

HumanIslets: An integrated platform for human islet data access and analysis

Jessica D. Ewald^{1,2}, Yao Lu¹, Cara E. Ellis^{3,4}, Jessica Worton^{3,5}, Jelena Kolic⁶, Shugo Sasaki^{7,8}, Dahai Zhang^{7,8}, Theodore dos Santos^{3,4}, Aliya F. Spigelman^{3,4}, Austin Bautista³, Xiao-Qing Dai^{3,4}, James G. Lyon³, Nancy P. Smith^{3,4}, Jordan M. Wong^{3,5}, Varsha Rajesh^{9,10}, Han Sun^{9,10}, Seth A. Sharp^{9,10}, Jason C. Rogalski¹¹, Renata Moravcova¹¹, Haoning H Cen⁶, Jocelyn E. Manning Fox^{3,4}, HI-DAS Consortium, Ella Atlas¹², Jennifer E. Bruin¹³, Erin E. Mulvihill^{14,15}, C. Bruce Verchere^{16,17,18}, Leonard J. Foster¹¹, Anna L. Gloyn^{9,10}, James D. Johnson⁶, Andrew R. Pepper^{3,5}, Francis C. Lynn^{7,8}, Jianguo Xia^{1,*}, Patrick E. MacDonald^{3,4,*}

¹Institute of Parasitology, McGill University, Montreal, QC

²Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA

³Alberta Diabetes Institute, University of Alberta, Edmonton, AB

⁴Department of Pharmacology, University of Alberta, Edmonton, AB

⁵Department of Surgery, University of Alberta, Edmonton, AB

⁶Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, Vancouver, BC

⁷Diabetes Research Group, BC Children's Hospital Research Institute, Vancouver, BC, Canada

⁸Department of Surgery, School of Biomedical Engineering, University of British Columbia, Vancouver, BC

⁹Department of Pediatrics, Division of Endocrinology, Stanford School of Medicine, Stanford, CA

¹⁰Stanford Diabetes Research Center, Stanford School of Medicine, Stanford, CA

¹¹Department of Biochemistry & Molecular Biology, Michael Smith Laboratories, Life Sciences Institute, University of British Columbia, Vancouver, BC

¹²Environmental Health Science and Research Bureau, Health Canada, Ottawa, ON

¹³Department of Biology & Institute of Biochemistry, Carleton University, Ottawa, ON

¹⁴Department of Biochemistry, Microbiology and Immunology, University of Ottawa, ON

¹⁵University of Ottawa Heart Institute, Ottawa, ON

¹⁶Department of Surgery, BC Children's Hospital Research Institute and University of British Columbia, Vancouver, BC

¹⁷Department of Pathology and Laboratory Medicine, BC Children's Hospital Research Institute and University of British Columbia, Vancouver, BC

¹⁸Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC

Correspondence to:

Patrick MacDonald
Alberta Diabetes Institute
LKS Centre, Rm 6-126
University of Alberta
Edmonton, AB, T6G 2R3
pmacdonald@ualberta.ca
www.bcell.org

Jianguo Xia
Office P107, Institute of Parasitology
McGill University, Macdonald Campus
2111 Lakeshore Road,
Sainte-Anne-de-Bellevue, QC H9X 3V9
jeff.xia@mcgill.ca
www.xialab.ca

SUMMARY

Comprehensive molecular and cellular phenotyping of human islets can enable deep mechanistic insights for diabetes research. We established the Human Islet Data Analysis and Sharing (HI-DAS) consortium to advance goals in accessibility, usability, and integration of data from human islets isolated from donors with and without diabetes at the Alberta Diabetes Institute (ADI) IsletCore. Here we introduce [HumanIslets.com](https://humanislets.com), an open resource for the research community. This platform, which presently includes data on 547 human islet donors, allows users to access linked datasets describing molecular profiles, islet function and donor phenotypes, and to perform various statistical and functional analyses at the donor, islet and single-cell levels. As an example of the analytic capacity of this resource we show a dissociation between cell culture effects on transcript and protein expression, and an approach to correct for exocrine contamination found in hand-picked islets. Finally, we provide an example workflow and visualization that highlights links between type 2 diabetes status, SERCA3b Ca^{2+} -ATPase levels at the transcript and protein level, insulin secretion and islet cell phenotypes. [HumanIslets.com](https://humanislets.com) provides a growing and adaptable set of resources and tools to support the metabolism and diabetes research community.

INTRODUCTION

Four decades following the development of methods for large-scale isolation of human pancreatic islets,^{1–3} the use of cadaveric islets has become a staple of diabetes research.^{4–7} As biomedical datasets continue to grow in size, there has been increased emphasis on improving the findability, accessibility, interoperability, and reusability (FAIR) of published data.^{8–10} This is highly relevant to human islet research⁶ since these are increasingly used as a comparator in the development of stem cell-derived products for transplantation,^{11–13} in large genomics studies at the islet^{14,15} and single-cell level,^{16,17} and in studies that seek to understand conventional type 1 diabetes (T1D) and type 2 diabetes (T2D).⁴ Furthermore, since islet tissue is not readily accessible it is particularly important to maximize the usage of existing datasets.

