A Single-Cell Transcriptomic Atlas of the Human Pancreas: Reanalysis and Expansion of GSE85241 data

Julia Szkóp

June 2025

Abstract

The pancreas plays a central role in metabolic regulation and glucose homeostasis, making it a critical organ in the context of diabetes and other metabolic diseases[1]. A detailed understanding of its cellular heterogeneity is essential for uncovering functional diversity within both its endocrine and exocrine compartments. This study replicates and extends the single-cell transcriptomic analysis published by Muraro et al.[2], using single-cell RNA sequencing data to identify and classify pancreatic cell types through dimensionality reduction and clustering techniques. Canonical markers were used to define major cell populations, including alpha, beta, delta, PP, acinar, ductal, and mesenchymal cells. Subclustering within alpha and beta populations revealed transcriptionally distinct subtypes, suggesting additional functional complexity.

Epsilon cells were not identified as a separate cluster, although expression of ϵ-cell markers was detected. A distinct cluster expressing macrophage-associated genes was also identified, indicating the presence of immune cells not addressed in the original study. Gene ontology enrichment analyses of both ϵ-cell and immune-related genes highlighted metabolic specialization and immune activation pathways, respectively. Age- and sex-specific analyses revealed notable differences in the expression of genes linked to inflammation, tissue repair, and sex chromosomes. Overall, the results confirm key aspects of the original atlas while providing new insights into immune cell involvement and cellular subtypes within the human pancreas.

1 Introduction

The human pancreas is a multifunctional organ essential for maintaining metabolic homeostasis. It performs both endocrine functions—through the secretion of hormones such as insulin, glucagon, somatostatin, and pancreatic polypeptide—and exocrine functions, which involve the production of digestive enzymes by acinar cells[3]. Pancreatic dysfunction is implicated in a wide range of diseases, including type 1 and type 2 diabetes, chronic pancreatitis, and pancreatic cancer[4]. A comprehensive understanding of the pancreas's cellular architecture and transcriptional landscape is therefore crucial for the advancement of both basic and translational biomedical research.

Anatomically, the pancreas is located in the upper abdominal cavity behind the stomach, spanning approximately 15–22 cm in length and weighing between 50–125 grams in adults. It is subdivided into four regions: the head, neck, body, and tail. The head lies adjacent to the duodenum, and the pancreatic duct releases digestive fluid into the small intestine at the ampulla of Vater[5]. The pancreas is composed of a heterogeneous mixture of highly specialized cell types. These include endocrine alpha, beta, delta, and PP cells within the islets of Langerhans; exocrine acinar and ductal cells; and non-epithelial components such as endothelial, mesenchymal, and immune cells. Each of these cell types performs distinct physiological roles, and their abundance and functional states can vary dynamically in response to developmental, metabolic, or pathological stimuli[6]. Precise characterization of these cellular populations is fundamental for understanding tissue function, identifying disease mechanisms, and develop-

ing targeted therapeutic strategies.

Traditional bulk RNA sequencing has limited resolution in distinguishing the contributions of individual cell types, as it reflects averaged gene expression across mixed populations[7]. In contrast, single-cell RNA sequencing (scRNA-seq) provides transcriptomic profiling at the resolution of individual cells, offering unprecedented insight into cellular identity, heterogeneity, and plasticity. This technology has revolutionized transcriptomic research, enabling the construction of high-resolution cellular atlases across a range of human tissues and organs[8 9].

Within this context, the study by Muraro et al. (2016) was a landmark in pancreatic biology. It delivered one of the first single-cell transcriptomic atlases of the human pancreas using scRNA-seq, successfully identifying the major pancreatic cell populations based on canonical marker genes[2]. However, the study had certain limitations. Rare or low-abundance populations, such as epsilon cells, were not deeply characterized, and the presence of non-parenchymal immune cells was largely unaddressed. Moreover, the analysis did not focus on inter-individual variability among donors.

Given the increasingly recognized role of immune cells in pancreatic homeostasis and disease, as well as the potential biological relevance of rare endocrine cell types, revisiting and expanding upon this dataset using more refined analytical approaches is both timely and necessary. The aim of this study is to reproduce the findings of the original atlas while extending the analysis to include a more detailed investigation of underrepresented cell types and inter-donor variability, particularly with respect to age and sex. Such an approach enables a more comprehensive understanding of the human pancreas at single-cell resolution and reveals previously overlooked biological features with potential clinical significance.

