Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Single-cell and pseudobulk transcriptomic profiling reveals immune-enriched disease- associated signatures in human pancreatic islets from type 2 diabetes

Mohammad Zahedi
Tehran University of Medical Sciences
Amrita Kakavand Hamidi
Tehran University of Medical Sciences
Farideh Jafari-Raddani
Tehran University of Medical Sciences
Maryam Amini Shirkouhi
Tehran University of Medical Sciences
Mahsa M. Amoli
Tehran University of Medical Sciences
Farideh Razi
f-razi@tums.ac.ir

Tehran University of Medical Sciences

Research Article

Keywords: Type 2 diabetes, human pancreatic islets, single-cell RNA sequencing, immune-enriched cells

Posted Date: June 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10040418/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Type 2 diabetes (T2D) is a multifactorial metabolic disorder characterized by impaired glucose homeostasis, insulin resistance, and progressive dysfunction of pancreatic islet cells. Given the cellular heterogeneity of human pancreatic islets, bulk transcriptomic approaches may obscure cell type-specific disease signals. Single-cell RNA sequencing provides an opportunity to resolve these alterations at higher resolution.

Method

We reanalyzed publicly available single-cell RNA-seq data from human pancreatic islets obtained from non-diabetic and T2D donors. After stringent quality control, integrated clustering, and marker-based annotation, we performed cell type-specific differential expression analysis and donor-aware pseudobulk analysis. Functional enrichment analysis was also conducted to interpret significant gene lists.

Results

Integrated analysis revealed substantial cellular heterogeneity in the pancreatic islet compartment, with no clear global separation between T2D and non-diabetic samples. Cell-level differential expression analysis did not identify robust disease-associated genes after multiple-testing correction in major cell populations. In pseudobulk analysis, the immune-enriched compartment showed an exploratory signal, with CXCL3 meeting the FDR < 0.05 and |log2FC| > 2 criterion, but not the more stringent FDR < 0.01 threshold. Functional enrichment analysis indicated signatures related to extracellular matrix, secreted proteins, and protein processing.

Conclusion

These findings suggest that transcriptional alterations in T2D human islets are subtle, compartment-specific, and best detected through donor-aware cell type-resolved analysis. The results support a model in which disease-related changes are distributed across specific cellular niches rather than reflected in a broad global transcriptomic shift.

Introduction

Type 2 diabetes (T2D) is a complex and multifactorial metabolic disease characterized by impaired glucose homeostasis, insulin resistance, and progressive dysfunction of pancreatic islet cells. Its global burden has continued to increase over recent decades, making diabetes one of the major public health challenges worldwide (1–3). This burden is particularly relevant in the Middle East and North Africa region, where diabetes prevalence is among the highest globally (4). In Iran, national STEPS surveys have similarly shown

a marked increase in diabetes prevalence among adults aged 25–65 years, rising from 8.4% in 2004 to 13.2% in 2016. These trends highlight the need for a deeper understanding of the molecular and cellular mechanisms underlying T2D development and progression (5).

Pancreatic islets are composed of multiple endocrine and non-endocrine cell populations, including beta, alpha, delta, pancreatic polypeptide, acinar, ductal, stromal, endothelial, and immune-associated cells (6). Because these populations have distinct transcriptional identities and biological functions, disease-associated changes may not be uniformly distributed across the whole islet. Instead, T2D-related alterations may be restricted to specific cell types or cellular states Therefore, bulk transcriptomic analysis of whole-islet preparations may mask important cell-type-specific signals, especially when changes are subtle or confined to less abundant cell populations.

Single-cell RNA sequencing offers an opportunity to resolve cellular heterogeneity in human pancreatic islets and examine transcriptional alterations at cell-type resolution (6). Previous human islet single-cell studies have identified distinct islet cell signatures and demonstrated that diabetes-associated gene expression changes can be cell-type dependent. More recent single-cell and multiomic studies have further underscored the importance of beta-cell heterogeneity, islet microenvironmental changes, and donor-level variation in interpreting T2D transcriptomic data. However, because public human islet datasets differ in donor composition, sequencing depth, quality-control strategy, and analytical design, careful reanalysis using integrated and donor-aware approaches remains valuable (7, 8).

In the present study, we reanalyzed publicly available single-cell RNA-seq data from human pancreatic islets obtained from non-diabetic and T2D donors. The study was designed to address the following question: how does T2D alter the cellular landscape and cell-type-specific transcriptional programs of human pancreatic islets at single-cell resolution? To answer this, we performed stringent quality control, integrated clustering, marker-based cell population annotation, cell-type-specific differential expression testing, donor-aware pseudobulk analysis, and functional enrichment analysis. Through this workflow, we aimed to build an integrated single-cell landscape of human islets, define transcriptionally distinct islet and non-islet cell populations, and identify candidate cell-type-specific transcriptional alterations associated with T2D.

Materials and Methods

Data source and study design

This study was conducted as a secondary analysis of publicly available single-cell RNA sequencing data from human pancreatic islets. The dataset was obtained from the Gene Expression Omnibus under accession number GSE221156. This dataset was originally generated to investigate cell-type-specific transcriptomic alterations in human pancreatic islets across non-diabetic, pre-diabetic, and type 2 diabetic states. In the present study, the analysis was restricted to non-diabetic and type 2 diabetic samples in order to evaluate disease-associated transcriptional differences in pancreatic islet cell populations. Raw count matrices, cell barcode files, and gene feature files were used as input for downstream analysis. Sample-level metadata were manually curated to assign each GSM accession to its corresponding islet sample identifier and disease group. Based on this classification, 37 samples were included in the analysis, comprising 20

type 2 diabetic and 17 non-diabetic islet samples. All downstream analyses were performed in R using Seurat-based single-cell RNA-seq workflows. After stringent filtering, 115404 high-quality cells were retained, including 66951T2D and 48453non-diabetic cells (Retention percentage: 44.44%).

Single-cell data preprocessing and quality control

Raw gene expression matrices were processed in R using the Seurat package. For each sample, the matrix file, cell barcode file, and gene feature file were imported, and an individual Seurat object was generated using CreateSeuratObject, retaining genes detected in at least three cells and cells with at least 200 detected features. Sample-level metadata, including GEO accession number, islet sample identifier, and disease status, were added to each object. The percentage of mitochondrial gene expression was calculated using genes matching the ^MT- pattern.

Quality control was performed independently for each sample based on the number of detected genes, total RNA counts, and mitochondrial gene percentage. Cells were retained if they had 500–10,000 detected genes, 1,000–50,000 UMIs, and <10% mitochondrial content. The filtered sample-level objects were subsequently merged into a single Seurat object for downstream analysis. Quality-control metrics were visualized across samples and disease groups to assess data distribution before downstream processing.

Normalization, dimensionality reduction, and data integration

After quality control, filtered cells from all samples were merged into a single Seurat object for downstream analysis. Gene expression data were normalized using the NormalizeData function, and the 2,000 most variable genes were identified using the variance-stabilizing transformation method implemented in FindVariableFeatures. The data were then scaled with ScaleData, and principal component analysis (PCA) was performed using RunPCA with 50 principal components. PCA plots were generated to assess the distribution of cells according to sample identity and disease status.