Several initiatives provide data resources and analysis pipelines for the exploration of human islet data, with differing degrees of openness. Some focus on deep human pancreas phenotyping and data curation,^{7,18–21} while others provide analysis tools for linking hormone production and gene expression,²² single-cell islet transcript visualization,^{23,24} and exploring the relationship between diabetes genetics and gene regulation.²⁵ While existing resources are valuable, there is an unmet need both for user-friendly analysis pipelines that are accessible to the general islet and diabetes research community, including non-specialists, and for improved access to analysis pipelines and minimally processed data for more customized analyses. Additionally, since there is often low concordance in findings from different datasets and analyses,⁴ a better understanding of how donor factors, technical aspects of organ processing, and cell culture impact these data is crucial for appropriate interpretation of results. Proper documentation, reporting, and integration of this information into transparent data collection and analysis pipelines can make important contributions to understanding islet data and diabetes research. We established the HumanIslets Data Analysis & Sharing (HI-DAS) Consortium to address these important gaps in the human islet research ecosystem.

The Alberta Diabetes Institute (ADI) IsletCore (www.isletcore.ca) isolates and distributes human islets specifically for research. Together with HI-DAS we present an online resource (HumanIslets.com) for the exploration, analysis, and download of data collected from human research islet preparations from organ donors with and without diabetes. This platform includes data (547 donors as of June 2024) and methods for analysis of bulk islet transcriptomics, bulk islet proteomics, and single-cell transcriptomics linked to measures of islet functions such as insulin secretion, oxygen consumption, electrophysiology, pancreas processing, and cell culture metadata. The HumanIslets.com platform offers intuitive support to directly analyze these data and metadata, including: *i*) exploration at the donor, islet and single-cell levels; *ii*) multi-omic

analysis of genes/proteins and pathways controlled for common factors that may influence results; and *iii*) download of data, results and analysis methods for transparent and reproducible re-analysis. Furthermore, we demonstrate the impact of correcting for culture time and exocrine contamination following cell type composition deconvolution using bulk proteomics data. This platform is under continuous growth as new donors and additional data types are added and is freely available for the diabetes and islet research community.

RESULTS

The [HumanIslets.com](https://humanislets.com) tool is a web-based platform for accessing, analyzing, and visualizing phenotyping data collected from human research islets isolated at the ADI IsletCore. The underlying database includes technical metadata on the islet isolation and cell culture process; multi-omics; functional data; and associated donor metadata that describes basic, de-identified demographic parameters (**Figure 1**). The tool is designed to be easily expanded with new data types, and regularly updated as data are collected from new donors.

Database contents. The *HumanIslets* platform presently contains data from 547 donors. These are 41.1% female, range in age from 10-weeks to 84 years, and include donors with no diabetes ($n=355$, 65%), pre-diabetes ($n=99$, 18%, defined as HbA1c of 5.7-6.5% and no clinical diagnosis of T2D), T2D ($n=78$, 14.3%, defined as a reported diagnosis of T2D or HbA1c>6.5%), and T1D ($n=15$, 2.7%, defined as a reported diagnosis of T1D, typically with validation by autoantibody screening, genetic risk scores, and/or biopsy immunostaining) (**Figure 2**). The donor metadata also contain technical details on pancreas characteristics and processing, islet isolation, and cell culture outcomes. There are extensive functional phenotyping data including static insulin secretion at various glucose concentrations, dynamic insulin secretion in response to glucose and other macronutrients, islet oxygen consumption in response to glucose and mitochondrial modulators, and single-cell electrophysiological outcomes for both α - and β -cells. Certain data are available for all donors, while others have been collected from more recent subsets or are not yet available (**Figure 2E**). In total, there are 74 distinct donor, technical, and functional outcomes summarized at the donor level, providing a rich context within which to analyze and interpret islet multi-omics data from the same donors (**Suppl Table 1**).

Bulk transcriptomics, bulk proteomics, and single-cell transcriptomics were collected from isolated islets of subsets of the donors. Two different platforms were used for the bulk transcriptomics: including RNA-seq ($n=117$ donors; whole transcriptome) and NanoString ($n=190$ donors; 156 genes). The single-cell transcriptomic data are available for analysis at the

single-cell or donor level. At the single-cell level, transcriptomes are linked with nine single-cell electrophysiology outcomes (Patch-seq data; $n \sim 4300$ cells collected from 96 donors) sub-setted by cell type (α -, β -cells) and glucose concentration (1, 5, and 10mM). For donor-level outcomes (ie. age, diabetes status), the data are summarized as a single α - and β -cell pseudobulk profile for each donor. Analysis of single-cell data with respect to donor-level outcomes like diabetes status can lead to artificially low p-values when each cell is treated as an independent replicate by not incorporating donor IDs into the statistical analysis (known as 'pseudoreplication').²⁶ Two main solutions have been proposed: analyzing pseudobulk profiles with respect to donor-level outcomes,^{27,28} or using linear mixed models with donor incorporated as a random effect.²⁶ Given there is no consensus on the best approach,²⁹ we use the pseudobulk profiles method to support interactive visual exploration. Users can download the data and corresponding analysis script to perform mixed model analysis, if desired.

Platform design. The [HumanIslets.com](https://humanislets.com) platform was developed using a design-thinking approach³⁰ to focus features on addressing the needs of the islet and diabetes research community. Within the HI-DAS Consortium, tool developers were in continual contact with islet biologists across career stages and with broad expertise across domains of islet biology and diabetes, enabling feedback on prototypes followed by multiple iterations of each feature. This process revealed a clear desire for support for simple queries and common analyses of well-organized and documented data via an interactive web interface, coupled with the ability to easily subset and export minimally processed data for more customized and sophisticated offline analyses. This focused the tool design on intuitive workflows using well established and widely understood statistical methods, providing clear documentation of all steps, and presenting results with familiar visualizations. The platform is designed to allow researchers to query statistical associations within the data by omics type (*Omics View*), by individual omics or metadata feature (*Feature View*), or by donor (*Donor View*) (**Figure 1C**). A visual walkthrough of the functions described below is presented in the **Supplementary Document 1**.

