*^high, *CRYBA2*^high, *PCSK2*^high, *TTR*^high) was subsequently subset into separate data. We next reassigned cell type identities back to the original numbering generated by the Louvain algorithm with a 0.8 resolution in Seurat's FindNeighbors() function. This yielded ten distinct clusters; however, five of them were composed of fewer than 10 cells. As such, we grouped these small clusters together with their nearest neighbors. Regarding differentially expressed genes (DEG), if some of the genes appear in two clusters, we substituted the next best DEG to avoid duplicated charts for different subclusters.

Reference-based projection

We annotated cell types in scRNA- and snRNA-seq of human islets in vitro according to the normalized gene expression level, referencing the canonical pancreatic cell type genes[17]. In the current study, we refined the cell types with five α- and three β-cell subpopulations. We subsequently established these updated annotations as a reference and projected all datasets analyzed in this study, including HPAP-HIRN (see below) adult non-diabetic and T2D human islet scRNA-seq data, onto this reference[40]. We also performed sample origin analysis of the scRNA-seq only data of human islets in vitro[17] using identical quality control and clustering algorithm with the same hyperparameters.

Pathway analysis

Single-cell level gene set enrichment analysis was performed to define the molecular and cellular processes using the escape package, which accesses the entire C2 and C5 library from Molecular Signature Database (v.7.0)[35,36]. The enrichment score for the entire C2 (6495 pathways) and C5 (15,937 pathways) were calculated for each cell. Then we

used getSignificance() function from escape package to rank the differentially enriched pathways among subpopulations using the ANOVA test for *fit* parameter due to the multiple groups. Additionally, we selected relevant pathways such as α-cell, glucagon, metabolites, and metabolism with qualifying statistics for analysis and visualization.

To obtain more updated pathway enrichment score for β-cell transcription factors and top expressed genes, we use the set of genes identified by Gurp et al.[78] and measured the module score from Seurat's AddModuleScore function on default hyperparameters (nbin=24, ctrl=100, k=FALSE, assay=NULL, seed=1, search=FALSE, data = "data.) For GURP_BETA_TOP MARKERS_MODULE, we added *INS, IAPP, G6PC2, ADCYAP1, ERO1B, DLK1, NPTX2, GSN, INS-IGF2, HADH* while for GURP_BETA_TRANSCRIPTION_FACTORS_MODULE, we added *PDX-1, GLIS3, NKX6.1, MAFA, MAFB, NR3C1, MLXIPL, ZNF395, SMAD9,* and *ZNF385D*[78].

For the stress module score, we used keyword search method on pathway enrichment score (UCell) using cellular stress, ER stress or stress-induced. Subsequently, we obtained single module score by rowMeans function from each pathway enrichment score matrix.

RNA velocity and PAGA analysis

RNA velocity analysis was conducted with integrated sample count tables generated by Matrix package. Velocyto was used to generate a Loom file referencing GRCh38 human genome[79]. UMAP coordinates, and α- and β-cell subtype labels were then extracted from the Seurat data to maintain identical UMAP location and cell type annotations. An AnnData file was created from the data using scanpy Package[80] and spliced/un-spliced counts were measured by scVelo package with proportions() function[39]. A stochastic model hyperparameter was used to estimate velocity, which was visualized with velocity streamline of α- and β-cell sub-clusters transitions. The scVelo algorithm was used to generate PAGA trajectories based on connectivity data[39,80] with the original UMAP coordinates extracted from the Seurat data object. RNA velocity and PAGA analysis were also performed in the scRNA-seq data only of human islets from non-diabetic donors[17] following the same approach as with the integrated sample.

RNA velocity confidence scores were derived from the integrated scRNA-seq and snRNA-seq or only the single-cell RNA sequencing data using scVelo's framework, which computes a confidence metric based on the similarity between a cell's RNA velocity vector and those of its local neighborhood. In brief, for each cell, the velocity confidence score represents the degree of consistency of its inferred future transcriptional state with that of its nearest neighbors, with higher scores indicating more robust and reliable predictions. Boxplot was generated using the Seaborn and Matplotlib library in Python to visually compare the distribution of velocity confidence scores between the "scRNA" and "integrated" groups.