To provide a harmonized representation of cells from different donors, aligned based on shared biology rather than batch effects, thereby facilitating joint clustering of T2D and ND cells; the RNA assay was split by sample identity and integrated using Seurat v5 IntegrateLayers with canonical correlation analysis-based integration corrected for inter-sample technical variation. The resulting integrated dimensional reduction, referred to as integrated.cca, was used for downstream clustering and visualization. Seurat v5 integration is designed to generate a shared low-dimensional representation that captures common biological variation across multiple layers and can subsequently be used for visualization and unsupervised clustering.

Clustering and visualization

Cell clustering was performed using the integrated CCA reduction. A shared nearest-neighbor graph was constructed with FindNeighbors using the first 30 integrated dimensions, followed by graph-based clustering with FindClusters at a resolution of 1.0. Two-dimensional visualization of the integrated single-cell transcriptomic landscape was performed using uniform manifold approximation and projection and t-distributed stochastic neighbor embedding, implemented with RunUMAP and RunTSNE, respectively. The

resulting embeddings were used to visualize the global cell landscape and demonstrate howto disease status, and donors are distributed across Seurat-defined clusters.

Cell-type marker-basedannotation

Cell-type identities were assessed using canonical marker genes for major pancreatic islet and non-islet cell populations. Endocrine cell populations were evaluated using INS, IAPP, PCSK1, and PCSK2 for beta cells; GCG, TTR, and ARX for alpha cells; SST for delta cells; GHRL for epsilon cells; and PPY for pancreatic polypeptide cells. Non-endocrine and contaminating cell populations were assessed using PRSS1, CPA1, and REG1A for acinar cells; KRT19 and CFTR for ductal cells; COL1A1, DCN, and RGS5 for mesenchymal/stromal cells; and PTPRC, TRAC, CD79A, NKG7, CD14, LYZ, CST3, and MS4A7 for immune cell populations. Marker expression patterns were visualized using feature plots (identifying the spatial location of major cells in the global embedding, which supports cluster-level cell type annotation), violin plots, dot plots, and heatmaps across Seurat-defined clusters.

Cluster-enriched genes were identified using FindAllMarkers with the Wilcoxon rank-sum test. Marker genes were initially detected using a log-fold change threshold of 0.2, a minimum expression fraction of 0.1, and a minimum difference in expression fraction of 0.2. For downstream visualization and prioritization, cluster markers were further filtered using an absolute average $\log_2$ fold-change threshold of at least 2 and an adjusted p-value threshold of less than or equal to 0.01. The top 10 marker genes per cluster were selected based on adjusted p-value and visualized using heatmaps and dot plots. The combined evaluation of canonical marker expression and cluster-specific marker profiles guided final cluster annotation. This provides cluster-specific gene expression signatures and supports the assignment of biological identities to each cluster.

Cell-type–specific subsetting

Cluster-to-cell type mapping, meaning mapping selected clusters to biological cell types, was done to allow subsequent analyses to be performed at the level of biologically meaningful cell types rather than pure unsupervised cluster IDs. Their distribution by sample and group, and re-identified variable features demonstrate that each cluster represents which classical cells and visualize potential differences in each cell marker expression between T2D and ND donors.

Cell-type-specific differential expression gene analysis

To identify disease-associated transcriptional alterations within specific pancreatic cell populations, differential expression analysis was performed between type 2 diabetic and non-diabetic cells within annotated cell-type subsets. For each selected cell population, cell identities were set according to disease status, and differentially expressed genes were identified using the `FindMarkers` function in Seurat with the Wilcoxon rank-sum test. Genes were tested using a log-fold change threshold of 0.2, a minimum expression fraction of 0.1, and a minimum difference in expression fraction of 0.2. Both upregulated and downregulated genes were retained by setting `only.pos = FALSE`. Differential expression results were

exported for downstream prioritization, and the most significant genes were visualized using violin plots across disease groups within specific cell types.

Pseudobulk differential expression gene analysis modeling

To account for biological replication and avoid treating individual cells from the same donor as independent observations, a pseudobulk differential expression strategy was additionally implemented. Cell-type labels were assigned based on the final cluster-to-cell-type mapping, and cells were grouped by sample identity, disease status, and cell type. For each cell type, raw RNA counts were aggregated at the sample level using `AggregateExpression`, generating one pseudobulk expression profile per sample and disease group. Pseudobulk count matrices were analyzed using DESeq2. For each cell type, genes with low total counts were removed by retaining genes with a summed count greater than 10 across pseudobulk samples. Differential expression between type 2 diabetic and non-diabetic samples was then tested using a negative binomial generalized linear model with disease group as the main variable in the design formula (`~ group`). Results were extracted using the contrast T2D versus ND, and cell-type-specific pseudobulk results were exported separately and combined for downstream interpretation. The adjusted p-value less than or equal to 0.01 and $\log_2$FC of at least 2 were set as the significant threshold cut-off.

Marker Overrepresentation Enrichment Analysis by DAVID Bioinformatics

Markers from each step, including the top 10 markers per cluster, cell type-specific, and pseudobulk significant DEGs according to each step threshold cut-off, were submitted to DAVID Bioinformatics (B.T. Sherman, M. Hao, J. Qiu, X. Jiao, M.W. Baseler, H.C. Lane, T. Imamichi and W. Chang. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Research. 23 March 2022. doi:10.1093/nar/gkac194.[PubMed]/// Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57. [PubMed]) to explore the related enriched terms and functional annotations.

Ethics statement

This study was performed as part of an approved research project at the Endocrinology and Metabolism Research Institute, Tehran University of Medical Sciences, Tehran, Iran. The study protocol was approved by the institutional ethics committee under approval code **IR.TUMS.EMRI.REC.1404.025**.

All analyses in the present study were conducted using publicly available, de-identified single-cell RNA-seq data retrieved from the Gene Expression Omnibus under accession number **GSE221156**. Since no new human subjects were recruited and no additional biological samples were collected, additional informed consent was not required for this secondary analysis. Data use complied with the terms of the public repository and the ethical approval obtained for the approved research project.

Statistical analysis and visualization

All single-cell analyses were performed in R. Cluster marker detection and cell-type-specific differential gene expression analyses were conducted using Wilcoxon rank-sum tests implemented in Seurat. Marker genes

were prioritized based on log-fold change, expression fraction, and adjusted p-values, as specified in each analysis step. For pseudobulk analysis, aggregated raw count matrices were analyzed using DESeq2, which models count data using negative binomial generalized linear models. Differential expression results were reported as $\log_2$ fold changes with corresponding adjusted p-valuesData visualization was performed using Seurat and ggplot2-based functions, including dimensional reduction plots, feature plots, violin plots, dot plots, and heatmaps. UMAP and t-SNE embeddings were used for visualization of cellular distributions across samples, disease groups, and clusters, whereas feature-level expression patterns were visualized using marker gene expression across annotated cell populations. Seurat provides dedicated visualization functions such as `FeaturePlot`, `VlnPlot`, `DotPlot`, and `DoHeatmap` for single-cell expression analysis.