Omics View allows researchers to find omics features and pathways associated with any of the 74 distinct donor, technical, and functional outcomes. The interface allows users to select an omics type, a metadata variable of interest, and any covariates that should be adjusted for. Researchers can perform the analysis on a subset of donors by applying different filters and can specify the p-value significance threshold. The association analysis results are displayed in an interactive volcano plot coupled with a detailed result table. Clicking on a point (in volcano plot) or result row (in table) will display a scatterplot or violin plot for the association/analysis of

interest which can be exported in PNG, PDF, or SVG format. Within these, the donor-level data points are interactive - clicking them will open the '*Donor View*' page query results for that donor, allowing researchers to quickly investigate outliers and to assess whether they are driven by technical or biological factors. The result table provides raw and adjusted p-values as well as estimates for cell type associations (*Associated Proportions*) based on deconvolution analysis of the bulk proteomics data. Links in this table take users to the relevant gene page at the National Center for Biotechnology Information ('NCBI') or Type 2 Diabetes Knowledge Portal ('T2DKP'),³¹ or generate a figure summarizing the cell-type specific gene expression across 14 cell types from 109 donors without diabetes, representing more than 140,000 individual cells ('SC'). Clicking on the feature name (gene/protein) will link directly to the *Feature View* page (see below) for that hit. Finally, using *Pathway Analysis*, data can be analyzed using either overrepresentation analysis (ORA) or gene set enrichment analysis (GSEA). Statistically significant pathways are displayed in an interactive ridgeline plot that allows comparison of relative positive and negative metadata associations across different pathways. Clicking on a pathway of interest will generate heatmaps of protein or gene expression for all features in a pathway. Pathway analysis results can also be viewed in table form (Results Table). This allows researchers to quickly perform and visualize standard differential expression, association, and pathway analyses, compare global signals across omics layers, and explore the influence of different covariates on the results. When users perform an analysis in the 'Omics View' page, the list of donors with completely paired sets of the omics data and metadata used in the analysis is made available for download. All associations and pathway results, along with the R scripts that generated them, can be downloaded directly via the 'Download Results' and 'Save Analysis' buttons, respectively.

Feature View consists of a search bar and database of pre-computed statistical associations describing over 3.7 million relationships, including pairwise combinations of all omics features and metadata variables (omics-metadata), while adjusting for influential technical and biological covariates (sex, age, BMI, culture time). Users can enter the name of any variable within the *HumanIslets* database to retrieve its associations with all other outcomes. If a gene name or symbol is searched, significant associations for both the protein and mRNA gene products will be included in the search results, which can be subsequently filtered by gene/protein symbol, metadata variable, or omics type. This page was designed for researchers who have a particular gene or outcome of interest, and would like to find out which proteins, mRNAs, or metadata variables are most correlated with it.

Donor View allows users to query a specific donor by ADI IsletCore DonorID or RRID, and returns that donor's metadata and functional outcomes, highlighted within the context of all other donors in the database. For example, searching donor 'R421' displays a histogram of HbA1c values from all donors, with donor R421's value of 5.4% highlighted with the percentile rank (0.38, or 38th percentile), and the same for age and BMI. The Donor View page has nine different tabs, each presenting a different category of data. Clicking through these will show, in addition to donor information: the properties and images of the islet isolation itself; insulin content results and static insulin secretion across glucose concentrations; dynamic insulin secretion responses to glucose, leucine and oleate/palmitate; oxygen consumption assessed via Seahorse assay; and electrophysiological profiles of α -cells and β -cells. Each of these data are shown in relation to the entire dataset, allowing researchers to understand how comparable a given donor is to the rest of the population represented within the database: this is a common question when a donor appears as an outlier in a particular analysis or when a research group receives islet samples from the ADI IsletCore from a small number of donors. Finally, the last tab presents a summary of omics data available in the tool for that donor.

Data Download allows researchers to download all data from the *HumanIslets* database using extensive filtering options to subset donors using common or advanced criteria, by uploading a list of donor IDs, or by dataset availability. For example, a researcher could filter the list to include only female donors with both insulin secretion and bulk RNA-seq data. The omics data are available for download in their unprocessed and processed forms. The unprocessed data refer to the original omics data table; for example, intensity values for proteomics data and counts for RNA-seq-based gene expression data. The processed omics data have undergone batch correction, missing value imputation, filtering, and normalization (see Methods for more details). Raw omics data (ie. proteomics MS spectra, sequencing FASTQ files) are not stored in the HumanIslets.com tool but are accessible in public repositories (links provided under 'Documentation' and in Data Accessibility below). The final data tables can be exported in either CSV, plain text, Excel format, or as an R object, together with a data dictionary outlining units and any formulas used to calculate each variable. Details on data collection and processing are available in the 'Documentation' (**Suppl Document 1**) link at the top of the web page.