2 Materials and Methods

Computational environment and software tools
All data processing and analyses were carried out in the R statistical environment (version 4.4.2) on a Linux workstation. Single-cell RNA–sequencing workflows were implemented using the Seurat framework[10], which provided a comprehensive pipeline for quality control, normalization, feature selection, dimensionality reduction, clustering, and visualization. Data manipulation and organization of both cell-level metadata and expression matrices relied on the `dplyr` package[11], while publication-ready graphics were produced with `ggplot2`[12]. Heatmaps with hierarchical clustering and annotation tracks were generated via the `pheatmap` package[13], and functional enrichment analyses were performed with `clusterProfiler`[14], using the `org.Hs.eg.db` database for gene identifier mapping.

Data import and initial filtering
Count data corresponding to the GSE85241 human pancreatic islet dataset were loaded directly into R and assembled into a Seurat object. To ensure robust downstream analysis, genes detected in fewer than three cells and cells expressing fewer than 500 genes were excluded, removing low-complexity cells and extremely rare features.

Normalization and feature selection
Library size differences across cells were corrected by scaling counts to a common total and log-transforming the data. Genes exhibiting high variability across the cell population were identified using a variance-stabilizing selection method, and the top 2,000 most variable genes were retained for downstream analyses. All selected features were centered and scaled to have mean zero and unit variance.

Dimensionality reduction and clustering
Principal component analysis (PCA) was applied to the scaled, highly variable genes to capture the main axes of transcriptional variation. The number of components retained was chosen by inspecting the elbow of the variance-explained plot. A shared nearest-neighbor graph was then constructed based on these components, and unsupervised clustering was performed across a range of resolution values (0.1–1.0) to identify cell populations at varying granularities. Low-dimensional embeddings were generated using t-distributed stochastic neighbor embedding (t-SNE)[15] and Uniform Manifold Approximation and Projection (UMAP) to facilitate visual exploration of cluster structure.

Cluster annotation and marker gene identification
Clusters were annotated by assigning cell-type labels according to canonical marker genes (e.g., INS for beta cells, GCG for alpha cells, SST for delta cells). Differential expression between each cluster and all others was assessed using a nonparametric rank-sum test, retaining genes with a minimum $\log_2$ fold-change of 0.25 and expression in at least 25% of cells in the cluster. The top markers for each cluster were visualized in heatmaps after scaling pseudobulk expression profiles to highlight expression gradients.

Inter-donor comparisons
To assess biological covariates such as donor sex and age, cells were subset by donor metadata and reprocessed independently through normalization, feature selection, PCA, and clustering. Differential expression between male and female donors, and between younger and older donors, was evaluated using adjusted p-value thresholds of 0.05. Significant genes were displayed in donor-structured heatmaps to illustrate inter-individual variation.

Subclustering of major populations
To explore heterogeneity within the major endocrine lineages, alpha and beta cells were isolated and each subset underwent a dedicated pipeline: normalization, variable gene detection, scaling, PCA, neighbor graph construction, clustering, and UMAP embedding. Subpopulations were then characterized by their marker gene expression, revealing finer granularity within alpha and beta cells.

Functional enrichment analysis
Marker gene lists were converted from HGNC symbols to Entrez IDs using the clusterProfiler mapping functions, and over-representation tests for Gene Ontology molecular function and biological process categories were performed with Benjamini–Hochberg correction (adjusted p-value$<$0.05). Enrichment results were summarized graphically using dot plots for concise biological interpretation.

3 Results

Identification of Pancreatic Cell Types by t-SNE
The analysis began with the import of the gene expression matrix (a CSV file containing UMI counts for each cell and the corresponding metadata). First, the data were normalized and the 2,000 most variable genes were selected. These genes were then used to perform PCA for dimensionality reduction. Based on the elbow plot, the first 15 principal components were selected for subsequent clustering steps.

Based on the first 15 PCs, Louvain clustering (resolution 0.2) identified nine clusters, which were clearly separated in the resulting t-SNE embedding.

To assess the impact of batch effects (systematic, non-biological differences between experimental runs), cells were colored by donor of origin. In the t-SNE plot (Figure 1A), no cluster comprised exclusively cells from a single donor, excluding a batch effect and confirming that the observed groupings reflect genuine pancreatic cell heterogeneity.

Genes with significant overexpression in each cluster relative to all other cells were identified, and for each cluster the ten markers exhibiting the highest mean log_2 fold-change were selected. Based on canonical marker expression patterns described by Muraro et al.[2], clusters were assigned to specific cell types e.g. GCG to α-cells, INS to β-cells, SST to δ-cells, PPY to PP-cells, PRSS1 to acinar cells, KRT to ductal cells, and COL to mesenchymal cells (Table 1). Despite thorough investigation, no cluster exhibited markers characteristic of ϵ-cells. The clusters were thereby annotated according to their respective cell types (Figure 1B).