Results

Quality control assessment of single-cell RNA-seq data

Initial quality-control assessment identified 259,709 cells across 37 pancreatic islet samples, including 20 T2D and 17 non-diabetic samples. The dataset contained 134,942 cells from T2D samples and 124,767 cells from non-diabetic samples. The median number of cells per sample was 6,741, with a range of 3,798 to 10,468. At the sample level, the median number of detected genes per cell ranged from 1,729 to 4,008.5, while median RNA counts ranged from 5,147 to 18,786. The overall median nFeature_RNA and nCount_RNA values were 2,544.5 and 11,202.5, respectively. Median mitochondrial gene percent ranged from 3.15% to 19.78% across samples, indicating sample-level variability in mitochondrial content and supporting the application of downstream cell-level QC filtering. Overall, both T2D and non-diabetic groups showed adequate cell recovery and transcriptomic complexity for subsequent integrated single-cell analysis (Table 1).

Table 1
Summary of sample-level quality control metrics before and after stringent cell filtering.

Metric	Overall	T2D	ND
	Before QC/After QC	Before QC/After QC	Before QC/After QC
Number of samples	37/37	20/20	17/17
Total cells	259,709/115404	134,942/66951	124,767/48453
Median cells per sample	6,741/2734	6,552/3105.5	7,961/2497
Cell range per sample	3,798–10,468/1026–6917	3,943–9,637/1104–6374	3,798–10,468/1026–6917
Median detected genes per cell	2,544.5/2948	2,569.25/3064.25	2,474/2776
Range of sample-level median detected genes	1,729–4,008.5/1503–4062	1,729–3,400/2059-3681.5	1,854.5–4,008.5/1503–4062
Median RNA counts per cell	11,202.5/12463	11,317.25/13163	10,850.5/10553
Range of sample-level median RNA counts	5,147–18,786/5005-17136.5	5,147–15,669/6541.5-16232	6,940–18,786/5005-17136.5
Median mitochondrial percentage	11.36%/5.96%	9.78%/5.92%	12.59%/6.06%
Range of sample-level median mitochondrial percentage	3.15–19.78%/3.02–7.63%	3.43–17.19%/3.3–7.63%	3.15–19.78%/3.02–7.63%

Stringent quality control filtering retained high-quality single-cell profiles

Following the initial sample-level quality assessment, stringent cell-level filtering was applied to remove low-quality cells and potential technical outliers. Cells were retained if they contained 500–10,000 detected genes, 1,000–50,000 RNA counts, and less than 10% mitochondrial transcript content, as described in the Methods. Because before QC, ND had a higher mitochondrial fraction (≈ 12.6%) than T2D (≈ 9.8%), proportionally more ND cells exceeded the cutoff and were removed. After filtering, the mitochondrial medians collapse to ~ 6% in both groups (T2D 5.92%, ND 6.06%). These criteria were applied to reduce the contribution of low-complexity cells, cells with excessive RNA content, and cells with elevated mitochondrial gene expression, which are commonly evaluated during scRNA-seq quality control.

After filtering, the retained cells showed QC metric distributions within the predefined thresholds. The post-filtering violin plots confirmed that nFeature_RNA values were restricted below the upper filtering boundary, nCount_RNA values were limited to the expected post-QC range, and mitochondrial transcript percentages were constrained below 10% in both non-diabetic and type 2 diabetic groups. Visual inspection of the disease-group-level QC distributions showed broadly comparable post-filtering profiles between T2D and non-diabetic cells, with no evident group-specific separation driven by quality-control metrics (Fig. 1).

Sample-level QC visualization further demonstrated that no single islet is a catastrophic outlier and the filtering procedure retained cells across all included islet samples while reducing extreme outliers in gene complexity, RNA counts, and mitochondrial content. Although inter-sample variability in gene detection and library size remained, the distributions were contained within the predefined QC boundaries across samples, supporting the suitability of the filtered dataset for downstream normalization, integration, clustering, and differential expression analyses (**Supplementary Fig. S1**).

Highly variable gene selection

Following stringent quality control, gene expression values were normalized, and the 2,000 most variable genes were identified using the variance-stabilizing transformation method. The mean–variance relationship showed that a subset of transcripts displayed higher standardized variance than expected across the filtered single-cell dataset. The top variable genes were highlighted on the variable feature plot (Fig. 2) and were used as input features for subsequent scaling, principal component analysis, integration, and clustering. The top 10 HGVs are highly characteristic of pancreatic acinar cells (PRSS1; PRSS2; CPA1; PLA2G1B), particularly acinar cells undergoing regeneration, inflammation, or acinar-to-ductal metaplasia (ADM) (REG1B; REG3A), and none of them are among mitochondrial, ribosomal, or classical stress genes (https://www.ncbi.nlm.nih.gov/gene/). This feature-selection step was performed to capture the major sources of biological heterogeneity in the dataset before downstream dimensionality reduction and cell population identification.

Principal component analysis did not show global disease-driven segregation

Following normalization and highly variable gene selection, PCA was performed to evaluate the major sources of transcriptomic variation in the QC-filtered dataset. Projection of cells onto the first six principal components showed substantial overlap between type 2 diabetic and non-diabetic cells across PC1/PC2, PC3/PC4, and PC5/PC6. The early principal components showed substantial overlap between T2D and ND cells, suggesting that disease status was not the dominant driver of global variance and there is no gross group-level batch effect that would sabotage integration. Instead, the observed variation likely reflected a combination of cellular heterogeneity, sample-level effects, and differences among pancreatic cell populations. Therefore, downstream analyses were continued using sample-aware integration, unsupervised clustering, and cell-type-specific differential expression analysis (**Supplementary Fig. S2**).

Integrated embedding revealed a shared single-cell landscape across T2D and non-diabetic samples

Seurat v5 IntegrateLayers with CCAIntegration was used to harmonize sample-specific variation. To reduce sample-specific variation and generate a shared transcriptional space, QC-filtered cells were integrated using CCA-based integration. The integrated CCA representation was subsequently used for graph-based clustering and non-linear dimensionality reduction. Visualization of the integrated dataset using UMAP and t-SNE revealed multiple transcriptionally distinct cellular regions, indicating substantial heterogeneity within the pancreatic islet-derived single-cell dataset. The Seurat v5 integration workflow is designed to generate a

shared low-dimensional representation, such as integrated.cca, that can be used for unsupervised clustering and visualization across multiple samples or layers.

When cells were colored by disease group, T2D and non-diabetic cells were broadly intermingled across the integrated UMAP space, with no complete global separation by disease status (**Supplementary Fig. 3)**. This observation suggests that disease status alone was not the dominant driver of the global single-cell structure after integration. Instead, the major transcriptional structure likely reflected differences among underlying pancreatic cell populations. Also, it confirms that no residual group-level batch artifact survived integration, and the pseudobulk-within-cell type approach is needed to detect the T2D signal, because it did not show up as separate clusters.

Coloring the integrated UMAP by sample identity showed that cells from different islet samples contributed to multiple regions of the embedding, rather than forming strongly sample-restricted clusters. This pattern supports the effectiveness of the integration step in reducing overt sample-driven segregation while preserving cellular heterogeneity for downstream clustering and marker-based annotation (**Supplementary Figs. 4 & 5**).