Gene commonality and splicing analysis

We extracted UMAP coordinates and cell type annotation from the Seurat data frame, then transformed it to the Monocle 3 data object. To calculate the pseudo-temporal gene expression association, we assigned the base according to the PAGA result, then subset the relevant clusters, and calculated the genes with high pseudo-temporal association with graph_test() function. Next, we defined the genes which share the common transitional direction. Subsequently, we obtained spliced/unspliced counts of genes from Anndata data object that contains RNA velocity results and calculated the regression by assigning unspliced count on x-axis, spliced count on y-axis.

To determine the impact of the common genes *SMOC1, PLCE1, PAPPA2, ZNF331, ALDH1A1, SLC30A8, BTG2, TM4SF4, NR4A1* and *PCSK2* on α-cell differentiation scores, we separated α-cells into gene positive and gene negative cells, calculated enrichment scores for GOBP

pathway of pancreatic α-cell differentiation and used ridge plots for visualization. Gene positive cells refer to the non-zero normalized expression level.

Analysis of non-diabetic and T2D donor islet scRNA-seq data from the Human Pancreas Analysis Program (HPAP) database

To evaluate α-β cell plasticity changes in T2D, we processed the publicly available scRNA-seq data from the Human Pancreas Analysis Program (HPAP) database (RRID:SCR_016202, https://hpap.pmacs.upenn.edu/) consortia under the Human Islet Research Network (HIRN) (RRID:SCR_014393)[40] obtained from islets of 13 adult non-diabetic and 13 adult T2D cadaveric donors using CellRanger (V7.1.0) on the 10X cloud platform, referencing GRCh38-2020-A transcriptome. Monocle 3 pseudotime plot for HIRN-HPAP data was also generated by assigning the root cells according to the PAGA results: the AB cluster for the human islets from non-diabetic donors and the β1 cluster for the human islets from T2D donors.

Single nucleus multi-omics of non-diabetic human islets

Four snRNA and snATAC raw sequencing datasets from adult human islets were downloaded from the Gene Expression Omnibus (GEO; accession code GSE1996363, samples GSM597968, GSM597969, GSM5979680, GSM5979681, GSM7096627, GSM7096628, GSM7096629, and GSM7096630) as part of Augsornworawat et al.'s study[34]. Data were processed using the cellranger-arc-2.0.2 pipeline with the GRCh38-2020-A-2.0.0 reference genome. After generating the count matrices and peak files, datasets were integrated and analyzed using Seurat (version 5.2.0) and Signac (version 1.14.0), applying quality control and integration (weighted nearest neighbor) parameters consistent with those used in the original publication.

Western blot of human islets and SMOC1 immunolabeling of human pancreas

Human islet protein extracts were analyzed by Western blotting with antibodies against SMOC1 (Abcam, Cat# ab313569, dilution:1:1000) and actin (Origene, Rockville, MD, Cat#AB0145-200, dilution:1:1000)[80]. Human pancreas specimens were fixed for 24 h at 4 °C in 4% paraformaldehyde and then washed with PBS and placed in 70% ethanol until paraffin embedding and sectioned, and the presence of SMOC1 in islet cells was detected by immunolabeling with antibodies against SMOC1 (Abcam, Cat# ab313569, dilution:1:1000), C-peptide (DSHB, CAT#GN-ID4, dilution:1:1000) and glucagon (Abcam, Cat# ab10988, dilution:1:1000) using a Zeiss confocal microscope (Zeiss LSM900) and image data were analyzed by ZEN3.1 (blue edition)[81].

