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 DLK1 for validation by immunostaining in human pancreas sections from healthy and T2D donors. DLK1 was heterogeneously expressed among β-cells and appeared depleted from T2D islets. 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-seq data have been combined and made searchable through web portals to browse datasets and compare cell types and marker genes [6-9]. To further utilize these types of integrated transcriptomic data and donor metadata,

53 we previously developed DEGAS as a flexible deep transfer learning framework that can be used to overlay
54 disease status, survival hazard, drug response, and other clinical information directly onto single cells. Machine
55 learning tools like DEGAS have been primarily used in cancer datasets, but not in human pancreatic islets until
56 now.

57 The reasons why some obese individuals succumb to T2D while other do not likely involves both genetic and
58 environmental factors, but the factors that underlie this transition are incompletely understood[10]. Analyses of
59 human genomics and islet transcriptomics indicate many β-cell genes have causal roles in T2D [11-15]. Subsets of
60 β-cells which are more resilient or susceptible to failure under the secretory pressure of insulin resistance may
61 be uncovered by combining transcriptomics with machine learning approaches like DEGAS. Here we have
62 implemented DEGAS to predict T2D- and obesity-associated subclusters of human pancreatic islet β-cells using
63 a combination of publicly available scRNA-seq and bulk RNA-seq human islet data and associated donor
64 metadata. Through this analysis we sought to identify novel and established genes implicated in T2D and obesity,
65 which were up- or down-regulated in subpopulations of β-cells identified by DEGAS, and to validate our findings
66 at the protein level using immunohistochemistry of pancreas tissue from non-diabetic and T2D organ donors.
67 Our current findings applying DEGAS to islet data have implications for β-cell heterogeneity in T2D and obesity.
68 The abundance of T2D-related factors and functional β-cell genes in our analysis validates applying DEGAS to
69 islet data to identify disease associated phenotypes and increase confidence in the novel candidates.
70

71 **DATA DESCRIPTION**

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).
85
86 ***Human islet single-cell data acquisition, filtering, and integration*** – We obtained read count tables for five
87 single-cell human islet datasets (GSE84133; GSE85241; E-MTAB-5061; GSE81608; GSE86469 (**Table S1**))
88 which were realigned to GRCh38.p5 (hg38) [8] (**Fig 1B**). Metadata were downloaded from GEO or obtained from
89 the supplemental information of the respective publications. These five datasets were previously integrated and
90 analyzed [8]. To prepare the data for our machine learning analyses, we repeated the integration and analysis of
91 these datasets in R using Seurat to exclude cells with low expressed genes and cells with over 20% mitochondrial
92 gene expression (low viability). The upper and lower limits varied between the datasets to account for differences
93 in library preparation and sequencing platform following guidelines in the Seurat documentation.
94 Datasets were integrated into a single dataset using Seurat version 4.1.3, the Seurat objects were normalized
95 using regularized negative binomial regression (SCTransform) to correct for batch effect and clustered using
96 default settings. The most variable features were used to identify integration anchors which are passed to the
97 *IntegrateData* function to return a Seurat object containing an integrated expression matrix of all cells.
98 The clusters were visualized using UMAP and expression of pancreatic hormone genes INS, GCG, SST,
99 PPY, GHRL were used to annotate β, α, delta, γ and ε cell clusters, respectively. Clustering analysis used the
100 Louvain algorithm [19] and labelling was performed to identify different cell types because our integrated dataset
101 contained a mixture of cells from pancreatic islets. In total, 22 clusters were identified from 17,273 pancreatic
102 cells and the clusters were classified into endocrine and non-endocrine cell types based on cell-specific marker
103 expression (**Fig 2A**). R scripts are available on GitHub (https://github.com/kalwatlab/Islet_DEGAS_v1). The
104 Seurat object containing all read count matrices and merged data and metadata is available on Mendeley[20].
105

106 **ANALYSES**

107 ***Differentially expressed genes from T2D bulk RNA-seq human islets show inflammatory signature and***
108 ***correlate with independent islet datasets*** – We selected bulk RNA-seq data from Marselli, et al. 2020 which
109 contains 58 non-diabetic and 27 T2D samples with associated metadata for BMI, age, and sex [16] (**Fig 1A-C**).
110 To determine the suitability of this data for DEGAS, we reanalyzed the read count matrix from GSE159984
111 using the edgeR likelihood ratio test to identify up- and down-regulated genes, visualized in both volcano (**Fig**
112 **1D**) and MD plot (**Fig 1E, Fig S1A**) formats to highlight fold-change versus significance and versus overall
113 expression levels, respectively (**Table S2**). We observed altered expression of genes well-known to be
114 dysregulated in T2D, including *IAPP* [21, 22], *PAX4* [23, 24], *SLC2A2* [25], *FFAR4*[26], and *ENTPD3* [27, 28] (**Fig 1D-E**).
115 Gene set enrichment analysis (GSEA) indicated that the gene expression profile of T2D human islets from this
116 dataset was associated with an inflammatory response phenotype (**Fig 1F**). To compare our analysis of
117 Marselli, et al. GSE159984 with an independent cohort, we selected a bulk RNA-seq human islet
118 transcriptomic analysis from Asplund et al. [17]. Because the raw read count data from Asplund et al. is not
119 accessible, we used the published supplementary table of $\log_2$-fold changes for differentially expressed genes
120 and compared it with the Marselli dataset (**Fig 1G**) (**Table S2**). Significant differentially expressed genes from
121 both Marselli and Asplund data included up-regulation of *SFRP4* and *PODN*, and down-regulation of *UNC5D*
122 and *FFAR4* (**Fig S1B**). We also found substantial agreement in the directional changes of most other
123 differentially expressed genes from both datasets, indicating that the Marselli islet data represents a suitable
124 cohort for DEGAS analysis. Finally, we compared up- and down-regulated genes from T2D islets in the
125 Marselli dataset with T2D effector genes from the T2D Knowledge Portal [29]. Down-regulated genes *PAX4* and
126 *SLC2A2* and up-regulated genes *APOE* each have evidence of a strong or causal role in T2D at the genetic
127 level, while *CYTIP* was up-regulated in T2D islets and has intron variant SNPs associated with T2D (e.g.
128 rs13384965) [30] (**Fig 1H**).

129

130 ***DEGAS revealed T2D-associated β-cells and marker genes within integrated human islet scRNA-seq***
131 ***data*** – We obtained five realigned scRNA-seq datasets [8] including Baron (GSE84133), Muraro (GSE85241),
132 Segerstolpe (E-MTAB-5061), Lawlor (GSE86469), and Xin (GSE81608), and integrated them using Seurat

133 4.3.0 to subtype each of the major cell types for use with DEGAS. All five datasets were successfully
134 integrated resulting in 17,273 cells (**Fig S2A**) and different cell types were clearly stratified from the scRNA-seq
135 data (**Fig 2A**). Distinct clusters of cell types were enriched for their respective markers including β-cells (*INS*),
136 α-cells (*GCG*), δ-cells (*SST*), ε-cells (*GHRL*), PP-cells (*PPY*), acinar cells (*REG1A*, *PRSS1*), ductal cells
137 (*KRT19*, *CFTR*), stellate cells (*COL6A1*), endothelial cells (*PLVAP*), and mesenchymal cells (*CD44*) (**Fig 2A**).
138 Next, we implemented DEGAS [31], a unique tool that leverages the increased sequencing depth, larger
139 number of clinical covariates and sample sizes of bulk sequencing data. We used the deep learning
140 architecture within DEGAS to project bulk islet sequencing data into the same latent representation as single
141 cell islet data via domain adaptation. From this common latent representation, clinical covariates from the bulk
142 sequencing data (e.g. T2D status and BMI) were projected onto single cells in a process called transfer
143 learning. This process generates unitless disease-association scores for each single cell. After mapping the
144 T2D-association scores to the single cell map (**Fig 3A**), we observed that β-cells had the lowest scores
145 compared to other cell types (**Fig 3B**). This may be interpreted as reflecting fewer β-cells or lower insulin
146 expression after onset of T2D [32]. We focused on differential T2D-association scores only among β-cells by
147 subsetting them based on *INS* expression (**Fig 4A**) and superimposed the T2D-association scores onto each
148 β-cell in the UMAP plot (**Fig 4B**). β-cells with higher scores (pink coloration) transcriptionally associate with
149 human islet T2D expression profiles more than the lower scoring (black) β-cells. Next, we set quantile
150 thresholds for the β-cell T2D-association scores and grouped the cells into high (upper 20%), medium and low
151 (lower 20%) categories (**Fig 4C**). We compared the high versus low $\beta^{T2D\text{-}DEGAS}$ groups to genes enriched in
152 either subset of β-cells **(Fig 4D, Table S3)**. High-scoring $\beta^{T2D\text{-}DEGAS}$ cells had elevated expression of *CDKN1C*
153 ($p57^{Kip2}$) which binds and inhibits G1 cyclin/CDK complexes and is also involved in cellular sensescence [33 34],
154 and reduced expression of *IAPP*, consistent with results from bulk T2D islet expression data (**Fig 1G**). These
155 findings show that the DEGAS can successfully identify diabetes-relevant genes enriched in specific
156 subclusters.
157 As an additional way to highlight genes that were identified by DEGAS, we excluded genes that were
158 differentially expressed between all non-diabetic (ND) and T2D β-cells (based on the merged scRNA-seq
159 donor metadata **Fig S4D, Table S3**). This process left over 90% of significant differentially expressed genes

160 remaining which were only identified through DEGAS analysis **(Fig 4E).** This emphasizes the advantage of
161 using DEGAS to generate continuous variables (disease-association scores) which enable thresholding and
162 prioritizing cellular subpopulations. Gene ontology (GO) analysis on high-scoring $\beta^{T2D\text{-}DEGAS}$ cells showed
163 enrichment in biological process categories including neutrophil degranulation, cell growth, and negative
164 regulation of cell development (**Fig 4F, Table S4**). These categories were driven in part by immune-related
165 genes (e.g. *HLA-B*, *PSAP*) and cell cycle regulator *BTG1*. Conversely, GO analysis on low-scoring $\beta^{T2D\text{-}DEGAS}$
166 cells showed enrichment of categories primarily related membrane protein targeting, translation, and
167 mitochondrial electron transport (**Fig 4G, Table S4**). The top six enriched GO categories were driven largely by
168 *RPS*/*RPL* genes. We also performed ranked-list GSEA on these same high- vs. low-scoring $\beta^{T2D\text{-}DEGAS}$
169 subpopulations which largely agreed with the GO analysis (**Fig 4H**). We observed significant positive
170 enrichment for hypoxia, TNFα signaling, estrogen response, and glycolysis (**Fig S3D**), while significant
171 negative enriched pathways included overlapping genes in oxidative phosphorylation and Myc-target pathways
172 (**Fig S3E**). Examples of genes that drive the GO and GSEA categories are show (**Fig 4I**) and heterogeneity in
173 the expression of these genes can be seen in β-cell UMAP plots (**Fig S3C**).
174
175 ***β-cell clusters with high BMI-association scores (high $\beta^{obese\text{-}DEGAS}$) show distinct differences between***
176 ***non-diabetic and T2D donors*** – Obesity and T2D are related, but not mutually inclusive. We surmised that
177 obesity-association scores in β-cells may highlight unique subpopulations compared to the $\beta^{T2D\text{-}DEGAS}$ cells.
178 Therefore, we categorized bulk RNA-seq donors by BMI (lean, <25; overweight, 25-30; obese >30) and ran the
179 DEGAS analysis based on these categories to calculate BMI-association scores similar to our approach with
180 T2D. Overlaying the BMI-DEGAS scores onto the islet single-cell data showed a distinct pattern of cell labeling
181 between the obesity-association score (**Fig 5A**) compared to lean- or overweight-association scores (**Fig**
182 **S5A,B**). It is interesting to note that the subpopulation of βlean-DEGAS cells appear to overlap with the
183 subpopulation of low-scoring $\beta^{T2D\text{-}DEGAS}$ cells (**Fig S4C**), suggesting that single β-cells share features of low
184 BMI and potentially lower T2D risk. The cell types with the highest obesity-association scores were β-cells,
185 acinar cells, and stellate cells (**Fig 5B**).

