Identification of type 2 diabetes- and obesity-associated human β-cells using deep transfer learning

Gitanjali Roy[1], Rameesha Syed[1†], Olivia Lazaro[1†], Sylvia Robertson[1], Sean D. McCabe[3], Daniela Rodriguez[1], Alex M. Mawla[4], Travis S. Johnson[1,3]*, Michael A. Kalwat[1,2]*

[1]Indiana Biosciences Research Institute, Indianapolis, IN, USA

[2]Center for Diabetes and Metabolic Diseases, Indiana University School of Medicine, Indianapolis, IN, USA.

[3]Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, IN, USA.

[4]Department of Neurobiology, Physiology and Behavior, College of Biological Sciences, University of California, Davis, Davis, CA

[†]Equal contributors.

*Correspondence:

mkalwat@indianabiosciences.org and

johnstrs@iu.edu

ABSTRACT

Diabetes affects >10% of adults worldwide and is caused by impaired production or response to insulin, resulting in chronic hyperglycemia. Pancreatic islet β-cells are the sole source of endogenous insulin and our understanding of β-cell dysfunction and death in type 2 diabetes (T2D) is incomplete. Single-cell RNA-seq data supports heterogeneity as an important factor in β-cell function and survival. However, it is difficult to identify which β-cell phenotypes are critical for T2D etiology and progression. Our goal was to prioritize specific disease-related β-cell subpopulations to better understand T2D pathogenesis and identify relevant genes for targeted therapeutics. To address this, we applied a deep transfer learning tool, DEGAS, which maps disease associations onto single-cell RNA-seq data from bulk expression data. Independent runs of DEGAS using T2D or obesity status identified distinct β-cell subpopulations. A singular cluster of T2D-associated β-cells was identified; however, β-cells with high obese-DEGAS scores contained two subpopulations derived largely from either non-diabetic or T2D donors. The obesity-associated non-diabetic cells were enriched for translation and

unfolded protein response genes compared to T2D cells. We selected CDKN1C and DLK1 for validation by immunostaining in human pancreas sections from healthy and T2D donors. Both CDKN1C and DLK1 were heterogeneously expressed among β-cells. CDKN1C was increased in β-cells from T2D donors, in agreement with the DEGAS predictions, while DLK1 appeared depleted from T2D islets of some donors. In conclusion, DEGAS has the potential to advance our holistic understanding of the β-cell transcriptomic phenotypes, including features that distinguish β-cells in obese non-diabetic or lean T2D states. Future work will expand this approach to additional human islet omics datasets to reveal the complex multicellular interactions driving T2D.

KEYWORDS: β-cells, ER stress, unfolded protein response, heterogeneity, single-cell RNA-seq, machine learning, deep transfer learning, diabetes, computational biology

BACKGROUND

The advent of high-throughput single-cell RNA sequencing (scRNA-seq) has enabled the generation of an array of single-cell atlases from pancreatic islets. These findings have expanded our understanding of the major cell types of the pancreas along with how they are implicated in both type 1 diabetes (T1D) and type 2 diabetes (T2D). Notably, these studies have: 1) identified multiple reliable transcriptomic markers for endocrine and exocrine pancreatic cell types; 2) provided insight into novel subtypes of cells; and 3) generated large cellular atlases spurring innovation in the development of single cell methods and analysis. scRNA-seq analysis has enabled a more robust characterization of islet cell heterogeneity which may underlie diversity in diabetes risk and drug response [1-5]. In published comparisons between the islet single-cell transcriptomes of humans versus mice and pigs, differences were observed in the relative proportion of major cell types as well as in many cell type-specific genes [6]. These findings further support a focus on integrating human islet transcriptomic data to delineate disease processes and identify therapeutic targets and biomarkers.

As bulk RNA-seq and scRNA-seq human islet datasets have become increasingly available, so too has the need for new computational tools to integrate the transcriptomics and donor metadata. Some sets of islet scRNA-

53 seq data have been combined and made searchable through web portals to browse datasets and compare cell
54 types and marker genes [6-9]. To further utilize these types of integrated transcriptomic data and donor metadata,
55 we previously developed DEGAS as a flexible deep transfer learning framework that can be used to overlay
56 disease status, survival hazard, drug response, and other clinical information directly onto single cells. Machine
57 learning tools like DEGAS have been primarily used in cancer datasets, but not in human pancreatic islets until
58 now.

59 The reasons why some obese individuals succumb to T2D while other do not likely involves both genetic and
60 environmental factors, but the factors that underlie this transition are incompletely understood[10]. Analyses of
61 human genomics and islet transcriptomics indicate many β-cell genes have causal roles in T2D [11-15]. Subsets of
62 β-cells which are more resilient or susceptible to failure under the secretory pressure of insulin resistance may
63 be uncovered by combining transcriptomics with machine learning approaches like DEGAS. Here we have
64 implemented DEGAS to predict T2D- and obesity-associated subclusters of human pancreatic islet β-cells using
65 a combination of publicly available scRNA-seq and bulk RNA-seq human islet data and associated donor
66 metadata. Through this analysis we sought to identify genes implicated in T2D and obesity which were up- or
67 down-regulated in subpopulations of β-cells identified by DEGAS and to validate our findings at the protein level
68 using immunohistochemistry of pancreas tissue from non-diabetic and T2D organ donors. Our findings applying
69 DEGAS to islet data have implications for β-cell heterogeneity in T2D and obesity.

70

71 **METHODS**

72 ***Human islet bulk transcriptomic dataset acquisition, processing, and analysis –*** In this study, human
73 islet bulk RNA-seq raw count data (aligned to GRCh38) and donor metadata from Marselli, et al. GSE159984
74 was downloaded from the Gene Expression Omnibus [16] (**Table S1**). The rationale for using this dataset was its
75 large sample size of both non-diabetic and T2D samples and the agreement of differentially-expressed genes
76 between the GSE159984 dataset and a similarly large independent dataset[17]. All genes in the read count data
77 were filtered based on the one-to-one identifier correspondences between gene symbol, Entrez gene identifier,
78 and Ensembl identifier from the Matched Annotation from NCBI and EBI table (MANE GRCh38 v1.1) [18].

79 Sample data were labeled and grouped by their RRID and disease status (ND, non-diabetes vs. T2D, type 2
80 diabetes). After filtering, the GSE159984 read count table contained 19,058 genes from N = 27 T2D samples
81 and N = 58 non-diabetes samples (**Fig 1A**). Differential gene expression analysis between ND and T2D groups
82 was performed using the edgeR likelihood ratio test and cutoffs were FDR < 0.05 and $|\log_2$ fold-change$| \geq 0.58$.
83 The processed and filtered read count table and edgeR results are provided in **Table S2** and the R script used
84 for processing is available on GitHub (https://github.com/kalwatlab/Islet_DEGAS_v1). Selected individual
85 genes are shown in Fig 1H as box plots to illustrate the quartiles across all donor samples.
86
87 ***Human islet single-cell data acquisition, filtering, and integration*** – We obtained read count tables for five
88 single-cell human islet datasets (GSE84133; GSE85241; E-MTAB-5061; GSE81608; GSE86469 (**Table S1**))
89 which were realigned to GRCh38.p5 (hg38) [8] (**Fig 1B**). Metadata were downloaded from GEO or obtained from
90 the supplemental information of the respective publications. These five datasets were previously integrated and
91 analyzed [8]. To prepare the data for our machine learning analyses, we repeated the integration and analysis of
92 these datasets in R using Seurat to exclude cells with low expressed genes and cells with over 20% mitochondrial
93 gene expression (low viability). The upper and lower limits varied between the datasets to account for differences
94 in library preparation and sequencing platform following guidelines in the Seurat documentation.
95 Datasets were integrated into a single dataset using Seurat version 4.1.3, the Seurat objects were normalized
96 using regularized negative binomial regression (SCTransform) to correct for batch effect and clustered using
97 default settings. The most variable features were used to identify integration anchors which are passed to the
98 *IntegrateData* function to return a Seurat object containing an integrated expression matrix of all cells.
99 The clusters were visualized using UMAP and expression of pancreatic hormone genes INS, GCG, SST,
100 PPY, GHRL were used to annotate β, α, delta, γ and ε cell clusters, respectively. Clustering analysis used the
101 Louvain algorithm [19] and labelling was performed to identify different cell types because our integrated dataset
102 contained a mixture of cells from pancreatic islets. In total, 22 clusters were identified from 17,273 pancreatic
103 cells and the clusters were classified into endocrine and non-endocrine cell types based on cell-specific marker
104 expression (**Fig 2A**). R scripts are available on GitHub (https://github.com/kalwatlab/Islet_DEGAS_v1). The
105 Seurat object containing all read count matrices and merged data and metadata is available on Mendeley[20].

106

107 ***Transfer learning using DEGAS on human islet transcriptomic data –*** The integrated single cell and bulk
108 datasets were processed with the DEGAS (v1.0) pipeline (https://github.com/tsteelejohnson91/DEGAS) [21] to
109 calculate disease risk scores associated with T2D status or BMI status. The bulk expression data were scaled
110 and normalized prior to DEGAS analysis using the *preprocessCounts* function. The merged scRNA-seq
111 expression matrix, the bulk expression matrix, and donor sample labels (matched with the bulk samples) were
112 used as input. The intersection of highly variable genes between scRNA-seq and bulk expression data were
113 used for further analysis. The DEGAS model was trained and predicted on the formatted data. Updated scripts
114 and instructions for running DEGAS are available on GitHub (version 1.0)
115 (https://github.com/tsteelejohnson91/DEGAS). DEGAS for T2D and BMI were run independently.
116 In DEGAS for T2D, the donor labels were "normal" vs "T2D." In DEGAS for BMI metadata, donors from bulk
117 RNA-seq were categorized and labeled as "lean" (<25 BMI), "overweight" (25-30 BMI) and "obese" (>30 BMI)
118 according to CDC guidelines. The models were trained to calculate T2D-association scores and BMI-association
119 scores for each of the respective categories. We identified differentially-expressed genes using *FindMarkers*
120 function in β-cells with high vs low T2D-association scores ($\beta^{T2D\text{-}DEGAS}$) or high obesity scores ($\beta^{obese\text{-}DEGAS}$) among
121 healthy or T2D donor single-cells.

122

123 ***Subclustering of β-cells and post-DEGAS scRNA-seq analysis*** – We isolated the β-cell cluster using *subset*
124 function and reclustered them seperately from the other islet cell types. The β-cells were further classified by
125 different thresholding parameters including median and quantiles. For quantile and median thresholds, we
126 generated high, medium and low, as well upper 50% and lower 50% of median. We used the *FindMarkers*
127 function to identify differentially-expressed genes in high vs. low (**Fig 4D**), high vs. medium+low, and median
128 (upper 50% vs lower 50%) **(Table S3)**.

129

130 ***Pathway analysis*** – To analyze differentially-expressed genes that were either up- or down-regulated between
131 clusters, enrichment analyses and plot generation were completed using ClusterProfiler, EnhancedVolcano, and
132 ggplot2 R packages in R version 4.2.2 and RStudio. Enrichment of gene sets within the the Biological Process

133 GO terms are shown and cluster profiler outputs are provided in **Table S4**. Venn diagrams for comparing DEGAS
134 and RePACT hit genes (**Table S5**) were created using https://molbiotools.com/listcompare.php. Gene set
135 enrichment analysis was run using GSEA software [22, 23] downloaded from www.gsea-msigdb.org/gsea and the
136 MSigDB Hallmark Gene Set [24] from the Broad Institute. For bulk RNA-seq analysis, the edgeR results containing
137 all expressed genes was used as input to run standard GSEA. For GSEA analysis of comparisons between
138 subpopluations of β-cells from scRNA-seq data, a ranked list based on log_2fold-change was generated without
139 a cutoff for adjusted P value. GSEAPreranked mode used the 'classic' enrichnment statistic for 1000
140 permutations.
141
142 ***Human pancreas tissue staining and microscopy*** – Formalin-fixed paraffin-embedded (FFPE) de-identified
143 human pancreas tissue were obtained through the Integrated Islet Distribution Program (IIDP) and the National
144 Disease Research Interchange (NDRI) (**Table S6**). IIDP provided 5 µm tissue sections on slides and NDRI tissue
145 blocks were processed into 5 µm sections and mounted on glass slides by the Indiana University School of
146 Medicine Histology Lab Service Core. Slides were deparaffinized by xylene and ethanol washes. Antigen
147 retrieval was performed by heating for 40 min in an Epitope Retrival Steamer with slides submerged in Epitope
148 Retrieval Solution (IHC-Tek). Subsequently, slides were placed onto disposable immunostaining coverplates
149 and inserted into the Sequenza slide rack (Ted Pella / EMS) for washing, blocking, and antibody incubations.
150 After three 10 min washes in IHC wash buffer (0.1% Triton X-100 and 0.01% sodium azide in PBS pH 7.4), slides
151 were blocked for 1 h at room temperature in normal donkey serum (NDS) block solution (5% donkey serum, 1%
152 bovine serum albumin, 0.3% Triton X-100, 0.05% Tween-20, and 0.05% sodium azide in PBS pH 7.4). Slides
153 were then incubated overnight at 4°C with primary antibodies diluted in NDS block solution (**Table S7**). After
154 three washes in IHC wash buffer 200 µL each), slides were incubated in secondary antibodies in NDS block
155 solution for 1 h at room temperature, followed by DAPI staining for 10 min and three additional washes. The
156 washed slides were mounted in polyvinyl alcohol (PVA) mounting medium (5% PVA, 10% glycerol, 50mM Tris
157 pH 9.0, 0.01% sodium azide) or Prolong Gold and imaged on Zeiss LSM710 confocal microscope equipped with
158 a Plan-Apochromat 20x/0.8 objective (#420650-9901). Images were processed in the Zeiss Zen software to add
159 scale bars, set coloration for channels, and generate merged images. Scale bars indicate 50 µm. Distribution of

160 DLK1 intensity within INS^+ areas and $CDKN1C^+$ β-cell nuclei were quantified using CellProfiler (v4.2.7). Data
161 were processed in R and plotted in GraphPad prism.
162
163 ***Statistical Analysis –*** In edgeR bulk RNAseq differential gene expression analysis, the likelihood ratio test
164 was applied. In the FindMarkers function in Seurat the default Wilcoxon Rank Sum test was used for
165 differential gene expression analysis. Cut-offs for signifcant differentially expressed genes were set at $|\log_2$fold-
166 change$|$ > 0.58 and adjusted p-value < 0.05.
167

168 **RESULTS**

169 ***Differentially expressed genes from T2D bulk RNA-seq human islets show inflammatory signature and***
170 ***correlate with independent islet datasets*** – We selected bulk RNA-seq data from Marselli, et al. 2020
171 (GSE159984) which contains 58 non-diabetic and 27 T2D samples with associated metadata for BMI, age, and
172 sex [16] (**Fig 1A-C**). To determine the suitability of this data for DEGAS, we reanalyzed the read count matrix
173 from GSE159984 using the edgeR likelihood ratio test to identify up- and down-regulated genes, visualized in
174 both volcano (**Fig 1D**) and MD plot (**Fig 1E**) formats to highlight fold-change versus significance and versus
175 overall expression levels, respectively (**Table S2**). We observed altered expression of genes well-known to be
176 dysregulated in T2D, including *IAPP* [25, 26], *PAX4* [27, 28], *SLC2A2* [29], *FFAR4*[30], and *ENTPD3* [31, 32] (**Fig 1D-E**).
177 Gene set enrichment analysis (GSEA) indicated that the gene expression profile of T2D human islets from this
178 dataset was associated with an inflammatory response phenotype (**Fig 1F**). To compare our analysis of
179 Marselli, et al. with an independent cohort, we selected a bulk RNA-seq human islet transcriptomic analysis
180 from Asplund et al. [17]. We compared the published supplementary table of $\log_2$-fold changes for differentially
181 expressed genes and to the Marselli dataset (**Fig 1G**) (**Table S2**). Significant differentially expressed genes
182 from both Marselli and Asplund data included up-regulation of *SFRP4* and *PODN* (**Fig S1A-B**), and down-
183 regulation of *UNC5D* and *FFAR4* (**Fig S1C-D**). We also found substantial agreement in the directional changes
184 of most other differentially expressed genes from both datasets, indicating that the Marselli islet data
185 represents a suitable cohort for DEGAS analysis. Finally, we compared up- and down-regulated genes from

186 T2D islets in the Marselli dataset with T2D effector genes from the T2D Knowledge Portal [33]. Down-regulated
187 genes *PAX4* and *SLC2A2* and up-regulated genes *APOE* each have evidence of a strong or causal role in T2D
188 at the genetic level, while *CYTIP* was up-regulated in T2D islets and has intron variant SNPs associated with
189 T2D (e.g. rs13384965) [34] (**Fig 1H**).
190
191 ***DEGAS revealed T2D- and obesity-associated β-cells and marker genes within integrated human islet***
192 ***scRNA-seq data*** – We obtained five realigned scRNA-seq datasets [8] including Baron (GSE84133), Muraro
193 (GSE85241), Segerstolpe (E-MTAB-5061), Lawlor (GSE86469), and Xin (GSE81608), and integrated them
194 using Seurat 4.3.0 to subtype each of the major cell types for use with DEGAS. All five datasets were
195 successfully integrated resulting in 17,273 cells (**Fig 2A-B**). Different cell types were clearly stratified from the
196 scRNA-seq data, as indicated by distinct clusters of cell types enriched for their respective markers including
197 β-cells (*INS*), α-cells (*GCG*), δ-cells (*SST*), ε-cells (*GHRL*), PP-cells (*PPY*), acinar cells (*REG1A*, *PRSS1*),
198 ductal cells (*KRT19*, *CFTR*), stellate cells (*COL6A1*), endothelial cells (*PLVAP*), and mesenchymal cells
199 (*CD44*) (**Fig 2C**). Other parameters including sex, BMI, age, and T2D status were represented in the merged
200 dataset (**Fig 2D-G**). Markers of T2D vs. ND for all cells and for only β-cells were also extracted for downstream
201 comparisons (**Fig S2A-B**). There was limited overlap between the differentially expressed genes in Marselli, et
202 al. compared to T2D vs. ND for all single cells, but included *IAPP*, *ENTPD3* and *FFAR4*.
203 Next, we implemented DEGAS [21], a unique tool that leverages the increased sequencing depth, larger
204 number of clinical covariates and sample sizes of bulk sequencing data. We used the deep learning
205 architecture within DEGAS to project bulk islet sequencing data into the same latent representation as single
206 cell islet data via domain adaptation. From this common latent representation, clinical covariates from the bulk
207 sequencing data (e.g. T2D status and BMI) were projected onto single cells in a process called transfer
208 learning. This process generates unitless disease-association scores for each single cell. In parallel, we
209 applied DEGAS to the merged scRNAseq data using either T2D or obesity as the scoring feature.
210 <u>T2D-DEGAS:</u> First, we next mapped the T2D-association scores onto the single cells (**Fig 3A**). β-cells had
211 a wide distribution of scores, possibly reflecting β-cell heterogeneity or altered β-cell gene expression after
212 onset of T2D (**Fig 3B**) [35]. We focused on differential T2D-association scores only among β-cells by subsetting

them based on *INS* expression and superimposed the T2D-association scores onto each β-cell in the UMAP plot (**Fig 4A**). β-cells with higher scores (pink coloration) transcriptionally associate with human islet T2D expression profiles more than the lower scoring (black) β-cells. Next, we explored median and quantile thresholds for the β-cell T2D-association scores (**Fig S3A-B**) and continued our analysis with the cells grouped into high (upper 20%), medium and low (lower 20%) categories (**Fig 4B**). We compared the high versus low $\beta^{T2D\text{-}DEGAS}$ groups to genes enriched in either subset of β-cells **(Fig 4C, Table S3)**. High-scoring $\beta^{T2D\text{-}DEGAS}$ cells had elevated expression of *CDKN1C* ($p57^{Kip2}$) which binds and inhibits G1 cyclin/CDK complexes and is also involved in cellular sensescence [36 37], and reduced expression of *IAPP*, consistent with results from bulk T2D islet expression data (**Fig 1G**). These findings show that the DEGAS can successfully identify diabetes-relevant genes enriched in specific subclusters.