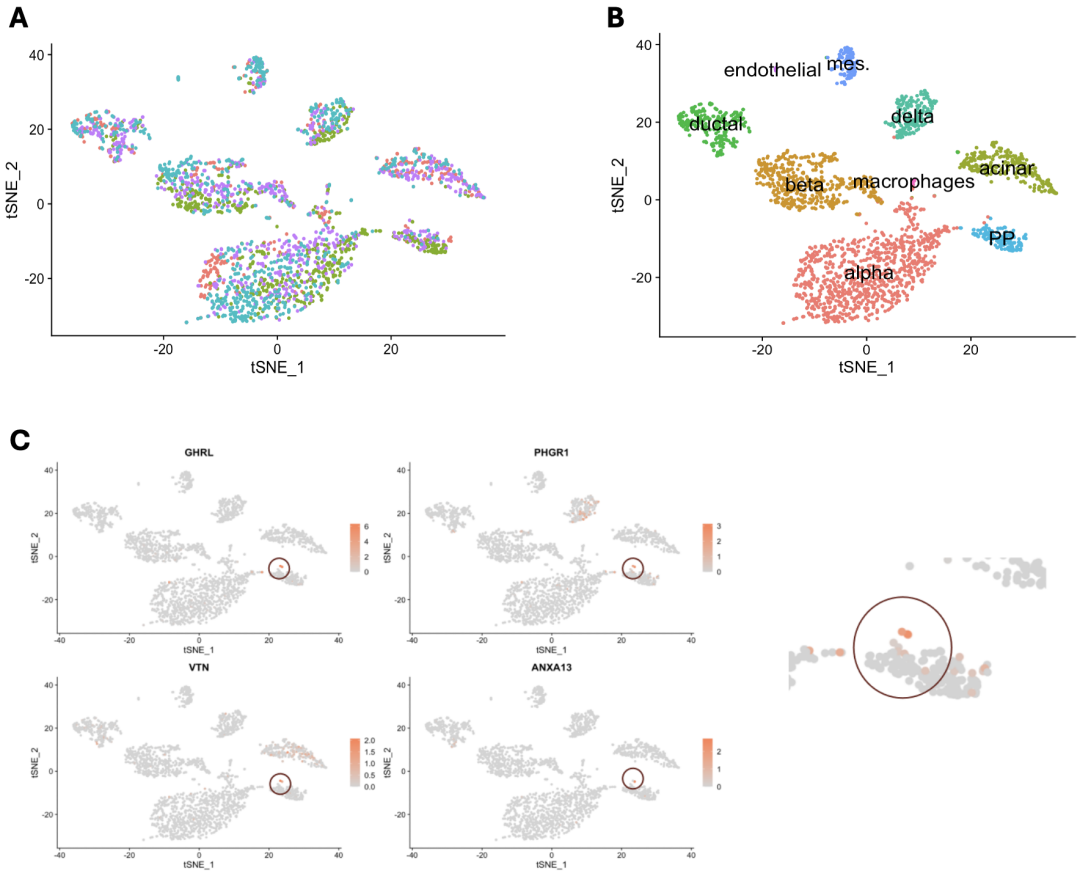

Figure 1: **Identification and characterization of pancreatic cell populations by single-cell RNA-seq. (A)** t-SNE embedding of human pancreatic cells, colored by donor of origin. Cells from each of the four donors (distinct colors) are interspersed across all clusters, indicating no batch effects and that cluster structure reflects biological heterogeneity. **(B)** Same t-SNE map annotated by major cell type. A previously unannotated cluster (macrophages) was identified by immunological markers. **(C)** t-SNE feature plots of representative ϵ-cell markers (GHRL, PHGR1, VTN, ANXA13). Each panel shows expression intensity across all cells (gray = low/absent, orange = high). The inset highlights the small group of cells co-expressing these markers, confirming the presence of ϵ-cells despite their low abundance.

Cluster 0	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Cluster 7**	**Cluster 8**
TMEM236	WSCD2	AMY2A	CEACAM7	GABRA1	PPY	POSTN	SELE	FPR3
PLCE1	HHATL	PRSS3P2	KRT23	NTNG1	NPFFR2	ITGA11	PLVAP	C1QA
LOXL4	DLK1	CPA2	FUT3	GHSR	SLITRK6	CDH11	NKX2-3	CCL3
PTPRT	INS	ALB	CEACAM6	PPFIA2	SERTM1	GLI2	ECSCR	C1QC
IRX2	MAFA	PNLIPRP2	APCDD1	SYT1	PTGFR	COL8A1	FCN3	LILRB2
KLHL41	LRRTM3	SERPINI2	SLC3A1	BCHE	GPC5-AS1	PTGIR	ROBO4	MS4A6A
DPP4	CAPN13	PNLIP	SLC34A2	FRZB	KCNG1	POM121L9P	SOX18	TREM2
CRYBA2	SIX3	PRSS1	ALDH1A3	LRFN5	THSD7A	HSD11B1	KDR	RNASE6
SPOCK3	ADCYAP1	CELA3A	NR1H4	HAP1	PCDH10	PDGFRB	EMCN	OLR1
GCG	IAPP	REG3G	VCAM1	CBLN4	CARD11	TBX2	ELTD1	MPEG1
α**-cells**	β**-cells**	**acinar**	**ductal**	δ**-cells**	**PP cells**	**mes. cells**	**end. cells**	**???**