Enhanced SMOC1 expression in human islet cells and EndoC-βH1 cells

Human islets (Supplementary Data 1) were dispersed into individual cells by Accutase (Corning, Cat# 25-058-CI) for 10 minutes at 37 °C. During this digestion, the islets were dispersed by gentle pipetting up and down every 5 minutes for 10 seconds. Single cells were transduced in a minimum volume of 0.1 ml RPMI 1640 supplemented with 2% FBS, with adenoviruses expressing CMV-GFP (Adv.GFP, Cat# AD-GFP) or SMOC1-GFP (Adv.SMOC1, Cat# ADV-223763) (100 MOI) (Vector Biolabs). After 1 h incubation at 37 °C, islet cells were washed with RPMI 1640 containing 5.5 mmol/L glucose and 1% penicillin/streptomycin (Thermo Fisher Scientific, Cat# 15140122) with 10% FBS (Millipore-Sigma, Cat# F4135-500ML) and placed in 24-well AggreWell 400 plates (STEMCELL Technologies, Cat#34415) with centrifugal-forced-aggregation. EndoC-βH1 cells (Human Cell Design, France) were cultured in DMEM low glucose supplemented with 2% bovine serum albumin fraction V (Sigma), 50 μM 2-mercaptoethanol, 10 mM nicotinamide (Sigma), 5.5 μg/mL transferrin (Roche Diagnostics), 6.7 ng/mL selenite (Sigma)[82] and transduced with Adv.SMOC1, Adv.GFP (200 MOI) or transfected with a plasmid containing hFOXO1 cDNA

(Addgene, cat# 17551). Cells were cultured for 48 h and harvested for further studies.

SMOC1 levels in the conditioned medium and in serum from non-diabetic and T2D subjects (City of Hope, Duarte, CA, and Medix Biochemica, Espoo, Finland) (Supplementary Data 5) were measured with SMOC1 ELISA from BIOMATIC (Cat# EKC35640-96T).

GSIS in human islets and EndoC-βH1 cells

Insulin release was measured in triplicate from human islets or EndoC-βH1 cells transduced with Adv.SMOC1, Adv.GFP or treated either with vehicle or recombinant human SMOC1 protein (R&D systems, Cat# 6074-SM-050) for 48 h[80]. Briefly, islets or EndoC-βH1 cells were pre-incubated in Krebs-Ringer bicarbonate buffer supplemented with 10 mM HEPES, 1% BSA, and 2.8 mM glucose for 1 h at 37 °C in a 5% CO_2 incubator. After washing once with the same solution, groups of 15 islet equivalents (IEQs) per condition were incubated in fresh Krebs-Ringer bicarbonate buffer plus 1% BSA and either normal glucose (human islets at 5.5 mM and EndoC-βH1 cells at 2.8 mM) or high glucose (human islets at 16.7 mM and EndoC-βH1 cells at 20 mM) for 30 min. Buffer was removed and frozen at −20 °C for insulin measurement by insulin ELISA (Mercodia, Cat# 10-1113-01). Islets and cells were then incubated with acid ethanol for insulin extraction. Insulin values are normalized to insulin content.

qPCR in human islets and EndoC-βH1 cells and calculation of scores

RNA was isolated from human islets and EndoC-βH1 cells and qPCR was performed[76]. Briefly, total RNA was extracted using the RNeasy Micro or Mini Kits (Qiagen cat# 74004 and 74104). cDNA synthesis was performed using 5X All-in-One RT Plus Mastermix (Lamda Biotech, Cat# G209). Gene expression was analyzed by real-time PCR performed using SYBR Green qPCR Master Mix (MedChemExpress, Cat# HY-K0523) on an ABI 7500 System. CYPA was used as the reference gene to normalize the gene expression. The experimental Ct was calculated using the algorithm enhancements provided by the equipment. Oligonucleotide primers used are shown in (Supplementary Data 6).