As an additional way to highlight genes that were identified by DEGAS, we excluded genes that were differentially expressed between all non-diabetic (ND) and T2D β-cells (based on the merged scRNA-seq donor metadata **Fig S4D, Table S3**). This process left over 90% of significant differentially expressed genes remaining which were only identified through DEGAS analysis **(Fig 4D).** This emphasized the advantage of using DEGAS to generate continuous variables (disease-association scores) which enable thresholding and prioritizing cellular subpopulations. Gene ontology (GO) analysis on high-scoring $\beta^{T2D\text{-}DEGAS}$ cells showed enrichment in biological process categories including neutrophil degranulation, cell growth, and negative regulation of cell development (**Fig 4E, Table S4**). These categories were driven in part by immune-related genes (e.g. *HLA-B*, *PSAP*) and cell cycle regulator *BTG1*. Conversely, GO analysis on low-scoring $\beta^{T2D\text{-}DEGAS}$ cells showed enrichment of categories primarily related membrane protein targeting, translation, and mitochondrial electron transport (**Fig 4F, Table S4**). The top six enriched GO categories were driven largely by *RPS*/*RPL* genes. We also performed ranked-list GSEA on these same high- vs. low-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations which largely agreed with the GO analysis (**Fig 4G**). We observed significant positive enrichment for hypoxia, TNFα signaling, estrogen response, and glycolysis (**Fig S3C**), while significant negative enriched pathways included overlapping genes in oxidative phosphorylation and Myc-target pathways (**Fig S3D**). Examples of genes that drive the GO and GSEA categories are shown (**Fig 4H**) and heterogeneity in the expression in a selection of these genes can be seen in β-cell UMAP plots (**Fig S3E**).

240

241 Obese-DEGAS: Second, we applied DEGAS using obesity as the disease feature. Obesity and T2D are
242 related, but not mutually inclusive. We surmised that obesity-association scores in β-cells may highlight unique
243 subpopulations compared to the $\beta^{T2D\text{-}DEGAS}$ cells. Therefore, we categorized bulk RNA-seq donors by BMI
244 (lean, <25; overweight, 25-30; obese >30) and ran the DEGAS analysis based on these categories to calculate
245 BMI-association scores similar to our approach with T2D. Overlaying the BMI-DEGAS scores onto the islet
246 single-cell data showed a distinct pattern of cell labeling betwee **Fig S4A-B**). The cell types with the highest
247 obesity-association scores were β-cells, acinar cells, and stellate cells (**Fig 5B**).

248 We next reclustered β-cells with overlaid lean, overweight, and obesity-association scores (**Fig S4C-D, 6A**).
249 It is notable that the subpopulation of $\beta^{lean\text{-}DEGAS}$ cells appear to overlap with the subpopulation of low-scoring
250 $\beta^{T2D\text{-}DEGAS}$ cells ( ), suggesting that single β-cells share features of low BMI and potentially lower T2D
251 association. $\beta^{overweight\text{-}DEGAS}$ scores did not highlight any subpopulation of β-cells (**Fig S4D**). However, we
252 observed two subpopulations of high-scoring $\beta^{obese\text{-}DEGAS}$ cells (**Fig 6A**), a feature that did not occur with $\beta^{T2D\text{-}}$
253 ^{DEGAS} cells (**Fig 4B**). We observed that a distinguishing difference between these two groups of obesity-
254 associated β-cells was the donor diabetes status (ND vs T2D) from within the single-cell metadata. We applied
255 a median quantile threshold to extract these two subpopulations (**Fig S4E**) and classified the cell clusters
256 based on their donor diabetes status (**Fig 6B, Fig S4F**) to create ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$
257 subpopulations (**Fig 6C**). We next determined differentially expressed genes between the ND-$\beta^{obese\text{-}DEGAS}$ and
258 T2D-$\beta^{obese\text{-}DEGAS}$ cells (**Fig 6D**). Gene ontology analysis of genes upregulated in the ND-$\beta^{obese\text{-}DEGAS}$ cluster
259 indicated an enrichment for unfolded protein response (UPR) processes like vesicle transport, translation, and
260 protein folding and stability (**Fig 6E**). Specifically, the ND-$\beta^{obese\text{-}\ DEGAS}$ cluster was enriched for expression of
261 adaptive (e.g. *MANF* and *HSPA5*) and maladaptive (e.g. *TRIB3*, *DDIT3*) UPR genes, as well as genes involved
262 in ER-associated degradation (e.g. *SEL1L*, *DERL2*, and *SDF2L1*). Gene ontology analysis of genes
263 upregulated in the T2D-$\beta^{obese\text{-}DEGAS}$ cluster showed enrichment for hormone transport and secretion pathways
264 (e.g. *IGFBP5*, *UCN3*, *G6PC2*) and inflammatory/immune-related pathways (e.g., *HLA-A/B/E*, *STAT3*) (**Fig 6F**).
265 In parallel, we performed ranked-list GSEA on the ND- and T2D-$\beta^{obese\text{-}\ DEGAS}$ subpopulations which was largely
266 consistent with gene ontology results (**Fig 6G**). Significantly enriched pathways for ND-$\beta^{obese\text{-}\ DEGAS}$ included the

267 UPR and proliferation (E2F and MTORC signaling hallmarks) (**Fig S4G**), while T2D-$\beta^{obese-DEGAS}$ enriched
268 pathways were similar to that of high-scoring $\beta^{T2D-DEGAS}$ cells and included hypoxia and glycolysis (**Fig S4H**).
269 Examples of genes driving the GO and GSEA categories are shown, as well as *CDKN1C* and *DLK1*, given
270 their differential enrichment in both $\beta^{T2D-DEGAS}$ and T2D-$\beta^{obese-DEGAS}$ cells (**Fig 6H**).
271 Combined, our results applying DEGAS using T2D and obesity as disease features support the idea that
272 the identified genes and pathways may underlie heterogeneity in β-cell resilience or susceptibility to obesity-
273 related stress in the context of T2D development. In particular, *DLK1* and *CDKN1C* ranked highly in both T2D-
274 DEGAS and obese-DEGAS analyses (**Fig S4I**), making them reasonable candidates for validation of this
275 approach.
276
277 ***Heterogeneous expression of DLK1 and CDKN1C in human pancreatic islets from diabetic and non-***
278 ***diabetic donors with lean, overweight, or obese BMI.*** – To support the DEGAS approach for identifying true
279 markers of β-cell subpopulations or heterogeneity associated with diabetes, we selected DLK1 and CDKN1C
280 as candidate genes for immunostaining in FFPE human pancreas sections from non-diabetic and T2D donors.
281 We chose Delta-Like 1 Homolog (DLK1) because of its strong associations with T1D [38] and T2D [39]. *DLK1*
282 expression overlapped with high scoring $\beta^{T2D-DEGAS}$ cells (**Fig 7A**) and with T2D-$\beta^{obese-DEGAS}$ cells (**Fig 7B**).
283 DLK1 immunostaining primarily colocalized with β- cells in non-diabetic human pancreas and exhibited
284 heterogeneity in its expression intensity (**Fig 7C, S5**). Some islets had DLK1/INS co-staining in many β-cells
285 while others only a few $DLK1^+$ β-cells and this pattern did not appear to be associated with BMI. In T2D
286 pancreas, DLK1 expression appeared to be reduced in some donors, however an analysis of the distribution of
287 DLK1 staining intensity across INS^+ cells indicated that there was no major difference between ND and T2D
288 (**Fig 7D**). We chose our second candidate, Cyclin-Dependent Kinase inhibitor 1C (CDKN1C), because it has
289 been linked to β-cell replication [40]. CDKN1C expression also overlapped with high scoring $\beta^{T2D-DEGAS}$ cells (**Fig**
290 **7E**) and with T2D-$\beta^{obese-DEGAS}$ cells (**Fig 7F**). CDKN1C staining was primarily nuclear and was heterogeneous
291 across β-cells in non-diabetic human pancreas (**Fig 7G**). In agreement with the DEGAS results, CDKN1C
292 staining was increased in β-cell nuclei from T2D donors as quantified by CellProfiler analysis (**Fig 7H**). BMI
293 status did not appear to correlate with CDKN1C expression in either non-diabetic or T2D donors.

294

295 **DISCUSSION**

296 ***Deep transfer learning is a useful approach to identify disease-associated genes in subsets of***
297 ***heterogeneous islet cells –*** The rapid advancement of deep learning has unveiled new opportunities in
298 diabetes research, specifically the ability to merge and analyze the large amount of publicly available omics data.
299 The DEGAS deep transfer learning framework provides an extremely powerful and versatile tool allowing for
300 time to event outcomes and classification outcomes to be transferred to individual cells. Notably, this direct
301 transfer to single cells avoids the cluster resolution problems that arise using deconvolution approaches, allowing
302 for subsets of cells within a cluster to be assigned different associations with clinical outcomes. In this way,
303 DEGAS allows the definition of essentially new cell subpopulations that would not necessarily be identified by
304 standard clustering algorithms and these subpopulations are rationally selected based on assigned disease
305 association scores. A key point from our study is that the disease-associated single cells identified by DEGAS
306 are those cells whose transcriptomic signature share complex patterns with patient transcriptomics that have
307 known clinical and metabolic attributes. Because of this, the genes identified as up- or down-regulated in subsets
308 of β-cells may be: 1) altered independent of disease due to genetic background of donors; 2) altered downstream
309 of the obese or diabetic states; or 3) representative of resilient or dysfunctional β-cells in the face of metabolic
310 syndrome. Through careful cohort inclusion criteria and batch effect removal, our analyses aim to limit these
311 caveats when possible.

312 Various deconvolution methods enable estimating relative proportions of cell types in bulk samples using
313 scRNA-seq data (e.g. BSEQ-sc [1], MuSiC [41], CIBERSORT [42], and DWLS [43]). These methods have been
314 instrumental in identifying changes to relative quantity of cell types within normal and disease tissues; however,
315 they do not provide detailed information about subsets of cell types which may be related to disease and are
316 limited to the resolution of the predefined cell types through clustering. Early attempts to address this include
317 unsupervised learning algorithms (e.g. RaceID) [44]. Now, there are multiple advanced computational approaches
318 to identify subsets of cells related to disease status, survival, drug response, and other disease metrics. These
319 tools can be broadly categorized as 'cell prioritization algorithms' and include: DEGAS [21], scAB [45], Scissor [46],

320 and scDEAL [47]. Scissor and scAB can assign survival or clinical information from bulk expression data to
321 disparate scRNA-seq datasets using regression and matrix factorization, respectively. scDEAL is a deep learning
322 framework specifically designed for the assignment of drug response information from bulk expression data to
323 subsets of single cells. The major advantages of DEGAS are 1) the ability to select subsets of β-cells at an
324 individual cell resolution based on their disease association; 2) unique qualities of the highly non-linear DEGAS
325 technology vs. linear models of other tools to detect complex molecular programs; and 3) speed and scalability
326 based on efficient implementation. DEGAS has the potential to be used on datasets containing well over 1 million
327 cells, unlike other methods.

328 Nevertheless, certain linear model-based approaches, like RePACT (regressing principal components for the
329 assembly of continuous trajectory), have successfully uncovered genes associated with β-cell heterogeneity in
330 T2D and obesity [5]. In that study, T2D and obesity 'trajectory' genes were identified. We compared the genes
331 enriched in our high-scoring $\beta^{T2D\text{-}DEGAS}$ and in our T2D-$\beta^{obese\text{-}DEGAS}$ cells with the corresponding T2D and obesity
332 trajectory genes from Fang et al., respectively (**Fig S6**, **Table S5**). We noted substantial agreement between
333 DEGAS and RePACT, for example the transcription factor *SIX3* was enriched in T2D-$\beta^{obese\text{-}DEGAS}$ cells and in
334 Fang et al.'s obese trajectory β-cells [5]. *SIX3* has been implicated as a diabetes risk gene and is important for β-
335 cell function [48, 49]. Additionally, both approaches highlighted the association of *DLK1* with obesity, but only
336 DEGAS identified *DLK1* association with T2D. Overall, the directionality of T2D or obesity gene associations
337 agreed between DEGAS and RePACT results. However, not all the same genes were identified by both
338 approaches, for example, *FXYD2* was identified via RePACT, but not in DEGAS, although *FXYD2* is a down-
339 regulated gene in T2D in Marselli et al. and Asplund et al. (**Fig 1G**). Interestingly, *FXYD2* has was recently shown
340 to be a downstream effector gene of *HNF1A* in single human β-cells [50]. Down-regulation of *HNF1A* in T2D β-
341 cells, and reduced *FXYD2* expression as a result, may contribute to membrane hyperpolarization and reduced
342 function [50]. Distinct from the application of RePACT in Fang, et al.[5], we did not observe any differences in *HNF1A*
343 expression in our DEGAS analyses and *FXYD2* was slightly enriched, but not significantly altered, in T2D-$\beta^{obese\text{-}}$
344 ^{DEGAS} cells. Taken together, it is necessary to apply multiple approaches to merged sets of publicly available data
345 because each approach will likely uncover unique and important subsets of β-cells or specific genes.

346

347 ***Differences between high- and low-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations*** – High-scoring $\beta^{T2D\text{-}DEGAS}$ cells had
348 enrichment of pathways like hypoxia and TNFα signaling which contain some overlapping genes like *BTG1*,
349 *BTG2*, and *BHLHE40*. *BHLHE40* was recently shown to cause hypoxia-induced β-cell dysfuction via interfering
350 with *MAFA* expression [51]. *BTG1* and *BTG2* are known as anti-proliferation factors. Given the expression of both
351 *BTG1* and *CDKN1C* in high-scoring $\beta^{T2D\text{-}DEGAS}$ cells, it is possible these cells have altered proliferative potential,
352 even in the context of β-cells which are well-known to have low proliferation. Although some genes like *CDKN1C*
353 can be identified as differentially expressed in simple comparisons of all T2D vs ND donor β-cells in scRNA-seq
354 data, DEGAS has the advantage of prioritization or ranking of candidates. For example, *CDKN1C* is within the
355 top ten enriched candidates (by adjusted P value and fold-change) in the $\beta^{T2D\text{-}DEGAS}$ cluster but is 286th when
356 simply comparing all T2D vs ND single β-cells (**Table S3**). Additionally, deletion of *CDKN1C* has been shown to
357 improve human islet function [40] and gain-of-function mutants are associated with diabetes development and
358 hyperinsulinism [52, 53]. In agreement with our results, we found CDKN1C was significantly increased in β-cells
359 from T2D donors (**Fig 7H**). From our analysis we cannot determine whether this is a response to the diabetic
360 milieu in those donors or if CDKN1C is increased early and prevents β-cell compensation, further contributing to
361 T2D pathogenesis. Understanding the differences between $CDKN1C^{+}$ and $CDKN1C^{-}$ human β-cells would be of
362 interest in future work to provide insights into the regulation of β-cell replication.

363 The low-scoring $\beta^{T2D\text{-}DEGAS}$ cells were enriched in pathways involving oxidative phosphorylation and Myc
364 targets. We also noted enrichment of many small and large ribosomal genes (*RPS*/*RPL*) in this subpopulation.
365 *RPS*/*RPL* ribosomal genes are known to be highly expressed in primary human β-cells [54], and glucose
366 stimulation induced translation of over 50 different *RPS*/*RPL* genes in human β-cell line EndoC-βH2 [55]. This
367 suggest that low-scoring $\beta^{T2D\text{-}DEGAS}$ cells may have relatively increased translation and metabolic activity. Taken
368 together, high-scoring $\beta^{T2D\text{-}DEGAS}$ cells may be reflective of dysfunctional β-cells in either a pre-T2D or T2D-onset
369 state, while low-scoring $\beta^{T2D\text{-}DEGAS}$ cells could potentially represent more resilient β-cells.

370

371 ***Differences between ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ subsets of β-cells*** – Our understanding of why
372 some individuals with obesity develop T2D while others do not is incomplete. A 2021 NHANES report indicated

373 a 41.9% prevalence for obesity in US adults aged 20 and over, with a 14.8% prevalence for diabetes [56], in line
374 with CDC reports [57]. This suggests that many individuals with obesity have either not progressed, or may never
375 progress, to T2D. This phenomenon may be related to those individuals categorized as metabolically healthy
376 obese as compared to metabolically unhealthy obese [58]. Expanded work with DEGAS may uncover specific
377 genes that underly this relationship. In our analysis we were able to stratify based on T2D status of the single
378 cell donors and overlay the relative obesity-association scores from DEGAS. That analysis allowed us to set
379 rational thresholds for grouping subsets of β-cells to compare ND- and T2D-$\beta^{obese\text{-}DEGAS}$ cells. Perhaps
380 counterintuitively, UPR genes were highly enriched in the ND-$\beta^{obese\text{-}DEGAS}$ cells as opposed to the T2D-$\beta^{obese\text{-}}$
381 ^{DEGAS} cells. It is possible these ND-$\beta^{obese\text{-}DEGAS}$ cells are in a stressed pre-diabetic state on the path to T2D.
382 Dominguez-Gutierrez et al. analyzed islet scRNA-seq data using pseudotime analysis and identified three major
383 β-cell states which were defined by their insulin expression and UPR level [59]. The authors speculate that β-cells
384 periodically pass through these states, which in scRNA-seq are visualized as subpopulations of cells. Previously,
385 a *FTH1* subpopulation of β-cells was identified with implications in the UPR [60]. In agreement with those results,
386 we found *FTH1* was upregulated in ND-$\beta^{obese\text{-}DEGAS}$ cells. Therefore, another possibility is that the UPR induced
387 in ND-$\beta^{obese\text{-}DEGAS}$ cells marks a population with lower functionality but higher proliferative potential to combat the
388 insulin resistant obese state, and the relatively lower UPR in T2D-$\beta^{obese\text{-}DEGAS}$ cells possibly correlates with
389 enhanced functionality in established T2D. Concordantly, T2D-$\beta^{obese\text{-}DEGAS}$ cells had enrichment of hormone
390 secretion genes (e.g. *SYT7, G6PC2, NEUROD1, UCN3*, *FFAR1*) in our pathway analysis (**Fig 6F**). In a study of
391 human, mouse, and pig islet scRNA-seq data, subpopulations of 'stressed' β-cells were identified that exhibited
392 enriched hallmark pathways similar to what we observe in ND-$\beta^{obese\text{-}DEGAS}$ cells [6]. Thus, multiple studies including
393 our current analysis support the existence of subpopulations of stressed β-cells, whether transient or stable, that
394 could in principle be targeted by therapeutics.

395 We also looked for enrichment of potential secreted biomarkers in high-scoring $\beta^{T2D\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$
396 cells and identified *LRPAP1* and *C1QL1*. *LRPAP1* encodes the LDL receptor-related protein-associated protein
397 1 (LRPAP) and is enriched in this subpopulation. LRPAP is ubiquitously expressed and is predicted to be an ER
398 resident protein [61]. LRPAP appears to be heterogeneously expressed among islet cells and was detected in
399 human plasma in the in the Human Protein Atlas [62]. *C1QL1* encodes a secreted peptide that is highly-expressed

400 in human islets [63] and was enriched in $\beta^{T2D\text{-}DEGAS}$ cells, although little else is known about its role in β-cells.
401 Increased C1QL1 release could potentially signal through its GPCR BAI3 to suppress insulin secretion of
402 surrounding β-cells, as has been shown for C1QL3 [64].