186 Reclustering β-cells with overlaid obesity-association scores enabled us to distinguish two major
187 subpopulations of high-scoring $\beta^{obese\text{-}DEGAS}$ cells (**Fig 6A**), a feature that did not occur with $\beta^{T2D\text{-}DEGAS}$ cells (**Fig**
188 **4B**). We observed that a distinguishing difference between these two groups of obesity-associated β-cells was
189 the donor diabetes status (ND vs T2D) in the single-cell metadata. We applied a median quantile threshold to
190 extract these two subpopulations and classified the cell clusters based on their donor diabetes status to create
191 ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ subpopulations (**Fig 6B**). We next determined differentially expressed
192 genes between the ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ cells (**Fig 6C**). Gene ontology analysis of genes
193 upregulated in the ND-$\beta^{obese\text{-}DEGAS}$ cluster indicated an enrichment for unfolded protein response (UPR)_
194 processes like vesicle transport, translation, and protein folding and stability (**Fig 6D**). Specifically, the ND-
195 $\beta^{obese\text{-}DEGAS}$ cluster was enriched for expression of adaptive (e.g. *MANF* and *HSPA5*) and maladaptive (e.g.
196 *TRIB3*, *DDIT3*) UPR genes, as well as genes involved in ER-associated degradation (e.g. *SEL1L*, *DERL2*, and
197 *SDF2L1*). Gene ontology analysis of genes upregulated in the T2D-$\beta^{obese\text{-}DEGAS}$ cluster showed enrichment for
198 hormone transport and secretion pathways (e.g. *IGFBP5*, *UCN3*, *G6PC2*) and inflammatory/immune-related
199 pathways (e.g., *HLA-A/B/E*, *STAT3*) (**Fig 6E**). In parallel, we performed ranked-list GSEA on the ND- and T2D-
200 $\beta^{obese\text{-}DEGAS}$ subpopulations which was largely consistent with gene ontology results (**Fig 6F**). Significantly
201 enriched pathways for ND-$\beta^{obese\text{-}DEGAS}$ included the UPR and proliferation (E2F and MTORC signaling
202 hallmarks) (**Fig S4I**), while T2D-$\beta^{obese\text{-}DEGAS}$ enriched pathways were similar to that of high-scoring $\beta^{T2D\text{-}DEGAS}$
203 cells and included hypoxia and glycolysis (**Fig S4J**). Examples of genes driving the GO and GSEA categories
204 are shown, as well as *CDKN1C* and *DLK1*, given their differential enrichment in both $\beta^{T2D\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}}$
205 ^{DEGAS} cells (**Fig 6G**). The transcript heterogeneity for genes highlighted in **Fig 6G** can be observed in single β-
206 cell expression plots (**Fig S4H**). Our results support the idea that these genes and pathways may underlie
207 heterogeneity in β-cell resilience or susceptibility to obesity-related stress in the context of T2D development.
208
209 ***DLK1 is a candidate identified in $\beta^{T2D\text{-}DEGAS}$ and ND-$\beta^{obese\text{-}DEGAS}$ cells and is heterogeneously expressed***
210 ***among diabetic and non-diabetic human pancreatic islets*** – To support the DEGAS approach for
211 identifying true markers of β-cell subpopulations or heterogeneity associated with diabetes, we selected a
212 candidate gene for immunostaining in FFPE human pancreas sections from non-diabetic and T2D donors. The

213 subpopulations of cells for both $\beta^{T2D\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ exhibited a subset of overlapping enriched
214 genes. One of the more interesting candidates we noted was *DLK1* (Delta-Like 1 Homolog), which is known to
215 have strong associations with T1D [35] and T2D [36]. *DLK1* expression highly overlapped with high scoring $\beta^{T2D\text{-}}$
216 ^{DEGAS} cells (**Fig 7A**) and with T2D-$\beta^{obese\text{-}DEGAS}$ cells (**Fig 7B**). DLK1 immunostaining primarily colocalized with β-
217 cells in non-diabetic human pancreas (**Fig 7C**). DLK1 showed heterogeneous expression within islets and
218 between islets within the same pancreas section, wherein some islets had DLK1/INS co-staining in most β-
219 cells and other islets had only a few DLK1$^+$ β-cells. In T2D pancreas, DLK1 staining was much less intense
220 and in fewer β-cells, yet DLK1$^+$/INS$^+$ cells were observed (**Fig 7C**). This contrasts with the relatively higher
221 *DLK1* gene expression seen in the β-cells from the $\beta^{T2D\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ subpopulations (**Fig 4D &**
222 **6C**) as highlighted in **Fig 7A,B**.
223

224 **DISCUSSION**

225 ***Deep transfer learning is a useful approach to identify disease-associated genes in subsets of***
226 ***heterogeneous islet cells –*** The rapid advancement of deep learning has unveiled new opportunities in
227 diabetes research, specifically the ability to merge and analyze the large amount of publicly available omics data.
228 The DEGAS deep transfer learning framework provides an extremely powerful and versatile tool allowing for
229 time to event outcomes and classification outcomes to be transferred to individual cells. Notably, this direct
230 transfer to single cells avoids the cluster resolution problems that arise using deconvolution approaches, allowing
231 for subsets of cells within a cluster to be assigned different associations with clinical outcomes. In this way,
232 DEGAS allows the definition of essentially new cell subpopulations that would not necessarily be identified by
233 standard clustering algorithms and these subpopulations are rationally selected based on assigned disease
234 association scores. A key point from our study is that the disease-associated single cells identified by DEGAS
235 are those cells whose transcriptomic signature share complex patterns with patient transcriptomics that have
236 known clinical and metabolic attributes. Because of this, the genes identified as up- or down-regulated in subsets
237 of β-cells may be: 1) altered independent of disease due to genetic background of donors; 2) altered downstream
238 of the obese or diabetic states; or 3) representative of resilient or dysfunctional β-cells in the face of metabolic

239 syndrome. Through careful cohort inclusion criteria and batch effect removal, our analyses aim to limit these
240 caveats when possible.

241 Various deconvolution methods enable estimating relative proportions of cell types in bulk samples using
242 scRNA-seq data (e.g. BSEQ-sc [1], MuSiC [37], CIBERSORT [38], and DWLS [39]). These methods have been
243 instrumental in identifying changes to relative quantity of cell types within normal and disease tissues; however,
244 they do not provide detailed information about subsets of cell types which may be related to disease and are
245 limited to the resolution of the predefined cell types through clustering. Early attempts to address this include
246 unsupervised learning algorithms (e.g. RaceID) [40]. Now, there are multiple advanced computational approaches
247 to identify subsets of cells related to disease status, survival, drug response, and other disease metrics. These
248 tools can be broadly categorized as 'cell prioritization algorithms' and include: DEGAS [31], scAB [41], Scissor [42],
249 and scDEAL [43]. Scissor and scAB can assign survival or clinical information from bulk expression data to
250 disparate scRNA-seq datasets using regression and matrix factorization, respectively. scDEAL is a deep learning
251 framework specifically designed for the assignment of drug response information from bulk expression data to
252 subsets of single cells. The major advantages of DEGAS are 1) the ability to select subsets of β-cells at an
253 individual cell resolution based on their disease association; 2) unique qualities of the highly non-linear DEGAS
254 technology vs. linear models of other tools to detect complex molecular programs; and 3) speed and scalability
255 based on efficient implementation. DEGAS has the potential to be used on datasets containing well over 1 million
256 cells, unlike other methods.

257 Nevertheless, certain linear model-based approaches, like RePACT (regressing principal components for the
258 assembly of continuous trajectory), have successfully uncovered genes associated with β-cell heterogeneity in
259 T2D and obesity [5]. In that study, T2D and obesity 'trajectory' genes were identified. We compared the genes
260 enriched in our high-scoring $\beta^{T2D\text{-}DEGAS}$ and in our T2D-$\beta^{obese\text{-}DEGAS}$ cells with the corresponding T2D and obesity
261 trajectory genes from Fang et al., respectively (**Fig S5**, **Table S5**). We noted substantial agreement between
262 DEGAS and RePACT, for example the transcription factor *SIX3* was enriched in T2D-$\beta^{obese\text{-}DEGAS}$ cells and in
263 Fang et al.'s obese trajectory β-cells [5]. *SIX3* has been implicated as a diabetes risk gene and is important for β-
264 cell function [44, 45]. Additionally, both approaches highlighted the association of *DLK1* with obesity, but only
265 DEGAS identified *DLK1* association with T2D. Overall, the directionality of T2D or obesity gene associations

266 agreed between DEGAS and RePACT results. However, not all the same genes were identified by both
267 approaches, for example, *FXYD2* was identified via RePACT, but not in DEGAS, although *FXYD2* is a down-
268 regulated gene in T2D in Marselli et al. and Asplund et al. (**Fig 1G**). Interestingly, *FXYD2* has was recently shown
269 to be a downstream effector gene of *HNF1A* in single human β-cells [46]. Down-regulation of *HNF1A* in T2D β-
270 cells, and reduced *FXYD2* expression as a result, may contribute to membrane hyperpolarization and reduced
271 function [46]. Distinct from the application of RePACT in Fang, et al.[5], we did not observe any differences in *HNF1A*
272 expression in our DEGAS analyses and *FXYD2* was slightly enriched, but not significantly altered, in T2D-$\beta^{obese\text{-}DEGAS}$
273 cells. Taken together, it is necessary to apply multiple approaches to merged sets of publicly available data
274 because each approach will likely uncover unique and important subsets of β-cells or specific genes.

275

276 ***Differences between high- and low-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations –***

277 High-scoring $\beta^{T2D\text{-}DEGAS}$ cells had enrichment of pathways like hypoxia and TNFα signaling which contain some
278 overlapping genes like *BTG1*, *BTG2*, and *BHLHE40*. *BHLHE40* was recently shown to cause hypoxia-induced
279 β-cell dysfuction via interfering with *MAFA* expression [47]. *BTG1* and *BTG2* are known as anti-proliferation factors.
280 Given the expression of both *BTG1* and *CDKN1C* in high-scoring $\beta^{T2D\text{-}DEGAS}$ cells, it is possible these cells have
281 very low proliferative potential, even in the context of β-cells which are well-known to have low proliferation.
282 Although some genes like *CDKN1C* can be identified as differentially expressed in simple comparisons of all
283 T2D vs ND donor β-cells in scRNA-seq data, DEGAS has the advantage of prioritization or ranking of candidates.
284 For example, *CDKN1C* is within the top ten enriched candidates (by adjusted P value and fold-change) in the
285 $\beta^{T2D\text{-}DEGAS}$ cluster but is 286th when simply comparing all T2D vs ND single β-cells (**Table S3**). Additionally,
286 deletion of *CDKN1C* has been shown to improve human islet function [48] and gain-of-function mutants are
287 associated with diabetes development and hyperinsulinism [49, 50].

288 The low-scoring $\beta^{T2D\text{-}DEGAS}$ cells were enriched in pathways involving oxidative phosphorylation and Myc
289 targets. We also noted enrichment of many small and large ribosomal genes (*RPS*/*RPL*) in this subpopulation.
290 *RPS*/*RPL* ribosomal genes are known to be highly expressed in primary human β-cells [51], and glucose
291 stimulation induced translation of over 50 different *RPS*/*RPL* genes in human β-cell line EndoC-βH2 [52]. This

292 suggest that low-scoring $\beta^{T2D\text{-}DEGAS}$ cells may have relatively increased translation and metabolic activity. Taken
293 together, high-scoring $\beta^{T2D\text{-}DEGAS}$ cells may be reflective of dysfunctional β-cells in either a pre-T2D or T2D-onset
294 state, while low-scoring $\beta^{T2D\text{-}DEGAS}$ cells could potentially represent more resilient β-cells.

295

296 ***Differences between ND-$\beta^{obese\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$ subsets of β-cells –*** Our understanding of why
297 some individuals with obesity develop T2D while others do not is incomplete. A 2021 NHANES report indicated
298 a 41.9% prevalence for obesity in US adults aged 20 and over, with a 14.8% prevalence for diabetes [53], in line
299 with CDC reports [54]. This suggests that many individuals with obesity have either not progressed, or may never
300 progress, to T2D. This phenomenon may be related to those individuals categorized as metabolically healthy
301 obese as compared to metabolically unhealthy obese [55]. Expanded work with DEGAS may uncover specific
302 genes that underly this relationship. In our analysis we were able to stratify based on T2D status of the single
303 cell donors and overlay the relative obesity-association scores from DEGAS. That analysis allowed us to set
304 rational thresholds for grouping subsets of β-cells to compare ND- and T2D-$\beta^{obese\text{-}DEGAS}$ cells. Perhaps
305 counterintuitively, UPR genes were highly enriched in the ND-$\beta^{obese\text{-}DEGAS}$ cells as opposed to the T2D-$\beta^{obese\text{-}}$
306 ^{DEGAS} cells. It is possible these ND-$\beta^{obese\text{-}DEGAS}$ cells are in a stressed pre-diabetic state on the path to T2D.
307 Dominguez-Gutierrez et al. analyzed islet scRNA-seq data using pseudotime analysis and identified three major
308 β-cell states which were defined by their insulin expression and UPR level [56]. The authors speculate that β-cells
309 periodically pass through these states, which in scRNA-seq are visualized as subpopulations of cells. Previously,
310 a *FTH1* subpopulation of β-cells was identified with implications in the UPR [57]. In agreement with those results,
311 we found *FTH1* was upregulated in ND-$\beta^{obese\text{-}DEGAS}$ cells. Therefore, another possibility is that the UPR induced
312 in ND-$\beta^{obese\text{-}DEGAS}$ cells marks a population with lower functionality but higher proliferative potential to combat the
313 insulin resistant obese state, and the relatively lower UPR in T2D-$\beta^{obese\text{-}DEGAS}$ cells possibly correlates with
314 enhanced functionality in established T2D. Concordantly, T2D-$\beta^{obese\text{-}DEGAS}$ cells had enrichment of hormone
315 secretion genes (e.g. *SYT7, G6PC2, NEUROD1, UCN3*, *FFAR1*) in our pathway analysis (**Fig 6E,F**). In a study
316 of human, mouse, and pig islet scRNA-seq data, subpopulations of 'stressed' β-cells were identified that
317 exhibited enriched hallmark pathways similarly to what we observe in ND-$\beta^{obese\text{-}DEGAS}$ cells [6]. Thus, multiple

318 studies including our current analysis support the existence of subpopulations of stressed β-cells, whether
319 transient or stable, that could in principle be targeted by therapeutics.