Transparency, reproducibility, and maintainability. The HumanIslets.com platform comes with extensive documentation on how each dataset was generated and the precise definition of each variable. A brief description is given for all statistical methods, and how they are combined into pipelines to generate the results. The HumanIsletsR Github repository

(<https://github.com/xia-lab/HumanIsletsR>) provides all R functions used to perform analyses within the tool, and all R scripts used to process each dataset. The processing scripts are designed to dynamically pull together data from different sources as various datasets are expanded and save the data at different stages of processing for download, analysis, and visualization within the tool. Some steps, such as outlier detection and omics data normalization, depend on statistics computed from all samples and therefore produce slightly different processed values as new donors are added (or in comparison to the original published analyses of individual datasets). This presents a reproducibility problem, as the same analysis on the omics View page will produce slightly different results after the data are updated, even if the donors are filtered to include the same list of IDs. Therefore, the “*Download Analysis*” feature was added to the tool, which allows researchers to download a time-stamped copy of the processed omics data, paired metadata outcomes, and R functions, allowing them to fully reproduce the analysis offline.

Important technical and biological associations. The omics data were analyzed with respect to each metadata variable to understand the relative signal strength associated with different functional outcomes, and to identify which donor and technical covariates were influential across multiple omics layers. These results informed the linear model used to generate the pre-computed association results in the ‘*Feature Search*’ page. Here we highlight overarching trends across the entire database and focus on technical covariates that are often overlooked.

The number of omics features statistically significantly associated with eight key donor and technical metadata are presented in **Table 1**, along with the dimensions of each omics data and whether there were significant batch effects present, details that are important to consider when interpreting the number of significant features across modalities. ‘Culture time’ and ‘percent purity’ are two metadata categories with the most significantly associated features in multiple omics datasets (**Table 1**). Both are technical parameters and should be carefully considered in the analysis and interpretation of data. The metadata comparison with the next highest number of significantly associated features across omics layers was T2D-ND. HbA1c (%) was associated with a high number of features in the proteomics data only. Age and sex also had a moderate number of associated features across multiple omics platforms. Most mRNAs and proteins significantly associated with sex are encoded by genes found on the X- and Y chromosomes, but do not show up in the NanoString data because they were not included in that gene panel; genes in the NanoString panel were mainly selected for their relevance to islet

function and development. This handful of explicitly sex chromosome-linked features is strongly associated ($\text{FDR} < 10^{-10}$) with sex, while most other mRNAs and proteins have no discernible difference between male and female donors.

There were many fewer, or no, pseudobulk mRNA features associated with any metadata variable except for sex, which makes sense given the lower sample size and because these profiles are derived from single-cell data that has greater noise, lower sequencing depth, and higher drop-out rates compared to bulk data. However, while the adjusted p-values may not be statistically significant at the feature level, analyzing the pseudobulk association results at the pathway level with GSEA produces statistically significant results that are consistent with known biology. For example, comparing pseudobulk mRNA expression in α -cells from donors with and without T1D highlights the upregulated Reactome pathway ‘Antigen Presentation: Folding, assembly and peptide loading of class I MHC’ as shown by others.³² An expanded table with the full association results for all metadata variables is available in the Supplementary Information (**Suppl Table 2**).

Culture time has less impact on protein expression than transcript expression. The distribution of culture time values is multimodal with each peak spaced roughly 24 hours apart (**Figure 3A**) because islets were typically removed from culture during business hours. Consistent with what we recently observed,³³ a large proportion of features were associated with culture time in the bulk RNA-seq (53%; **Figure 3B**) and NanoString (32%) data, but not the bulk proteomics data ($\sim 0.01\%$; **Figure 3C**). Gene products tend to have consistent relationships across the bulk RNA-seq and NanoString data, but not with the proteomics data. For example, *FGF2* (fibroblast growth factor 2) is positively associated with culture time in both the bulk RNA-seq ($\text{FDR} = 5.3\text{e-}9$, **Figure 3D**) and NanoString ($\text{FDR} = 4.1\text{e-}6$, **Figure 3E**) data, but not in the proteomics data ($\text{FDR} = 0.99$, **Figure 3F**). Pathway analysis of the bulk RNA-seq data with GSEA (features ranked by T-statistic) using the KEGG, Reactome, and Gene Ontology Biological Process libraries consistently identified proteasome, spliceosome, and ribosome-related gene sets as the most significantly enriched with culture time, along with evidence for altered immune pathways, cell cycle pathways, oxidative phosphorylation, and fatty acid metabolism (**Figure 3G**; **Suppl Table 3**). Consistent with the idea that these changes in gene expression have not (yet) translated into alterations in protein expression and functionality, analysis of insulin secretion from these donors show no impact of culture time on stimulation index within this time frame.³⁴

Control for culture time in data analysis. Culture time is largely determined by the time of day and week that the ADI IsletCore receives donor organs, and is not significantly associated with any donor metadata, technical metadata other than other measures of processing, or functional outcomes (correlation for continuous metadata, ANOVA for discrete metadata, FDR<0.05). Users can control for culture time by including 'culture time' (under 'Cell Culture Outcomes') as a covariate when performing analysis on the '*Omics Analysis*' page. Accounting for culture time when analyzing the bulk transcriptomics data improves the strength of the relationships between gene expression and nearly every other variable, resulting in decreased p-values for associated transcripts and a higher number of transcripts that pass the significance threshold (**Figure 3H**). For example, the number of transcripts with significant differential expression between donors with and without T2D increases from 418 to 542 (RNA-seq) and from 25 to 30 (NanoString) when culture time is included as a covariate.