403

404 ***Potential for DEGAS in identifying β-cell heterogeneity markers –*** Subsets or subpopulations of β-cells are
405 an emergent property of β-cell heterogeneity. Two of the most widely used scRNA-seq datasets have also
406 identified significant heterogeneity within pancreatic cell types, notably β-cells [7, 60]. Various markers have been
407 identified to define these β-cell subpopulations, including expression of ST8SIA1 and CD9 [65] or Flattop
408 expression [66]. Additionally, UCN3 marks mature β cells [67] and $RBP4^{+}$ β-cells correlated with reduced function
409 [68]. Functional heterogeneity has also been described as in the case of leader/first-responder β cells [69, 70] or hub
410 β cells [71], and has been linked to transcription factors including PDX1 [72] and HNF1A [50]. Nevertheless, cellular
411 heterogeneity within a single cell type is complicated and there is uncertainty in the single cell analytics field
412 about what constitutes a stable cell type versus a transient cell state [73]. Tying transcriptionally distinct clusters
413 to function or morphology may be key to making this distinction [74]. By inferring disease outcomes in high
414 resolution cellular subtypes, application of DEGAS can help to inform this debate.

415 In our analysis, the expression of *DLK1* and *CDKN1C* was enriched in high-scoring $\beta^{T2D\text{-}DEGAS}$ and in T2D-
416 $\beta^{obese\text{-}DEGAS}$ cells. Recent *in vitro* work has also highlighted DLK1 and CDKN1C as having increased expression
417 in human pancreatic islets treated with 30mM glucose for 48 h [75]. In agreement with scRNAseq data, we also
418 observed heterogeneous DLK1 and CDKN1C immunostaining of β-cells from human pancreas sections. *DLK1*
419 was previously noted to be enriched in subpopulations of β-cells [1, 2, 4, 8, 76], however its heterogeneity in
420 transcript and protein abundance in non-diabetic versus T2D human islet β-cells had not been described until
421 now. DLK1 is a maternally-imprinted gene[77] with described roles in Notch signaling[78-80] and glucose
422 homeostasis[81]. Supporting an active role β-cell disease, a recent preprint reported DLK1 was required for
423 proper maturation, function, and stress resilience of β-like cells differentiated from human embryonic stem cells
424 [82]. Our results combined with the published and available preprint data suggest the role of DLK1 in β-cell
425 development and function may be nuanced, given that we did not observe a consistent up or down regulation

426 at the protein level in human islets. However, DLK1 may serve as a surface marker of, or be released from,
427 stressed or at-risk β-cells, but increased sample sizes and in-depth analyses of human islet samples will be
428 required to better describe these possible roles. Our DEGAS results were confirmed however with CDKN1C
429 expression being increased in T2D β-cells. Other work has shown that suppression of CDKN1C expression in
430 hyperglycemic, immunodeficient mice increased β-cell replication by three fold and those newly replicated cells
431 retained mature β-cell functionality, indicating that manipulating CDKN1C expression can cause B-cell
432 expansion in T2D patients [40]. Additionally, loss-of-function mutations in CDKN1C can cause focal congenital
433 hyperinsulinism, wherein increased β-cell proliferation and IGF-II expression are observed [83].

434 Applying tools like DEGAS to publicly available data will increase our understanding of the underlying β-cell
435 features associated with progression to T2D or with an obese non-T2D state. Expanded DEGAS analyses will
436 be needed to include scRNA-seq data from islets subjected to various models of T1D/T2D (e.g. cytokines,
437 glucolipotoxicity). Major questions include whether the gene candidates identified by DEGAS are protective,
438 disease-causing, or simply markers in β-cell subpopulations. In the future, these questions may be addressed
439 by functional validation studies or observations in genome-wide association studies.

440

441 ***Limitations of the study and future directions*** – We think the approach of joining islet omics data to deep
442 learning and artificial intelligence is an area in need of increased attention; however, we appreciate that our
443 current study has limitations. First, our analysis utilized a single bulk RNA-seq human islet dataset, although the
444 Marselli study contains a relatively large number of islet donors compared to most available datasets. Second,
445 we merged only five scRNA-seq human islet datasets for this application of DEGAS. The datasets were chosen
446 because of the extensive meta-analyses to which they have already been subjected [8], and their use as
447 benchmarking data in many scRNA-seq analysis tools [84-86]. scRNA-seq has a limitation on the number of
448 detectable genes in each cell, as compared to bulk RNA-seq samples. While these limitations may eventually
449 be mitigated by improved technologies, we can begin to overcome these by merging a larger number of human
450 islet bulk and scRNA-seq studies which contain more donors. Increased sample sizes and study variables come
451 with their own challenges, however the future gene candidates identified by DEGAS will be of even higher

452 confidence. Additionally, scRNA-seq is a snapshot in time and it may be tenuous to claim that a particular cell
453 state is transient or persistent. To support a claim that a given subset/subtype of β-cells represents an actual
454 stable population in scRNA-seq data requires at least finding that the cells comprising the candidate population
455 occur across multiple individual donors. In future studies, our use of the diabetes field's considerable investment
456 in islet transcriptomics data and state-of-the-art cell prioritization tools will enable simultaneous identification of
457 disease-associated cellular subtypes and associated biomarkers of function and dysfunction. Additionally,
458 increasing the number of human pancreas donor tissue samples for high-content image analysis will improve
459 confidence in candidate gene validation.

460 **CONCLUSIONS**

461 DEGAS has the potential to be applied to even larger mergers of single cell data (> 1 million cells) using the
462 newly developed DEGAS atlas implementations. A vast amount of human (and other species) islet single cell
463 transcriptomics data is publicly available, but often requires substantial reformatting of metadata and realigning
464 of reads to be harmonized. As our endeavors and those of others proceed, DEGAS will lead to even higher
465 confidence predictions of β-cell subpopulations. It is also apparent from our DEGAS analyses that other non-β-
466 cells are highlighted in the islet single cell map for associations with T2D and obesity. Although outside the
467 scope of this work, further exploration of our current DEGAS analyses, or of analyses using larger single cell
468 integrations, will have implications for the other major islet and pancreas cell types, including α-cells, δ-cells
469 and PP-cells. For example, PP cells have been shown to have a role in pancreatogenic diabetes as opposed
470 to T2D [87], and there appear to be subpopulations of high and low scoring $PP^{T2D\text{-}DEGAS}$ cells within the DEGAS
471 analysis (**Fig 3A**). Our successful application of deep transfer learning in human islet data opens the possibility
472 of predicting subtypes of β-cells in other diseases, like T1D and congenital hyperinsulinism, to find potential
473 biomarkers and therapeutic candidates. In some rare diseases, like congenital hyperinsulinism, single cell data
474 is not available, however spatial transcriptomics (ST) could be applied to FFPE sections. DEGAS has already
475 been implemented to analyze ST data in prostate cancer and diabetic pancreas [88-90], and will be increasingly
476 applicable to human pancreas samples as ST technology continues to advance. It is important to note that
477 DEGAS is just one of many machine learning approaches to analyze transcriptomic data. It will be important to

include DEGAS in combination with both linear and other non-linear models to capture as many relevant gene candidates as possible.

List of abbreviations

Abbreviation	Definition
ND	non-diabetic
T1D	type 1 diabetes
T2D	type 2 diabetes
RNA-seq	RNA sequencing
scRNA-seq	single cell RNA sequencing
FFPE	formalin-fixed paraffin-embedded
DEGAS	Diagnostic Evidence GAuge of Single cells
DLK1	Delta Like Non-Canonical Notch Ligand 1
MANE	Matched Annotation from NCBI and EBI
RRID	Research Resource identifer
GSEA	gene set enrichment analysis
MSigDB	Molecular signatures database
BMI	body mass index
UMAP	uniform manifold approximation and projection
$\beta^{T2D\text{-}DEGAS}$	T2D-DEGAS disease assocation score for beta cells
$\beta^{obese\text{-}DEGAS}$	obese-DEGAS disease assocation score for beta cells
ND-$\beta^{obese\text{-}DEGAS}$	obese-DEGAS disease assocation score for beta cells from ND donors
T2D-$\beta^{obese\text{-}DEGAS}$	obese-DEGAS disease assocation score for beta cells from T2D donors
GO	gene ontology
RePACT	regressing principle components for the assembly of continuous trajectory
NDS	normal donkey serum
PBS	phosphate buffered saline

DECLARATIONS

Consent for publication

Human islet transcriptomic data used in this project are all published datasets and the donors are deidentified. Human pancreatic tissue sections were obtained from NDRI and were deidentified. Studies on deidentified human tissue or publicly available human transcriptomics datasets do not constitute human research as defined in 45 CFR 46.102(f).

DATA AVAILABILITY

The raw data sets supporting the results of this article are available in the NCBI GEO and the ArrayExpress repositories under the following persistent identifiers (also shown in **Table S1**): bulk RNA-seq: GSE159984; scRNA-seq: GSE84133, GSE85241, GSE86469, E-MTAB-5061, GSE81608.

AVAILABILITY OF SOURCE CODE AND REQUIREMENTS

- Project name: Islet DEGAS
- Project home page: https://github.com/kalwatlab/Islet_DEGAS_v1
- Operating system(s): Platform independent
- Programming language: R and Python
- Other requirements: not applicable
- License: ©

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by internal funds at the Indiana Biosciences Research Institute (M.A.K. and T.S.J.), AnalytixIN (T.S.J.), Indiana University Precision Health Initiative (T.S.J), 1R01GM148970 (T.S.J), 1R21CA264339 (T.S.J).

AUTHORS' CONTRIBUTIONS

G.R., R.S., O.L., S.R., S.D.M., A.M.M., T.S.J., and M.A.K. performed data analyses. G.R. and D.R. performed experiments. M.A.K. and T.S.J. conceived and initiated this project. M.A.K. and T.S.J. supervised the project. A.M.M. provided processed data files. G.R., T.S.J. and M.A.K. wrote the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGEMENTS

513 We thank Dr. Mark Huising for helpful conversations and providing access to processed data. Thank you to Dr.
514 Andrew Templin for helpful conversations and critical review of the manuscript. Thank you to Dr. Anthony Piron
515 at Université Libre de Bruxelles for assistance with metadata mapping for GSE159984. This work was
516 supported by the Histology Core of the Indiana Center for Musculoskeletal Health at IU School of Medicine and
517 the Bone and Body Composition Core of the Indiana Clinical Translational Sciences Institute (CTSI). A specific
518 thanks to Drew Brown in the Histology Lab Service Core for assisting with pancreas sectioning.

FIGURES AND LEGENDS

Figure 1. Data acquisition and workflow to train DEGAS using human pancreatic islet transcriptomic data for prediction of T2D-associated cells. A) Read count data for human islet bulk RNA-seq from GSE138748 was downloaded, processed, and matched with donor metadata. The dataset included 58 non-diabetic and 27 type 2 diabetic (T2D) samples. **B)** Human islet single-cell RNA-seq (scRNA-seq) count matrices were generated by realignment of reads and the datasets were integrated in Seurat. **C)** DEGAS transfers the bulk donor expression data and clinical trait information to individual cells in the single cell matrix for the purpose of prioritizing cells. This allows cells to be assigned scores which can then be thresholded for downstream analysis. **D,E)** Bulk RNA-seq data from Marselli et al. was analyzed by edgeR and displayed as both (**D**) volcano plot and (E) MA plot. Differentially expressed genes (DEGs) are those with $p<0.05$ and >1.5-fold changed and are highlighted in red. The MA plot in (**E**) provides a sense of relative transcript abundance among DEGs. **F)** Gene set enrichment analysis (GSEA) results for Hallmark_Inflammatory_Response show an enrichment for DEGs in T2D human islets. **G)** A simple comparison of up- and down-regulated genes in T2D human islets in RNA-seq data between Marselli et al. and Asplund et al. shows consistent findings. Significant DEGs from Marselli et al are outlined in black. **H)** Selected significant genes from panel **1D** that were also found on the Type 2 Diabetes Knowledge Portal database of known diabetes effector genes. The data for selected genes across all donors are displayed in box and whiskers plots. Whiskers extend to 1.5 times the interquartile range.

Figure 2. Validation of merged scRNA-seq datasets from five human islet studies. A) Seurat clustering of merged human islet scRNA-seq data from Fig 1B. **B)** Study identifiers overlaid on single cell plot. **C)** Gene expression plots for the major endocrine and exocrine pancreas genes to demonstrate successful clustering of specific cell types. Metadata was overlaid onto single cell UMAP plots for sex (**D**), BMI (**E**), age (**F**), and diabetes status (**G**).

Figure 3. DEGAS analysis based on T2D status. A) DEGAS T2D association scores (T2D-DEGAS) for each cell was overlaid onto the single cell UMAP plot. Higher scores in pink/red indicate a strong positive association and negative scores in darker black indicate a negative association of those cells with T2D. **B)** Violin plot displaying the aggregate T2D-DEGAS scores per cell type.

Figure 4. Identification of differentially-expressed genes in high- and low-scoring $\beta^{T2D\text{-}DEGAS}$ cell populations. . **A)** DEGAS scores for T2D association ($\beta^{T2D\text{-}DEGAS}$) were overlaid onto the cell plot. **B)** β-cells were classified as low, medium and high-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations for downstream analysis. **C)** Differential expression analysis of high vs. low $\beta^{T2D\text{-}DEGAS}$ scores. Genes with $p<0.05$ and >1.5-fold change are highlighted in red. **D)** Genes from (**C**) were filtered to remove DEGs that could be identified by comparing β-cells of T2D vs non-diabetic donors in the single cell data (**Fig S2B**). Gene ontology (GO) analysis of genes enriched in the high (**E**) and low (**F**) scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations. **G)** GSEA results for high- vs. low-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations. **H)** Bubble plot highlighting genes driving GO and GSEA categories.

Figure 5. DEGAS analysis based on obesity status. A) DEGAS obesity association scores (obese-DEGAS) for each cell was overlaid onto the single cell UMAP plot. Higher scores in pink/red indicate a strong positive association and negative scores in darker black indicate a negative association of those cells with obesity. **B)** Violin plot displaying the aggregate obese -DEGAS scores per cell type.

Figure 6. Two subpopulations of high $\beta^{obese\text{-}DEGAS}$ cells defined by underlying single cell donor T2D status uncovers stress signature in non-diabetic high-scoring obese-DEGAS β-cells. A) β-cells were subsetted from Fig 5A based on *INS* expression and obese-DEGAS scores overlaid onto the single cells. **B)** Donor diabetes status overlaid onto the β-cells. **C)** β-cells were further subsetted based on both their high obese-DEGAS scores and by the underlying T2D status of the single cell donors, leading to two major subpopulations, ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ cells. **D)** Differential expression analysis comparing ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ subpopulations. Genes with $p<0.05$ and >1.5-fold change are highlighted in red. Gene ontology (GO) analysis of genes enriched in the ND-$\beta^{obese\text{-}DEGAS}$ (**E**) and T2D-$\beta^{obese\text{-}DEGAS}$ (**F**)

subpopulations. **G)** Gene set enrichment analysis (GSEA) of ND- vs. T2D-$\beta^{obese\text{-}DEGAS}$ subpopulations. **H)** Bubble plot highlighting genes driving GO and GSEA categories.

Figure 7. DLK1 is a β-cell gene heterogeneously expressed between cells and islets of non-diabetic and T2D humans. A) DLK1 expression and $\beta^{T2D\text{-}DEGAS}$ scores overlaid onto the β-cell cluster. Yellow dots indicate cells with high DLK1 expression and high $\beta^{T2D\text{-}DEGAS}$ score, as shown in the adjacent overlay heatmap. **B)** DLK1 expression and $\beta^{obese\text{-}DEGAS}$ scores overlaid onto the β-cell cluster. Yellow dots indicate cells with high DLK1 expression and high $\beta^{obese\text{-}DEGAS}$ score. **C)** Formalin-fixed paraffin-embedded (FFPE) sections of human pancreas from non-diabetic (ND) and T2D donors were stained for INS and DLK1. Representative images are shown from three different ND and T2D donors. **D)** CellProfiler analysis showing distribution of DLK1 in ND vs T2D donors. UMAP plot overlaying CDKN1C expression and $\beta^{T2D\text{-}DEGAS}$ (**E**) or $\beta^{obese\text{-}DEGAS}$ (**F**) scores. **G)** Formalin-fixed paraffin-embedded (FFPE) sections of human pancreas from non-diabetic (ND) and T2D donors were stained for INS and CDKN1C. Representative images are shown from three different ND and T2D donors. **H)** Increased CDKN1C expression observed in β-cells from T2D patients. *, $P < 0.05$ by Student's T-test.

Supplemental Information

Figure S1. Islet Gene View results of DEGs found in both Marselli et al. and Asplund et al. *SFRP4* (**A**), *PODN* (**B**), *FFAR4* (**C**), and *UNC5D* (**D**) were each significantly regulated in T2D vs. non-T2D human islet RNA-seq data in Marselli et al. and Asplund et al. Plots shown were generated using Islet Gene View browser (https://mae.crc.med.lu.se/IsletGeneView/). *FFAR4* and *UCN5D* were both down-regulated in T2D and *SFRP4* and *PODN* were both up-regulated.

Figure S2. Human islet single cell RNA-seq clusters and diabetes marker genes for all cells and β-cells. A) FindClusterMarkers between all single cells from T2D donors vs. non-diabetic (ND) donors. **B)** β-cells reclustered and **(C)** labeled based on donor T2D status. **D)** FindClusterMarkers between β-cells from T2D vs. ND donor single cell data.

Figure S3. Thresholding analysis for $\beta^{T2D\text{-}DEGAS}$ scores and differential gene expression analysis and plots. A) Quantile thresholding and volcano plot comparing the upper 20% of $\beta^{T2D\text{-}DEGAS}$ scores to the lower 80%. **B)** Median thresholding and volcano plot comparing the upper 50% vs lower 50% of $\beta^{T2D\text{-}DEGAS}$ scoring cells. **C)** GSEA results for high βT2D-DEGAS scoring cells. **D)** GSEA results of low-scoring $\beta^{T2D\text{-}DEGAS}$ scoring cells. **E)** Gene expression overlays on β-cells of selected genes that drive enrichment of GO terms and GSEA categories.

Figure S4. Score overlay on all islet cells for $\beta^{lean\text{-}DEGAS}$ **(A)** and $\beta^{overweight\text{-}DEGAS}$ **(B). C,D)** Similar score overlays are also shown for β-cells alone. **E)** Median thresholding of $\beta^{obese\text{-}DEGAS}$ score. **F)** Subsetting highest-scoring cells based on $\beta^{obese\text{-}DEGAS}$ score. **G)** GSEA results for ND-$\beta^{obese\text{-}DEGAS}$ cells. **H)** GSEA results for T2D-$\beta^{obese\text{-}DEGAS}$ cells. **I)** Venn diagram showing number of genes in $\beta^{T2D\text{-}DEGAS}$ and $\beta^{obese\text{-}DEGAS}$.

Figure S5. Formalin-fixed paraffin-embedded (FFPE) sections of human pancreas from non-diabetic (ND) and T2D donors were stained for INS and DLK1. Representative images are shown from two different ND and T2D donors.

Figure S6. **Overlap of RePACT trajectory genes with genes identified by DEGAS in human islet scRNA-seq data.** Genes from βT2D-DEGAS (**A**), or ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ (**B**) that had adjusted p-values < 0.05 and $|\log_2 \text{fold-changed}| > 0.58$ were selected for comparison to Fang et al. **Table S5** T2D and obesity trajectory genes, respectively.