Table 1: **Top 10 markers for each of the 9 clusters (according to average $\log_2$ FC).**

Analysis of gene expression characteristic of ϵ-cells

For the most common genes of ϵ-cells, an analysis of their individual expression was performed (Fig. 1C). All examined markers showed a consistent and repeatable signal distribution in the same small group of cells, confirming the existence of a population of ϵ-cells. However, modification of the clustering parameters (changing the number of principal components and the resolution value) did not lead to the isolation of a separate cluster for these cells, reflecting their very low percentage in the total population.

In addition, the UMAP method was used to reduce dimensionality and visualize the individual expression of selected ϵ-cell genes (Fig. 2A). Unlike the t-SNE projection, UMAP mapping did not clearly identify a distinct expression cluster—the signals are poorly defined and scattered within the population, reflecting the very low proportion of ϵ-cells in the sample (Fig. 2B).

Finally, a GO:MF functional enrichment analysis was performed for the most common genes in the ϵ-cell population (Fig. 3A). Despite the limited number of markers, the plot reveals a small but coherent set of molecular functions, dominated by regulators of kinase activity (both activators and negative regulators), lipid-metabolizing enzymes (e.g. arachidonate-CoA and long-chain fatty acid–CoA ligases), and specific binding activities (carbohydrate and phosphatidylglycerol binding). This suggests that, even though ϵ-cells are defined by relatively few markers, they may uniquely influence local pancreatic lipid metabolism and signal transduction through kinase modulation.

Additional Public Pancreas scRNA-seq

To verify the presence and location of ϵ-cells in samples with a higher proportion of this cell type, the publicly available single-cell RNA-seq dataset of human islet preparations from donors with documented elevated ϵ-cell proportions (van Gurp et al.[16]) was used. These data were subjected to a similar processing and visualization procedure: after normalization and extraction of the same list of markers, dimensionality reduction was performed using the t-SNE method, maintaining similar parameters (Fig. 2C).

The resulting t-SNE plot clearly showed a cluster in which the ghrelin marker (GHRL) and other ϵ-cell genes (e.g., PAX4) showed the highest expression (Fig. 2D). These signals were consolidated in a single, well-separated region of the map, which contrasts with the scattered and poorly visible signal in the analysis of the original data. This result confirms that with a sufficiently high representation of ϵ-cells and consistent t-SNE methodology, it is possible to clearly identify them as a distinct population.

Characteristics of the unannotated cluster

As shown in Table 1, the markers identified in cluster 8 did not correspond to the expression of any previously described pancreatic cell type. In order to establish their identity, a search of the Human Protein Atlas database was performed, which confirmed that the set of these genes – including C1QA, C1QC, CCL3, CD68 and RNASE6 – is characteristic of tissue macrophages. This finding is important, because in the original work by Muraro et al. (2016) the presence of immune cells in the analyzed material was not reported.

To characterize the molecular functions of the immune-cell–enriched cluster, a GO:MF enrichment analysis was performed on genes overexpressed in cluster 8 (Fig. 3B). This revealed strong enrichment for diverse binding activities (amyloid-β, amide, peptide, high/low-density lipoproteins) alongside multiple receptor activities (immune receptor, inhibitory MHC I, complement, semaphorin). These functions point to a role for the macrophage population in ligand recognition and immune signaling within the pancreatic microenvironment—roles that have been underappreciated in prior studies.

Donor-level differential analysis

In order to limit the influence of donor age and gender on the transcriptomic profile, two heat maps were prepared. The first one compared the expression of key genes in donors aged 59 and 23 (Fig. 3C).In older donors, genes associated with inflammation(including CRP, CFB, LGALS2) and oxidative stress (DUOX2) were upregulated, wwhile in younger cells their expression was almost absent. At the same time, tissue markers (OLFM4, REG3A) were silenced in the 23-year-old, but strongly active in the 59-year-old donor.