Impacts of cell type composition. The purity of the islet preparation (percent purity) is estimated by the isolation research technician,^{35,36} leading to more frequent instances of whole numbers (**Figure 4A**). Purity is significantly associated with many transcripts and proteins in the bulk omics data (**Figure 4B**). This initially seems surprising since the islets were hand-picked to nearly 100% purity prior to acquiring these omics data. However, purity is negatively correlated with the 'percentage trapped' islets (islets embedded within acinar fragments, $r = -0.28$, p-value = $1.88e-11$, Pearson's correlation) and is significantly lower in donors with diabetes compared to those without (two-sided t-test, p-value = $1.39e-7$ for T2D versus no diabetes and p-value = $3.93e-17$ for T1D versus no diabetes). Single-cell gene expression distributions show that bulk RNA-seq and proteomics features with the strongest positive and negative associations with purity have biased expression in endocrine and non-endocrine cell types respectively (**Suppl Figures 1,2**), suggesting that their relationship with purity is explained by differential amounts of non-endocrine cells that get trapped on islet surfaces and that persist after hand picking. Thus, lower initial islet purity could be associated with more acinar tissue (even in hand-picked islets) as well as islet dysfunction. It may therefore be difficult to distinguish features that are direct reflections of lower purity preparations from those with perturbed biological function. To address this, we used the abundance of known endocrine marker proteins to estimate the relative amounts of endocrine (β -, α -, δ -, and pancreatic polypeptide (PP) cells) and non-endocrine cells in the samples with proteomics analysis (**Figure 4D**). The estimated non-endocrine proportion is negatively correlated with purity ($r = -0.54$, p-value = $1.13e-11$, Pearson correlation), and has the highest number of significantly associated features across both the bulk proteomics and

transcriptomics data (**Table 2, Suppl Table 4**). This is especially interesting because no non-endocrine markers were used in the deconvolution analysis. Instead, the non-endocrine proportion is the remainder signal left over after estimating the signal strength from each major endocrine cell type. Examination of the features most positively associated with non-endocrine proportion show highly biased expression in various non-endocrine cell types, across all omics layers. For example, *GALK1* in proteomics (FDR = 1.6e-42; **Figure 4E, F**) is most expressed in acinar and ductal cells, *RBPM5* in RNA-seq (FDR = 0.00195; **Suppl Fig 3A**) is most expressed in ductal and stellate cells, and *MYC* in NanoString (FDR = 7.7e-6; **Suppl Fig 3B**) is most expressed in stellate cells.

The α - and β -cell proportions are positively associated with many features in all three bulk omics datasets, most of which are not marker genes that were used in the deconvolution (**Table 2, Suppl Table 4**). Features with the most significant positive associations with the estimated α - and β -cell proportions have distinctive gene expression in the associated cell type, even for non-marker genes. For example, the non-marker gene features with the most significant positive associations with α -cell proportions are *DOP1B* (proteomics) and *LAT2* (bulk RNA-seq) and with β -cell proportions are *SLC27A2* (proteomics) and *HHATL* (bulk RNA-seq). Examination of the single cell distributions for these genes confirm the cell type-specific expression (**Suppl Figures 4 and 5**). Overall, there were more proteomics features associated with the estimated proportions than transcripts, and with lower p-values, even after excluding features corresponding to the marker genes used in the deconvolution analysis. This is likely because the proteomics data was used for the deconvolution, and therefore had the highest number of overlapping cell type proportion estimates and omics profiles ($n = 134$ for proteomics versus $n = 90$ and 91 for bulk RNA-seq and NanoString respectively), giving this analysis higher statistical power. Additionally, each omics dataset was from separate islet preparations. While the cell type composition should be similar across samples from the same donor, this likely contributes to the fewer significant associations observed for the bulk RNA-seq and NanoString data. Finally, while many features were significantly associated with the proportion of delta and PP cells in the proteomics data (**Table 2**), there were no significant associations with these cell types in either the bulk RNA-seq or NanoString gene expression data. However, examination of the single-cell gene expression distributions of the transcripts most significantly associated with the estimated delta and PP cell proportions (raw p-values < 0.05) shows that many of these do have distinct expression patterns in appropriate cell types. For example, in the bulk RNA-seq data, *F5*, *LRFN5*, and *SST* were the most associated with δ -cells and *FGB* and *CACNA2D3* were two of the four features most associated with PP cells (**Suppl Figures 6 and 7**).

Control for purity in data analysis. Including the ‘non-endocrine percentage’ as a covariate greatly reduces the number of bulk omics features identified as statistically significantly different between donors with T2D and those with no diabetes, which is consistent with previous findings³⁷ (**Figure 4G**). The significant feature count decreases from 1519 to 281 in the proteomics data, 418 to 43 in the RNA-seq data, and 25 to 10 in the NanoString data. Importantly, adjusting for the non-endocrine proportion eliminates or greatly reduces statistically significant features with non-endocrine biased expression from the significant RNA-seq and proteomics results (**Figure 4H**), suggesting that this approach successfully filters out signals driven by differential amounts of trapped non-endocrine tissue. However, this correction also likely decreases the statistical power for detecting features that are truly reflective of perturbed islet function and could therefore be overly stringent.

Mini Case-study: SERCA pump expression, genetic risk for diabetes, and islet function.

To illustrate a potential analysis workflow that considers the different metadata factors, we first examined proteomic data comparing ‘Corrected diabetes status’ of ‘Type 2’ vs ‘No Diabetes’. As opposed to ‘Diabetes diagnosis’ which is based on diagnosis information provided upon organ procurement, the ‘Corrected diabetes status’ additionally accounts for donor HbA1c. Donors with reported T2D remain in that category, donors with no reported T2D, but HbA1c>6.5 are now considered as T2D, and donors with no diabetes and HbA1c between 5.7-6.5 are considered ‘pre-T2D’. With no covariate correction, this comparison yields 1636 differentially expressed proteins, while correcting for age, sex, and BMI yields 1478 (FDR<0.05). Because islets isolated from donors with T2D typically have lower purity as noted above,^{38,39} we further adjusted for the calculated ‘non-endocrine cell proportion’. This resulted in 284 significantly different proteins (138 up, 146 down; FDR<0.05; **Figure 5A**).