Unsupervised clustering revealed distinct pancreatic cell populations

Following CCA-based integration, graph-based clustering was performed on the integrated single-cell transcriptomic space. This analysis identified 35 Seurat-defined clusters, indicating substantial transcriptional heterogeneity across the pancreatic islet-derived dataset. Visualization on the integrated UMAP showed that these clusters occupied distinct regions of the embedding, supporting the presence of multiple pancreatic and non-pancreatic cell populations within the dataset (**Supplementary Fig. 6**).

To biologically interpret the unsupervised clusters, canonical marker genes for major pancreatic cell populations were examined across the integrated UMAP space and across Seurat clusters. Endocrine-associated markers, including *INS, IAPP, PCSK1, PCSK2, GCG, TTR, SST*, and PPY, showed expression across broad islet-related regions of the embedding. In contrast, acinar markers such as *PRSS1, CPA1*, and *REG1A*, ductal markers such as *KRT19* and *CFTR*, stromal/mesenchymal markers such as *COL1A1, DCN*, and *RGS5*, and immune-associated markers such as *PTPRC, TRAC, CD79A, NKG7, CD14, LYZ, CST3*, and *MS4A7* showed more restricted expression patterns across specific cluster regions. Together, these marker distributions supported the presence of endocrine, acinar, ductal, stromal/mesenchymal, and immune-enriched cell populations in the integrated dataset (**Supplementary Fig. 7**).

Cluster-level violin plots further confirmed heterogeneous expression of canonical markers across Seurat-defined clusters. However, because several endocrine markers showed broad or partially overlapping expression patterns, final cell-type annotation was not based on a single marker alone. Instead, downstream annotation was guided by the combined evaluation of canonical marker expression, cluster-specific marker genes, and the spatial organization of clusters in the integrated UMAP.

The relative abundance of Seurat clusters was next examined across disease groups and individual islet samples. At the disease-group level, the overall cluster composition appeared broadly comparable between

non-diabetic and type 2 diabetic cells, with no obvious dominance of a single cluster in either group (**Supplementary Fig. 8**). At the sample level, cluster proportions showed inter-donor variability, which is expected in human islet single-cell datasets (**Supplementary Fig. 9**). Nevertheless, multiple clusters were represented across several samples, suggesting that the main clustering structure was not solely driven by a single donor. These analyses provided an overview of cluster distribution across biological groups and supported subsequent cell-type-specific analyses.

Refined clustering of the islet-enriched compartment resolved multiple pancreatic cell populations

After global clustering and canonical marker assessment, an islet-enriched subset was generated to focus subsequent analyses on pancreatic cell populations of interest. This subset included cells expressing at least one of the selected endocrines, exocrine, ductal, stromal, or immune-associated marker genes, as defined in the analytical workflow. The subset was then reprocessed and reclustered to resolve finer transcriptional heterogeneity within the pancreatic compartment.

Reclustering of the islet-enriched subset identified 21 Seurat-defined subclusters. UMAP and t-SNE visualization showed that these subclusters occupied distinct regions of the low-dimensional embedding, indicating substantial heterogeneity within the selected pancreatic cell population (**Supplementary Figs. 10 & 11**). Canonical marker expression across the reclustered subset supported the presence of multiple pancreatic cell compartments, including endocrine-like, acinar-like, ductal-like, stromal/mesenchymal-like, and immune-enriched populations.

Feature-level visualization showed that endocrine-associated markers, including INS, IAPP, PCSK1, PCSK2, GCG, TTR, SST, and PPY, were distributed across broad regions of the islet-enriched embedding. In contrast, exocrine and non-endocrine markers showed more spatially restricted expression patterns, including PRSS1, CPA1, and REG1A for acinar-enriched regions; KRT19 and CFTR for ductal-enriched regions; COL1A1, DCN, and RGS5 for stromal/mesenchymal-enriched regions; and immune-associated markers such as PTPRC, TRAC, CD79A, NKG7, CD14, LYZ, CST3, and MS4A7 in smaller immune-enriched regions (**Supplementary Fig. 12**).

The relative distribution of islet-enriched subclusters was also examined across disease groups and individual samples **(Supplementary Figs. 13 &14)**. Overall cluster composition appeared broadly comparable between non-diabetic and type 2 diabetic groups, without complete disease-specific segregation of any major subcluster. At the sample level, cluster proportions varied across donors, highlighting inter-individual heterogeneity within the human islet dataset. These results support the need for subsequent cluster marker identification and cell-type-specific differential expression analyses rather than relying on global disease-level comparisons alone.

Cluster-enriched marker genes supported the biological interpretation of global and islet-enriched clusters

To further characterize the transcriptionally defined clusters, cluster-enriched marker genes were identified using differential gene expression testing across Seurat-defined clusters. In the global integrated dataset,

13,679 cluster-associated marker genes were initially detected across 35 clusters. After applying stringent filtering criteria, including an average $\log_2$ fold-change of at least 2 and an adjusted p-value threshold of less than or equal to 0.01, 1,840 high-confidence marker genes were retained. The top marker genes per cluster were subsequently selected for visualization (The Excel file including the significant markers per cluster is provided as **Supplementary Table 1**).

DotPlot and heatmap visualizations of the top cluster-enriched genes showed distinct cluster-specific expression patterns across the global dataset. Several clusters displayed marker signatures consistent with major pancreatic and non-pancreatic cell populations, including endocrine-like, acinar-enriched, ductal/epithelial-enriched, stromal/mesenchymal-enriched, endothelial-enriched, and immune-enriched populations. For example, endocrine-associated signatures included markers such as IAPP, PCSK1, HADH, MEG3, and NPTX2, whereas acinar-enriched clusters showed expression of genes such as PRSS1, PRSS2, CTRB1, CTRB2, CPA1, and REG1A. Ductal/epithelial-enriched clusters were characterized by genes including KRT19, KRT7, TACSTD2, MMP7, and CFTR, while stromal and endothelial signatures were represented by markers such as COL1A1, COL1A2, COL6A1, BGN, SPARC, PECAM1, PLVAP, FLT1, and ESAM.

The same marker discovery workflow was then applied to the islet-enriched subset. In this subset, 16,095 marker genes were initially identified across 21 reclustered subclusters, of which 2,014 passed the stringent filtering criteria. Visualization of the top marker genes revealed distinct transcriptional programs across islet-enriched subclusters, supporting the presence of multiple pancreatic cell compartments within the refined subset. These marker profiles provided the basis for downstream cell-type annotation and cell-type-specific differential expression analyses (The Excel file including the significant markers per cluster is provided as **Supplementary Table 2**).

Top 10 markers per cluster in the global and the subset of clusters

Filtering the 1840 significant markers to the top 10 resulted in 221 genes across all 35 clusters, although some clusters did not have 10 significant markers. The results are provided in **Supplementary Table 3** in Excel format. Heatmap and dot plot showing the top 10 significant markers are provided as **Supplementary Figs. 15 & 16**.

Also, the filtering of 2,014 markers of the subset cluster to the top 10 resulted in 139 genes across all 21 clusters, although some clusters did not have 10 significant markers. The results are provided in **Supplementary Table 4** in Excel format. Heatmap and dot plot showing the top 10 significant markers are provided as **Supplementary Figs. 17 & 18**.