Combined score of T2D α-like cells and β-like cells

The gene sets that define T2D β-like-cells include *RFX6, RFX7, FOXO1, PPARg, PPARa, RB1, POU5F1, NANOG,* and *MYCL*[41]. The T2D α-like genes include *IRX2, ZNHIT1, ZFPL1, PAX6,* and *DRAP1*[41]. As specific weights for each gene were not provided in the reference, we assigned a weight of 1 to all genes. The combined score for the T2D likelihood was calculated using the following formula: Combined Score $= \frac{\sum_{i=1}^{n}(E_i \cdot W_i)}{n}$ where Ei represents the normalized expression of each gene, Wi is the weight assigned to each gene (set to 1 for all genes), and n is the total number of genes. This approach assumes that each gene contributes equally to the induction of T2D in α- and β-cells. This equation calculates the average expression level of these genes across all conditions. To obtain the combined expression level, sum up the expression levels of each gene across all conditions and divide by the total number of conditions.

Ca²⁺ flux and intracellular cAMP measurements

EndoC-βH1 cells were plated at a uniform density on glass-bottom dishes (MatTek, Cat# P35G-1.0-14-C) and transduced with Ad-GCaMP5G, generated with the pCMV-GCaMP5G plasmid (Addgene, Cat# 31788), at an MOI of 100. After 24 h, cells were treated with or without soluble human SMOC1 (R&D Systems, Cat# 6074-SM-050) for 48 h. For calcium flux imaging, cells were transferred to an in-house recording chamber with Krebs-Ringer bicarbonate (KRB) buffer at 37 °C and imaged using a Zeiss AxioZoom.V16 microscope (Zeiss, Germany) at 20x magnification.

Glucose stimulation was performed with 2 mM glucose for 3 minutes followed by 20 mM glucose for 7 minutes to elicit calcium responses. Image sequences were analyzed using FIJI software for cell detection, background subtraction, and region of interest (ROI) intensity analysis. Fluorescence intensity (F) was normalized to the baseline intensity (F0) at the start of recording, and results were expressed as F/F0. The area under the curve (AUC) of the F/F0 vs. time curve was calculated to quantify the cumulative calcium response under different experimental conditions, allowing for a comparison of SMOC1 treatment effects on β-cell calcium dynamics.

For the intracellular cAMP measurements, 4×10^5 EndoC-βH1 cells were washed with ice-cold PBS and extracted with 0.2 ml of 0.1 M HCl. Then, cell extracts were centrifuged, and supernatants collected. Intracellular cAMP levels were measured using a direct cAMP ELISA kit (Enzo Life Sciences, Cat# ADI-901-066A) following the manufacturer's instructions. Protein concentrations in the same samples were determined using the Bradford protein assay. Data were normalized to the respective protein content.

Electron microscopy and quantitation of mature and immature insulin granules

Fresh human islets were treated with or without soluble human SMOC1 peptide at a concentration of 1000 ng/mL for 48 h. The islets were fixed with 2.5% glutaraldehyde, 4% paraformaldehyde in 0.15 M Cacodylate buffer (Na(CH3)2AsO2 ·3H2O), pH7.4, at 4 °C. Standard sample preparation for TEM was followed including post-fixation with osmium tetroxide, serial dehydration with ethanol, and embedment in Eponate[83]. Ultra-thin sections (70 nm thick) were acquired by ultra-microtomy, post-stained, and examined on an FEI Tecnai 12 transmission electron microscope equipped with a Gatan OneView CMOS camera. Mature and immature insulin granules were quantified per total cytoplasmic area from more than 20 electron micrographs per group (*n* = 3 biological replicates, >1.0 mm² total cytoplasmic area).