Many of the top hits, some of which were identified in the initial analysis of these data,³³ appear enriched in endocrine cells (calculated from the proteomics deconvolution as above; **Figure 5B**). The results and analysis pipeline for this Feature View can be downloaded directly via the ‘Download Results’ and ‘Save Analysis’ buttons. This endocrine enrichment can be visualized at the single-cell RNA-seq level by clicking on ‘SC’ (shown for *ATP2A3*, **Figure 5C**), and information from the NCBI gene database for each hit can be accessed via the ‘NCBI’ link. For example, *ATP2A3* (ATPase sarcoplasmic/endoplasmic reticulum Ca^{2+} transporting 3 protein) is also known as SERCA3b, a Ca^{2+} -ATPase involved in clearing cytosolic Ca^{2+} and has been suggested to play a role in regulated insulin secretion.^{40–42}

We visualized the downregulation of ATP2A3 protein expression in islets from donors with T2D (as opposed to no diabetes or pre-T2D) by clicking on the appropriate row in the 'Results Table' (**Figure 5D**). Although this protein is downregulated in T2D, a possible outlier with very low expression can be seen in the T2D group. Clicking on this outlier will open a new tab showing the 'Donor view' page for this donor, for inspection. This donor (R241) is a 65 year old male with T2D and elevated HbA1c (9.9%) but has a very small pancreas (7th percentile; seen in the 'Isolation Information' tab), and the islet preparation has high exocrine content and very low purity (**Figure 5E**). The computed cell composition for this donor suggests a high proportion of β -cells, which we confirmed by immunostaining of banked paraffin embedded biopsy from this donor (**Figure 5E, inset**). To analyze ATP2A3 protein expression data with this potential outlier removed, we: 1) From the 'Data Download' page filtered donors by availability of proteomics data; 2) Selected 'Donor information' and 'Proteomics' as data to download in .xlsx format, and clicked 'download'; 3) Analyzed the data by 'Corrected diabetes status' ('diagnosis.computed' in the downloaded file), using including outlier removal by ROUT (Q = 5%) which removed one outlier in each of the 'Pre-T2D' and 'T2D' groups, including R241. Analysis following removal of these outliers confirmed a significant reduction in ATP2A3 expression in the T2D group (**Figure 5F**).

Via the 'Pathway Analysis' tab we performed gene-set enrichment analysis (GSEA) on the results shown in **Figure 5A** using T-statistic for weighting with an FDR cutoff of 0.1 and the KEGG library. This revealed 54 significantly enriched pathways that appear either up- or down-regulated in T2D islets. Pathways include the citrate cycle (TCA cycle), synaptic vesicle cycle, insulin secretion, lysosome, and antigen processing/presentation. To narrow potentially overlapping pathways the GSEA can be re-run with collapse redundant sets checked, which filters out gene sets with gene lists and enrichment patterns that are statistically indistinguishable from other gene sets in the library. Re-running GSEA with this setting reduced the number of significant pathways to 36 (**Figure 5G**, selected pathways shown). The most significant of these ($\text{FDR}=2.0 \times 10^{-7}$) is 'valine, leucine and isoleucine degradation', as denoted by the dark shade of green in the ridge plot and the adjusted p-value in the 'Results Table' tab. The individual features in this pathway can be explored by clicking on the ridge plot to generate a heatmap of all proteins in this pathway (**Figure 5G**). As for the protein features, the results and analysis pipeline for this Pathway Analysis can be downloaded directly via the 'Download Results' and 'Save Analysis' buttons.

To explore the evidence from human genetics for a role of variation in genes in these pathways in metabolic control, the 'T2DKP' link associated with each hit can be clicked. This

opens the relevant page in the Type 2 Diabetes Knowledge Portal to allow examination of common and rare variant associations in the gene and Human Genetic Evidence (HuGE) scores,⁴³ tissue specific gene expression, and other genetic information as described recently.³¹ Several of the endocrine-enriched hits above appear to have genetic support for a role in metabolism, with HuGE scores ranging from ‘Moderate’ (>3) to ‘Compelling’ (>350) (**Figure 5H**). *ATP2A3* for example presents “*compelling*” evidence that it is the gene mediating genetic associations at the locus with T2D (HuGE score 641.77), along with evidence for common variant associations with T2D adjusted for BMI, HbA1c, and T1D. This is in line with candidate genes studies of variants associated with T2D which have been replicated in larger genome-wide studies.⁴⁴

To explore whether *ATP2A3* is associated with other features, the feature name in the associated results table can be clicked, or, after navigating to the ‘Feature View’ page, *ATP2A3* can be directly searched. This reveals statistically significant associations of potential interest at both the protein and transcript levels, including a negative association with HbA1c (adjusted p-value = 4.4×10^{-5}); down-regulation of *ATP2A3* transcript in T2D (adjusted p-value = 0.036); a positive association with β -cell size in Patch-seq data (adjusted p-value = 4.3×10^{-12}); and positive associations with various measures of insulin secretion, such as the area under the curve for insulin secretion at 15 mM glucose (adjusted p-value = 7.6×10^{-7}) (**Figure 5I**).