Marker-guided subsetting and reclustering resolved heterogeneity within major pancreatic cell compartments

To refine the analysis of major pancreatic cell populations, marker-guided cell-type-enriched subsets were generated and reclustered separately. Alpha-, delta-, epsilon-, and PP-cell populations were not further subsetted because these populations contained insufficient cluster-specific marker resolution and/or

limited cell representation after filtering; four major compartments were evaluated, including beta-like, acinar-like, ductal-like, and stromal/mesenchymal-like cell populations. Each subset was reprocessed and visualized using UMAP to assess within-compartment transcriptional heterogeneity.

In the beta-like subset, expression of canonical beta-cell-associated genes, including INS, IAPP, PCSK1, and PCSK2, supported the identity of the selected population **(Supplementary Fig. 19)**. Reclustering of this subset resolved nine beta-like subclusters, indicating transcriptional heterogeneity within the beta-enriched compartment **(Supplementary Fig. 20).** Visualization of marker expression across disease groups showed that beta-associated transcripts were detectable in both non-diabetic and type 2 diabetic cells **(Supplementary Fig. 21).**

The acinar-like subset was characterized by strong expression of acinar-associated markers, including PRSS1, CPA1, and REG1A **(Supplementary Fig. 22).** Reclustering identified eight acinar-like subclusters, suggesting additional heterogeneity within the acinar-enriched population **(Supplementary Figs. 23 & 24)**. Similarly, the ductal-like subset showed expression of ductal markers KRT19 and CFTR and was resolved into nine subclusters **(Supplementary Figs. 25–27)**. The stromal/mesenchymal-like subset was supported by expression of COL1A1, DCN, and RGS5, and reclustering identified twelve subclusters within this compartment **(Supplementary Figs. 28–30)**.

Across all evaluated subsets, marker expression patterns supported the biological identity of the selected cell populations and provided a basis for downstream cell-type-specific differential gene expression analysis. Disease-group violin plots were used as exploratory visualizations of marker expression distributions in non-diabetic and type 2 diabetic cells; however, statistical interpretation of disease-associated expression changes was reserved for differential gene expression testing.

Cell-type-specific differential gene expression analysis did not identify robust disease-associated genes after multiple-testing correction

To investigate disease-associated transcriptional changes within selected pancreatic cell compartments, differential gene expression analysis was performed between type 2 diabetic and non-diabetic cells within all types of cells, including beta, acinar, ductal, and mesenchymal cells that remained from the previous step. However, there were no identified DEGs within any of the cell types.

Pseudobulk differential gene expression analysis identified some exploratory disease-associated transcriptional changes

To obtain a donor-aware assessment of disease-associated transcriptional alterations, pseudobulk differential gene expression analysis was performed across annotated cell-type compartments. For each cell type, raw counts were aggregated at the sample level, and differential gene expression between type 2 diabetic and non-diabetic samples was tested using DESeq2. This strategy was used to reduce the risk of

treating individual cells from the same donor as independent observations and to provide a more conservative assessment of cell-type-specific disease-associated gene expression changes.

Across beta-like, acinar-like, and ductal-like compartments, no genes reached statistical significance after multiple-testing correction. In beta-like cells, 18,404 genes were tested; however, none passed the adjusted p-value threshold of 0.01. Similarly, no significant differentially expressed genes were identified in acinar-like cells among 17,452 tested genes or in ductal-like cells among 18,009 tested genes.

In contrast, at a different level of exploratory significance ($|\log_2 FC > 2|$ & adjusted p-value < 0.05), firstly, the immune-enriched compartment showed evidence of disease-associated transcriptional remodeling. Among 17,586 tested genes, CXCL3 was upregulated in T2D, with a $\log_2$ fold change of 2.521 and an adjusted p-value of 0.03854 (Table 2). Also, when the exploratory significance level of p-value < 0.01 was considered, some additional genes became significant in all cell types, including beta, acinar, ductal and immune compartments More raw signals in beta cells than acinar or ductal is biologically sensible since beta cells are where T2D acts. But none survive multiple-testing correction. The interpretation is this dataset doesn't have the statistical power to call individual genes significant after FDR correction with ~ 20 vs 17 donors, large between-donor variability, and a subtle within-cell-type disease effect, the per-gene test is underpowered. Therefore, these genes were not considered high-confidence differentially expressed genes (Table 3).

Table 2
The results of pseudobulk DEG identification within each cell type

Cell type	**Genes tested**	**FDR < 0.01 & $\|\log_2 FC > 2\|$**	**FDR < 0.05 & |log2FC > 2|**	**Main conclusion**
Beta-like	18,404	0	0	No significant pseudobulk DEGs
Acinar-like	17,452	0	0	No significant pseudobulk DEGs
Ductal-like	18,009	0	0	No significant pseudobulk DEGs
Immune-enriched	17,586	0	1(**CXCL3**; FDR:0.03854; log2FC: 2.521)	Near desirable significant pseudobulk DEGs

Table 3
Exploratory significant pseudobulk differentially expressed genes in the different cell types.

cell_type	gene	padj	pvalue	log2FoldChange	baseMean
Beta cells	CHI3L1	0.697971852	0.004649196	2.034680765	4.357367924
	GAD1	0.406734609	0.000994797	2.126207857	19.9326185
	LINC01117	0.774687726	0.009092184	2.280210094	1.662850184
	PCOLCE2	0.36513651	0.000683164	2.217716326	7.417786842
	ST6GAL1	0.317978834	0.000406716	2.304867113	18.83389275
	CAMK4	0.586735385	0.002660672	2.793107275	2.196492299
	FABP6	0.359080066	0.000585329	2.485078944	3.124082252
	ELFN1	0.551214758	0.001776767	3.785568249	2.110136983
	RAMP3	0.653619156	0.003764596	3.887309535	8.115362268
	AC090197.1	0.74514913	0.008025646	2.23013971	3.94736959
	AC027031.2	0.565763533	0.002267391	2.047033669	28.44913916
	ITIH5	0.681976598	0.004335539	2.044552185	6.914885601
	CPXM2	0.705514981	0.004803704	3.207673895	2.071056761
	DKK3	0.182944994	6.38E-05	2.68151157	72.47468174
	SYT1	0.743369939	0.006339166	2.180256668	8.236547554
	TMEM132C	0.607967719	0.003236971	2.343992794	2.348813665
	JPH4	0.551214758	0.001954754	2.204603232	4.716191234
	TSHR	0.36513651	0.000697109	2.646857229	3.287026606
	AC009041.2	0.248582683	0.000243126	2.525754268	1.124619226
	IRX3	0.406734609	0.001016616	2.59571898	4.095262186
	TBX2-AS1	0.551214758	0.001764302	4.260360993	9.457098233
	TBX2	0.321694337	0.00043699	2.694513591	3.429853315
	LINC01497	0.317978834	0.000402183	2.680439237	22.70338082
	IL11	0.182944994	5.63E-05	2.323297002	17.27799088
	APOL4	0.406734609	0.000985425	3.008198621	9.394678975
Acinar cells	APOA1	0.999964984	0.004701182	-2.315727405	2.480851743
	MIOX	0.999964984	0.00375017	-2.025675034	0.940229637
Ductal cells	KDR	0.999464425	0.001866331	-2.347217001	1.398399396
	LINC00520	0.999464425	0.00600938	-2.44303348	1.231811581