Additional Files

Table S1. List of bulk and single cell transcriptomic datasets.
Table S2. Bulk RNA-seq metadata, analysis, and comparisons to publicly-available datasets.
Table S3. Differential gene expression in DEGAS-identified β-cell subpopulations.
Table S4. Pathway analysis results.
Table S5. Comparison of DEGAS T2D and obesity genes to Fang et al. RePACT trajectory genes..
Table S6. FFPE human pancreas metadata.
Table S7. Antibodies.

REFERENCES

1. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst. 2016;3(4):346-60 e4. doi: 10.1016/j.cels.2016.08.011. PMID: 27667365. PMCID: PMC5228327.
2. Li J, Klughammer J, Farlik M, Penz T, Spittler A, Barbieux C, et al. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. EMBO Rep. 2016;17(2):178-87. doi: 10.15252/embr.201540946. PMID: 26691212. PMCID: PMC4784001.
3. Wang YJ, Schug J, Won KJ, Liu C, Naji A, Avrahami D, et al. Single-Cell Transcriptomics of the Human Endocrine Pancreas. Diabetes. 2016;65(10):3028-38. doi: 10.2337/db16-0405. PMID: 27364731. PMCID: PMC5033269.
4. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, Sivakamasundari V, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res. 2017;27(2):208-22. doi: 10.1101/gr.212720.116. PMID: 27864352. PMCID: PMC5287227.
5. Fang Z, Weng C, Li H, Tao R, Mai W, Liu X, et al. Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key beta-Cell-Specific Disease Genes. Cell Rep. 2019;26(11):3132-44 e7. doi: 10.1016/j.celrep.2019.02.043. PMID: 30865899. PMCID: PMC6573026.
6. Tritschler S, Thomas M, Böttcher A, Ludwig B, Schmid J, Schubert U, et al. A transcriptional cross species map of pancreatic islet cells. Molecular Metabolism. 2022. doi: 10.1016/j.molmet.2022.101595.
7. Segerstolpe A, Palasantza A, Eliasson P, Andersson EM, Andreasson AC, Sun X, et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes.

Cell Metab. 2016;24(4):593-607. doi: 10.1016/j.cmet.2016.08.020. PMID: 27667667. PMCID: PMC5069352.

8. Mawla AM, Huising MO. Navigating the Depths and Avoiding the Shallows of Pancreatic Islet Cell Transcriptomes. Diabetes. 2019;68(7):1380-93. doi: 10.2337/dbi18-0019. PMID: 31221802. PMCID: PMC6609986.

9. Elgamal RM, Kudtarkar P, Melton RL, Mummey HM, Benaglio P, Okino ML, et al. An integrated map of cell type-specific gene expression in pancreatic islets. bioRxiv. 2023. doi: 10.1101/2023.02.03.526994. PMID: 36778506. PMCID: PMC9915747.

10. Fabbrini E, Yoshino J, Yoshino M, Magkos F, Tiemann Luecking C, Samovski D, et al. Metabolically normal obese people are protected from adverse effects following weight gain. J Clin Invest. 2015;125(2):787-95. doi: 10.1172/JCI78425. PMID: 25555214. PMCID: PMC4319438.

11. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579-89. doi: 10.1038/ng.609. PMID: 20581827. PMCID: PMC3080658.

12. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007;316(5829):1331-6. doi: 10.1126/science.1142358. PMID: 17463246.

13. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007;316(5829):1341-5. doi: 10.1126/science.1142382. PMID: 17463248. PMCID: PMC3214617.

14. Rottner AK, Ye Y, Navarro-Guerrero E, Rajesh V, Pollner A, Bevacqua RJ, et al. A genome-wide CRISPR screen identifies CALCOCO2 as a regulator of beta cell function influencing type

2 diabetes risk. Nat Genet. 2023;55(1):54-65. doi: 10.1038/s41588-022-01261-2. PMID: 36543916. PMCID: PMC9839450.

15. Kim H, Westerman KE, Smith K, Chiou J, Cole JB, Majarian T, et al. High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. Diabetologia. 2022. doi: 10.1007/s00125-022-05848-6. PMID: 36538063.

16. Marselli L, Piron A, Suleiman M, Colli ML, Yi X, Khamis A, et al. Persistent or Transient Human beta Cell Dysfunction Induced by Metabolic Stress: Specific Signatures and Shared Gene Expression with Type 2 Diabetes. Cell Rep. 2020;33(9):108466. doi: 10.1016/j.celrep.2020.108466. PMID: 33264613.

17. Asplund O, Storm P, Chandra V, Hatem G, Ottosson-Laakso E, Mansour-Aly D, et al. Islet Gene View-a tool to facilitate islet research. Life Sci Alliance. 2022;5(12). doi: 10.26508/lsa.202201376. PMID: 35948367. PMCID: PMC9366203.

18. Morales J, Pujar S, Loveland JE, Astashyn A, Bennett R, Berry A, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022;604(7905):310-5. doi: 10.1038/s41586-022-04558-8. PMID: 35388217. PMCID: PMC9007741 P.F. is a member of the scientific advisory boards of Fabric Genomics, Inc., and Eagle Genomics, Ltd. All other authors declare no competing interests.

19. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment. 2008;2008(10):P10008.

20. Kalwat MA. Islet DEGAS v1. Mendeley Data. 2024. doi: 10.17632/3sdxv5tzbd.1.

21. Johnson TS, Yu CY, Huang Z, Xu S, Wang T, Dong C, et al. Diagnostic Evidence GAuge of Single cells (DEGAS): a flexible deep transfer learning framework for prioritizing cells in relation to disease. Genome Med. 2022;14(1):11. doi: 10.1186/s13073-022-01012-2. PMID: 35105355. PMCID: PMC8808996.

22. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34(3):267-73. doi: 10.1038/ng1180. PMID: 12808457.

23. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. doi: 10.1073/pnas.0506580102. PMID: 16199517. PMCID: PMC1239896.

24. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25. doi: 10.1016/j.cels.2015.12.004. PMID: 26771021. PMCID: PMC4707969.

25. Chen YC, Taylor AJ, Verchere CB. Islet prohormone processing in health and disease. Diabetes Obes Metab. 2018;20 Suppl 2:64-76. doi: 10.1111/dom.13401. PMID: 30230179.

26. Folli F, La Rosa S, Finzi G, Davalli AM, Galli A, Dick EJ, Jr., et al. Pancreatic islet of Langerhans' cytoarchitecture and ultrastructure in normal glucose tolerance and in type 2 diabetes mellitus. Diabetes Obes Metab. 2018;20 Suppl 2:137-44. doi: 10.1111/dom.13380. PMID: 30230173.

27. Lorenzo PI, Fuente-Martin E, Brun T, Cobo-Vuilleumier N, Jimenez-Moreno CM, I GHG, et al. PAX4 Defines an Expandable beta-Cell Subpopulation in the Adult Pancreatic Islet. Sci Rep. 2015;5:15672. doi: 10.1038/srep15672. PMID: 26503027. PMCID: PMC4622080.

28. Collombat P, Mansouri A, Hecksher-Sorensen J, Serup P, Krull J, Gradwohl G, et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev. 2003;17(20):2591-603. doi: 10.1101/gad.269003. PMID: 14561778. PMCID: PMC218152.

29. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010;42(2):105-16. doi: 10.1038/ng.520. PMID: 20081858. PMCID: PMC3018764.

30. Wu CT, Hilgendorf KI, Bevacqua RJ, Hang Y, Demeter J, Kim SK, et al. Discovery of ciliary G protein-coupled receptors regulating pancreatic islet insulin and glucagon secretion. Genes Dev. 2021. doi: 10.1101/gad.348261.121. PMID: 34385262.

31. Docherty FM, Riemondy KA, Castro-Gutierrez R, Dwulet JM, Shilleh AH, Hansen MS, et al. ENTPD3 Marks Mature Stem Cell-Derived beta-Cells Formed by Self-Aggregation In Vitro. Diabetes. 2021;70(11):2554-67. doi: 10.2337/db20-0873. PMID: 34380694. PMCID: PMC8564403.

32. Syed SK, Kauffman AL, Beavers LS, Alston JT, Farb TB, Ficorilli J, et al. Ectonucleotidase NTPDase3 is abundant in pancreatic beta-cells and regulates glucose-induced insulin secretion. Am J Physiol Endocrinol Metab. 2013;305(10):E1319-26. doi: 10.1152/ajpendo.00328.2013. PMID: 24085034.

33. Costanzo MC, von Grotthuss M, Massung J, Jang D, Caulkins L, Koesterer R, et al. The Type 2 Diabetes Knowledge Portal: An open access genetic resource dedicated to type 2 diabetes and related traits. Cell Metab. 2023;35(4):695-710 e6. doi: 10.1016/j.cmet.2023.03.001. PMID: 36963395. PMCID: PMC10231654.

34. Forgetta V, Jiang L, Vulpescu NA, Hogan MS, Chen S, Morris JA, et al. An effector index to predict target genes at GWAS loci. Hum Genet. 2022;141(8):1431-47. doi: 10.1007/s00439-022-02434-z. PMID: 35147782.

35. Weir GC, Gaglia J, Bonner-Weir S. Inadequate β-cell mass is essential for the pathogenesis of type 2 diabetes. The Lancet Diabetes & Endocrinology. 2020;8(3):249-56. doi: 10.1016/s2213-8587(20)30022-x.

36. Stampone E, Caldarelli I, Zullo A, Bencivenga D, Mancini FP, Della Ragione F, et al. Genetic and Epigenetic Control of CDKN1C Expression: Importance in Cell Commitment and Differentiation, Tissue Homeostasis and Human Diseases. Int J Mol Sci. 2018;19(4). doi: 10.3390/ijms19041055. PMID: 29614816. PMCID: PMC5979523.

37. Ou K, Yu M, Moss NG, Wang YJ, Wang AW, Nguyen SC, et al. Targeted demethylation at the CDKN1C/p57 locus induces human beta cell replication. J Clin Invest. 2019;129(1):209-14. doi: 10.1172/JCI99170. PMID: 30352048. PMCID: PMC6307972.

38. Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. Nat Genet. 2010;42(1):68-71. doi: 10.1038/ng.493. PMID: 19966805. PMCID: PMC2820243.

39. Kameswaran V, Golson ML, Ramos-Rodriguez M, Ou K, Wang YJ, Zhang J, et al. The Dysregulation of the DLK1-MEG3 Locus in Islets From Patients With Type 2 Diabetes Is Mimicked by Targeted Epimutation of Its Promoter With TALE-DNMT Constructs. Diabetes. 2018;67(9):1807-15. doi: 10.2337/db17-0682. PMID: 30084829. PMCID: PMC6110314.

40. Avrahami D, Li C, Yu M, Jiao Y, Zhang J, Naji A, et al. Targeting the cell cycle inhibitor p57Kip2 promotes adult human beta cell replication. J Clin Invest. 2014;124(2):670-4. doi: 10.1172/JCI69519. PMID: 24430183. PMCID: PMC3904605.

41. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun. 2019;10(1):380. doi: 10.1038/s41467-018-08023-x. PMID: 30670690. PMCID: PMC6342984.

42. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773-82. doi: 10.1038/s41587-019-0114-2. PMID: 31061481. PMCID: PMC6610714.

43. Tsoucas D, Dong R, Chen H, Zhu Q, Guo G, Yuan GC. Accurate estimation of cell-type composition from gene expression data. Nat Commun. 2019;10(1):2975. doi: 10.1038/s41467-019-10802-z. PMID: 31278265. PMCID: PMC6611906.

44. Herman JS, Sagar, Grun D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat Methods. 2018;15(5):379-86. doi: 10.1038/nmeth.4662. PMID: 29630061.

45. Zhang Q, Jin S, Zou X. scAB detects multiresolution cell states with clinical significance by integrating single-cell genomics and bulk sequencing data. Nucleic Acids Res. 2022;50(21):12112-30. doi: 10.1093/nar/gkac1109. PMID: 36440766. PMCID: PMC9757078.

46. Sun D, Guan X, Moran AE, Wu LY, Qian DZ, Schedin P, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. Nat Biotechnol. 2022;40(4):527-38. doi: 10.1038/s41587-021-01091-3. PMID: 34764492. PMCID: PMC9010342.

47. Chen J, Wang X, Ma A, Wang QE, Liu B, Li L, et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. Nat Commun. 2022;13(1):6494. doi: 10.1038/s41467-022-34277-7. PMID: 36310235. PMCID: PMC9618578.

48. Broadaway KA, Yin X, Williamson A, Parsons VA, Wilson EP, Moxley AH, et al. Loci for insulin processing and secretion provide insight into type 2 diabetes risk. Am J Hum Genet. 2023. doi: 10.1016/j.ajhg.2023.01.002. PMID: 36693378.

49. Bevacqua RJ, Lam JY, Peiris H, Whitener RL, Kim S, Gu X, et al. SIX2 and SIX3 coordinately regulate functional maturity and fate of human pancreatic beta cells. Genes Dev. 2021;35(3-4):234-49. doi: 10.1101/gad.342378.120. PMID: 33446570. PMCID: PMC7849364.

50. Weng C, Gu A, Zhang S, Lu L, Ke L, Gao P, et al. Single cell multiomic analysis reveals diabetes-associated beta-cell heterogeneity driven by HNF1A. Nat Commun. 2023;14(1):5400. doi: 10.1038/s41467-023-41228-3. PMID: 37669939.

51. Tsuyama T, Sato Y, Yoshizawa T, Matsuoka T, Yamagata K. Hypoxia causes pancreatic beta-cell dysfunction and impairs insulin secretion by activating the transcriptional repressor BHLHE40. EMBO Rep. 2023:e56227. doi: 10.15252/embr.202256227. PMID: 37341148.

52. Kerns SL, Guevara-Aguirre J, Andrew S, Geng J, Guevara C, Guevara-Aguirre M, et al. A novel variant in CDKN1C is associated with intrauterine growth restriction, short stature, and early-adulthood-onset diabetes. J Clin Endocrinol Metab. 2014;99(10):E2117-22. doi: 10.1210/jc.2014-1949. PMID: 25057881. PMCID: PMC4184067.

53. Brioude F, Netchine I, Praz F, Le Jule M, Calmel C, Lacombe D, et al. Mutations of the Imprinted CDKN1C Gene as a Cause of the Overgrowth Beckwith-Wiedemann Syndrome: Clinical Spectrum and Functional Characterization. Hum Mutat. 2015;36(9):894-902. doi: 10.1002/humu.22824. PMID: 26077438.

54. Augsornworawat P, Maxwell KG, Velazco-Cruz L, Millman JR. Single-Cell Transcriptome Profiling Reveals beta Cell Maturation in Stem Cell-Derived Islets after Transplantation. Cell Rep. 2020;32(8):108067. doi: 10.1016/j.celrep.2020.108067. PMID: 32846125.

55. Bulfoni M, Bouyioukos C, Zakaria A, Nigon F, Rapone R, Del Maestro L, et al. Glucose controls co-translation of structurally related mRNAs via the mTOR and eIF2 pathways in human pancreatic beta cells. Front Endocrinol (Lausanne). 2022;13:949097. doi: 10.3389/fendo.2022.949097. PMID: 35992129. PMCID: PMC9388909.

56. Stierman B, Afful J, Carroll M, Chen T-C, Davy O, Steven F, et al. National Health and Nutrition Examination Survey 2017–March 2020 prepandemic data files—Development of files

and prevalence estimates for selected health outcomes. National Health Statistics Reports. National Center for Health Statistics. 2021;no. 158. doi: 10.15620/cdc:106273.

57. Centers for Disease Control and Prevention. National Diabetes Statistics Report website. [cited 12/26/2022]; Available from: https://www.cdc.gov/diabetes/data/statistics-report/index.html.

58. Smith GI, Mittendorfer B, Klein S. Metabolically healthy obesity: facts and fantasies. J Clin Invest. 2019;129(10):3978-89. doi: 10.1172/JCI129186. PMID: 31524630. PMCID: PMC6763224 funding from Merck Sharp & Dohme Corp. and Janssen Pharmaceuticals, and has served as a consultant for Pfizer, Novo Nordisk, and Merck Sharp & Dohme Corp.

59. Dominguez-Gutierrez G, Xin Y, Gromada J. Heterogeneity of human pancreatic beta-cells. Mol Metab. 2019;27S(Suppl):S7-S14. doi: 10.1016/j.molmet.2019.06.015. PMID: 31500834. PMCID: PMC6768494.

60. Muraro MJ, Dharmadhikari G, Grun D, Groen N, Dielen T, Jansen E, et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. 2016;3(4):385-94 e3. doi: 10.1016/j.cels.2016.09.002. PMID: 27693023. PMCID: PMC5092539.

61. Bu G, Geuze HJ, Strous GJ, Schwartz AL. 39 kDa receptor-associated protein is an ER resident protein and molecular chaperone for LDL receptor-related protein. EMBO J. 1995;14(10):2269-80. doi: 10.1002/j.1460-2075.1995.tb07221.x. PMID: 7774585. PMCID: PMC398334.

62. Uhlen M, Karlsson MJ, Hober A, Svensson AS, Scheffel J, Kotol D, et al. The human secretome. Sci Signal. 2019;12(609). doi: 10.1126/scisignal.aaz0274. PMID: 31772123.

63. Atanes P, Ruz-Maldonado I, Hawkes R, Liu B, Zhao M, Huang GC, et al. Defining G protein-coupled receptor peptide ligand expressomes and signalomes in human and mouse islets. Cell

Mol Life Sci. 2018;75(16):3039-50. doi: 10.1007/s00018-018-2778-z. PMID: 29455414. PMCID: PMC6061145.

64. Gupta R, Nguyen DC, Schaid MD, Lei X, Balamurugan AN, Wong GW, et al. Complement 1q like-3 protein inhibits insulin secretion from pancreatic beta-cells via the cell adhesion G protein-coupled receptor BAI3. J Biol Chem. 2018. doi: 10.1074/jbc.RA118.005403. PMID: 30228187.

65. Dorrell C, Schug J, Canaday PS, Russ HA, Tarlow BD, Grompe MT, et al. Human islets contain four distinct subtypes of beta cells. Nat Commun. 2016;7:11756. doi: 10.1038/ncomms11756. PMID: 27399229. PMCID: PMC4942571 (HPi2/HIC1-2B4, HPx1/HIC0-3B3, HPd3/DHIC5-4D9); C.D., P.R.S and M.G. are inventors of these antibodies. This potential conflict of interest has been reviewed and managed by OHSU. The remaining authors declare no competing financial interest.

66. Bader E, Migliorini A, Gegg M, Moruzzi N, Gerdes J, Roscioni SS, et al. Identification of proliferative and mature beta-cells in the islets of Langerhans. Nature. 2016;535(7612):430-4. doi: 10.1038/nature18624. PMID: 27398620.

67. van der Meulen T, Xie R, Kelly OG, Vale WW, Sander M, Huising MO. Urocortin 3 marks mature human primary and embryonic stem cell-derived pancreatic alpha and beta cells. PLoS One. 2012;7(12):e52181. doi: 10.1371/journal.pone.0052181. PMID: 23251699. PMCID: PMC3522648.

68. Camunas-Soler J, Dai XQ, Hang Y, Bautista A, Lyon J, Suzuki K, et al. Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. Cell Metab. 2020;31(5):1017-31 e4. doi: 10.1016/j.cmet.2020.04.005. PMID: 32302527. PMCID: PMC7398125.

69. Kravets V, Dwulet JM, Schleicher WE, Hodson DJ, Davis AM, Pyle L, et al. Functional architecture of the pancreatic islets: first responder cells drive the first- phase [Ca2+] response. bioRxiv. 2021. doi: 10.1101/2020.12.22.424082.

70. Salem V, Silva LD, Suba K, Georgiadou E, Neda Mousavy Gharavy S, Akhtar N, et al. Leader beta-cells coordinate Ca(2+) dynamics across pancreatic islets in vivo. Nat Metab. 2019;1(6):615-29. doi: 10.1038/s42255-019-0075-2. PMID: 32694805.