DISCUSSION

The past decade has produced major efforts to generate large amounts and diverse types of data from human pancreas that can be leveraged for insights into diabetes pathophysiology and potential therapeutic targets. Access to, and integration of, these large datasets remain a challenge. Three main approaches have emerged to facilitate integrated human pancreas and islet datasets: 1) Portals that aggregate minimally processed datasets and images,^{7,18,20,21,45,46} 2) Provision of islet quality control and donor demographic data, often associated with biobanking activity,^{19,47} and 3) Implementation of pipelines for data analysis and visualization, largely in the genomic space.^{22,25,48,49} Each of these approaches are tremendously valuable in understanding islet and pancreas (patho)physiology, but naturally are subject to key limitations. Current resources occasionally include low numbers of donors, little or no donor or technical data, limited accessibility to underlying data, narrow breadth of data types, or conversely broad and deep datasets that require advanced expertise for analysis. The ADI IsletCore (www.isletcore.ca) is a large single-center biobanking program that has provided human islets for research since 2011.³⁹ We have, over more than a decade, increased the

amount of data collected from each research islet isolation through a collaborative 'distributed phenotyping network',^{14,33,39,50,51} and have now leveraged this to develop HumanIslets.com as an open resource for the community.

Our goal is to provide a flexible, open, and user-friendly platform for exploring the relationship between donor, isolation, and islet phenotypes and molecular datasets from many donors. This should be useful to the broader diabetes and islet research community via built-in analysis pipelines with standardized options and outputs, and for users who are more experienced in data processing and analysis by providing ease of access to minimally processed data in a variety of formats when possible. To balance data access, research benefits, and requirements for privacy and compliance, all data are de-identified at the source of organ procurement and sensitive data such as RNA sequencing, is provided only in a processed form while raw data remains accessible via controlled access repositories. Although the depth of pancreas phenotyping here does not presently match that provided by detailed case-control initiatives, notably the Human Pancreas Analysis Program (HPAP),^{7,18} our goal is to provide access to as broad a range of donors as possible, facilitating an understanding of both biological variation and diabetes pathophysiology. We also provide more detailed metadata describing not only donor phenotypes, but also organ processing parameters and cell culture outcomes which are not commonly included with publicly available datasets elsewhere.

Confounding relationships can be quickly viewed with other variables and across different omics modalities to inform hypotheses about causality. Visual analytics can help researchers assess potential relationships beyond the summary statistics. The technical parameters, especially, should be commonly available since they add no identifiable information. Indeed, our exploration of potential confounding technical variables led us to include culture time as covariate when generating precomputed results for the Feature View page. We did not include purity as a covariate since it is only a rough estimate of the amount of non-endocrine tissue in the analyzed samples, particularly following hand-picking, and often confounds measures of islet function. We also did not include the computed non-endocrine proportion, mainly because the estimates are only available for donors with proteomics data. This can nonetheless be informative, and we include a column in the results table that indicates whether each feature is positively associated ($FDR < 0.05$) with either the estimated endocrine or non-endocrine proportions, to alert researchers to cases where a significant relationship could be explained by variable cell type composition. Also, the cell type proportions are available under the covariates drop-down menu in the omics Analysis page, so researchers can employ it in their own customized analyses if they choose.

This resource can replicate well-established findings from us and others. We highlighted commonly accepted pathophysiology above in **Figure 5**, for example reduced oxidative phosphorylation in T2D islets has previously been shown both at the transcript and functional levels.^{52,53} Analysis of single- α -cell patch-seq data via the web tool also replicates our previous findings⁵⁴ that the expression of lineage marker transcripts (e.g. *ISL1*, *NKX2-2*, *NEUROD1*, *RFX6*) all negatively correlate with α -cell function in T2D, and with numerous developmental pathways (e.g. GSEA for Go Biological Process ‘Central Nervous System Development’, FDR=6.0x10⁻⁴; n=620 cells from 24 donors). It is important to note that some results obtained with this platform may not match exactly published findings using the same or overlapping datasets, likely due to the updated datasets and/or slightly different analysis pipelines.

In summary, [HumanIslets.com](https://www.humanislets.com) is a valuable resource for the diabetes research community, containing data and utility both for wet lab scientists, with easy analysis and visualization of data at the donor-islet-cellular level, and for computational biologists, providing straightforward downloading of minimally processed datasets. This resource continues to evolve through the ongoing collection of datasets outlined in this report, and by the addition of new data types for additional analyses. Examples of the latter, include donor genotyping; genetically predicted ancestry, and genetic risk scores for diabetes,^{55,56} prohormone processing,^{57,58} metabolomics;⁵⁹ and the pancreatic accumulation of environmental contaminants^{60,61} and lipids,⁶² among other data types that will be added in the future. In time, we hope to allow all ADI IsletCore islet and tissue recipients to provide/upload data for integration. We welcome feedback for future refinement of this resource (see ‘Feedback’ under the ‘About’ tab - <https://www.humanislets.com/#/about>), which we hope will support exploration and hypothesis generation by the diabetes, transplant, and metabolism research community.