cell_type	gene	padj	pvalue	log2FoldChange	baseMean
	HOXB6	0.999464425	0.005558151	-2.084971951	1.303149628
	GATD3A	0.999464425	0.002551189	2.306285895	7.24042673
Immune cells	CHI3L1	0.341156394	0.000455187	2.313338394	323.537342
	ERMN	0.54878472	0.001798927	2.298590194	0.758074989
	CXCL3	0.038538505	5.96E-06	2.520876946	55.16068585
	IL33	0.054202377	1.23E-05	2.226673989	42.17992808
	AL606807.1	0.341156394	0.000469248	2.416177283	1.620063288
	TSPAN18	0.185354171	0.000126478	-2.35015148	1.596708167
	CEMIP	0.48121781	0.000839095	2.160979361	9.275147068
	OTOA	0.268096314	0.000208655	-2.162328652	1.50532476
	FAM20A	0.171915109	0.000102749	2.016254698	7.405199076
	VEGFD	0.632651199	0.002770052	2.208478878	1.784670435

Significant Markers Overrepresentation Enrichment Analysis Results

To further interpret the biological relevance of the significant marker gene lists, functional overrepresentation enrichment analysis was performed using DAVID Bioinformatics Resources. The enriched annotations were ranked based on Benjamini-adjusted p-values, and the top significant terms were summarized in Table 4. DAVID analysis of the significant marker gene set revealed a dominant extracellular and secretory signature. The most enriched terms included extracellular matrix, extracellular region, secreted proteins, protein digestion and absorption, signal peptide-containing proteins, activation peptide, zymogen, and extracellular matrix structural constituent terms. These findings indicate that the significant marker set was mainly associated with extracellular matrix organization, collagen-related structural components, secreted proteins, and pancreatic digestive/protease-related functions.

In addition, functional annotation of the immune-enriched exploratory gene list showed significant enrichment for Secreted and extracellular region terms, as summarized in Table 5. These terms were driven by genes including IL33, OTOA, CEMIP, CHI3L1, VEGFD, FAM20A, and CXCL3. Together, the DAVID enrichment results support the interpretation that the transcriptional signatures identified in this study are mainly related to extracellular matrix remodeling, secreted protein activity, acinar/protease-associated programs, and immune/extracellular inflammatory signaling.

Table 4
Top 10 significantly enriched functional annotation terms among significant marker genes identified by DAVID.

Term	Category	Count	Fold enrichment	Benjamini-adjusted p-value
extracellular matrix	GOTERM_CC_DIRECT	29	10.50	2.20E-18
extracellular region	GOTERM_CC_DIRECT	57	3.08	3.57E-14
Secreted	UP_KW_CELLULAR_COMPONENT	47	3.36	6.36E-13
Protein digestion and absorption	KEGG_PATHWAY	14	19.91	1.37E-11
Signal	UP_KW_DOMAIN	62	2.09	4.74E-10
Liver Cirrhosis, Experimental	DISGENET	32	4.09	4.68E-09
Extracellular matrix	UP_KW_CELLULAR_COMPONENT	15	8.44	3.87E-08
PROPEP: Activation peptide	UP_SEQ_FEATURE	10	23.89	2.02E-07
Zymogen	UP_KW_PTM	13	7.96	1.77E-06
extracellular matrix structural constituent	GOTERM_MF_DIRECT	11	13.77	1.81E-06

Table 5
DAVID enrichment analysis of the immune-enriched exploratory gene list

Term	Category	Count	Fold enrichment	Benjamini-adjusted p-value	Contributing genes
Secreted	UP_KW_CELLULAR_COMPONENT	7	6.34	0.000915	IL33, OTOA, CEMIP, CHI3L1, VEGFD, FAM20A, CXCL3
extracellular region	GOTERM_CC_DIRECT	7	4.96	0.010	IL33, OTOA, CEMIP, CHI3L1, VEGFD, FAM20A, CXCL3

Discussion

This study reanalyzed publicly available single-cell RNA-seq data from human pancreatic islets to define transcriptional differences between type 2 diabetes and non-diabetic samples at the cell-type level. Using

stringent quality control, integrated clustering, marker-based annotation, and donor-aware pseudobulk analysis, we found that a clear global transcriptomic separation did not characterize the diabetic signal in this dataset; instead, it emerged as limited, compartment-specific changes within a heterogeneous islet landscape. These findings suggest that disease-related transcriptional alterations in human T2D islets are subtle and best interpreted within specific cellular compartments rather than as a uniform disease-wide program (6, 9, 10).

At the global level, diabetic and non-diabetic cells were broadly intermixed in the integrated embedding, indicating that cell identity and inter-donor heterogeneity were the main sources of structure in the dataset (11). This pattern underscores the need for cell-type-resolved analysis, because whole-islet comparisons can obscure biologically meaningful signals restricted to particular endocrine or non-endocrine populations. Prior human islet single-cell studies similarly showed that cellular programs are highly structured by cell identity and that diabetes-associated differences are often cell-type dependent rather than globally uniform (6, 7, 12, 13). In this respect, the present analysis reinforces the view that the transcriptional phenotype of T2D in human islets is distributed across cellular context rather than concentrated in a single dominant signature.

Cell-level differential expression analysis did not identify robust disease-associated genes after multiple-testing correction in the beta-like, acinar-like, or ductal-like compartments. This result should be interpreted cautiously, as it does not imply biological equivalence between groups; rather, it likely reflects modest effect sizes, substantial donor-to-donor variability, and the limited statistical power typical of human single-cell studies. The pseudobulk framework provided a more conservative donor-level assessment and yielded broadly consistent findings, suggesting that any disease effect in these compartments is likely small and context-dependent. In this setting, the absence of strongly FDR-significant genes is more informative than it may first appear, as it argues against broad, easily detectable transcriptional reprogramming in the examined compartments (14–16).

Among the exploratory pseudobulk findings, the overall distribution of nominal signals was biologically coherent. The beta-like and immune-enriched compartments, which are directly relevant to T2D pathophysiology through beta-cell dysfunction and islet inflammation, showed the most notable disease-associated signals, whereas acinar-like and ductal-like compartments showed weaker evidence of differential expression. Although most comparisons did not yield robust high-confidence DEGs under the most stringent thresholds, this pattern suggests that the nominal signal was not randomly distributed across cell types, but instead followed biologically plausible T2D-relevant compartments (16–18). Within that framework, an immune-enriched chemokine signal may reflect localized inflammatory communication within the islet niche (18–20).

In the immune-enriched compartment, pseudobulk analysis identified STK38L, CXCL3, and IMPA2 as FDR-significant genes under the exploratory threshold. Among these, CXCL3 was the most biologically interpretable candidate because it also showed a large effect size and is linked to chemokine-mediated inflammatory signaling. Importantly, CXCL3 met the exploratory criterion of FDR < 0.05 with |log2FC| > 2, but it did not reach the more stringent FDR < 0.01 threshold used for high-confidence calls. Therefore, CXCL3 should be interpreted as a candidate disease-associated signal rather than a definitive T2D marker.