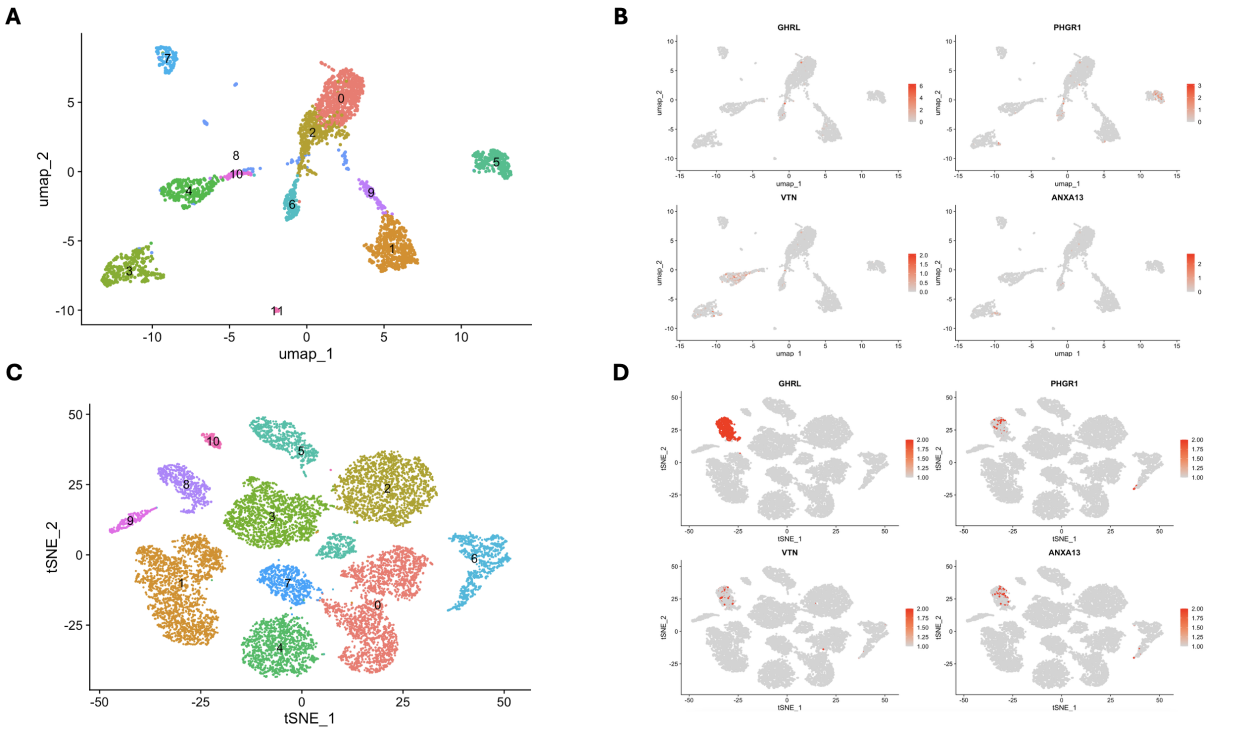

Figure 2: **Visualization of ϵ-cell marker expression by UMAP and t-SNE. (A)** UMAP embedding of pancreatic cells showing nine clusters (0–8). **(B)** UMAP feature plots for representative ϵ-cell genes (GHRL, PHGR1, VTN, ANXA13; gray = low, orange = high), illustrating dispersed, low-frequency expression. **(C)** t-SNE embedding of an enriched ϵ-cell dataset (van Gurp et al.), revealing eleven clusters. **(D)** t-SNE feature plots of the same markers in the enriched dataset, showing a clearly defined ϵ-cell cluster with high marker co-expression.

The second heat map (Fig. 3D) reveals clear sex-specific expression patterns. Cells from the male donor (D30) show strong expression of Y-chromosome genes (RPS4Y1, DDX3Y, ZFY), which are virtually absent in the female donor (D29). In contrast, XIST—marker of X-chromosome inactivation—is highly expressed in female cells (D29) but low in male cells (D30). Pancreatic markers such as REG3A and CFB are expressed in both donors, although with donor-specific intensity.

Sub-clustering of major cell populations

The last step of the analysis was the subclustering of the main populations of α- and β-cells. Each of them was subjected to independent dimensionality reduction using the UMAP method, which allowed for the separation of four subclusters within both the α (Fig. 4A) and β (Fig. 4B) populations. In the next step, ten genes with the highest overexpression were identified for each subcluster, and their expression profiles were visualized against the background of the entire population to assess whether their activity was strictly limited to a specific cluster or diffuse. In the α-cell population, a clear focus of XIST gene expression was observed in one of the subclusters, while in the β-cell subclusters, the CRYM gene was the most distinctive, the expression of which was localized exclusively in this subset (Fig. 4C and 4D).

4 Discussion

Robustness of Dimensionality Reduction and Clustering

The first step of the analysis—selecting the 2000 most variable genes and reducing dimensions using PCA—is a common strategy in scRNA-seq analysis[17]. The selection of the first 15 principal components was based on an "elbow" plot, which allows for balancing biological information and technical noise.