71. Johnston NR, Mitchell RK, Haythorne E, Pessoa MP, Semplici F, Ferrer J, et al. Beta Cell Hubs Dictate Pancreatic Islet Responses to Glucose. Cell Metab. 2016;24(3):389-401. doi: 10.1016/j.cmet.2016.06.020. PMID: 27452146. PMCID: PMC5031557.

72. Weidemann BJ, Marcheva B, Kobayashi M, Omura C, Newman MV, Kobayashi Y, et al. Repression of latent NF-kappaB enhancers by PDX1 regulates beta cell functional heterogeneity. Cell Metab. 2024;36(1):90-102 e7. doi: 10.1016/j.cmet.2023.11.018. PMID: 38171340.

73. Morris SA. The evolving concept of cell identity in the single cell era. Development. 2019;146(12). doi: 10.1242/dev.169748. PMID: 31249002.

74. Cell Systems V. What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism? Cell Syst. 2017;4(3):255-9. doi: 10.1016/j.cels.2017.03.006. PMID: 28334573.

75. Yang T, Yan C, Yang L, Tan J, Jiang S, Hu J, et al. Identification and validation of core genes for type 2 diabetes mellitus by integrated analysis of single-cell and bulk RNA-sequencing. Eur J Med Res. 2023;28(1):340. doi: 10.1186/s40001-023-01321-1. PMID: 37700362. PMCID: PMC10498638.

76. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, et al. RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. Cell Metab. 2016;24(4):608-15. doi: 10.1016/j.cmet.2016.08.018. PMID: 27667665.

77. Kobayashi S, Wagatsuma H, Ono R, Ichikawa H, Yamazaki M, Tashiro H, et al. Mouse Peg9/Dlk1 and human PEG9/DLK1 are paternally expressed imprinted genes closely located to the maternally expressed imprinted genes: mouse Meg3/Gtl2 and human MEG3. Genes Cells. 2000;5(12):1029-37. doi: 10.1046/j.1365-2443.2000.00390.x. PMID: 11168589.

78. Huang D, Han Y, Tang T, Yang L, Jiang P, Qian W, et al. Dlk1 maintains adult mice long-term HSCs by activating Notch signaling to restrict mitochondrial metabolism. Exp Hematol Oncol. 2023;12(1):11. doi: 10.1186/s40164-022-00369-9. PMID: 36653853. PMCID: PMC9850540.

79. Grassi ES, Pietras A. Emerging Roles of DLK1 in the Stem Cell Niche and Cancer Stemness. J Histochem Cytochem. 2022;70(1):17-28. doi: 10.1369/00221554211048951. PMID: 34606325. PMCID: PMC8721543.

80. Nueda ML, Naranjo AI, Baladron V, Laborda J. The proteins DLK1 and DLK2 modulate NOTCH1-dependent proliferation and oncogenic potential of human SK-MEL-2 melanoma cells. Biochim Biophys Acta. 2014;1843(11):2674-84. doi: 10.1016/j.bbamcr.2014.07.015. PMID: 25093684.

81. Abdallah BM, Ditzel N, Laborda J, Karsenty G, Kassem M. DLK1 Regulates Whole-Body Glucose Metabolism: A Negative Feedback Regulation of the Osteocalcin-Insulin Loop. Diabetes. 2015;64(9):3069-80. doi: 10.2337/db14-1642. PMID: 25918236.

82. Zhao Z, D'Oliveira Albanus R, Taylor H, Tang X, Han Y, Orchard P, et al. An integrative single-cell multi-omics profiling of human pancreatic islets identifies T1D associated genes and regulatory signals. Res Sq. 2023. doi: 10.21203/rs.3.rs-3343318/v1. PMID: 37886586. PMCID: PMC10602166.

83. Kassem SA, Ariel I, Thornton PS, Hussain K, Smith V, Lindley KJ, et al. p57(KIP2) expression in normal islet cells and in hyperinsulinism of infancy. Diabetes. 2001;50(12):2763-9. doi: 10.2337/diabetes.50.12.2763. PMID: 11723059.

84. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96. doi: 10.1038/s41592-019-0619-0. PMID: 31740819. PMCID: PMC6884693.

85. Wang T, Johnson TS, Shao W, Lu Z, Helm BR, Zhang J, et al. BERMUDA: a novel deep transfer learning method for single-cell RNA sequencing batch correction reveals hidden high-resolution cellular subtypes. Genome Biol. 2019;20(1):165. doi: 10.1186/s13059-019-1764-6. PMID: 31405383. PMCID: PMC6691531.

86. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol. 2019;37(6):685-91. doi: 10.1038/s41587-019-0113-3. PMID: 31061482. PMCID: PMC6551256.

87. Hart PA, Kudva YC, Yadav D, Andersen DK, Li Y, Toledo FGS, et al. A Reduced Pancreatic Polypeptide Response is Associated With New-onset Pancreatogenic Diabetes Versus Type 2 Diabetes. J Clin Endocrinol Metab. 2023;108(5):e120-e8. doi: 10.1210/clinem/dgac670. PMID: 36404274. PMCID: PMC10306084.

88. Alsaleh L, Li C, Couetil JL, Ye Z, Huang K, Zhang J, et al. Spatial Transcriptomic Analysis Reveals Associations between Genes and Cellular Topology in Breast and Prostate Cancers. Cancers (Basel). 2022;14(19). doi: 10.3390/cancers14194856. PMID: 36230778. PMCID: PMC9562681.

89. Couetil JL, Liu Z, Alomari AK, Zhang J, Huang K, Johnson TS. Diagnostic Evidence Gauge of Spatial Transcriptomics (DEGAS): Using transfer learning to map clinical data to spatial transcriptomics in prostate cancer. bioRxiv. 2023. doi: 10.1101/2023.04.21.537852.

90. Chatterjee D, Couetil JL, Liu Z, Dong T, Huang K, Chen C, et al. Identification of High-Risk Cells in Single-Cell Spatially Resolved Transcriptomics Data Using Deep Transfer Learning. bioRxiv. 2025. doi: 10.1101/2025.01.30.635803.

Figure 1

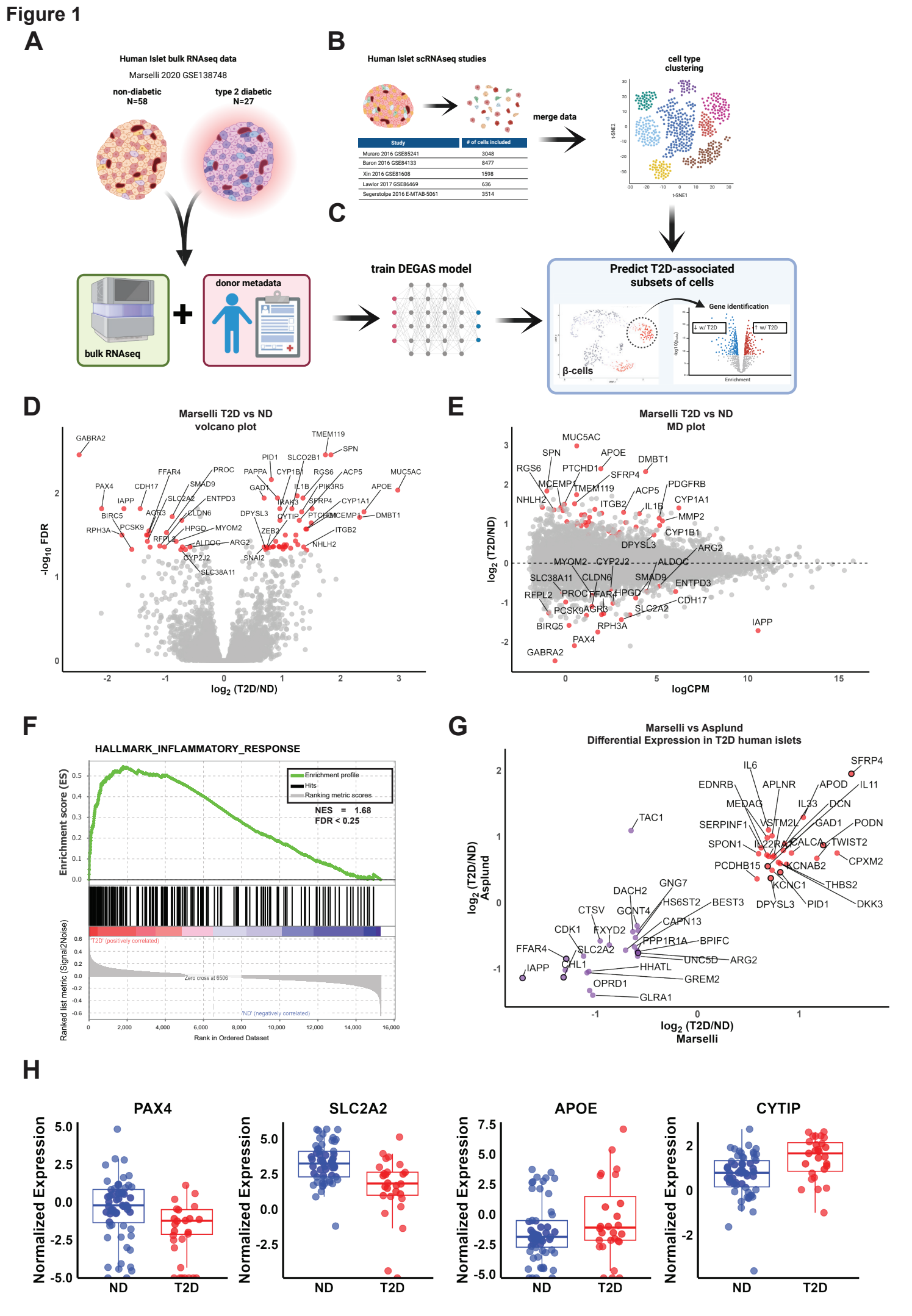

A
Human Islet bulk RNAseq data
Marselli 2020 GSE138748
non-diabetic N=58
type 2 diabetic N=27
bulk RNAseq
donor metadata
B
Human Islet scRNAseq studies
Study | # of cells included
Muraro 2016 GSE85241 | 3048
Baron 2016 GSE84133 | 8477
Xin 2016 GSE81608 | 1598
Lawlor 2017 GSE86469 | 636
Segerstolpe 2016 E-MTAB-5061 | 3514
merge data
cell type clustering
t-SNE1
t-SNE2
C
train DEGAS model
Predict T2D-associated subsets of cells
β-cells
Gene identification
↓ w/ T2D
↑ w/ T2D
Enrichment
D
Marselli T2D vs ND volcano plot
-log10 FDR
log2 (T2D/ND)
E
Marselli T2D vs ND MD plot
log2 (T2D/ND)
logCPM
F
HALLMARK_INFLAMMATORY_RESPONSE
Enrichment score (ES)
Enrichment profile
Hits
Ranking metric scores
NES = 1.68
FDR < 0.25
'T2D' (positively correlated)
Zero cross at 6506
'ND' (negatively correlated)
Ranked list metric (Signal2Noise)
Rank in Ordered Dataset
G
Marselli vs Asplund Differential Expression in T2D human islets
log2 (T2D/ND) Asplund
log2 (T2D/ND) Marselli
H
PAX4
SLC2A2
APOE
CYTIP
Normalized Expression
ND
T2D

Figure 2

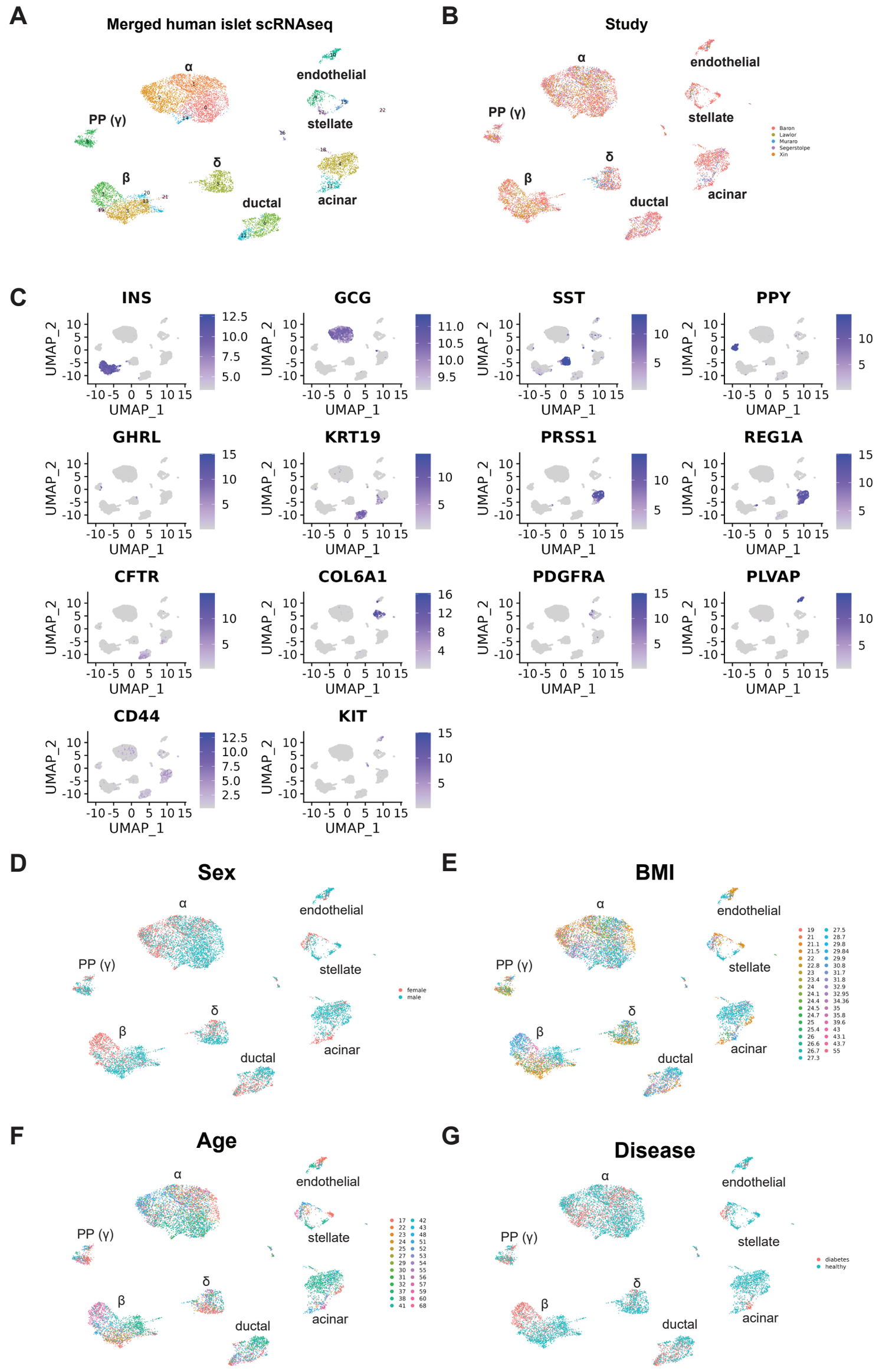
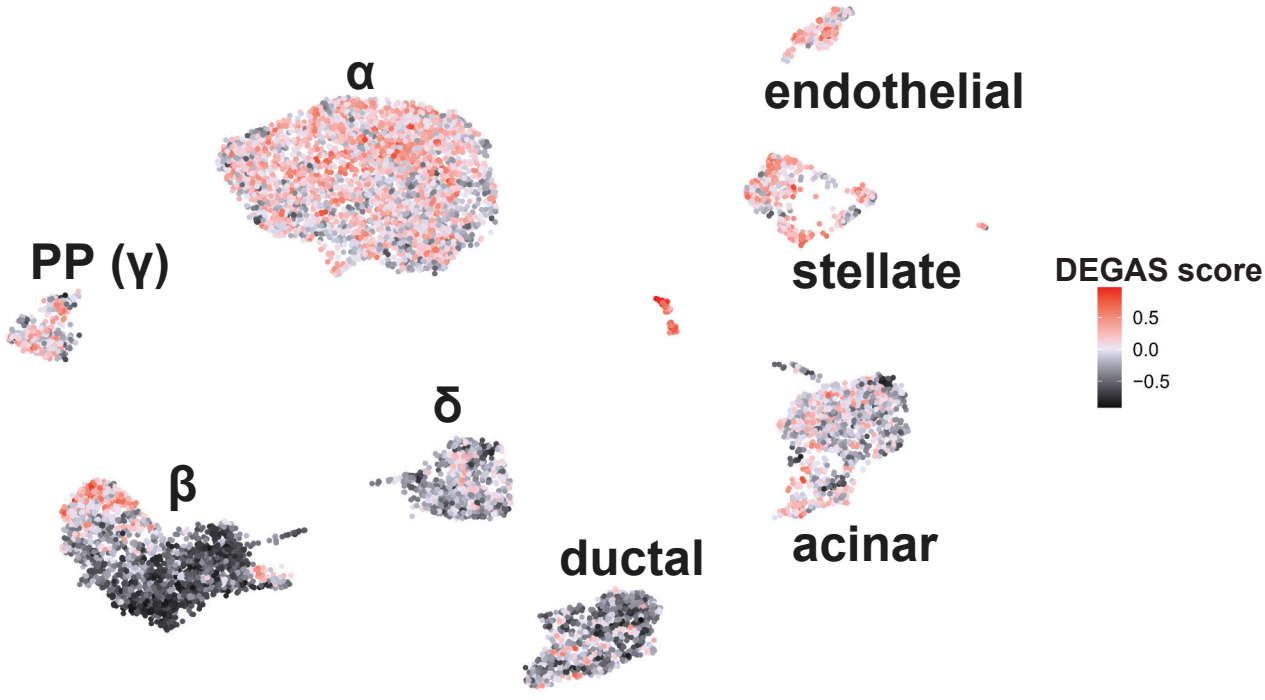

A
Merged human islet scRNAseq
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
B
Study
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
Baron
Lawlor
Muraro
Segerstolpe
Xin
C
INS
GCG
SST
PPY
GHRL
KRT19
PRSS1
REG1A
CFTR
COL6A1
PDGFRA
PLVAP
CD44
KIT
UMAP_1
UMAP_2
D
Sex
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
female
male
E
BMI
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
F
Age
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
G
Disease
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
diabetes
healthy

Figure 3

A


DEGAS:
T2D-association score
α
endothelial
PP (γ)
stellate
DEGAS score
0.5
0.0
−0.5
δ
β
ductal
acinar