Nevertheless, its direction and biological context are plausible, since chemokine signaling and innate immune activation have long been implicated in islet inflammation and T2D pathophysiology.

Functional enrichment analysis using DAVID highlighted gene sets related to extracellular matrix, secreted proteins, protein processing, and proteolytic activity. These results should be interpreted as functional signatures of the marker gene lists, not as direct evidence of disease causality. Nevertheless, the enrichment pattern is informative because it points to tissue structure, secretion, and biochemical processing programs that are highly relevant to pancreatic islet biology. Any disease-related interpretation of these enrichments should therefore remain cautious, with the strongest disease-relevant signal reserved for the exploratory immune-enriched CXCL3 finding rather than the broader annotation output (21, 22).

The present findings are broadly consistent with earlier human single-cell studies, showing that islet cell identity and cell-type-specific transcriptional programs are central to understanding diabetes-related changes. Segerstolpe and colleagues demonstrated that human pancreatic islets contain distinct endocrine programs and that single-cell analysis reveals cellular heterogeneity that bulk profiling cannot capture. Subsequent studies likewise showed that diabetes-associated expression changes are often cell-type specific and may be missed in whole-islet analyses. Our results fit this framework by showing that the disease signal is subtle, compartment-restricted, and better detected through donor-aware, cell-type-level modeling than through global comparisons alone (7, 12).

Several strengths and limitations should be noted. A major strength of this study is the combination of rigorous quality control, unbiased clustering, marker-guided annotation, and pseudobulk inference, which together provide a more reliable interpretation of single-cell data than cell-level testing alone. However, the analysis is limited by its secondary nature, as it relies on a public dataset not specifically generated for the questions addressed here. Additional limitations include the lack of experimental validation, limited power to detect modest disease effects, and incomplete adjustment for possible donor-level covariates such as age, sex, body mass index, and treatment exposure. Some immune and stromal annotations may also remain imperfect because transcriptomic identities in dissociated islet preparations can overlap across related cell states.

Overall, this analysis suggests that type 2 diabetes in human pancreatic islets is associated with limited but biologically coherent transcriptional changes rather than broad global reprogramming. The most notable candidate signal, CXCL3, points to a possible inflammatory component in the immune-enriched compartment, while the enrichment results indicate functional programs related to secretion, matrix organization, and proteolysis. These findings support a model in which disease-related effects are distributed across specific cellular niches and become visible only when analyzed at the appropriate biological resolution. Larger cohorts, independent datasets, and orthogonal validation will be necessary to determine which of these exploratory signals represent reproducible disease mechanisms.

Conclusion

In conclusion, this study provides a cell type-resolved reanalysis of publicly available single-cell RNA-seq data from human pancreatic islets in type 2 diabetic and non-diabetic donors. The integrated analysis did

not reveal a broad global transcriptomic separation between groups, but rather indicated subtle, compartment-specific signals that were better evaluated through donor-aware pseudobulk analysis than through cell-level comparisons alone. Although no robust differentially expressed genes were identified in the major beta-like, acinar-like, or ductal-like compartments under stringent thresholds, the immune-enriched compartment showed an exploratory CXCL3 signal, suggesting a possible contribution of localized inflammatory communication within the islet microenvironment. Functional enrichment analysis highlighted extracellular matrix, secretory, and proteolytic programs among marker genes, pointing to biologically relevant features of the islet niche. Overall, these findings support a model in which T2D-associated transcriptional alterations in human pancreatic islets are distributed across specific cellular compartments rather than reflected by a single uniform disease program. Larger independent cohorts, spatial transcriptomic profiling, and experimental validation will be needed to determine whether these exploratory signals represent reproducible mechanisms of human T2D islet pathology.

Declarations

Data Availability Statement

The single-cell RNA sequencing dataset analyzed in this study is publicly available from the Gene Expression Omnibus (GEO) under accession number GSE221156. All data generated or analyzed during the current study are included in this published article and its supplementary information files. Processed results and additional materials are available from the corresponding authors upon reasonable request.

Competing interests

The authors declare no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions

M.Z. and A.K.H. contributed equally as co-first authors. F.R. and M.M.A. contributed equally as co-corresponding authors. M.A.S. and F.J.R. contributed to data collection, interpretation, and critical revision of the manuscript. M.M.A. and F.R. supervised the study, provided overall guidance, and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final manuscript.

References

1. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nature Reviews Disease Primers. 2015;1(1):15019.
2. Galicia-Garcia U, Benito-Vicente A, Jebari S, Larrea-Sebal A, Siddiqi H, Uribe KB, et al. Pathophysiology of Type 2 Diabetes Mellitus. Int J Mol Sci. 2020;21(17).

3. Goyal R, Singhal M, Jialal I. Type 2 diabetes. StatPearls [Internet]. 2023.
4. Reed J, Bain S, Kanamarlapudi V. A Review of Current Trends with Type 2 Diabetes Epidemiology, Aetiology, Pathogenesis, Treatments and Future Perspectives. Diabetes Metab Syndr Obes. 2021;14:3567-602.
5. Cornell S. Continual evolution of type 2 diabetes: an update on pathophysiology and emerging treatment options. Ther Clin Risk Manag. 2015;11:621-32.
6. 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.
7. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, Sun X, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell metabolism. 2016;24(4):593-607.
8. Grenko CM, Taylor HJ, Bonnycastle LL, Xue D, Lee BN, Weiss Z, et al. Single-cell transcriptomic profiling of human pancreatic islets reveals genes responsive to glucose exposure over 24 h. Diabetologia. 2024;67(10):2246-59.
9. 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 metabolism. 2016;24(4):608-15.
10. Wang G, Chiou J, Zeng C, Miller M, Matta I, Han JY, et al. Integrating genetics with single-cell multiomic measurements across disease states identifies mechanisms of beta cell dysfunction in type 2 diabetes. Nat Genet. 2023;55(6):984-94.
11. Weng C, Gu A, Zhang S, Lu L, Ke L, Gao P, et al. Single cell multiomic analysis reveals diabetes-associated β-cell heterogeneity driven by HNF1A. Nature Communications. 2023;14(1):5400.
12. Prasad RB, Groop L. Single-Cell Sequencing of Human Pancreatic Islets-New Kids on the Block. Cell Metab. 2016;24(4):523-4.
13. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A single-cell transcriptome atlas of the human pancreas. Cell systems. 2016;3(4):385-94. e3.
14. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nature communications. 2021;12(1):5692.
15. Zimmerman KD, Espeland MA, Langefeld CD. A practical solution to pseudoreplication bias in single-cell studies. Nature communications. 2021;12(1):738.
16. Maestas MM, Ishahak M, Augsornworawat P, Veronese-Paniagua DA, Maxwell KG, Velazco-Cruz L, et al. Identification of unique cell type responses in pancreatic islets to stress. Nature Communications. 2024;15(1):5567.
17. Westwell-Roper C, Nackiewicz D, Dan M, Ehses JA. Toll-like receptors and NLRP3 as central regulators of pancreatic islet inflammation in type 2 diabetes. Immunology & Cell Biology. 2014;92(4):314-23.
18. Böni-Schnetzler M, Meier DT. Islet inflammation in type 2 diabetes. Semin Immunopathol. 2019;41(4):501-13.
19. Collier JJ, Sparer TE, Karlstad MD, Burke SJ. Pancreatic islet inflammation: an emerging role for chemokines. J Mol Endocrinol. 2017;59(1):R33-r46.