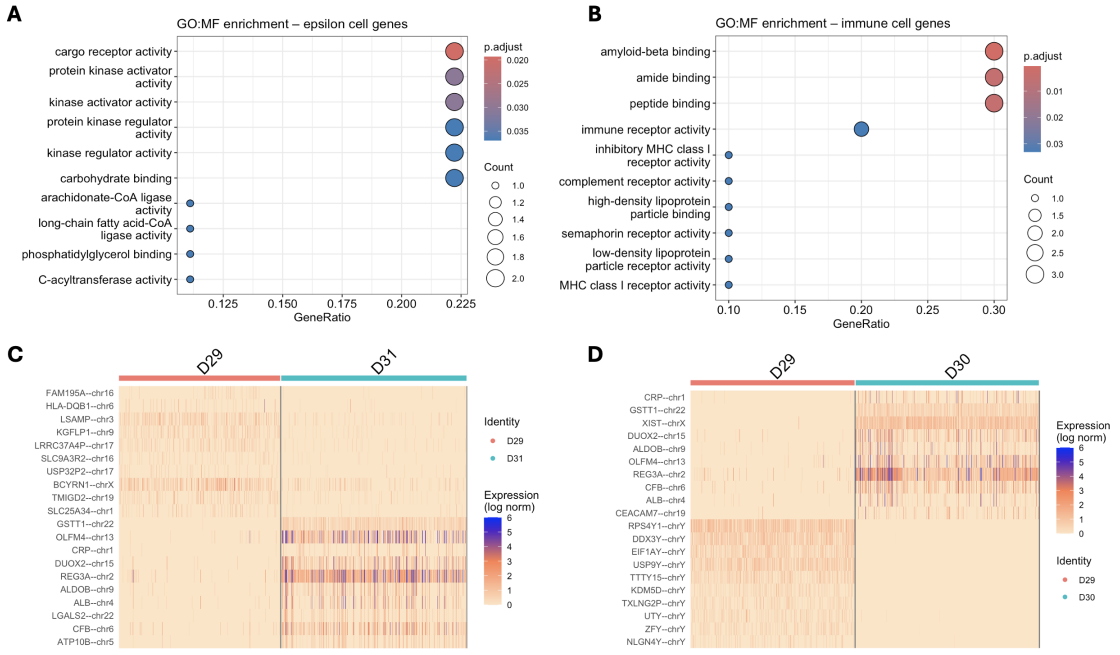

Figure 3: **Functional enrichment and donor-specific expression.** **(A)** GO:MF of ϵ-cells: kinase regulator/activator, carbohydrate and phosphatidylglycerol binding, long-chain fatty acid–CoA ligase. **(B)** GO:MF of immune cluster: amyloid-β, peptide and lipoprotein binding; immune/MHC I/complement/semaphorin receptors. **(C)** Donor D29 vs D31 heat map: younger enriched for proliferation/tissue markers; older for inflammatory/oxidative-stress genes. **(D)** Sex-specific heat map: XIST in female, Y-chr genes in male; shared pancreatic markers with differential intensity.