320 We also looked for enrichment of potential secreted biomarkers in high-scoring $\beta^{T2D\text{-}DEGAS}$ and T2D-$\beta^{obese\text{-}DEGAS}$
321 cells and identified *LRPAP1* and *C1QL1*. *LRPAP1* encodes the LDL receptor-related protein-associated protein
322 1 (LRPAP) and is enriched in this subpopulation. LRPAP is ubiquitously expressed and is predicted to be an ER
323 resident protein [58]. LRPAP appears to be heterogeneously expressed among islet cells and was detected in
324 human plasma in the in the Human Protein Atlas [59]. *C1QL1* encodes a secreted peptide that is highly-expressed
325 in human islets [60] and was enriched in $\beta^{T2D\text{-}DEGAS}$ cells, although little else is known about its role in β-cells.
326 Increased C1QL1 release could potentially signal through its GPCR BAI3 to suppress insulin secretion of
327 surrounding β-cells, as has been shown for C1QL3 [61].

328

329 ***Potential for DEGAS in identifying β-cell heterogeneity markers –***

330 Subsets or subpopulations of β-cells are an emergent property of β-cell heterogeneity. Two of the most widely
331 used scRNA-seq datasets have also identified significant heterogeneity within pancreatic cell types, notably β-
332 cells [7, 57]. Various markers have been identified to define these β-cell subpopulations, including expression of
333 ST8SIA1 and CD9 [62] or Flattop expression [63]. Additionally, UCN3 marks mature β cells [64] and $RBP4^{+}$ β-cells
334 correlated with reduced function [65]. Functional heterogeneity has also been described as in the case of
335 leader/first-responder β cells [66, 67] or hub β cells [68], and has been linked to transcription factors including PDX1
336 [69] and HNF1A [46]. Nevertheless, cellular heterogeneity within a single cell type is complicated and there is
337 uncertainty in the single cell analytics field about what constitutes a stable cell type versus a transient cell state
338 [70]. Tying transcriptionally distinct clusters to function or morphology may be key to making this distinction [71]. By
339 inferring disease outcomes in high resolution cellular subtypes, application of DEGAS can help to inform this
340 debate.

341 In our analysis, *DLK1* expression was enriched in high-scoring $\beta^{T2D\text{-}DEGAS}$ and in T2D-$\beta^{obese\text{-}DEGAS}$ cells. We
342 also observed heterogeneous DLK1 staining of β-cells from human pancreas sections. *DLK1* was previously
343 noted to be enriched in subpopulations of β-cells [1, 2, 4, 8, 72], however its heterogeneity in transcript and protein

344 abundance in non-diabetic versus T2D human islet β-cells had not been described until now. DLK1 is a
345 maternally-imprinted gene[73] with described roles in Notch signaling[74-76] and glucose homeostasis[77]. Supporting
346 an active role β-cell disease, a recent preprint reported DLK1 was required for proper maturation, function, and
347 stress resilience of β-like cells differentiated from human embryonic stem cells [78]. Our results combined with
348 the published data indicate DLK1 as a significant player in β-cell function in diabetes. DLK1 may serve as a
349 surface marker of, or be released from, stressed or at-risk β-cells. However, increased sample sizes and in-
350 depth analyses of both non-diabetic, T1D, and T2D human islets will be required to better describe the
351 differences seen in transcript and protein abundance of DLK1 between β-cells within the same islet, between
352 islets of the same and different donors, and across different stages of diabetes development.

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

359

360 ***Limitations of the study and future directions*** – We think the approach of joining islet omics data to deep
361 learning and artificial intelligence is an area in need of increased attention; however, we appreciate that our
362 current study has limitations. First, our analysis utilized a single bulk RNA-seq human islet dataset, although the
363 Marselli study contains a relatively large number of islet donors compared to most available datasets. Second,
364 we merged only five scRNA-seq human islet datasets for this application of DEGAS. The datasets were chosen
365 because of the extensive meta-analyses to which they have already been subjected [8], and their use as
366 benchmarking data in many scRNA-seq analysis tools [79-81]. scRNA-seq has a limitation on the number of
367 detectable genes in each cell, as compared to bulk RNA-seq samples. While these limitations may eventually
368 be mitigated by improved technologies, we can begin to overcome these by merging a larger number of human
369 islet bulk and scRNA-seq studies which contain more donors. Increased sample sizes and study variables come

370 with their own challenges, however the future gene candidates identified by DEGAS will be of even higher
371 confidence. Additionally, scRNA-seq is a snapshot in time and it may be tenuous to claim that a particular cell
372 state is transient or persistent. To support a claim that a given subset/subtype of β-cells represents an actual
373 stable population in scRNA-seq data requires at least finding that the cells comprising the candidate population
374 occur across multiple individual donors. In future studies, our use of the diabetes field's considerable investment
375 in islet transcriptomics data and state-of-the-art cell prioritization tools will enable simultaneous identification of
376 disease-associated cellular subtypes and associated biomarkers of function and dysfunction. Additionally,
377 increasing the number of human pancreas donor tissue samples for high-content image analysis will improve
378 confidence in candidate gene validation.

379 **POTENTIAL IMPLICATIONS**

380 DEGAS has the potential to be applied to even larger mergers of single cell data (> 1 million cells) using the
381 newly developed DEGAS atlas implementations. A vast amount of human (and other species) islet single cell
382 transcriptomics data is publicly available, but often requires substantial reformatting of metadata and realigning
383 of reads to be harmonized. As our endeavors and those of others proceed, DEGAS will lead to even higher
384 confidence predictions of β-cell subpopulations. It is also apparent from our DEGAS analyses that other non-β-
385 cells are highlighted in the islet single cell map for associations with T2D and obesity. Although outside the
386 scope of this work, further exploration of our current DEGAS analyses, or of analyses using larger single cell
387 integrations, will have implications for the other major islet and pancreas cell types, including α-cells, δ-cells
388 and PP-cells. For example, PP cells have been shown to have a role in pancreatogenic diabetes as opposed
389 to T2D [82], and there appear to be subpopulations of high and low scoring $PP^{T2D\text{-}DEGAS}$ cells within the DEGAS
390 analysis (**Fig 3A**). Our successful application of deep transfer learning in human islet data opens the possibility
391 of predicting subtypes of β-cells in other diseases, like T1D and congenital hyperinsulinism, to find potential
392 biomarkers and therapeutic candidates. In some rare disease, like congenital hyperinsulinism, single cell data
393 is not available, however spatial transcriptomics (ST) could be applied to FFPE sections. DEGAS has already
394 been implemented to analyze ST data in prostate cancer [83], and will be applicable to human pancreas as ST
395 technology continues to advance. It is important to note that DEGAS is just one of many machine learning

396 approaches to analyze transcriptomic data. It will be important to include DEGAS in combination with both
397 linear and other non-linear models to capture as many relevant gene candidates as possible.
398

399 **METHODS**

400 ***Transfer learning using DEGAS on human islet transcriptomic data –*** The integrated single cell and bulk
401 datasets were processed with the DEGAS (v1.0) pipeline (https://github.com/tsteelejohnson91/DEGAS) [31] to
402 calculate disease risk scores associated with T2D status or BMI status. The bulk expression data were scaled
403 and normalized prior to DEGAS analysis using the *preprocessCounts* function. The merged scRNA-seq
404 expression matrix, the bulk expression matrix, and donor sample labels (matched with the bulk samples) were
405 used as input. The intersection of highly variable genes between scRNA-seq and bulk expression data were
406 used for further analysis. The DEGAS model was trained and predicted on the formatted data. Updated scripts
407 and instructions for running DEGAS are available on GitHub (version 1.0)
408 (https://github.com/tsteelejohnson91/DEGAS). DEGAS for T2D and BMI were run independently.
409 In DEGAS for T2D, the donor labels were "normal" vs "T2D." In DEGAS for BMI metadata, donors from bulk
410 RNA-seq were categorized and labeled as "lean" (<25 BMI), "overweight" (25-30 BMI) and "obese" (>30 BMI)
411 according to CDC guidelines. The models were trained to calculate T2D-association scores and BMI-association
412 scores for each of the respective categories. We identified differentially-expressed genes using *FindMarkers*
413 function in β-cells with high vs low T2D-association scores ($\beta^{T2D\text{-}DEGAS}$) or high obesity scores ($\beta^{obese\text{-}DEGAS}$) among
414 healthy or T2D donor single-cells.
415

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

422

423 ***Pathway analysis*** – To analyze differentially-expressed genes that were either up- or down-regulated between
424 clusters, enrichment analyses and plot generation were completed using ClusterProfiler, EnhancedVolcano, and
425 ggplot2 R packages in R version 4.2.2 and RStudio. Enrichment of gene sets within the the Biological Process
426 GO terms are shown and cluster profiler outputs are provided in **Table S4**. Venn diagrams for comparing DEGAS
427 and RePACT hit genes (**Table S5**) were created using https://molbiotools.com/listcompare.php. Gene set
428 enrichment analysis was run using GSEA software [84, 85] downloaded from www.gsea-msigdb.org/gsea and the
429 MSigDB Hallmark Gene Set [86] from the Broad Institute. For bulk RNA-seq analysis, the edgeR results containing
430 all expressed genes was used as input to run standard GSEA. For GSEA analysis of comparisons between
431 subpopluations of β-cells from scRNA-seq data, a ranked list based on log_2fold-change was generated without
432 a cutoff for adjusted P value. GSEAPreranked mode used the 'classic' enrichnment statistic for 1000
433 permutations.

434

435 ***Human pancreas tissue staining and microscopy*** – Formalin-fixed paraffin-embedded (FFPE) de-identified
436 human pancreas tissue were obtained through the NDRI (**Table S6**). Tissue was processed into 5 µm sections
437 and mounted on glass slides by the Indiana University School of Medicine Histology Lab Service Core. Slides
438 were deparaffinized by xylene and ethanol washes. Antigen retrieval was performed by heating for 40 min in an
439 Epitope Retrival Steamer with slides submerged in Epitope Retrieval Solution (IHC-Tek). Subsequently, slides
440 were placed onto disposable immunostaining coverplates and inserted into the Sequenza slide rack (Ted Pella
441 / EMS) for washing, blocking, and antibody incubations. After three 10 min washes in IHC wash buffer (0.1%
442 Triton X-100 and 0.01% sodium azide in PBS pH 7.4), slides were blocked for 1 h at room temperature in normal
443 donkey serum (NDS) block solution (5% donkey serum, 1% bovine serum albumin, 0.3% Triton X-100, 0.05%
444 Tween-20, and 0.05% sodium azide in PBS pH 7.4). Slides were then incubated overnight at 4°C with primary
445 antibodies diluted in NDS block solution (**Table S7**). After three washes in IHC wash buffer 200 µL each), slides
446 were incubated in secondary antibodies in NDS block solution for 1 h at room temperature. The washed slides
447 were mounted in polyvinyl alcohol (PVA) mounting medium (5% PVA, 10% glycerol, 50mM Tris pH 9.0, 0.01%
448 sodium azide) with DAPI (300 nM) added and imaged on Zeiss LSM710 confocal microscope equipped with a

449 Plan-Apochromat 20x/0.8 objective (#420650-9901). Images were processed in the Zeiss Zen software to add
450 scale bars, set coloration for channels, and generate merged images. Scale bars indicate 50 µm.
451
452 ***Statistical Analysis –*** In edgeR bulk RNAseq differential gene expression analysis, the likelihood ratio test
453 was applied. In the FindMarkers function in Seurat the default Wilcoxon Rank Sum test was used for
454 differential gene expression analysis. Cut-offs for signifcant differentially expressed genes were set at $|\log_2$fold-
455 change$|$ > 0.58 and adjusted p-value < 0.05.
456

457 **AVAILABILITY OF SOURCE CODE AND REQUIREMENTS**

458 - Project name: Islet DEGAS
459 - Project home page: https://github.com/kalwatlab/Islet_DEGAS_v1
460 - Operating system(s): Platform independent
461 - Programming language: R and Python
462 - Other requirements: not applicable
463 - License: ©
464

465 **DATA AVAILABILITY**

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

469 **DECLARATIONS**

470 **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

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).

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

We thank Dr. Mark Huising for helpful conversations and providing access to processed data. Thank you to Dr. Andrew Templin for helpful conversations and critical review of the manuscript. Thank you to Dr. Anthony Piron at Université Libre de Bruxelles for assistance with metadata mapping for GSE159984. This work was supported by the Histology Core of the Indiana Center for Musculoskeletal Health at IU School of Medicine and the Bone and Body Composition Core of the Indiana Clinical Translational Sciences Institute (CTSI). A specific 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)** DEGs from Marselli et al. that were also found on the Type 2 Diabetes Knowledge Portal database of known diabetes effector genes.