Statistical analysis

Data are presented as means ± SE in bar graphs, violin plots, scatter-plots, and text. Statistical significance analysis was performed using Wilcoxon rank-sum test or Student's t test for comparison between groups as indicated in the figure legends. $P < 0.05$ was considered statistically significant. The simplified asterisk statistical significance annotation followed conventional criteria of 0.05, 0.01, 0.001, and 0.0001 for increment number of asterisks.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw sequencing reads for the scRNA-seq and snRNA-seq data generated from isolated human islets for this study are compiled and available from NCBI under the GEO accession code: GSE217837. The publicly available multiome data of snRNA-seq/snATAC-seq datasets on human islets that were used in this manuscript are from Gene Expression Omnibus (GEO; accession code GSE1996363 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE199636], samples GSM5979678 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5979678], GSM5979679 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5979679], GSM5979680 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5979680], GSM5979681 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5979681], GSM7096627 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM7096627], GSM7096628 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM7096628], GSM7096629 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM7096629], and

GSM7096630 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM7096630]) as part of Augsornworawat et al.'s study[34]. Raw FASTQ data from the Human Pancreas Analysis Program (HPAP) database, consortia under the Human Islet Research Network (HIRN) were obtained from https://hpap.pmacs.upenn.edu/ with provided bash script from HIRN-HPAP website. All single cell RNA sequencing data were aligned with CellRanger 8.0 or higher with GRCh-38-2020A reference genome. Source data are provided with this paper.

Code availability

R/Shell/Python code for data handling and analysis for the main data is publicly available from Github repository (https://github.com/randystyle21/scRNA-snRNA-Marker/blob/main/README.md). Any novel analysis of data in this study is published separately in https://github.com/randystyle21/Alpha_cell_SMOC1.