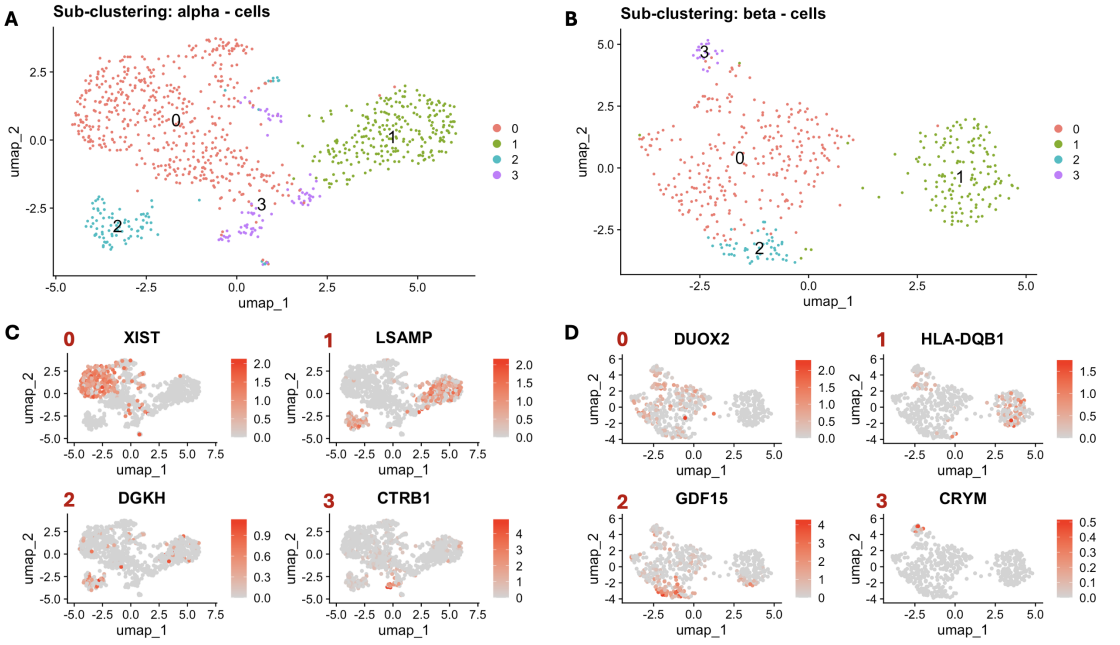

Figure 4: **Subclustering of α- and β-cells reveals transcriptional heterogeneity.** **(A)** UMAP embedding of α-cells showing four subclusters. **(B)** UMAP embedding of β-cells showing four subclusters. **(C)** UMAP feature plots of representative α-cell markers—XIST (cluster 0), LSAMP (1), DGKH (2), CTRB1 (3)—demonstrating subcluster-restricted expression. **(D)** UMAP feature plots of representative β-cell markers—DUOX2 (0), HLA-DQB1 (1), GDF15 (2), CRYM (3)—highlighting their exclusive localization in individual subclusters.

However, the literature recommends a range of 15–7 PCs to capture the main sources of variance without overfitting the model[18].

Louvain clustering at 0.2 resolution generated 9 clusters with clear separation in t-SNE space. The lack of concentration of cells from a single donor in a single cluster excluded a batch effect[19], confirming that proper scaling and normalization of the data minimized technical artifacts.

Detection and characterization of rare ϵ-cells
Although ϵ-cells constitute a small fraction of the population, their presence was confirmed by the correlated expression of markers such as GHRL and PHGR1 in the same cell group[16]. This highlights the sensitivity of scRNA-seq analysis in detecting micro-populations even when they do not form distinct clusters with standard settings.

GO:MF functional enrichment analysis revealed that ϵ-cells may play specific roles in the regulation of kinase activity and lipid metabolism. These results are consistent with the hypothesis that ϵ-cells influence the local environment of the pancreatic islets by secreting ghrelin[20].

Comparison of t-SNE and UMAP projections shows that in sparse cells, t-SNE may have an advantage in locally resolving micro-populations, while UMAP better preserves the global structure but scatters the ϵ-cell signal.

Data analysis[16] confirmed that as the percentage of ϵ-cells increases, the t-SNE cluster becomes more distinct. This confirms that the detectability of rare populations increases with their percentage in the sample — a key issue to consider when designing future studies.

The comparative methodology, maintaining nearly identical t-SNE parameters, eliminates the possibility that the differences are an artifact of the analysis, while also revealing the influence of sample composition on the result.

Identification of the pancreatic macrophage population
Cluster 8 was assigned immunological markers (C1QA, CD68, CCL3) typical of tissue macrophages, which was not mentioned in the original studies[2]. This finding indicates that the biological material contained a small number of immune system cells.

GO:MF function analysis showed enrichment in immune receptor activity and amyloid-β binding, suggesting that pancreatic macrophages may participate in the removal of protein aggregation products and modulate inflammation. Similar conclusions have been described in studies of macrophages residing in pancreatic islets, which have demonstrated their ability to recognize and remove protein complexes and modulate the inflammatory response[21].

Genomic signatures dependent on donor age
A comparison of gene expression in cells from a 59-year-old and a 23-year-old revealed increased expression of pro-inflammatory genes (CRP, CFB, LGALS2) and genes associated with oxidative stress (DUOX2) in samples from the older donor. Similar changes, indicative of chronic, low-grade inflammation, were observed in the pancreatic islets of aging rats[22].

In the younger donor, there was almost no expression of inflammatory genes and increased activity of proliferation markers (OLFM4, REG3A in the older donor), suggesting that aging affects the ability to regenerate tissues.

Extending the study to include a larger number of donors in different age groups will allow for statistical capture of trends and separation of the influence of age from individual variability.

Gender-specific gene expression patterns
It is interesting to note that some islet markers (REG3A, CFB) showed different expression levels between the sexes, which may indicate subtle differences in the functioning of pancreatic islets depending on sex hormones. Other studies of sex differences in pancreatic islet cell dysfunction have shown that in women, estrogens increase insulin secretion and protect β-cells from oxidative stress, while in men, androgens modulate the insulin response through androgen receptors, leading to different metabolic adaptations[23].