Figure 2. Validation of merged scRNA-seq datasets from five human islet studies. A) After integrating the five scRNA-seq datasets from Fig 1B, gene expression plots were created 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 show cex (**B**), BMI (**C**), age (**D**), and diabetes status (**E**).

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) β-cells were subsetted from all other cells based on *INS* expression and reclustered. **B)** DEGAS scores for T2D association ($\beta^{T2D\text{-}DEGAS}$) were overlaid onto the cell plot. **C)** β-cells were classified as low, medium and high-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations for downstream analysis. **D)** 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. **E)** Genes from (**D**) 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 S1D**). Gene ontology (GO) analysis of genes enriched in the high (**F**) and low (**G**) scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations. **H)** GSEA results for high- vs. low-scoring $\beta^{T2D\text{-}DEGAS}$ subpopulations. **I)** 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)** 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. **C)** 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-$\beta^{obese\text{-}DEGAS}$ (**D**) and T2D-$\beta^{obese\text{-}DEGAS}$ (**E**) subpopulations. **F)** Gene set enrichment analysis (GSEA) of ND- vs. T2D-$\beta^{obese\text{-}DEGAS}$ subpopulations. **G)** 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) Formalin-fixed paraffin-embedded (FFPE) sections of human pancreas from non-diabetic (ND) and T2D donors was stained for INS and DLK1. Representative images are shown from three

different ND and T2D donors. Arrows indicate $INS^+/DLK1^+$ cells. **B)** UMAP plot overlaying DLK1 expression and βT2D-DEGAS (**i**) or βobese-DEGAS (**ii**) scores.

Supplemental Information

Figure S1. Islet Gene View results of DEGs found in both Marselli et al. and Asplund et al. *FFAR4* (**A**), *UCN5D* (**B**), *SFRP4* (**C**), and *PODN* (**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) Seurat clustering of merged human islet scRNA-seq data. **B)** Study identifiers overlaid on single cell plot. **C)** Differential gene expression analysis comparing all cells from T2D donors vs. non-diabetic (ND) donors. **D)** β-cells reclustered and labeled based on donor T2D status. **E)** Differential expression analysis of β-cells from T2D vs. ND donors.

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)** Gene expression overlays on β-cells of selected genes that drive enrichment of GO terms and GSEA categories. **D)** GSEA results for high βT2D-DEGAS scoring cells. **E)** GSEA results of low-scoring $\beta^{T2D\text{-}DEGAS}$ scoring cells.

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)** Differential expression analysis for high vs low $\beta^{obese\text{-}DEGAS}$ scores based on median threshold. **G)** Subsetting highest-scoring cells based on $\beta^{obese\text{-}DEGAS}$ score. **H)** Gene expression overlays on β-cell UMAP. **I)** GSEA results for ND-$\beta^{obese\text{-}DEGAS}$ cells. **J)** GSEA results for T2D-$\beta^{obese\text{-}DEGAS}$ cells.

Figure S5. 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 $|\log_2$fold-changed$|$ > 0.58 were selected for comparison to Fang et al. **Table S4** 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. FFPE human pancreas metadata.

Table S6. Antibodies.

Table S7. Comparison of DEGAS T2D and obesity genes to Fang et al. RePACT trajectory genes.

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.
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 of AstraZeneca, S.K.T. is now an employee of Vertex Pharmaceuticals and A.W-A. is now an employee of Genomics plc. D.E. and E.N-G are now also affiliated with XCellomics. All experimental work by A.K.R., S.K.T. and A.W-A. was carried out under employment at the University of Oxford. A.K.R. holds stocks in Astra Zeneca. A.L.G.'s spouse holds stock options in Roche. The other authors declare no competing interests.

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.

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. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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. doi: 10.1016/j.cmet.2023.03.001. PMID: 36963395.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. 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.

64. 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 following conflicts: OGK is an employee of Viacyte, receiving salary and/or stock options. The paper describes or references intellectual property owned by Viacyte. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials. The other authors declare that no competing interests exist.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

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

71. 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.

72. 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. doi: 10.1016/j.cmet.2016.08.018. PMID: 27667665.

73. 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.

74. 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.

75. 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.

76. 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.

77. 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.

78. 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 of Vesalius Therapeutics. The other authors declare no competing interest.

79. 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.

80. 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.

81. 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.

82. 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.

83. 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.

84. 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.

85. 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.

86. 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.

Figure 1

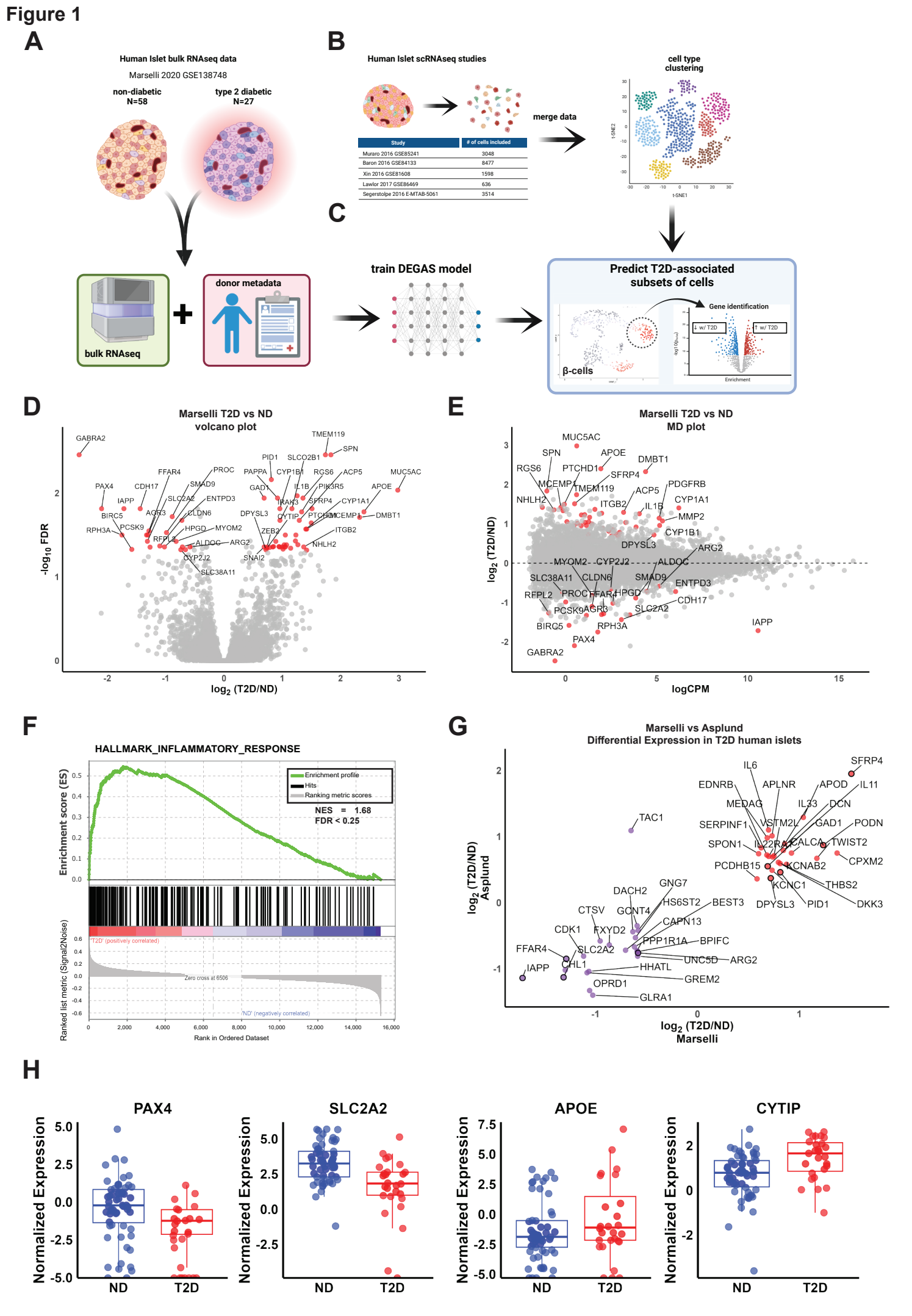

A
Human Islet bulk RNAseq data
Marselli 2020 GSE138748
non-diabetic N=58
type 2 diabetic N=27
bulk RNAseq
donor metadata
B
Human Islet scRNAseq studies
Study
# of cells included
Muraro 2016 GSE85241 3048
Baron 2016 GSE84133 8477
Xin 2016 GSE81608 1598
Lawlor 2017 GSE86469 636
Segerstolpe 2016 E-MTAB-5061 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

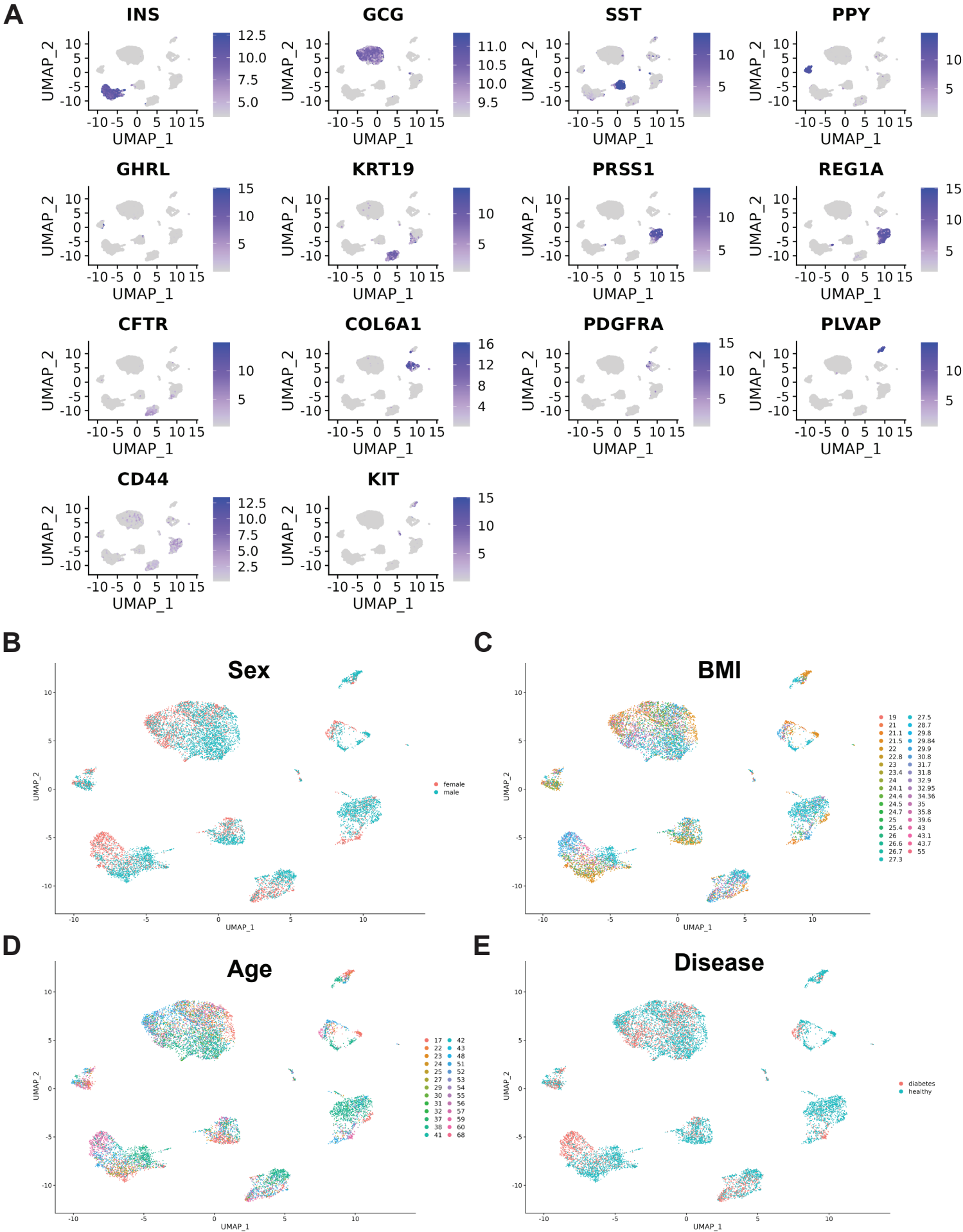
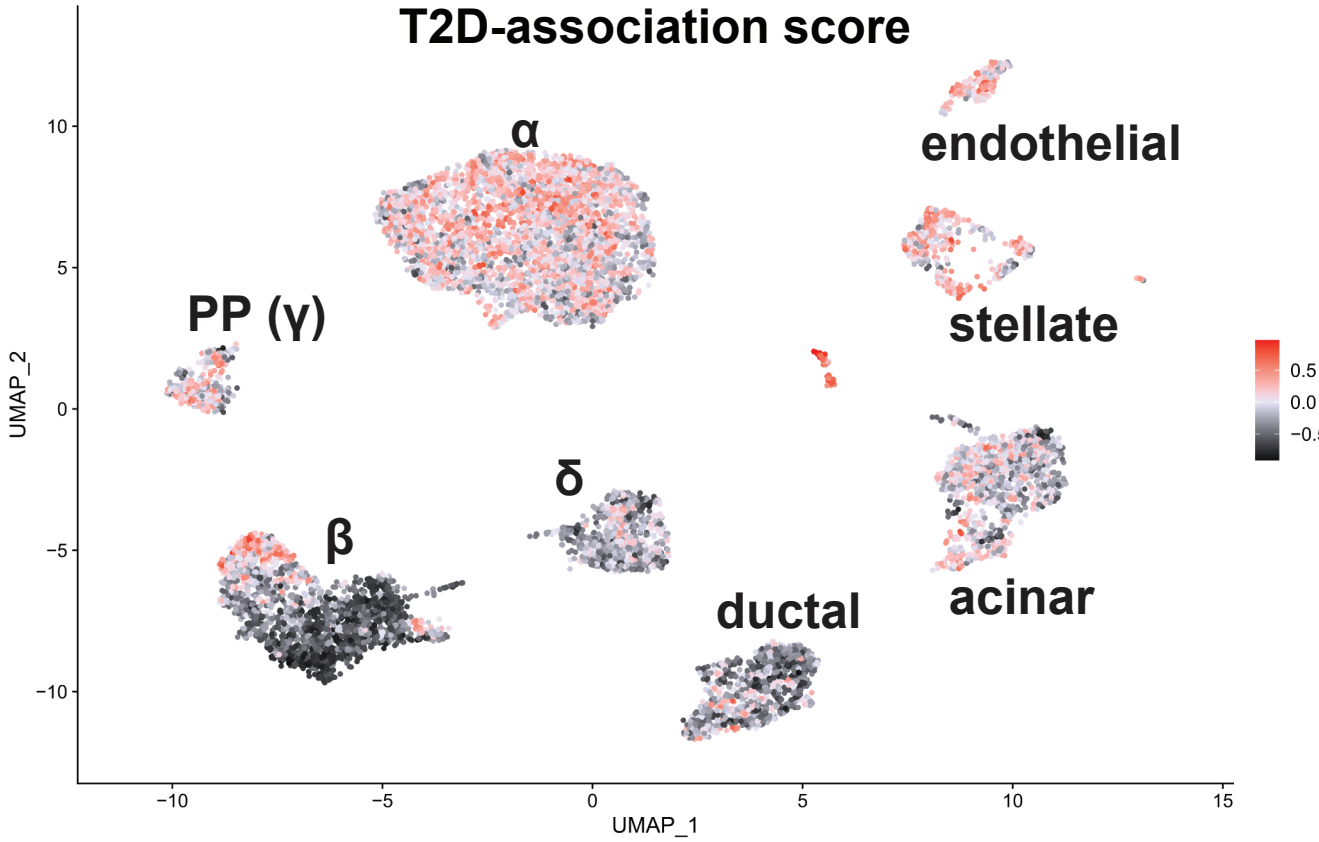