References

1. Sutherland, E. W. & De Duve, C. Origin and distribution of the hyperglycemic-glycogenolytic factor of the pancreas. *J. Biol. Chem.* **175**, 663–674 (1948).
2. Pettway, Y. D., Saunders, D. C. & Brissova, M. The human α cell in health and disease. *J. Endocrinol.* **258**, e220298 (2023).
3. Gerich, J. E., Langlois, M., Noacco, C., Karam, J. H. & Forsham, P. H. Lack of glucagon response to hypoglycemia in diabetes: evidence for an intrinsic pancreatic alpha cell defect. *Science* **182**, 171–17 (1973).
4. Sakurai, H., Dobbs, R. & Unger, R. H. Somatostatin-induced changes in insulin and glucagon secretion in normal and diabetic dogs. *J. Clin. Invest.* **54**, 1395–1402 (1974).
5. Unger, R. H., Aguilar-Parada, E., Müller, W. A. & Eisentraut, A. M. Studies of pancreatic alpha cell function in normal and diabetic subjects. *J. Clin. Invest.* **49**, 837–848 (1970).
6. Unger, R. H. & Cherrington, A. D. Glucagonocentric restructuring of diabetes: a pathophysiologic and therapeutic makeover. *J. Clin. Invest.* **122**, 4–12 (2012).
7. Ghazvini, Z. et al. ZIGIR, a Granule-Specific Zn2+ Indicator, Reveals Human Islet α Cell Heterogeneity. *Cell Rep.* **32**, 107904 (2020).
8. Dai, X. Q. et al. Heterogenous impairment of α cell function in type 2 diabetes is linked to cell maturation state. *Cell Metab.* **34**, 256–268.e5 (2022).
9. Campbell, S. A. et al. Human islets contain a subpopulation of glucagon-like peptide-1 secreting α cells that is increased in type 2 diabetes. *Mol. Metab.* **39**, 101014 (2020).
10. Smith, D. K. et al. Elevated Serum Amino Acids Induce a Subpopulation of Alpha Cells to Initiate Pancreatic Neuroendocrine Tumor Formation. *Cell Rep. Med.* **1**, 100058 (2020).
11. Clarke, Z. A. et al. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat. Protoc.* **16**, 2749–2764 (2021).
12. Segerstolpe, Å et al. Single-cell Transcriptome profiling of human pancreatic islets in health and Type 2 diabetes. *Cell Metab.* **24**, 593–607 (2016).
13. Tritschler, S., Theis, F. J., Lickert, H. & Böttcher, A. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas. *Mol. Metab.* **6**, 974–990 (2017).
14. Augsornworawat, P. & Millman, J. R. Single-cell RNA sequencing for engineering and studying human islets. *Curr. Opin. Biomed. Eng.* **16**, 27–33 (2020).
15. Shrestha, S. et al. Combinatorial transcription factor profiles predict mature and functional human islet α and β cells. *JCI Insight* **6**, e151621 (2021).
16. Weng, C. et al. Single cell multiomic analysis reveals diabetes-associated β-cell heterogeneity driven by HNF1A. *Nat. Commun.* **14**, 5400 (2023).
17. Kang, R. B. et al. Single-nucleus RNA sequencing of human pancreatic islets identifies novel gene sets and distinguishes β-cell subpopulations with dynamic transcriptome profiles. *Genome Med.* **15**, 30 (2023).
18. Dorrell, C. et al. Human islets contain four distinct subtypes of β cells. *Nat. Commun.* **7**, 11756 (2016).
19. Dominguez-Gutierrez, G., Xin, Y. & Gromada, J. Heterogeneity of human pancreatic β-cells. *Mol. Metab.* **27S**, S7–S14 (2019).
20. Hunter, C. S. & Stein, R. W. Evidence for loss in identity, de-differentiation, and trans-differentiation of islet beta-cells in type 2 diabetes. *Front. Genet.* **8**, 35 (2017).
21. Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. *Cell* **150**, 1223–1234 (2012).
22. Vannahme, C. et al. Characterization of SMOC-1, a novel modular calcium-binding protein in basement membranes. *J. Biol. Chem.* **277**, 37977–37986 (2002).
23. Wang, Y. et al. SPARC-related modular calcium binding 1 regulates aortic valve calcification by disrupting BMPR-II/p-p38 signaling. *Cardiovasc. Res.* **118**, 913–928 (2022).
24. Delgado Lagos, F. et al. Secreted modular calcium-binding protein 1 binds and activates thrombin to account for platelet hyperreactivity in diabetes. *Blood* **137**, 1641–1651 (2021).
25. Lyubchenko, T., Collins, H. K., Vang, K. A., Leung, D. Y. M. & Goleva, E. SMOC1 and IL-4 and IL-13 Cytokines Interfere with Ca2+ Mobilization in Primary Human Keratinocytes. *J. investig. Dermatol.* **141**, 1792–1801.e5 (2021).
26. Takahata, Y. et al. Smoc1 and Smoc2 regulate bone formation as downstream molecules of Runx2. *Commun. Biol.* **4**, 1199 (2021).
27. Okada, I. et al. SMOC1 is essential for ocular and limb development in humans and mice. *Am. J. Hum. Genet* **88**, 30–41 (2011).
28. Montgomery, M. K. et al. SMOC1 is a glucose-responsive hepatokine and therapeutic target for glycemic control. *Sci. Transl. Med.* **12**, eaaz8048 (2020).
29. Koltes, J. E. et al. A gene expression network analysis of the pancreatic islets from lean and obese mice identifies complement 1q like-3 secreted protein as a regulator of β-cell function. *Sci. Rep.* **9**, 10119 (2019).
30. Yoneda, S. et al. Predominance of β-cell neogenesis rather than replication in humans with an impaired glucose tolerance and newly diagnosed diabetes. *J. Clin. Endocrinol. Metab.* **98**, 2053–2061 (2013).
31. Mezza, T. et al. Insulin resistance alters islet morphology in non-diabetic humans. *Diabetes* **63**, 994–1007 (2014).
32. Mezza, T. et al. β-Cell Glucose Sensitivity Is Linked to Insulin/Glucagon Bihormonal Cells in Nondiabetic Humans. *J. Clin. Endocrinol. Metab.* **101**, 470–475 (2016).
33. Karakose, E. et al. Cycling alpha cells in regenerative drug-treated human pancreatic islets may serve as key beta cell progenitors. *Cell Rep. Med.* **5**, 101832 (2024).
34. Augsornworawat, P. et al. Single-nucleus multi-omics of human stem cell-derived islets identifies deficiencies in lineage specification. *Nat. Cell. Biol.* **25**, 904–916 (2023).
35. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
36. Borcherding, N. & Andrews, J. Package escape: easy single cell analysis platform for enrichment. *R. package version* **1**, 0 (2022).
37. Cigliola, V., Thorel, F., Chera, S. & Herrera, P. L. Stress-induced adaptive islet cell identity changes. *Diab., Obes. Metab.* **18**, 87–96 (2016).
38. Saikia, M. et al. GLP-1 receptor signaling increases PCSK1 and β cell features in human α cells. *JCI Insight* **6**, e141851 (2021).
39. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