B

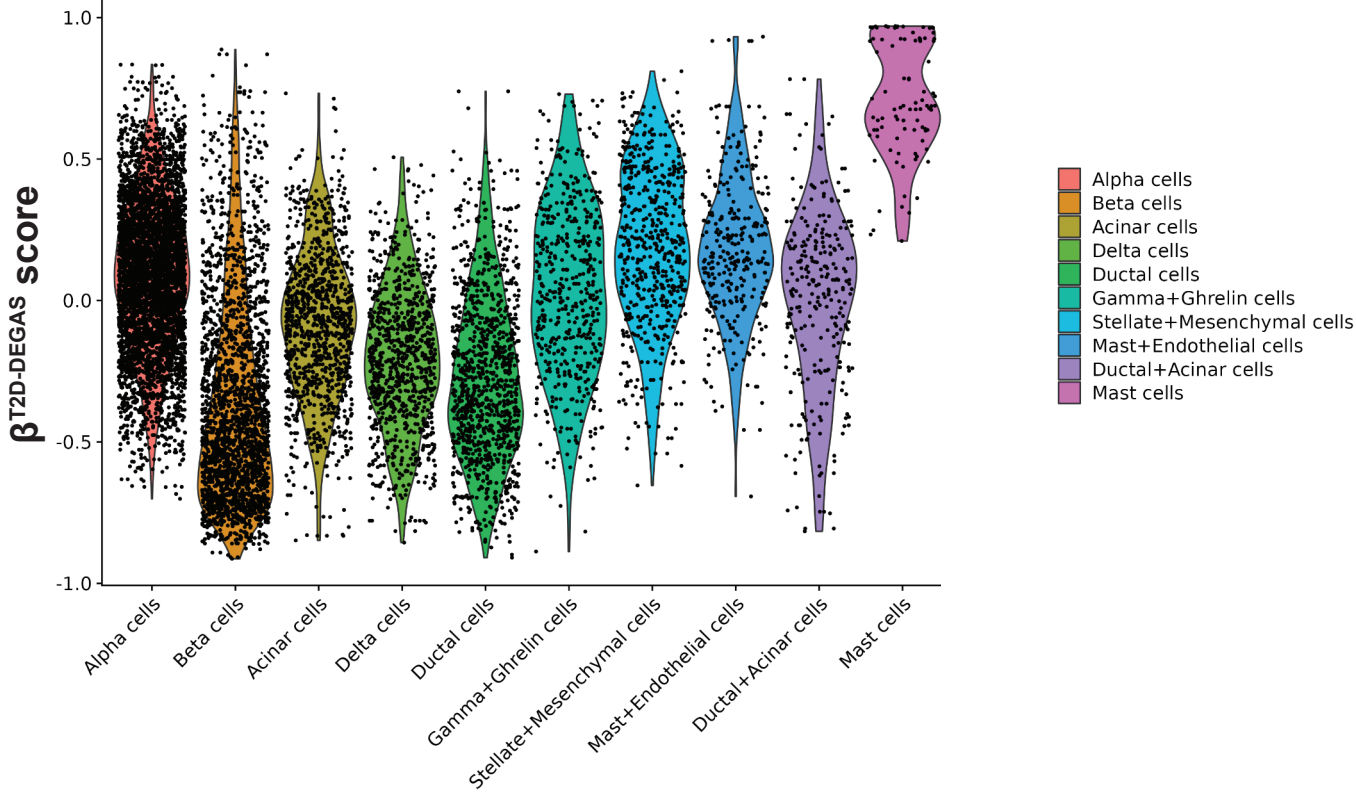

β T2D-DEGAS score
1.0
0.5
0.0
-0.5
-1.0
Alpha cells
Beta cells
Acinar cells
Delta cells
Ductal cells
Gamma+Ghrelin cells
Stellate+Mesenchymal cells
Mast+Endothelial cells
Ductal+Acinar cells
Mast cells

Figure 4

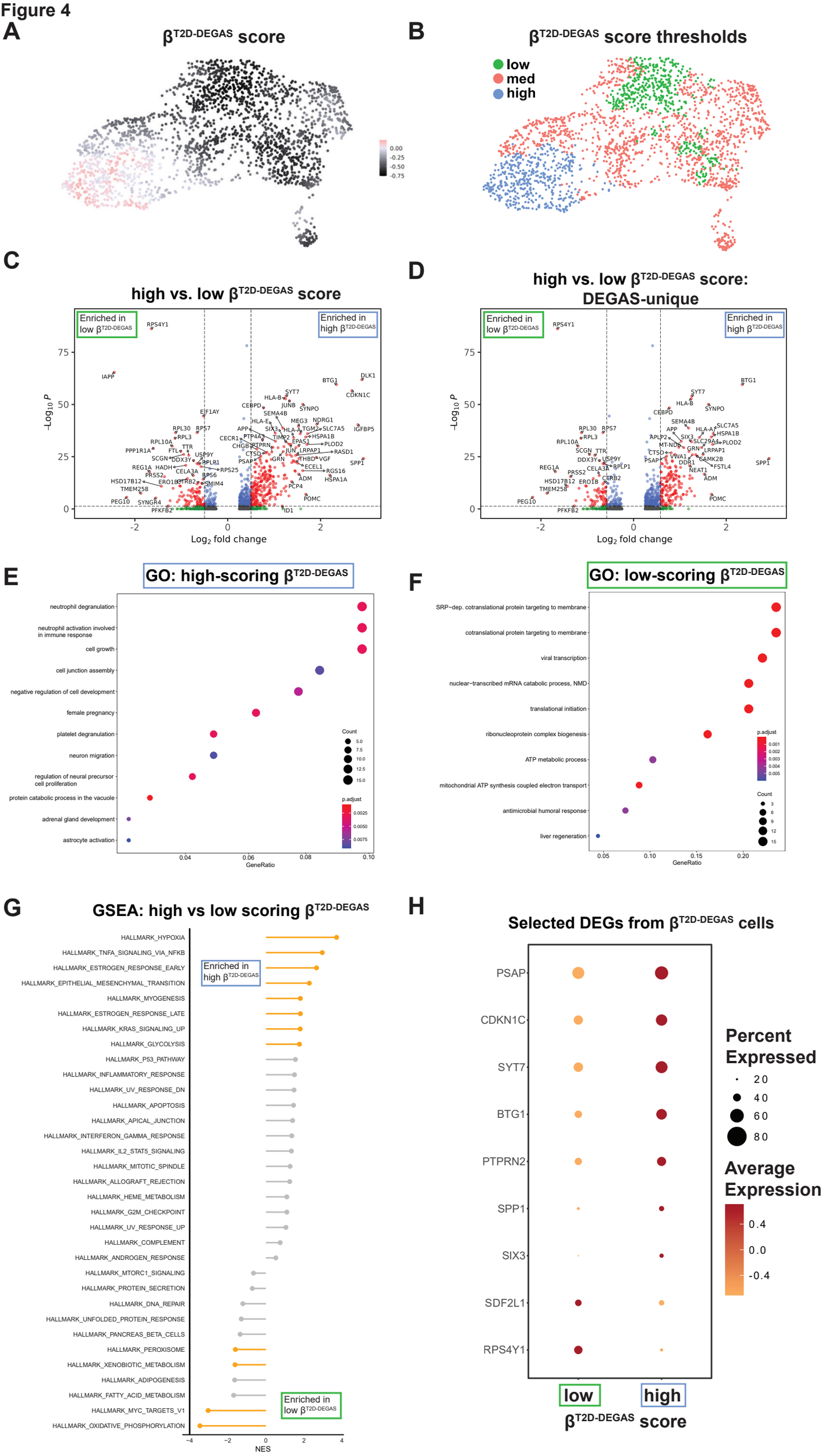
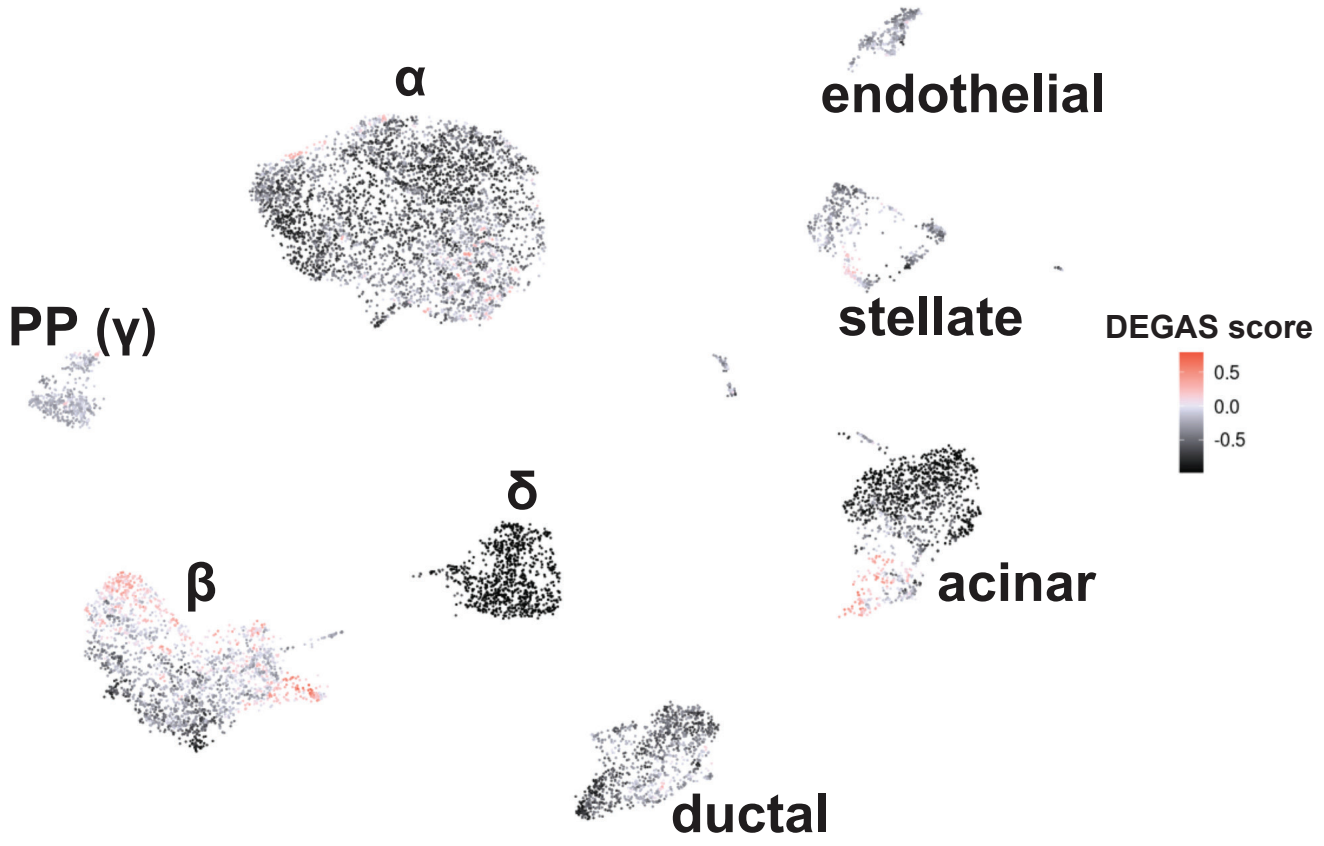

A
βT2D-DEGAS score
0.00
-0.25
-0.50
-0.75
B
βT2D-DEGAS score thresholds
low
med
high
C
high vs. low βT2D-DEGAS score
Enriched in low βT2D-DEGAS
Enriched in high βT2D-DEGAS
−Log10 P
Log2 fold change
D
high vs. low βT2D-DEGAS score: DEGAS-unique
Enriched in low βT2D-DEGAS
Enriched in high βT2D-DEGAS
−Log10 P
Log2 fold change
E
GO: high-scoring βT2D-DEGAS
neutrophil degranulation
neutrophil activation involved in immune response
cell growth
cell junction assembly
negative regulation of cell development
female pregnancy
platelet degranulation
neuron migration
regulation of neural precursor cell proliferation
protein catabolic process in the vacuole
adrenal gland development
astrocyte activation
GeneRatio
Count
p.adjust
F
GO: low-scoring βT2D-DEGAS
SRP−dep. cotranslational protein targeting to membrane
cotranslational protein targeting to membrane
viral transcription
nuclear−transcribed mRNA catabolic process, NMD
translational initiation
ribonucleoprotein complex biogenesis
ATP metabolic process
mitochondrial ATP synthesis coupled electron transport
antimicrobial humoral response
liver regeneration
GeneRatio
p.adjust
Count
G
GSEA: high vs low scoring βT2D-DEGAS
Enriched in high βT2D-DEGAS
Enriched in low βT2D-DEGAS
HALLMARK_HYPOXIA
HALLMARK_TNFA_SIGNALING_VIA_NFKB
HALLMARK_ESTROGEN_RESPONSE_EARLY
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_MYOGENESIS
HALLMARK_ESTROGEN_RESPONSE_LATE
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_GLYCOLYSIS
HALLMARK_P53_PATHWAY
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_UV_RESPONSE_DN
HALLMARK_APOPTOSIS
HALLMARK_APICAL_JUNCTION
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_IL2_STAT5_SIGNALING
HALLMARK_MITOTIC_SPINDLE
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_HEME_METABOLISM
HALLMARK_G2M_CHECKPOINT
HALLMARK_UV_RESPONSE_UP
HALLMARK_COMPLEMENT
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_MTORC1_SIGNALING
HALLMARK_PROTEIN_SECRETION
HALLMARK_DNA_REPAIR
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_PANCREAS_BETA_CELLS
HALLMARK_PEROXISOME
HALLMARK_XENOBIOTIC_METABOLISM
HALLMARK_ADIPOGENESIS
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_MYC_TARGETS_V1
HALLMARK_OXIDATIVE_PHOSPHORYLATION
NES
H
Selected DEGs from βT2D-DEGAS cells
PSAP
CDKN1C
SYT7
BTG1
PTPRN2
SPP1
SIX3
SDF2L1
RPS4Y1
low
high
βT2D-DEGAS score
Percent Expressed
20
40
60
80
Average Expression
0.4
0.0
-0.4

Figure 5

A


DEGAS:
obese-association score
α
endothelial
PP (γ)
stellate
DEGAS score
0.5
0.0
-0.5
δ
β
acinar
ductal

B

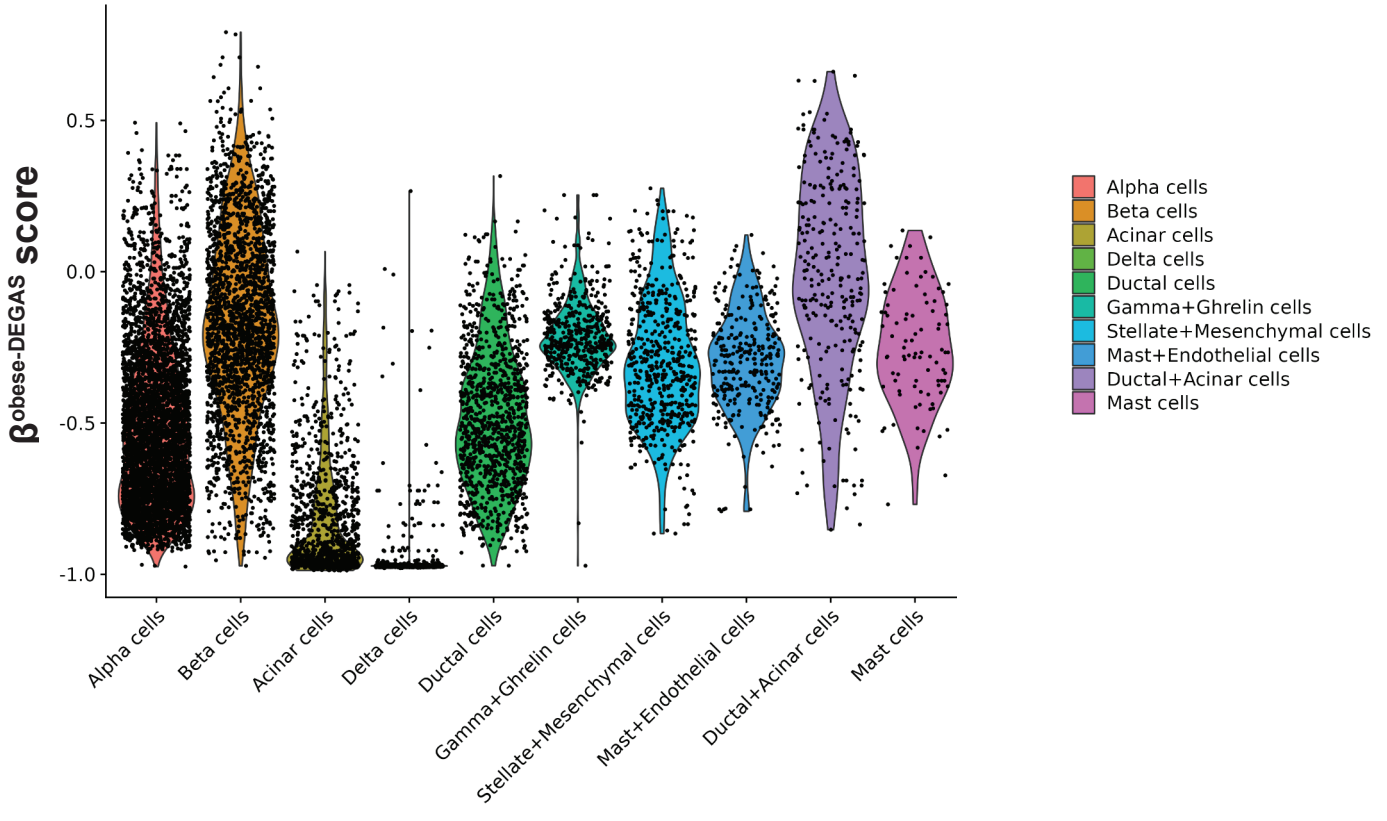

β obese-DEGAS score
0.5
0.0
-0.5
-1.0
Alpha cells
Beta cells
Acinar cells
Delta cells
Ductal cells
Gamma+Ghrelin cells
Stellate+Mesenchymal cells
Mast+Endothelial cells
Ductal+Acinar cells
Mast cells