A
INS
GCG
SST
PPY
GHRL
KRT19
PRSS1
REG1A
CFTR
COL6A1
PDGFRA
PLVAP
CD44
KIT
UMAP_1
UMAP_2
B
Sex
female
male
C
BMI
D
Age
E
Disease
diabetes
healthy

Figure 3

A


DEGAS:
T2D-association score
α
endothelial
PP (γ)
stellate
δ
β
ductal
acinar
0.5
0.0
−0.5
UMAP_2
UMAP_1
10
5
0
−5
−10
15

B

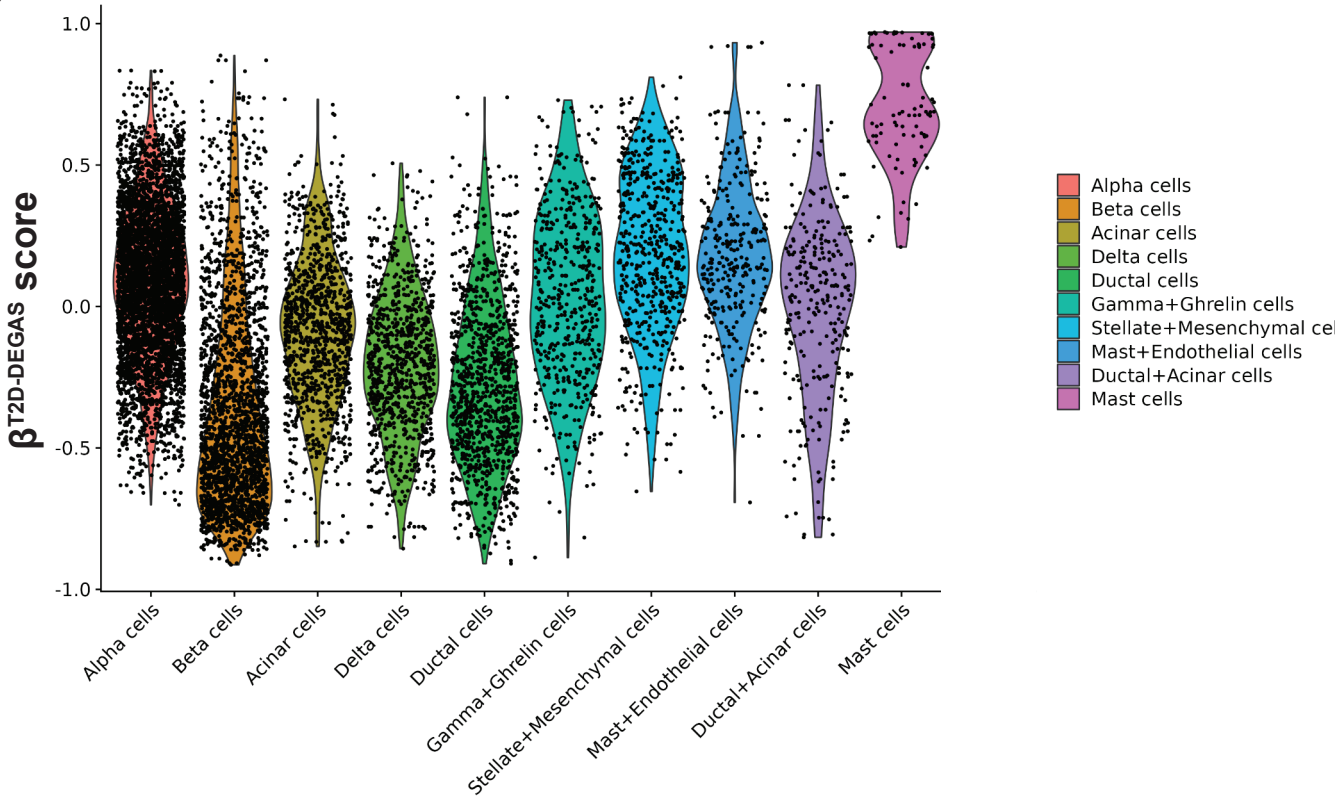

β T2D-DEGAS score
1.0
0.5
0.0
-0.5
-1.0
Alpha cells
Beta cells
Acinar cells
Delta cells
Ductal cells
Gamma+Ghrelin cells
Stellate+Mesenchymal cells
Mast+Endothelial cells
Ductal+Acinar cells
Mast cells

Figure 4

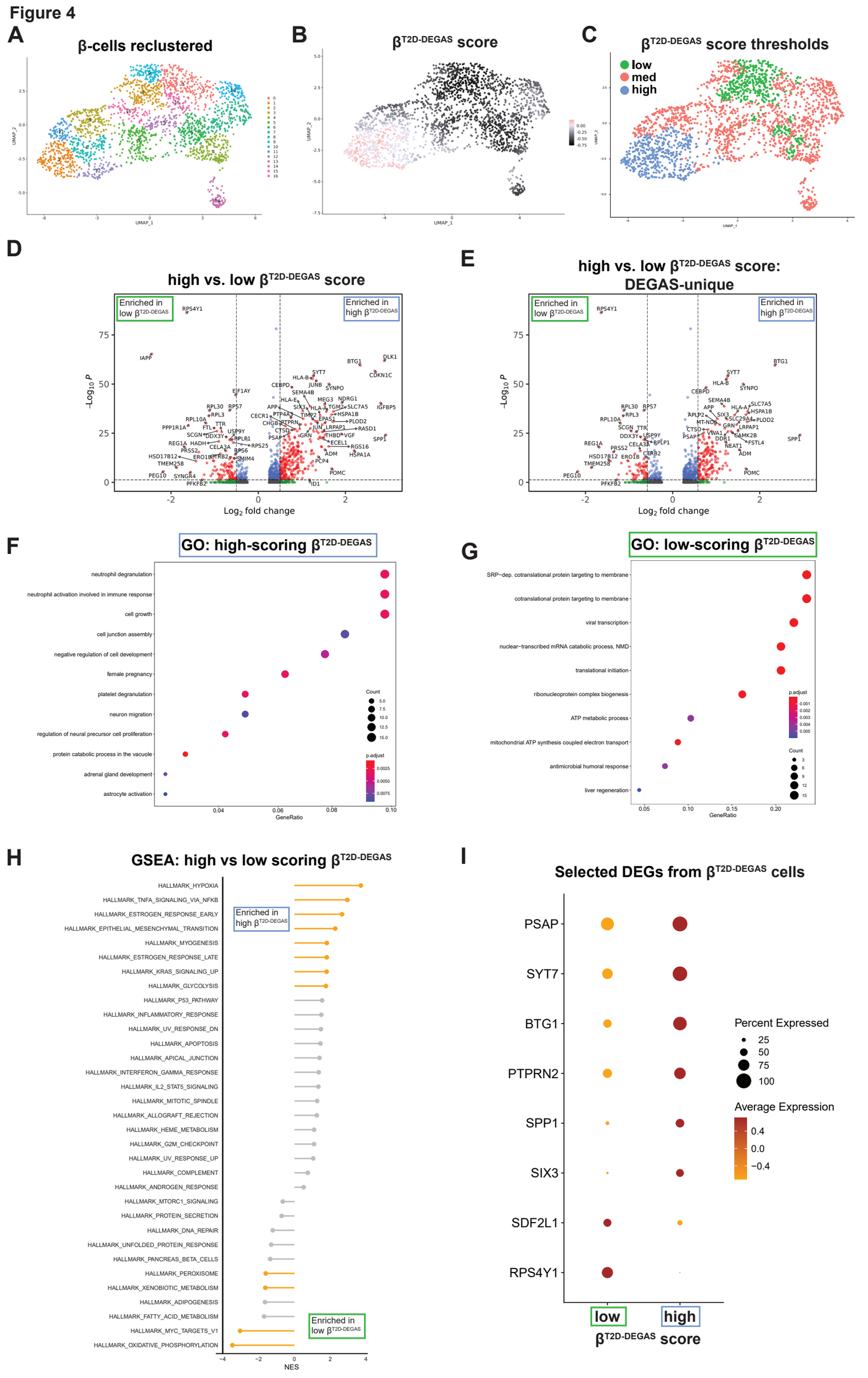
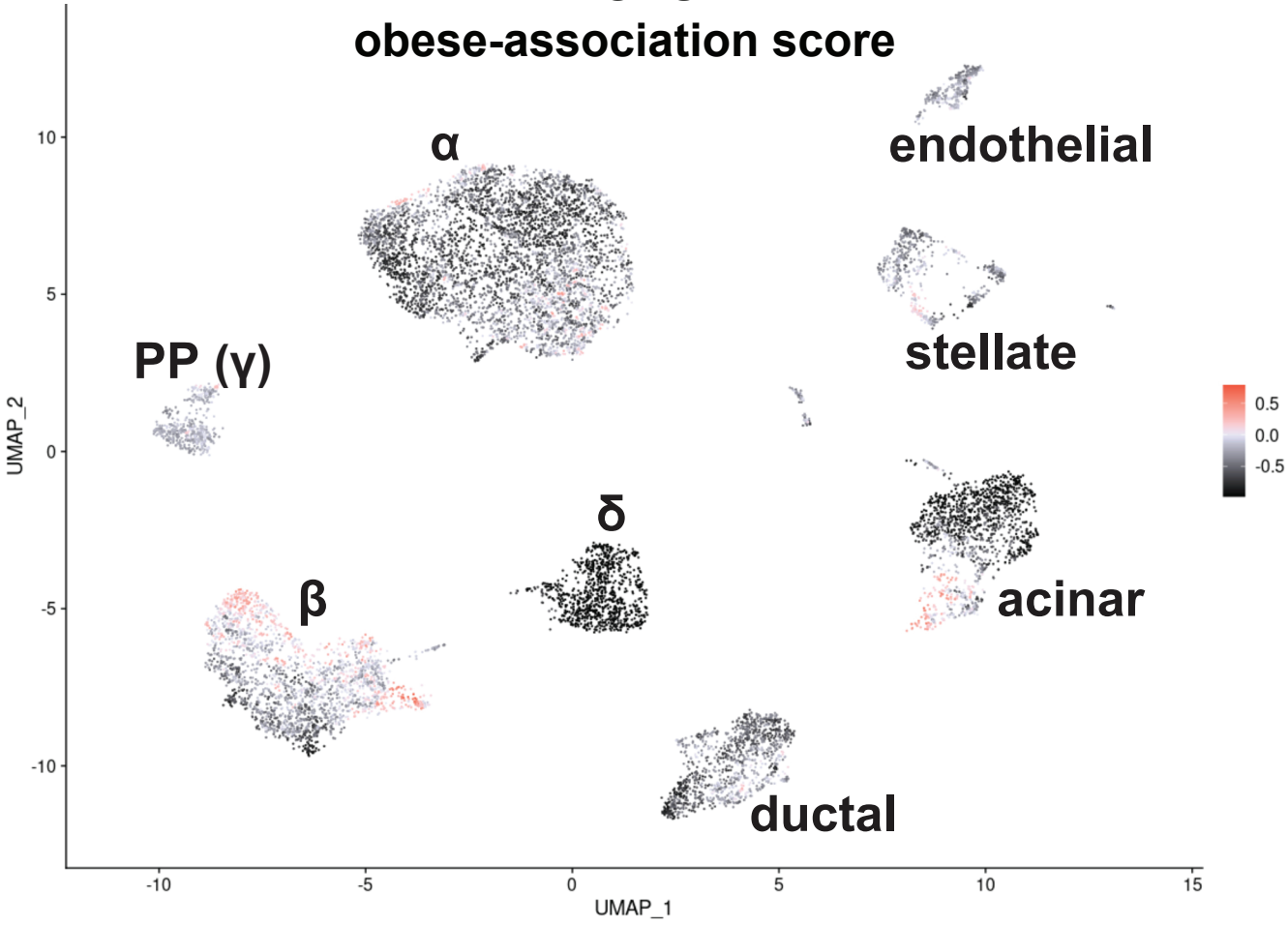