Further research into hormonal regulators of gene expression in α and β cells is suggested to better understand the clinical implications for personalized ther-

apies.

Heterogeneity within major populations: subclustering of α and β cells
Subclustering of α- and β-cells using UMAP revealed four subgroups in each line, indicating an additional layer of heterogeneity. In α cells, one of the subclusters strongly expressed the XIST gene, which may reflect epigenetic differences associated with X chromosome inactivation.

In the β-cell population, a subcluster expressing the CRYM gene was identified, suggesting that some β-cells may perform a specific metabolic or stress function.

Comparison with the analyzed publication
These observations largely confirm the key findings of the publication by Muraro et al. ("A Single-Cell Transcriptome Atlas of the Human Pancreas"), which used the SORT-seq protocol to obtain a high number of unique transcripts per cell, which significantly increased the sensitivity of cell subpopulation detection. Similarly, based on t-SNE and UMAP analysis, nine major clusters corresponding to α, β, δ, PP, exocrine, and endothelial cells were identified, and the initial lack of a separate ϵ-cell population was associated with their low representation in the sample.

Unlike the original publication, which reported only heterogeneity among β cells associated with ER and oxidative stress, subclustering revealed additional subgroups in both the α- and β-cell populations, differing in the expression of XIST and CRYM genes, suggesting the existence of epigenetic and metabolic differences within these major cell types. Importantly, the scope of the analysis was extended to identify a macrophage cluster, represented by markers such as C1QA and CD68, indicating the presence of immune populations in pancreatic materials—an aspect omitted in the original atlas.

Furthermore, donor-specific differential analysis revealed changes in the expression of pro-inflammatory and X/Y inactivation-related genes, dependent on the age and sex of the donor, further deepening our understanding of the biological heterogeneity of the human pancreas[24].

References

[1] Röder, P. V., Wu, B., Liu, Y. & Han, W. Pancreatic regulation of glucose homeostasis (2016).

[2] Muraro, M. J. *et al.* A single-cell transcriptome atlas of the human pancreas. *Cell Systems* **3**, 385–394.e3 (2016).

[3] Slack, J. M. W. Developmental biology of the pancreas. *Development* **121**, 1569–1580 (1995).

[4] Hu, J.-X. *et al.* Pancreatic cancer: A review of epidemiology, trend, and risk factors. *World Journal of Gastroenterology* **27**, 4298–4321 (2021).

[5] Russell, T. B. & Aroori, S. The pancreas from a surgical perspective: an illustrated overview. *Art of Surgery* **6**, 1–1 (2022).

[6] Cabrera, O. *et al.* The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proceedings of the National Academy of Sciences* **103**, 2334–2339 (2006).

[7] Wang, Z., Gerstein, M. & Snyder, M. Rna-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57–63 (2009). URL https://www.nature.com/articles/nrg2484.

[8] Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).

[9] Kolodziejczyk, A., Kim, J. K., Svensson, V., Marioni, J. & Teichmann, S. The technology and biology of single-cell rna sequencing. *Molecular Cell* **58**, 610–620 (2015).

[10] Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411–420 (2018).

[11] Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. *dplyr: A Grammar of Data Manipulation* (2025). URL https://dplyr.tidyverse.org. R package version 1.1.4.

[12] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016). URL https://ggplot2.tidyverse.org.

[13] Kolde, R. *pheatmap: Pretty Heatmaps* (2025). URL https://github.com/raivokolde/pheatmap. R package version 1.0.13.

[14] Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).

[15] van der Maaten, L. & Hinton, G. Viualizing data using t-sne. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

[16] van Gurp, L. *et al.* Generation of human islet cell type-specific identity genesets. *Nature Communications* **13**, 2020 (2022).

[17] Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411–420 (2018).

[18] Luecken, M. D. & Theis, F. J. Current best practices in single-cell rna-seq analysis: a tutorial. *Molecular Systems Biology* **15** (2019).

[19] Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data (2010).

[20] Kim, S. W. *et al.* Ghrelin stimulates proliferation and differentiation and inhibits apoptosis in osteoblastic mc3t3-e1 cells. *Bone* **37**, 359–369 (2005).

[21] Calderon, B. *et al.* The pancreas anatomy conditions the origin and properties of resident macrophages. *Journal of Experimental Medicine* **212**, 1497–1512 (2015).

[22] Sandovici, I. *et al.* Ageing is associated with molecular signatures of inflammation and type 2 diabetes in rat pancreatic islets. *Diabetologia* **59**, 502–511 (2016).

[23] Gannon, M., Kulkarni, R. N., Tse, H. M. & Mauvais-Jarvis, F. Sex differences underlying pancreatic islet biology and its dysfunction (2018).

[24] OpenAI. ChatGPT: language model GPT-4. https://chat.openai.com/ (2025).