Figure 6

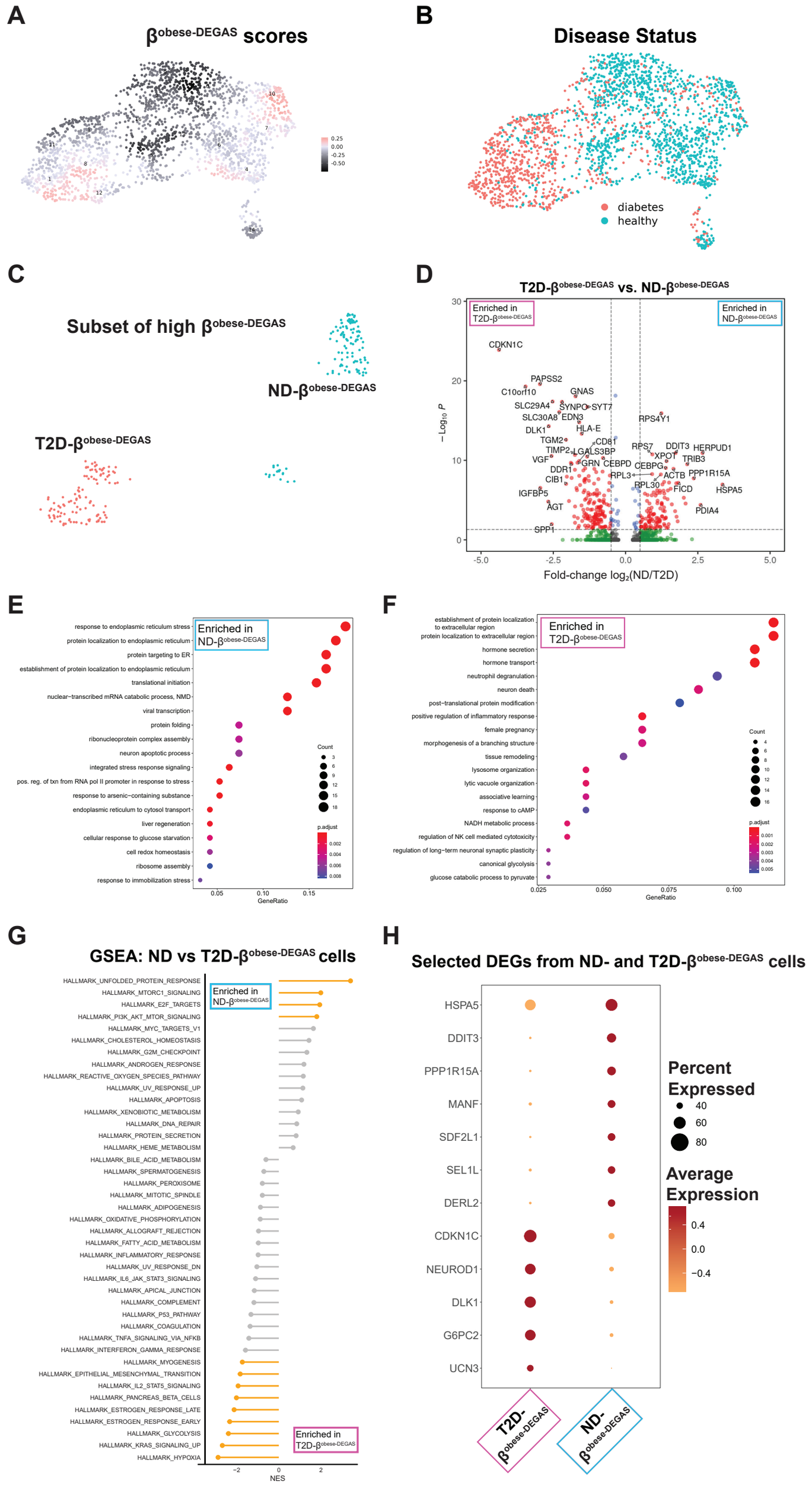

A
βobese-DEGAS scores
B
Disease Status
diabetes
healthy
C
Subset of high βobese-DEGAS
ND-βobese-DEGAS
T2D-βobese-DEGAS
D
T2D-βobese-DEGAS vs. ND-βobese-DEGAS
Enriched in T2D-βobese-DEGAS
Enriched in ND-βobese-DEGAS
CDKN1C
PAPSS2
C10orf10
GNAS
SLC29A4
SYNPO
SYT7
SLC30A8
EDN3
DLK1
HLA-E
TGM2
CD81
TIMP2
LGALS3BP
VGF
GRN
CEBPD
DDR1
CIB1
IGFBP5
AGT
SPP1
RPS4Y1
RPS7
DDIT3
HERPUD1
XPOT
TRIB3
CEBPG
RPL3
ACTB
PPP1R15A
RPL30
FICD
HSPA5
PDIA4
−Log10 P
Fold-change log2(ND/T2D)
E
Enriched in ND-βobese-DEGAS
response to endoplasmic reticulum stress
protein localization to endoplasmic reticulum
protein targeting to ER
establishment of protein localization to endoplasmic reticulum
translational initiation
nuclear−transcribed mRNA catabolic process, NMD
viral transcription
protein folding
ribonucleoprotein complex assembly
neuron apoptotic process
integrated stress response signaling
pos. reg. of txn from RNA pol II promoter in response to stress
response to arsenic−containing substance
endoplasmic reticulum to cytosol transport
liver regeneration
cellular response to glucose starvation
cell redox homeostasis
ribosome assembly
response to immobilization stress
Count
p.adjust
GeneRatio
F
Enriched in T2D-βobese-DEGAS
establishment of protein localization to extracellular region
protein localization to extracellular region
hormone secretion
hormone transport
neutrophil degranulation
neuron death
post−translational protein modification
positive regulation of inflammatory response
female pregnancy
morphogenesis of a branching structure
tissue remodeling
lysosome organization
lytic vacuole organization
associative learning
response to cAMP
NADH metabolic process
regulation of NK cell mediated cytotoxicity
regulation of long−term neuronal synaptic plasticity
canonical glycolysis
glucose catabolic process to pyruvate
Count
p.adjust
GeneRatio
G
GSEA: ND vs T2D-βobese-DEGAS cells
Enriched in ND-βobese-DEGAS
Enriched in T2D-βobese-DEGAS
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_MTORC1_SIGNALING
HALLMARK_E2F_TARGETS
HALLMARK_PI3K_AKT_MTOR_SIGNALING
HALLMARK_MYC_TARGETS_V1
HALLMARK_CHOLESTEROL_HOMEOSTASIS
HALLMARK_G2M_CHECKPOINT
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY
HALLMARK_UV_RESPONSE_UP
HALLMARK_APOPTOSIS
HALLMARK_XENOBIOTIC_METABOLISM
HALLMARK_DNA_REPAIR
HALLMARK_PROTEIN_SECRETION
HALLMARK_HEME_METABOLISM
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_SPERMATOGENESIS
HALLMARK_PEROXISOME
HALLMARK_MITOTIC_SPINDLE
HALLMARK_ADIPOGENESIS
HALLMARK_OXIDATIVE_PHOSPHORYLATION
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_UV_RESPONSE_DN
HALLMARK_IL6_JAK_STAT3_SIGNALING
HALLMARK_APICAL_JUNCTION
HALLMARK_COMPLEMENT
HALLMARK_P53_PATHWAY
HALLMARK_COAGULATION
HALLMARK_TNFA_SIGNALING_VIA_NFKB
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_MYOGENESIS
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_IL2_STAT5_SIGNALING
HALLMARK_PANCREAS_BETA_CELLS
HALLMARK_ESTROGEN_RESPONSE_LATE
HALLMARK_ESTROGEN_RESPONSE_EARLY
HALLMARK_GLYCOLYSIS
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_HYPOXIA
NES
H
Selected DEGs from ND- and T2D-βobese-DEGAS cells
HSPA5
DDIT3
PPP1R15A
MANF
SDF2L1
SEL1L
DERL2
CDKN1C
NEUROD1
DLK1
G6PC2
UCN3
T2D-βobese-DEGAS
ND-βobese-DEGAS
Percent Expressed
40
60
80
Average Expression
0.4
0.0
−0.4

Figure 7

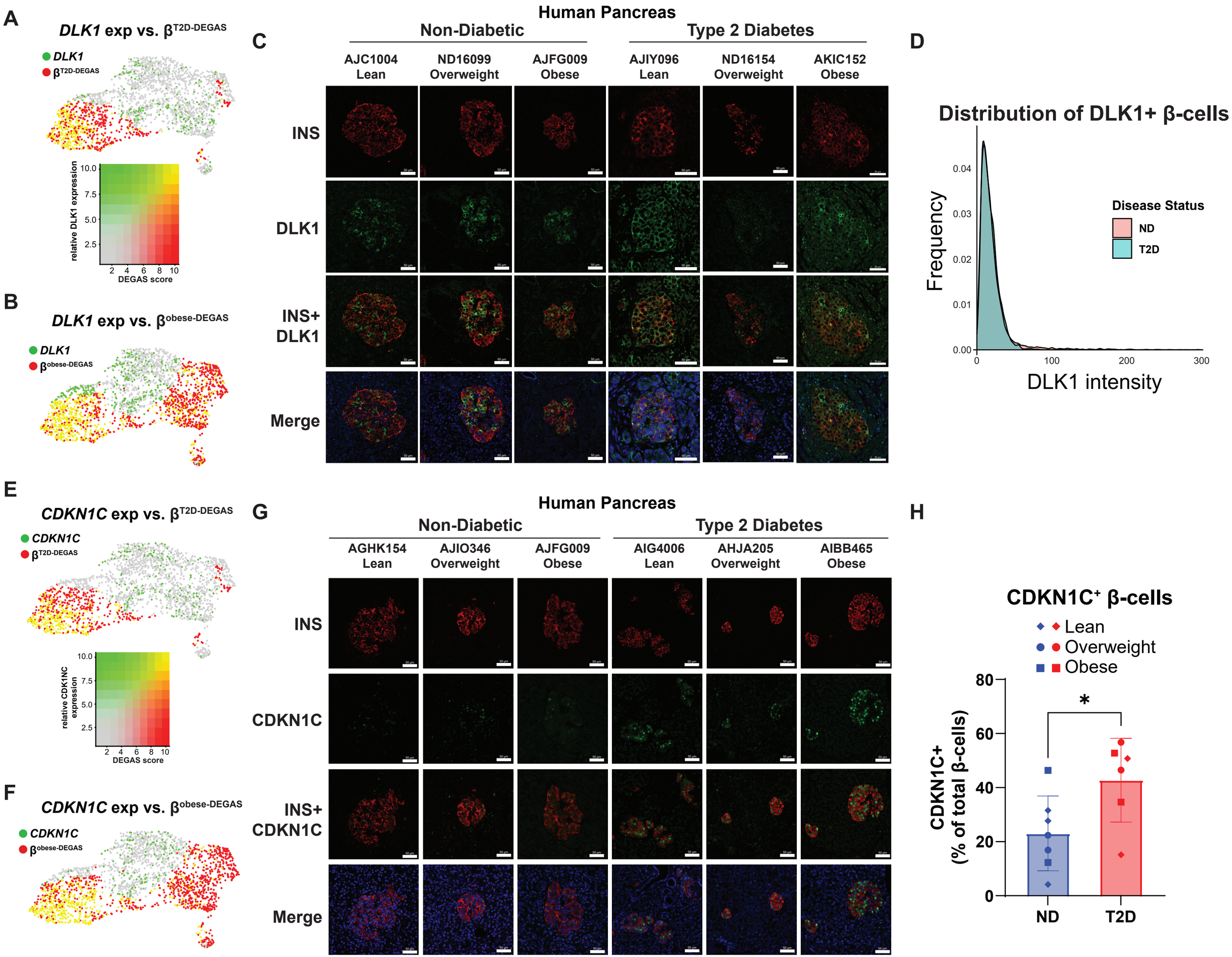

A
DLK1 exp vs. βT2D-DEGAS
DLK1
βT2D-DEGAS
relative DLK1 expression
10.0
7.5
5.0
2.5
2 4 6 8 10
DEGAS score
B
DLK1 exp vs. βobese-DEGAS
DLK1
βobese-DEGAS
C
Human Pancreas
Non-Diabetic
Type 2 Diabetes
AJC1004 Lean
ND16099 Overweight
AJFG009 Obese
AJIY096 Lean
ND16154 Overweight
AKIC152 Obese
INS
DLK1
INS+ DLK1
Merge
D
Distribution of DLK1+ β-cells
Frequency
0.04
0.03
0.02
0.01
0.00
0
100
200
300
DLK1 intensity
Disease Status
ND
T2D
E
CDKN1C exp vs. βT2D-DEGAS
CDKN1C
βT2D-DEGAS
relative CDK1NC expression
10.0
7.5
5.0
2.5
2 4 6 8 10
DEGAS score
F
CDKN1C exp vs. βobese-DEGAS
CDKN1C
βobese-DEGAS
G
Human Pancreas
Non-Diabetic
Type 2 Diabetes
AGHK154 Lean
AJIO346 Overweight
AJFG009 Obese
AIG4006 Lean
AHJA205 Overweight
AIBB465 Obese
INS
CDKN1C
INS+ CDKN1C
Merge
H
CDKN1C+ β-cells
Lean
Overweight
Obese
CDKN1C+ (% of total β-cells)
80
60
40
20
0
*
ND
T2D

Figure S1

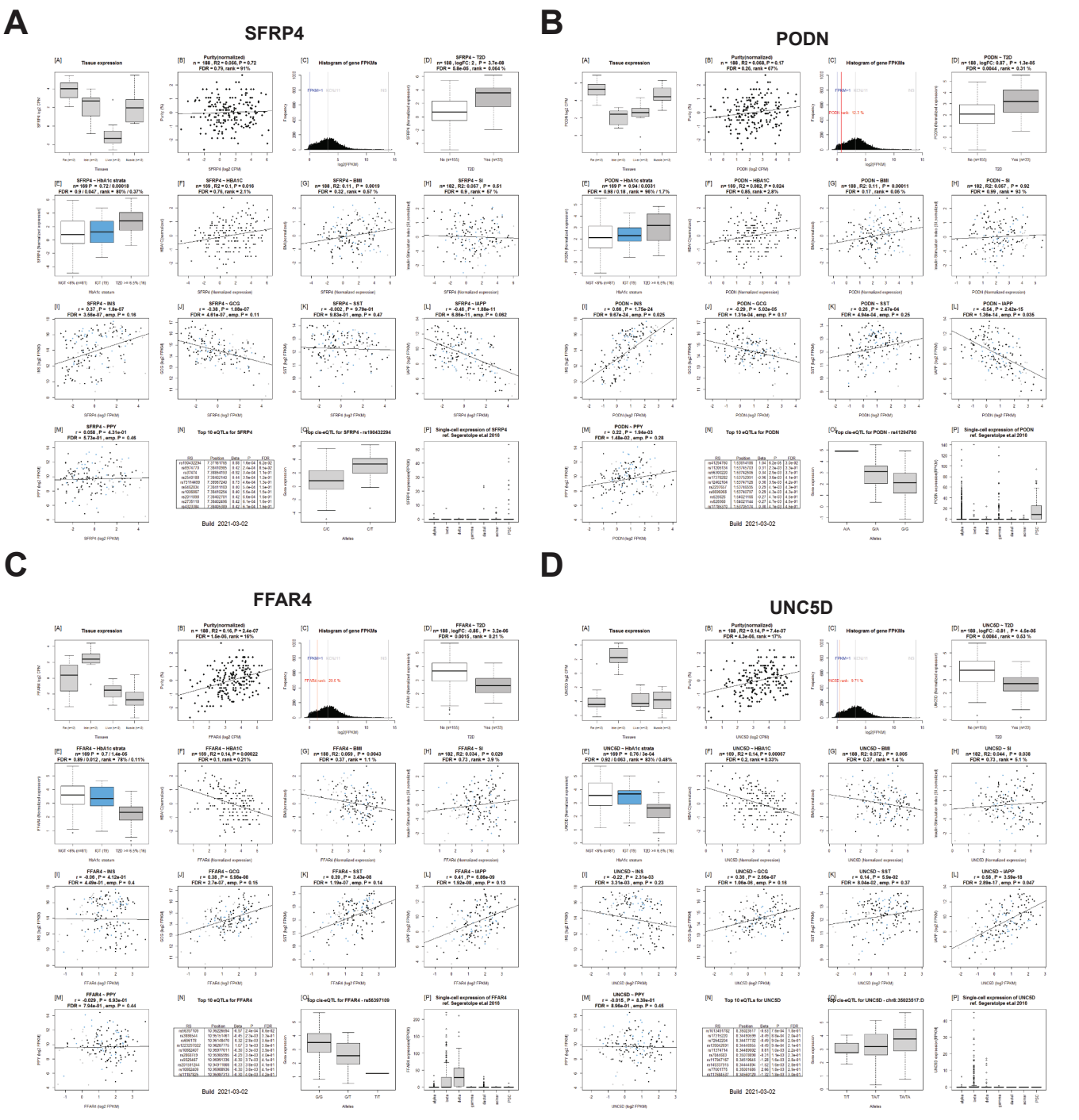
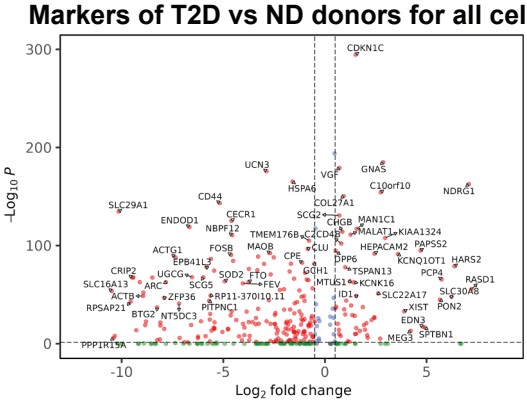


Figure S2

A


Markers of T2D vs ND donors for all cells
−Log10 P
Log2 fold change
300
200
100
0
-10
-5
0
5
CDKN1C
UCN3
VGF
GNAS
HSPA6
C10orf10
NDRG1
CD44
COL27A1
SLC29A1
CFCR1
SCG2
CHGB
MAN1C1
ENDOD1
NBPF12
TMEM176B
C2CD4B
MALAT1
KIAA1324
ACTG1
FOSB
MAOB
CLU
HEPACAM2
PAPSS2
EPB41L3
CPE
DPP6
KCNQ1OT1
HARS2
CRIP2
UGCG
SOD2
FTO
GDH1
TSPAN13
PCP4
SLC16A13
ARC
SCG5
FEV
MTUS1
KCNK16
RASD1
ACTB
ZFP36
RP11-370I10.11
ID1
SLC22A17
SLC30A8
RPSAP21
PITPNC1
XIST
PON2
BTG2
NT5DC3
EDN3
MEG3
SPTBN1
PPP1R15A

B

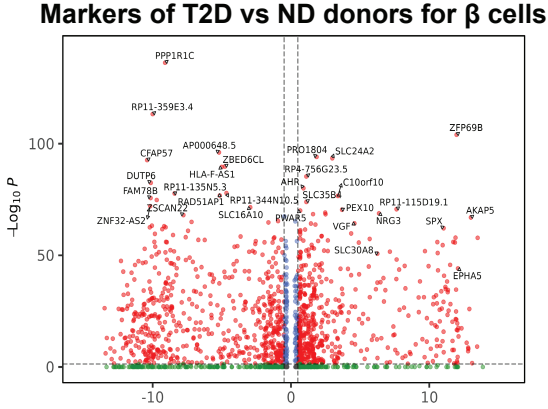
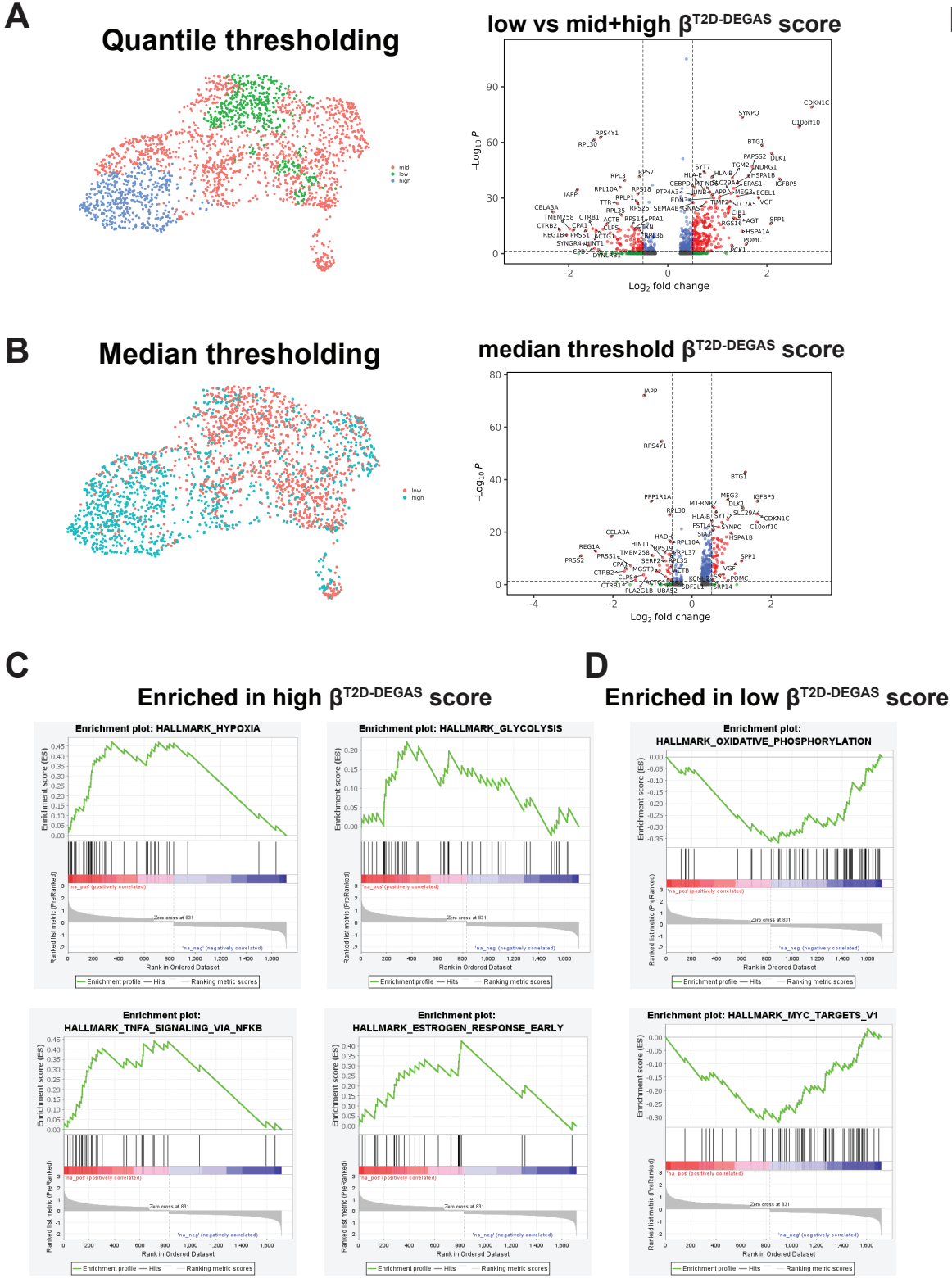