A
β-cells reclustered
UMAP_1
UMAP_2
B
βT2D-DEGAS score
C
βT2D-DEGAS score thresholds
low
med
high
D
high vs. low βT2D-DEGAS score
Enriched in low βT2D-DEGAS
Enriched in high βT2D-DEGAS
−Log10 P
Log2 fold change
E
high vs. low βT2D-DEGAS score: DEGAS-unique
F
GO: high-scoring βT2D-DEGAS
neutrophil degranulation
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
G
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
H
GSEA: high vs low scoring 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
Enriched in high βT2D-DEGAS
Enriched in low βT2D-DEGAS
NES
I
Selected DEGs from βT2D-DEGAS cells
PSAP
SYT7
BTG1
PTPRN2
SPP1
SIX3
SDF2L1
RPS4Y1
Percent Expressed
25
50
75
100
Average Expression
0.4
0.0
−0.4
low
high
βT2D-DEGAS score

Figure 5

A


DEGAS:
obese-association score
α
endothelial
PP (γ)
stellate
δ
β
acinar
ductal
UMAP_2
UMAP_1
10
5
0
-5
-10
-10
-5
0
5
10
15
0.5
0.0
-0.5

B

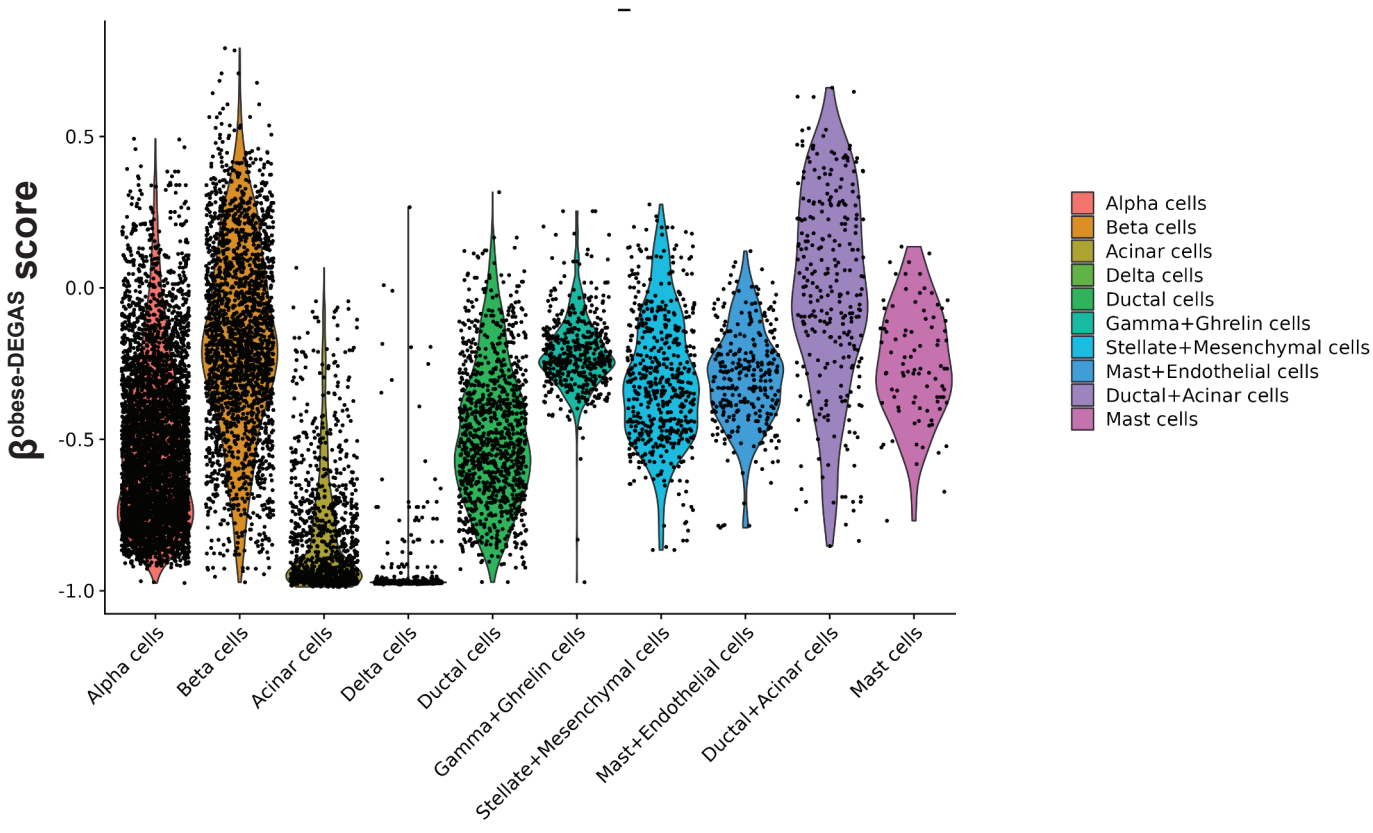

β obese-DEGAS score
0.5
0.0
-0.5
-1.0
Alpha cells
Beta cells
Acinar cells
Delta cells
Ductal cells
Gamma+Ghrelin cells
Stellate+Mesenchymal cells
Mast+Endothelial cells
Ductal+Acinar cells
Mast cells

Figure 6

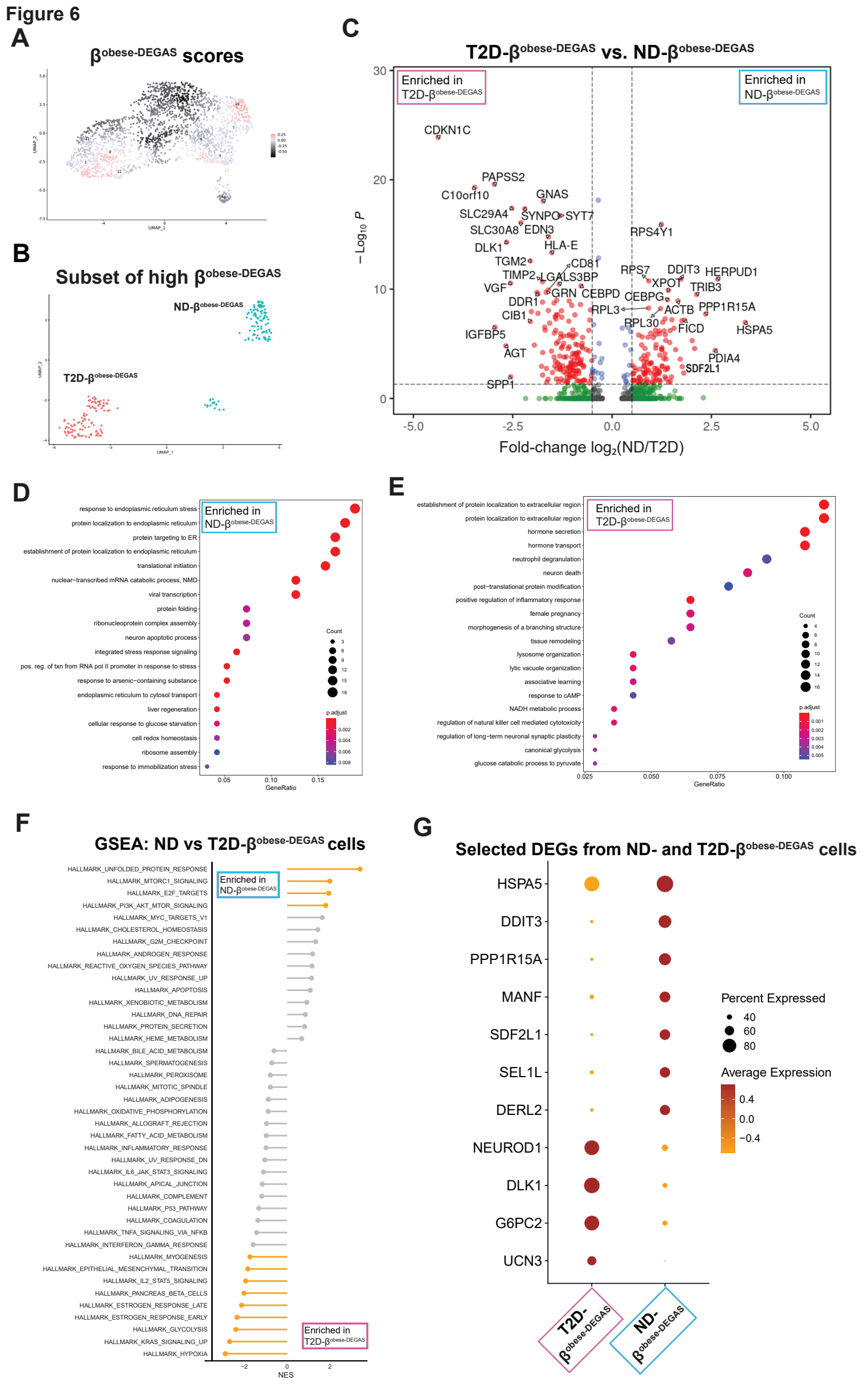

A
βobese-DEGAS scores
UMAP_1
UMAP_2
B
Subset of high βobese-DEGAS
ND-βobese-DEGAS
T2D-βobese-DEGAS
C
T2D-βobese-DEGAS vs. ND-βobese-DEGAS
Enriched in T2D-βobese-DEGAS
Enriched in ND-βobese-DEGAS
CDKN1C
PAPSS2
C10orf10
GNAS
SLC29A4
SYNPO
SYT7
SLC30A8
EDN3
RPS4Y1
DLK1
HLA-E
TGM2
CD81
TIMP2
LGALS3BP
RPS7
DDIT3
HERPUD1
XPOT
TRIB3
VGF
GRN
CEBPD
CEBPG
DDR1
RPL3
ACTB
PPP1R15A
CIB1
RPL30
FICD
HSPA5
IGFBP5
AGT
PDIA4
SDF2L1
SPP1
−Log10 P
Fold-change log2(ND/T2D)
D
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
E
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 natural killer cell mediated cytotoxicity
regulation of long−term neuronal synaptic plasticity
canonical glycolysis
glucose catabolic process to pyruvate
Count
p.adjust
GeneRatio
F
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
G
Selected DEGs from ND- and T2D-βobese-DEGAS cells
HSPA5
DDIT3
PPP1R15A
MANF
SDF2L1
SEL1L
DERL2
NEUROD1
DLK1
G6PC2
UCN3
Percent Expressed
40
60
80
Average Expression
0.4
0.0
−0.4
T2D-βobese-DEGAS
ND-βobese-DEGAS