20. Eguchi K, Nagai R. Islet inflammation in type 2 diabetes and physiology. The Journal of clinical investigation. 2017;127(1):14-23.
21. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic acids research. 2022;50(W1):W216-W21.
22. Craig-Schapiro R, Li G, Chen K, Gomez-Salinero JM, Nachman R, Kopacz A, et al. Single-cell atlas of human pancreatic islet and acinar endothelial cells in health and diabetes. Nature Communications. 2025;16(1):1338.

Figures

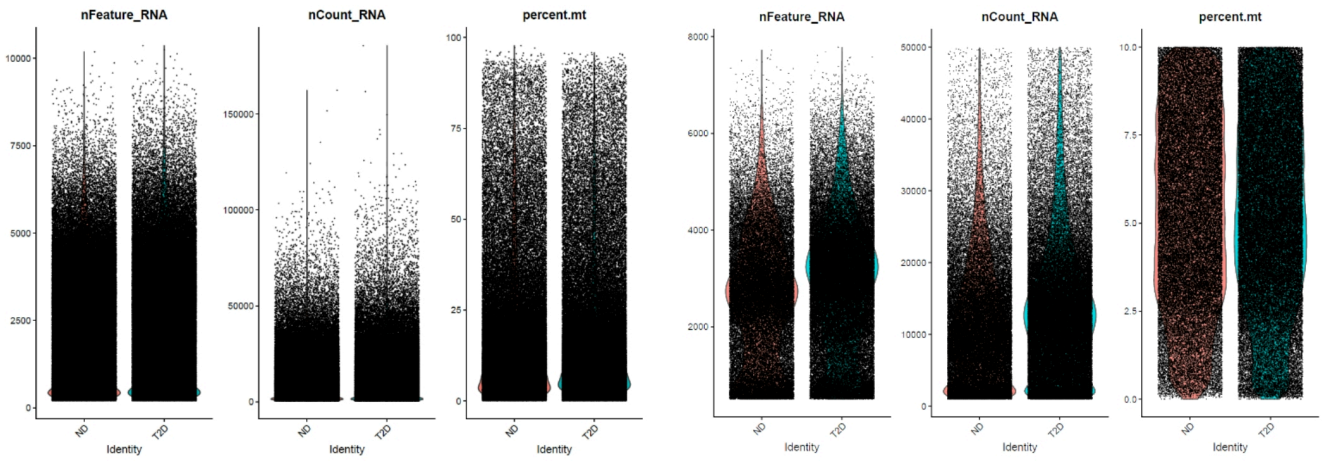


Figure 1

The number of detected genes, RNA, and mitochondrial percentage before (left) and after (right) data quality control (QC).

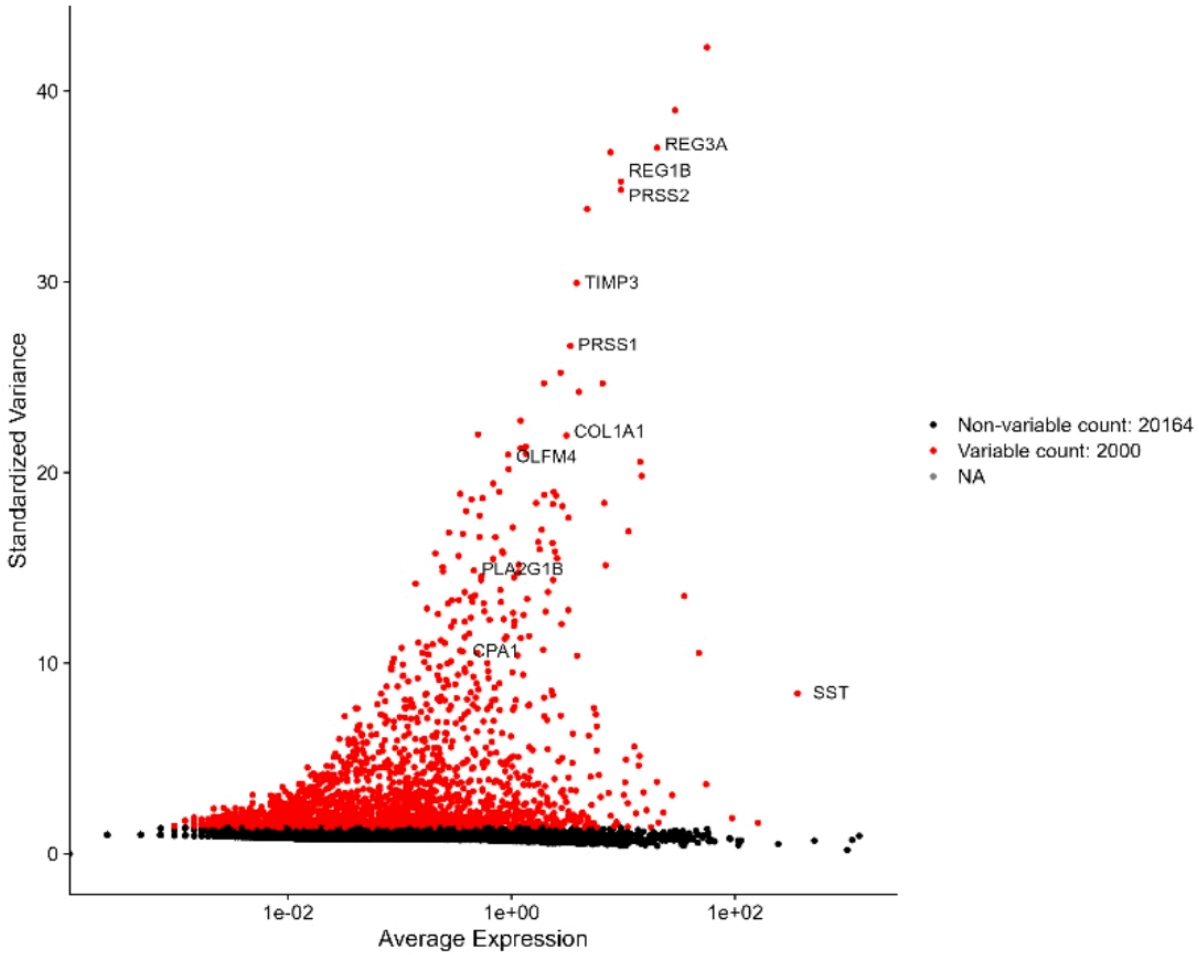


Figure 2

Top 10 variable features. The x-axis represents the mean gene expression across all cells, and the y-axis represents standardized variance. Red dots indicate the 2000 highly variable genes (HVGs) across samples, and black dots indicate the non-variable genes. The top 10 HVGs are also named, and their expression varies far more than expected for their mean.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplemetaryTable4top10markersperclusterlog2FCge2.csv.xlsx
- SupplementaryTable3top10markersperclusterlog2FCge2.csv.xlsx
- SupplementaryTable220clustermarkers.xlsx

- SupplementaryTable134clustermarkers.xlsx
- supplementaryFigures.docx