Markers of T2D vs ND donors for β cells
−Log10 P
Log2 fold change
100
50
0
-10
0
10
PPP1R1C
RP11-359E3.4
ZFP69B
AP000648.5
PRO1804
SLC24A2
CFAP57
ZBED6CL
RP4-756G23.5
HLA-F-AS1
DUTP6
AHR
C10orf10
FAM78B
RP11-135N5.3
SLC35B4
RAD51AP1
RP11-344N10.5
RP11-115D19.1
ZSCAN42
SLC16A10
PEX10
AKAP5
ZNF32-AS2
PWAR5
NRG3
SPX
VGF
SLC30A8
EPHA5

Figure S3


A
Quantile thresholding
low vs mid+high β^T2D-DEGAS score
mid
low
high
-Log10 P
Log2 fold change
B
Median thresholding
median threshold β^T2D-DEGAS score
low
high
-Log10 P
Log2 fold change
C
Enriched in high β^T2D-DEGAS score
Enrichment plot: HALLMARK_HYPOXIA
Enrichment plot: HALLMARK_GLYCOLYSIS
Enrichment plot: HALLMARK_TNFA_SIGNALING_VIA_NFKB
Enrichment plot: HALLMARK_ESTROGEN_RESPONSE_EARLY
D
Enriched in low β^T2D-DEGAS score
Enrichment plot: HALLMARK_OXIDATIVE_PHOSPHORYLATION
Enrichment plot: HALLMARK_MYC_TARGETS_V1
Enrichment score (ES)
Ranked list metric (PreRanked)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

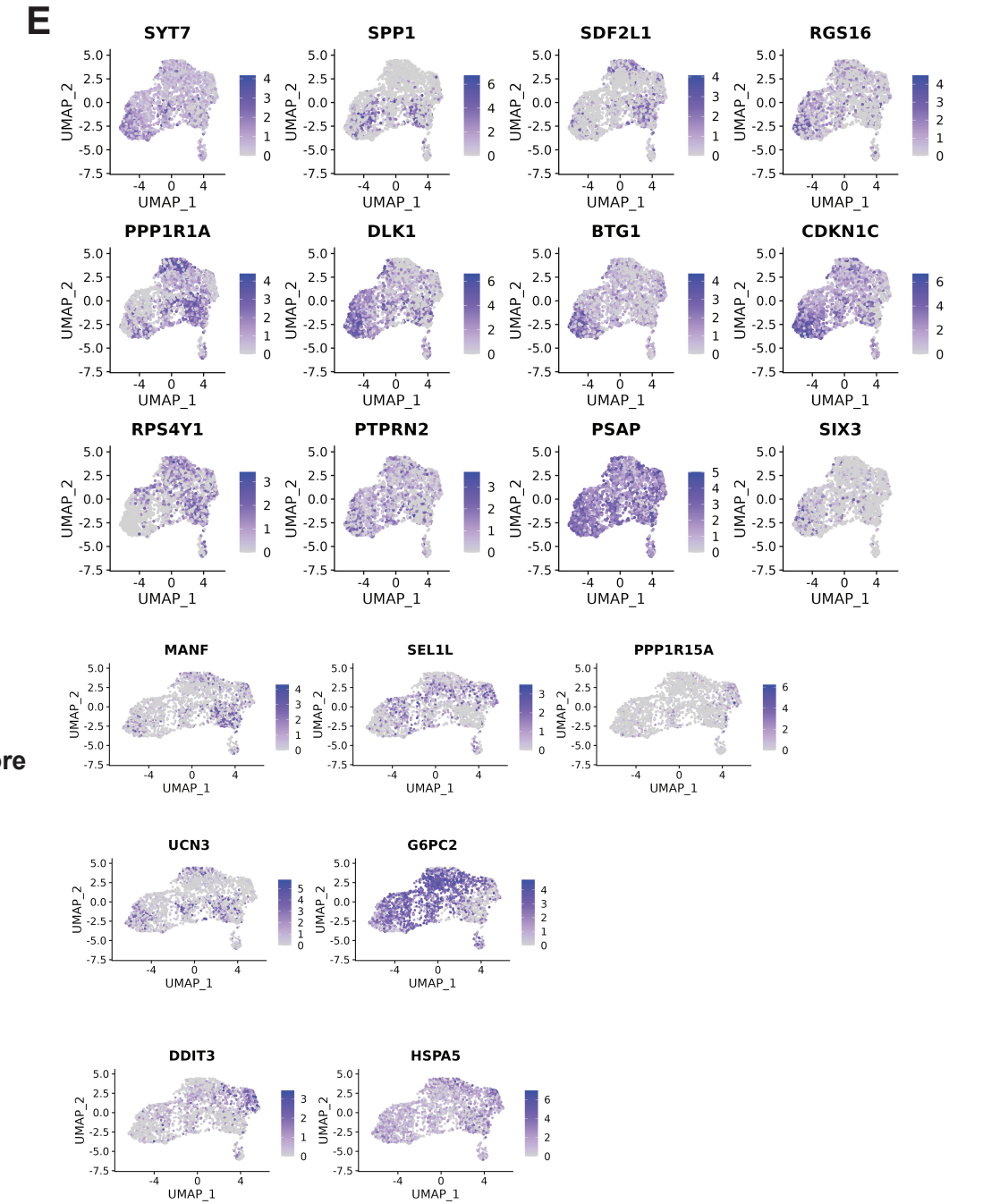

E
SYT7
SPP1
SDF2L1
RGS16
PPP1R1A
DLK1
BTG1
CDKN1C
RPS4Y1
PTPRN2
PSAP
SIX3
MANF
SEL1L
PPP1R15A
UCN3
G6PC2
DDIT3
HSPA5
UMAP_1
UMAP_2

Figure S4

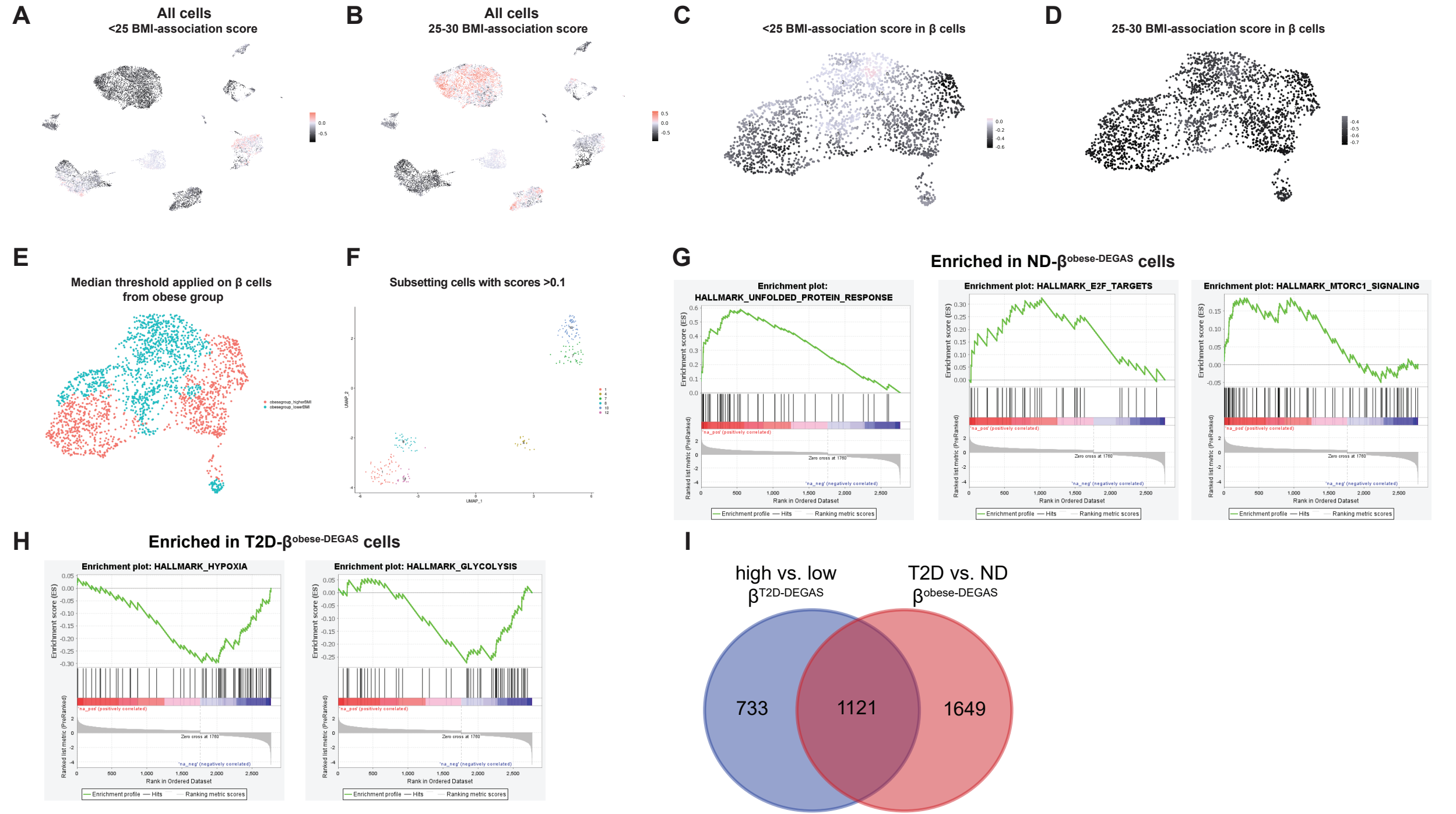

A
All cells
<25 BMI-association score
B
All cells
25-30 BMI-association score
C
<25 BMI-association score in β cells
D
25-30 BMI-association score in β cells
E
Median threshold applied on β cells
from obese group
obesegroup_higherBMI
obesegroup_lowerBMI
F
Subsetting cells with scores >0.1
UMAP_1
UMAP_2
G
Enriched in ND-βobese-DEGAS cells
Enrichment plot:
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
Enrichment plot: HALLMARK_E2F_TARGETS
Enrichment plot: HALLMARK_MTORC1_SIGNALING
Enrichment score (ES)
Ranked list metric (PreRanked)
'na_pos' (positively correlated)
Zero cross at 1760
'na_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
H
Enriched in T2D-βobese-DEGAS cells
Enrichment plot: HALLMARK_HYPOXIA
Enrichment plot: HALLMARK_GLYCOLYSIS
I
high vs. low
βT2D-DEGAS
T2D vs. ND
βobese-DEGAS
733
1121
1649

Figure S5

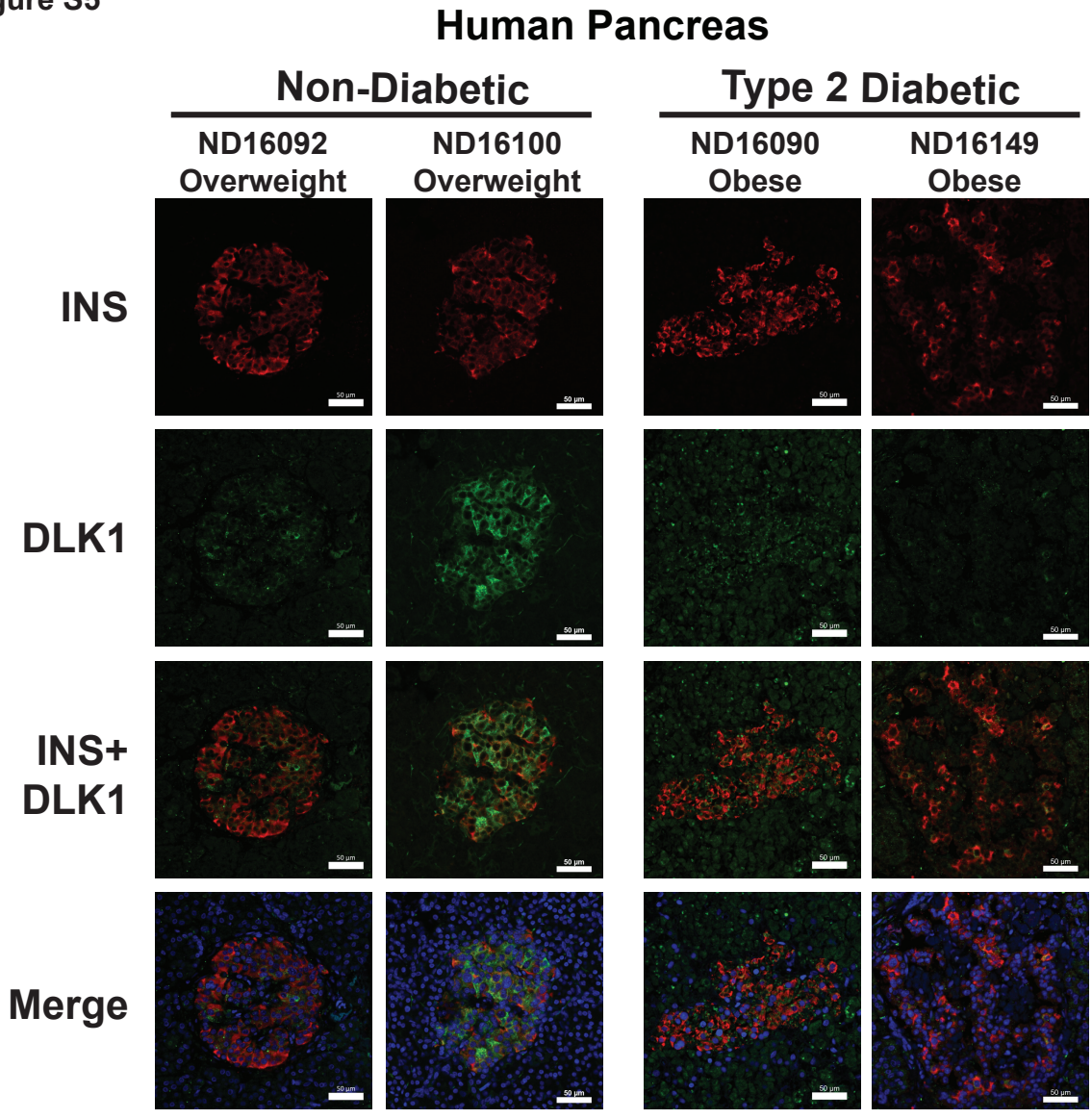
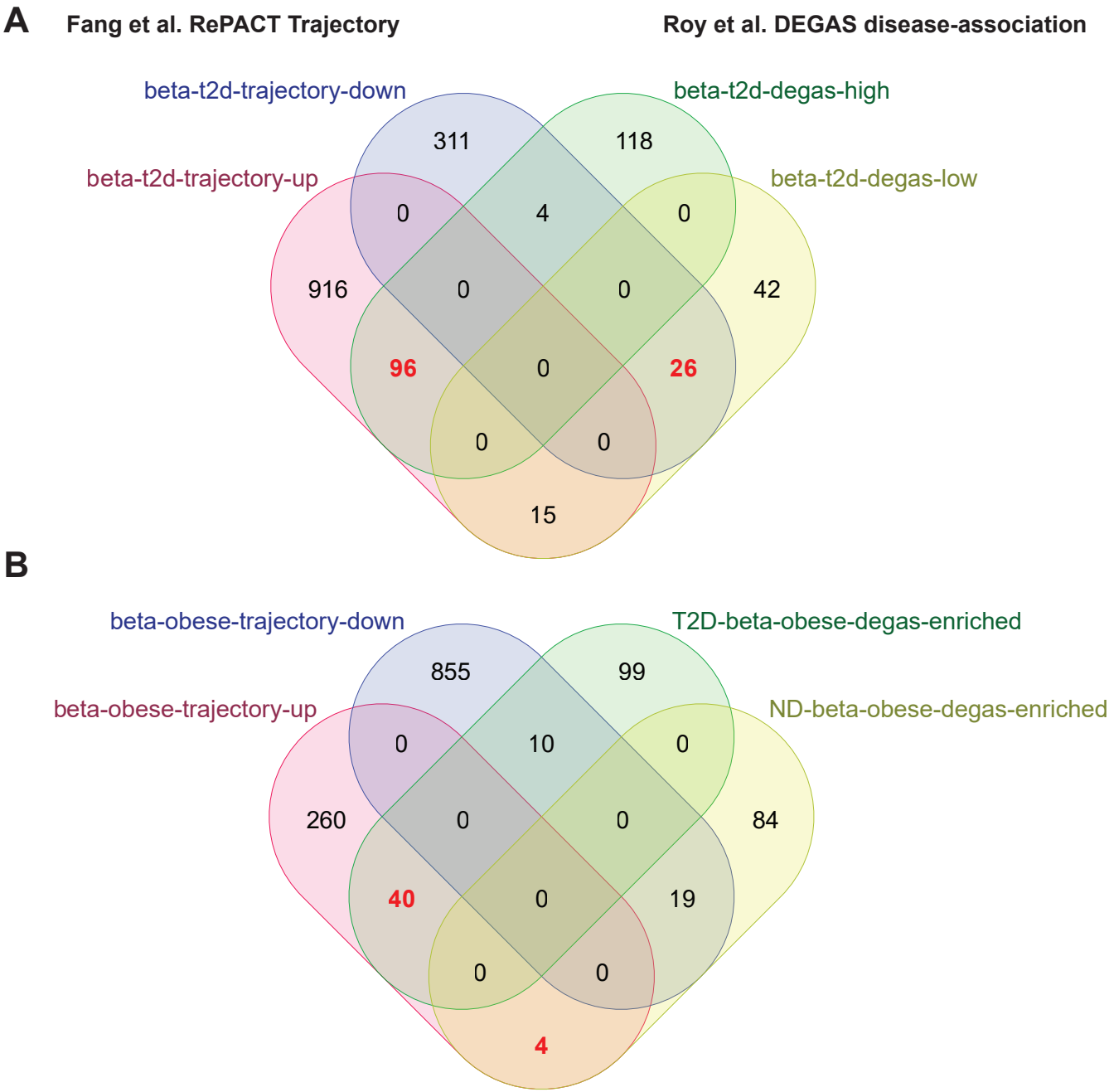

Human Pancreas
Non-Diabetic
Type 2 Diabetic
ND16092
Overweight
ND16100
Overweight
ND16090
Obese
ND16149
Obese
INS
DLK1
INS+
DLK1
Merge

Figure S6


A
Fang et al. RePACT Trajectory
Roy et al. DEGAS disease-association
beta-t2d-trajectory-down
beta-t2d-degas-high
beta-t2d-trajectory-up
beta-t2d-degas-low
311
118
0
4
0
916
0
0
42
96
0
26
0
0
15
B
beta-obese-trajectory-down
T2D-beta-obese-degas-enriched
beta-obese-trajectory-up
ND-beta-obese-degas-enriched
855
99
0
10
0
260
0
0
84
40
0
19
0
0
4