Figure 7

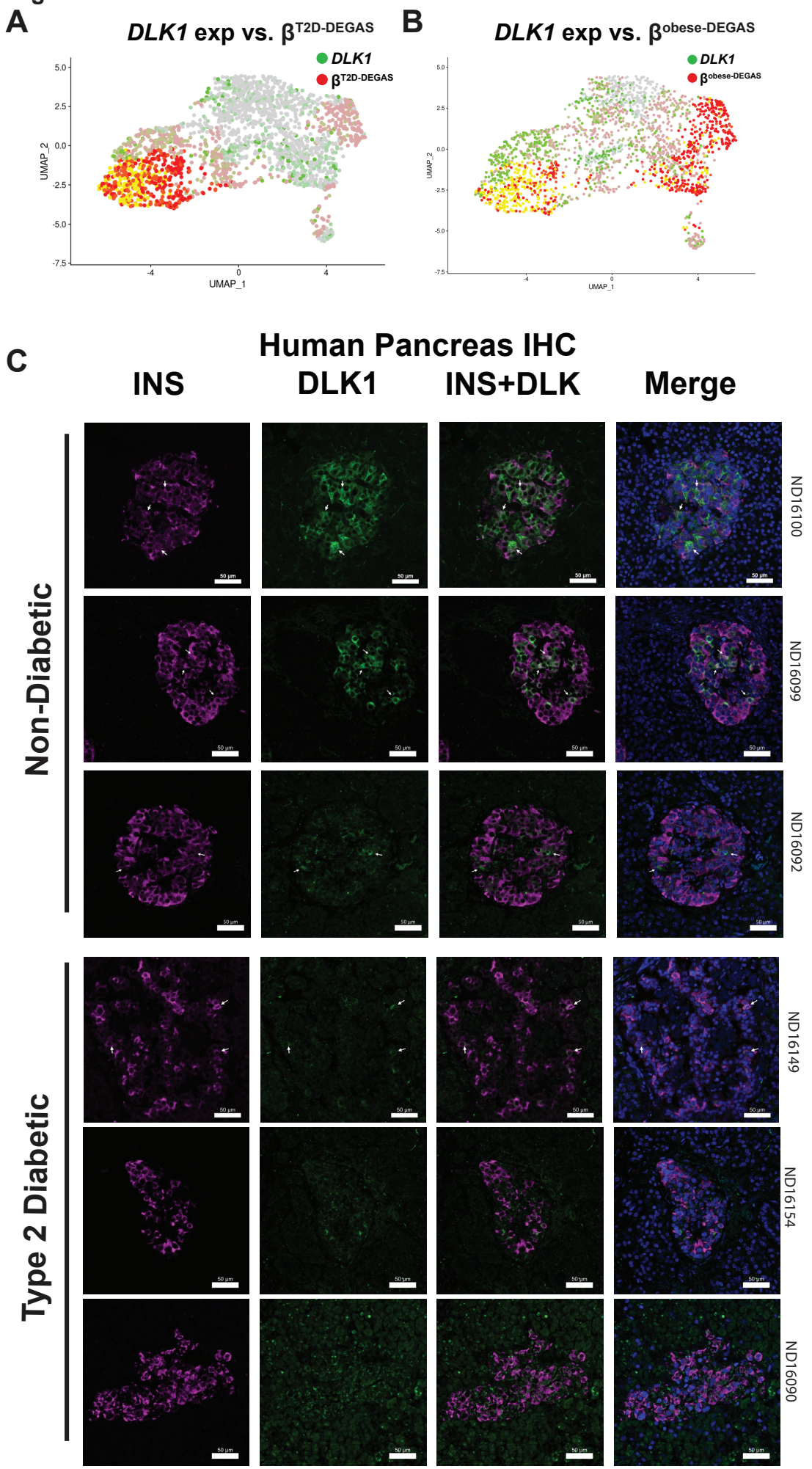

A
DLK1 exp vs. βT2D-DEGAS
DLK1
βT2D-DEGAS
UMAP_2
UMAP_1
B
DLK1 exp vs. βobese-DEGAS
DLK1
βobese-DEGAS
UMAP_2
UMAP_1
C
Human Pancreas IHC
INS
DLK1
INS+DLK
Merge
Non-Diabetic
Type 2 Diabetic
ND16100
ND16099
ND16092
ND16149
ND16154
ND16090
50 µm

Figure S1

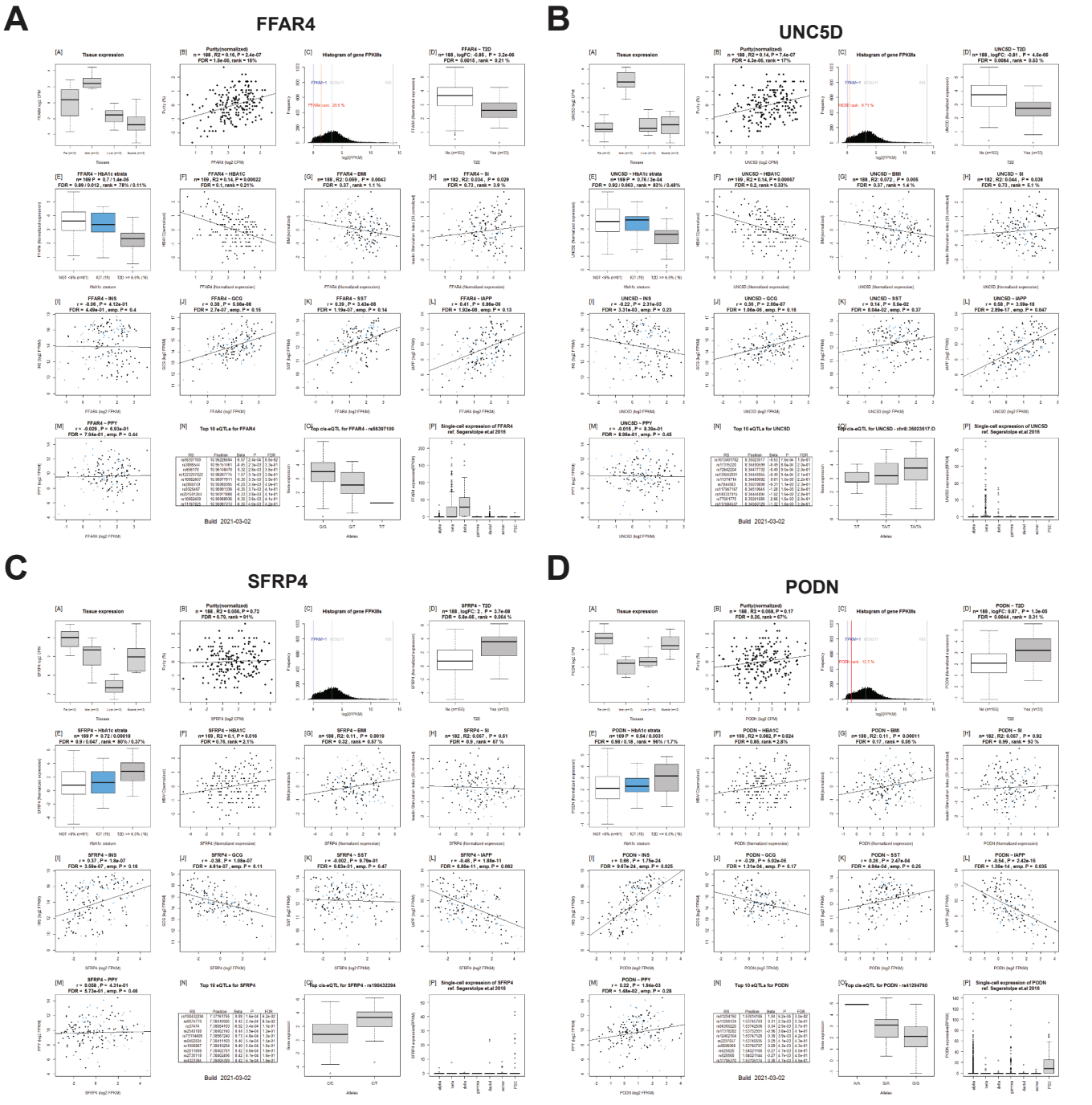


Figure S2

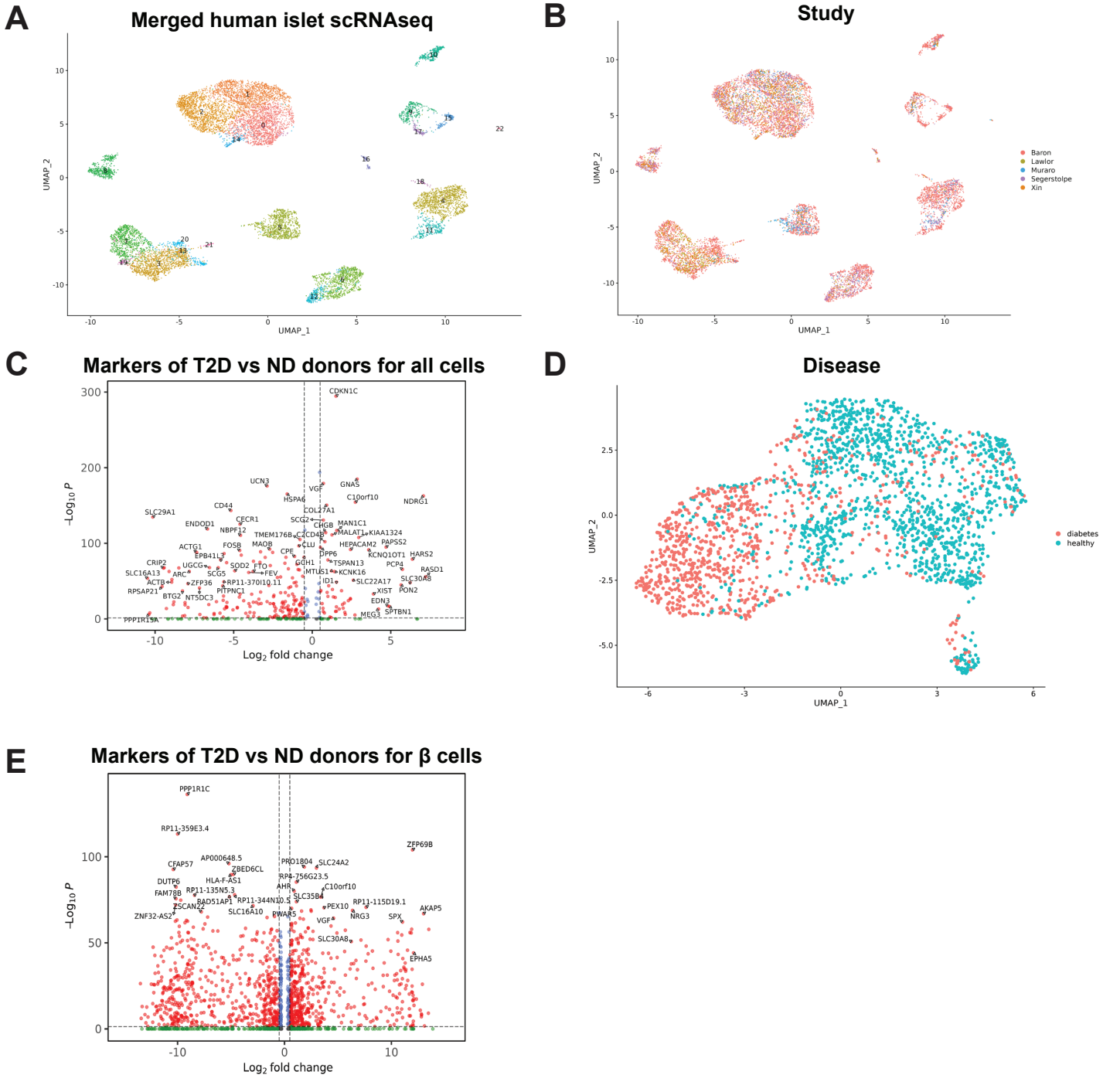

A
Merged human islet scRNAseq
UMAP_1
UMAP_2
B
Study
Baron
Lawlor
Muraro
Segerstolpe
Xin
C
Markers of T2D vs ND donors for all cells
Log2 fold change
−Log10 P
CDKN1C
UCN3
VGF
GNAS
HSPA6
C10orf10
NDRG1
SLC29A1
CD44
COL27A1
CECR1
SCG2
CHGB
MAN1C1
ENDOD1
NBPF12
TMEM176B
C2CD4B
MALAT1
KIAA1324
MAOB
CLU
HEPACAM2
PAPSS2
ACTG1
FOSB
CPE
DPP6
KCNQ1OT1
HARS2
EPB41L3
CRIP2
UGCG
SOD2
FTO
GCH1
TSPAN13
PCP4
RASD1
SLC16A13
ARC
SCG5
FEV
MTUS1
KCNK16
SLC30A8
ACTB
ZFP36
RP11-370I10.11
ID1
SLC22A17
XIST
PON2
RPSAP21
PITPNC1
BTG2
NT5DC3
EDN3
SPTBN1
MEG3
PPP1R15A
D
Disease
diabetes
healthy
UMAP_1
UMAP_2
E
Markers of T2D vs ND donors for β cells
Log2 fold change
−Log10 P
PPP1R1C
RP11-359E3.4
ZFP69B
AP000648.5
PRO1804
SLC24A2
CFAP57
ZBED6CL
RP4-756G23.5
DUTP6
HLA-F-AS1
AHR
C10orf10
FAM78B
RP11-135N5.3
SLC35B4
RAD51AP1
RP11-344N10.5
RP11-115D19.1
AKAP5
ZSCAN22
SLC16A10
PEX10
ZNF32-AS2
PWAR5
NRG3
SPX
VGF
SLC30A8
EPHA5