40. Kaestner, K. H., Powers, A. C., Naji, A. & Atkinson, M. A. NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in Type 1 diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes* **68**, 1394–1402 (2019).

41. Son, J. et al. BACH2 inhibition reverses β cell failure in type 2 diabetes models. *J. Clin. Invest.* **131**, e153876 (2021).

42. Teaney, N. A. & Cyr, N. E. FoxO1 as a tissue-specific therapeutic target for type 2 diabetes. *Front Endocrinol.* **14**, 1286838 (2023).

43. Thorel, F. et al. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* **464**, 1149–1154 (2010).

44. Chakravarthy, H. et al. Converting adult pancreatic islet α cells into β cells by targeting both Dnmt1 and Arx. *Cell Metab.* **25**, 622–634 (2017).

45. Xiao, X. et al. Endogenous reprogramming of Alpha cells into Beta cells, induced by viral gene therapy, reverses autoimmune diabetes. *Cell Stem Cell* **22**, 78–90.e4 (2018).

46. Saleh, M., Gittes, G. K. & Prasadan, K. Alpha-to-beta cell trans-differentiation for treatment of diabetes. *Biochem. Soc. Trans.* **49**, 2539–2548 (2021).

47. Oropeza, D. & Herrera, P. L. Glucagon-producing α-cell transcriptional identity and reprogramming towards insulin production. *Trends Cell Biol.* **23**, S0962–S8924 (2023).

48. Dominguez-Gutierrez, G. et al. Gene signature of proliferating human pancreatic α cells. *Endocrinology* **159**, 3177–3186 (2018).

49. Elgamal, R. M. et al. An integrated map of cell type-specific gene expression in pancreatic islets. *Diabetes* **72**, 1719–1728 (2023).

50. Bonnycastle, L. L. et al. Single-cell transcriptomics from human pancreatic islets: sample preparation matters. *Biol. Methods Protoc.* **5**, bpz019 (2020).

51. Tritschler, S. et al. A transcriptional cross species map of pancreatic islet cells. *Mol. Metab.* **66**, 101595 (2022).

52. Yu, J. et al. Zinc-finger protein 331, a novel putative tumor suppressor, suppresses growth and invasiveness of gastric cancer. *Oncogene* **32**, 307–317 (2013).

53. Dzhura, I. et al. Phospholipase C-ε links Epac2 activation to the potentiation of glucose-stimulated insulin secretion from mouse islets of Langerhans. *Islets* **3**, 121–128 (2011).

54. Xin, Y. et al. RNA sequencing of single human islet cells reveals Type 2 Diabetes genes. *Cell Metab.* **24**, 608–615 (2016).

55. Chung, J. Y. et al. Pancreatic islet cell type-specific transcriptomic changes during pregnancy and postpartum. *iScience* **26**, 106439 (2023).

56. Lu, J. et al. IGFBP1 increases β-cell regeneration by promoting α- to β-cell transdifferentiation. *EMBO J.* **35**, 2026–2044 (2016).

57. Lewitt, M. S. et al. Insulin-like growth factor-binding protein-1 in the prediction and development of type 2 diabetes in middle-aged Swedish men. *Diabetologia* **51**, 1135–1145 (2008).