Figure S3

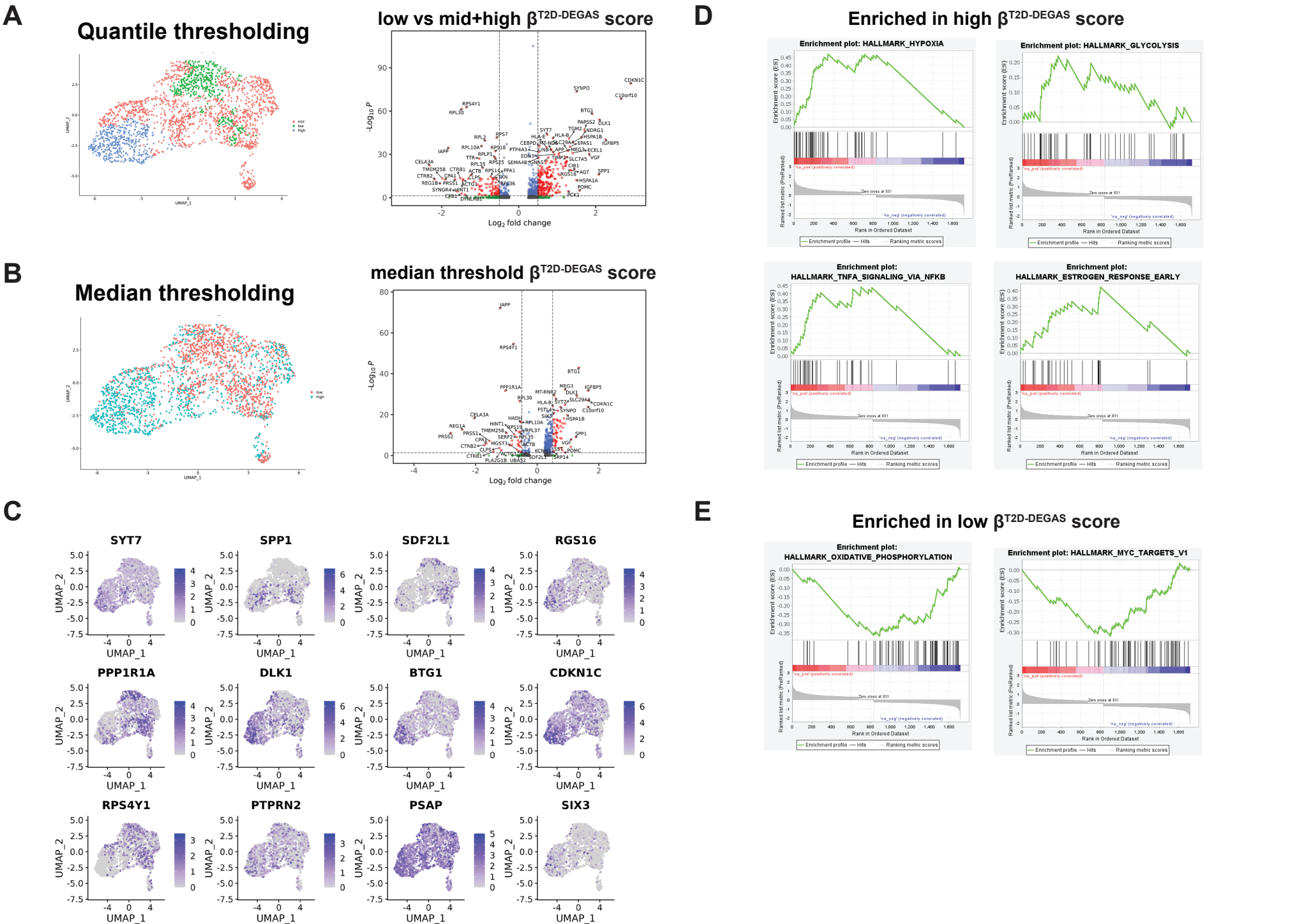

A
Quantile thresholding
low vs mid+high β T2D-DEGAS score
B
Median thresholding
median threshold β T2D-DEGAS score
C
SYT7
SPP1
SDF2L1
RGS16
PPP1R1A
DLK1
BTG1
CDKN1C
RPS4Y1
PTPRN2
PSAP
SIX3
D
Enriched in high β T2D-DEGAS score
Enrichment plot: HALLMARK_HYPOXIA
Enrichment plot: HALLMARK_GLYCOLYSIS
Enrichment plot: HALLMARK_TNFA_SIGNALING_VIA_NFKB
Enrichment plot: HALLMARK_ESTROGEN_RESPONSE_EARLY
E
Enriched in low β T2D-DEGAS score
Enrichment plot: HALLMARK_OXIDATIVE_PHOSPHORYLATION
Enrichment plot: HALLMARK_MYC_TARGETS_V1

Figure S4

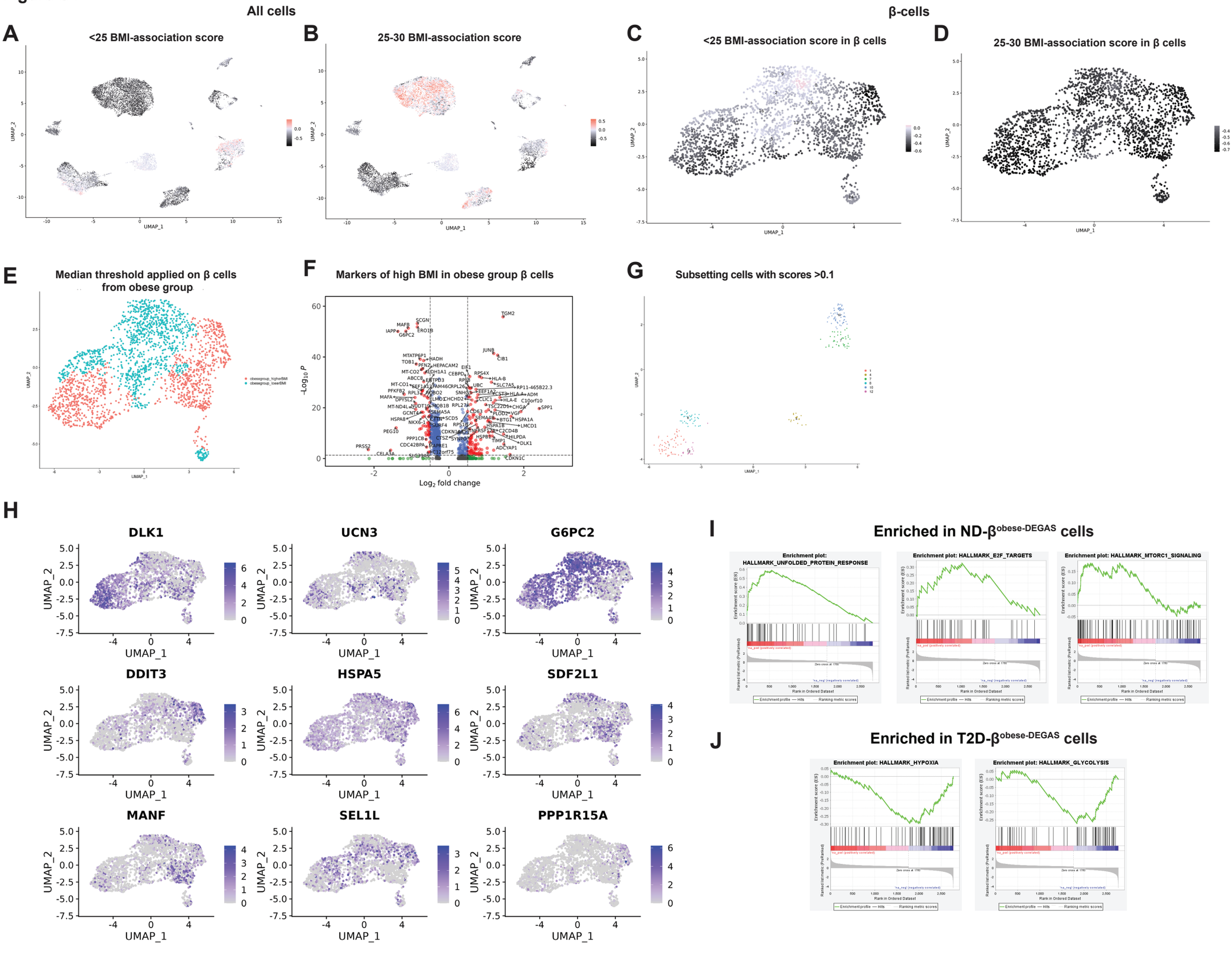
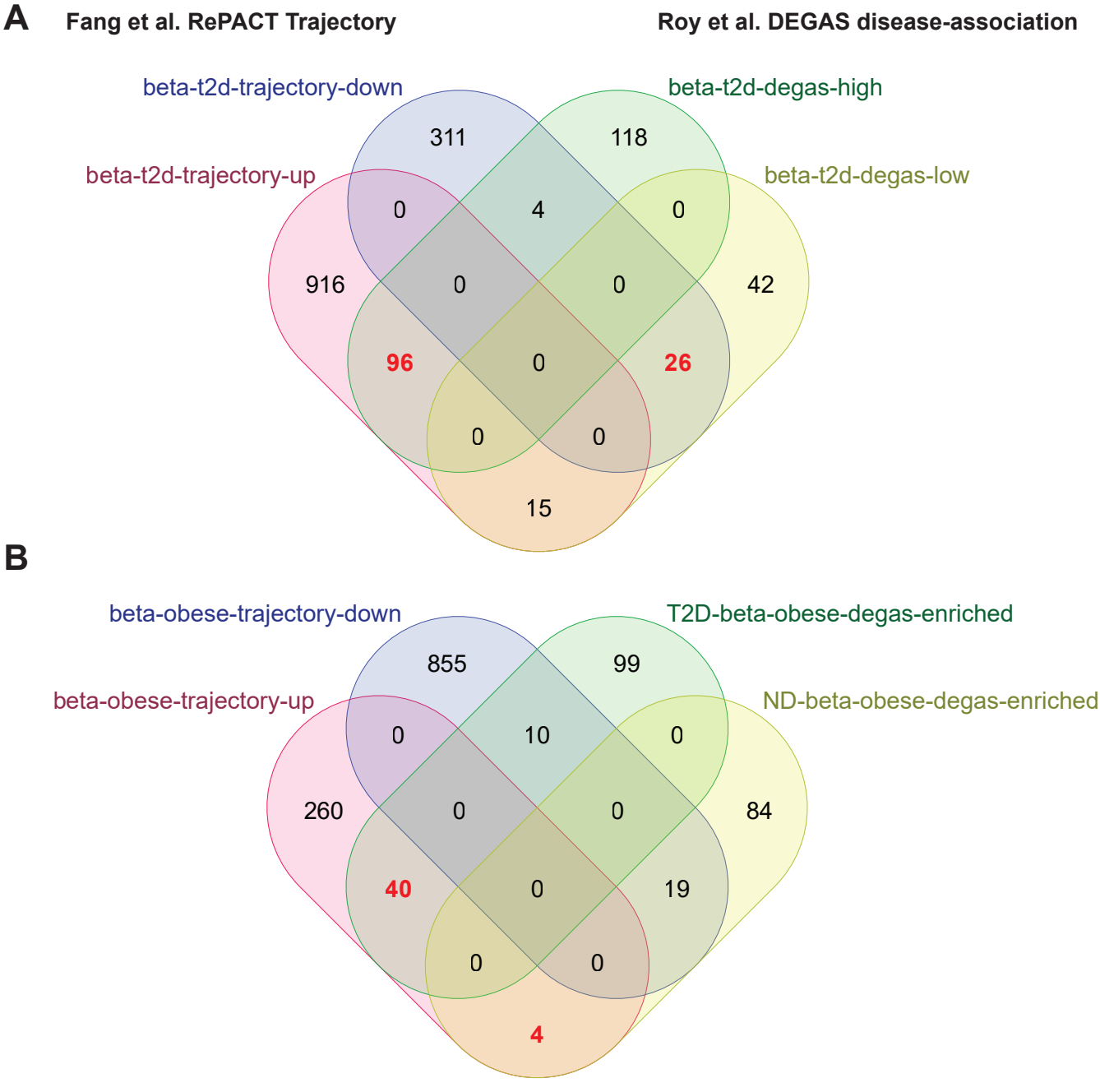

All cells
β-cells
A
<25 BMI-association score
B
25-30 BMI-association score
C
<25 BMI-association score in β cells
D
25-30 BMI-association score in β cells
E
Median threshold applied on β cells from obese group
F
Markers of high BMI in obese group β cells
G
Subsetting cells with scores >0.1
H
DLK1
UCN3
G6PC2
DDIT3
HSPA5
SDF2L1
MANF
SEL1L
PPP1R15A
UMAP_1
UMAP_2
I
Enriched in ND-β obese-DEGAS cells
Enrichment plot: HALLMARK_UNFOLDED_PROTEIN_RESPONSE
Enrichment plot: HALLMARK_E2F_TARGETS
Enrichment plot: HALLMARK_MTORC1_SIGNALING
J
Enriched in T2D-β obese-DEGAS cells
Enrichment plot: HALLMARK_HYPOXIA
Enrichment plot: HALLMARK_GLYCOLYSIS

Figure S5


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