58. Muraro, M. J. et al. A single-cell Transcriptome Atlas of the human pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).

59. Anderson, K. R. et al. The L6 domain tetraspanin Tm4sf4 regulates endocrine pancreas differentiation and directed cell migration. *Development* **138**, 3213–3224 (2011).

60. Reynolds, M. S. et al. β-Cell deletion of Nr4a1 and Nr4a3 nuclear receptors impedes mitochondrial respiration and insulin secretion. *Am. J. Physiol. Endocrinol. Metab.* **311**, E186–E201 (2016).

61. Close, A. F., Dadheech, N., Villela, B. S., Rouillard, C. & Buteau, J. The orphan nuclear receptor Nor1/Nr4a3 is a negative regulator of β-cell mass. *J. Biol. Chem.* **294**, 4889–4897 (2019).

62. Kooptiwut, S. et al. Estradiol prevents high glucose-induced β-cell apoptosis by decreased BTG2 Expression. *Sci. Rep.* **8**, 12256 (2018).

63. Butler, A. E. et al. Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* **52**, 102–110 (2003).

64. Solomou, A. et al. The Zinc Transporter Slc30a8/ZnT8 is required in a subpopulation of pancreatic α-cells for hypoglycemia-induced glucagon secretion. *J. Biol. Chem.* **290**, 21432–21442 (2015).

65. Flannick, J. et al. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.* **46**, 357–363 (2014).

66. Ramzy, A., Asadi, A. & Kieffer, T. J. Revisiting Proinsulin processing: evidence that human β-cells process Proinsulin with Prohormone Convertase (PC) 1/3 but Not PC2. *Diabetes* **69**, 1451–1462 (2020).

67. Cinti, F. et al. Evidence of β-cell dedifferentiation in human Type 2 Diabetes. *J. Clin. Endocrinol. Metab.* **101**, 1044–1054 (2016).

68. Son, J. et al. Genetic and pharmacologic inhibition of ALDH1A3 as a treatment of β-cell failure. *Nat. Commun.* **14**, 558 (2023).

69. Gersdorff, N., Muller, M., Schall, A. & Miosge, N. Secreted modular calcium-binding protein-1 localization during mouse embryogenesis. *Histochem. Cell Biol.* **126**, 705–712 (2006).

70. Hou, J. C., Min, L. & Pessin, J. E. Insulin granule biogenesis, trafficking and exocytosis. *Vitam. Horm.* **80**, 473–506 (2009).

71. Sabatini, P. V., Speckmann, T. & Lynn, F. C. Friend and foe: β-cell Ca²⁺ signaling and the development of diabetes. *Mol. Metab.* **21**, 1–12 (2019).

72. Urizar, A. I. et al. Beta cell dysfunction induced by bone morphogenetic protein (BMP)–2 is associated with histone modifications and decreased NeuroD1 chromatin binding. *Cell Death Dis.* **14**, 399 (2023).

73. Scharp, D. W., Arulmoli, J., Morgan, K., Sunshine, H. & Hao, E. Advances in human islet processing: manufacturing steps to achieve predictable islet outcomes from research pancreases. *OBM Transplant.* **3**, 1901052 (2019).

74. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **77**, 1888–1902.e21 (2019).

75. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience.* **9**, giaa151 (2020).

76. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).

77. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

78. van Gurp, L. et al. Generation of human islet cell type-specific identity genesets. *Nat. Commun.* **13**, 2020 (2022).

79. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

80. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

81. Wang, P. et al. A high-throughput chemical screen reveals that harmine-mediated inhibition of DYRK1A increases human pancreatic beta cell replication. *Nat. Med.* **21**, 383–388 (2015).

82. Ravassard, P. et al. A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Investig.* **121**, 3589–3597 (2011).

83. Mascorro, J. A. & Bozzola, J. J. Processing biological tissues for ultrastructural study. *Methods Mol. Biol.* **369**, 19–34 (2007).

Acknowledgements

Author contributions

Competing interests

Additional information