Limitations. We attempted to balance sophisticated *versus* intuitive analysis pipelines, to serve both inexperienced and experienced users and allow meaningful analysis within the computing restrictions of a web server. Some detailed analyses will require off-line computation, and while the data processing pipeline was designed for maximum comparability across omics types there may be more optimal analyses for each individual dataset. This should be facilitated, however, through the addition of a donor/data filtering and download tool. While the ability to easily add new and expanded datasets is an important advantage of this platform, it is important to note that some results will change as new data are added (i.e. pre-computed statistics will be different after they are re-computed with more samples). As such, version tracking and reporting by users will be important. While some datasets are limited by low total donor numbers or lack

some overlap between measures (i.e. there is currently little overlap between oxygen consumption measurements and RNA-seq or proteomics results, as these remain in analyses pipelines), the continuing addition of data when available will enhance the overall functionality of the resource. Additionally, while this dataset aims to include as much technical and donor metadata as possible, some donor metadata that may be important are not available. This may include information on donor gender or reproductive status, ethnicity, and exposure history, among potentially many others. Some of our ongoing new dataset collection explicitly designed to fill these gaps include plasma measurement of sex hormones, determination of genetic ancestry, and pancreas accumulation of environmental contaminants. Finally, long-term maintenance of academic software and databases are challenging. However, our team has extensive experience in tool maintenance^{63,64} and tissue biobanking sustainability that should benefit the long-term evolution of [HumanIslets.com](https://humanislets.com) in support of diabetes and islet research.

Consortium

Members of the HumanIslets Data Analysis and Sharing (HI-DAS) Consortium are Ella Atlas (Health Canada), Austin Bautista (Alberta), Jennifer E Bruin (Carleton), Alice L Carr (Alberta), Haoning H Cen (UBC), Yi-chun Chen (UBC), Angela Ching (Carleton), Xiao-Qing Dai (Alberta), Tina Dafoe (Alberta), Theodore dos Santos (Alberta), Cara E Ellis (Alberta), Jessica D Ewald (Broad Institute), Leonard J Foster (UBC), Anna L Gloyn (Stanford), Aryana Hossein (UBC), Myriam P Hoyeck (Carleton), James D Johnson (UBC), Jelena Kolic (UBC), Yao Lu (McGill), Francis Lynn (UBC), James G Lyon (Alberta), Patrick E MacDonald (Alberta), Jocelyn E Manning Fox (Alberta), Renata Moravcova (UBC), Nadya M Morrow (Ottawa), Erin E Mulvihill (Ottawa), Andrew Pepper (Alberta), Lindsay Pallo (UBC), Varsha Rajesh (Stanford), Jason Rogalski (UBC), Shugo Sasaki (UBC), Seth A Sharp (Stanford), Nancy P Smith (Alberta), Aliya F Spigelman (Alberta), Han Sun (Stanford), Swaraj Thaman (Stanford), C Bruce Verchere (UBC), Jane Velghe (UBC), Charles Viau (McGill), Kyle van Allen (Carleton), Jessica Worton (Alberta), Jordan Wong (Alberta), Jianguo Xia (McGill), Dahai Zhang (UBC).

Methods

A general overview of methods and usage of the [HumanIslets.com](https://humanislets.com) web tool can be found in the online documentation (<http://doc.humanislets.com>) and as **Supplementary Document 2**. Subsets of all data types have been published elsewhere, with extensive details on the protocols used to perform experiments and collect the data. Here, we provide a brief overview of

the entire process, along with references to the previous descriptions of complete data collection methodology.

Islet isolation and culture. Islets were isolated from pancreas of cadaveric human donors at the Alberta Diabetes Institute IsletCore using standard procedures.^{39,65,66} The isolated islets were cultured prior to experimental use or distribution in CMRL 1066 (Corning) supplemented with 0.5% BSA (Equitech-Bio), 1% insulin-transferrin-selenium (Corning), 100 U/mL penicillin/streptomycin (Life Technologies), and L-glutamine (Sigma-Aldrich) at 22°C with 5% CO₂ at a typical seeding density of 225 IEQ/cm². Following release from the ADI IsletCore, islets were typically cultured in DMEM (11885, Gibco) supplemented with L-glutamine, 110 mg/l sodium pyruvate, 10% fetal bovine serum (FBS) (12483, Gibco), and 100 U/ml penicillin/streptomycin (15140, Gibco) at 37°C with 5% CO₂.

All research was performed with approval of the Human Research Ethics Board at the University of Alberta (Pro00013094; Pro00001754), the University of British Columbia (UBC) Clinical Research Ethics Board (H13-01865), the University of Oxford's Oxford Tropical Research Ethics Committee (OxTREC Ref.2-15), the Oxfordshire Regional Ethics Committee B (REC reference: 09/H0605/2), or by the Stanford Center for Biomedical Ethics (IRB Protocol: 57310). All donors' families gave informed consent for the use of pancreatic tissue in research.

Insulin secretion. Static glucose-stimulated insulin secretion was performed as described.⁶⁷ Briefly, on the day of sample collection islets were hand-picked and cultured overnight. Fifteen hand-picked islets, in triplicate, from each donor were pre-incubated at low glucose for two 1-hour periods, consecutively, and then sequentially treated with low glucose followed by high glucose 1 hour each. Experiments were performed with pairings of glucose concentrations as follows: 1→10mM; 1→16.7mM; and 2.8→16.7 mM. Collected supernatants, and insulin content collected by acid ethanol lysis of the pellet, were measured by ELISA (ALPCO STELLUX Chemiluminescence). For in-tool omics Analysis, a single donor-level measurement was computed as the median of the three replicate measures, reducing the impact of replicate-level outliers. Then, donor-level values were filtered to remove those that were more than 4 standard deviations away from the mean of all donors on a log10 scale. This removed values that were many orders of magnitude different from the population. Data in the Data Download tool are provided at the replicate level (before outlier removal).

Dynamic insulin secretion responses were assessed as described⁶⁸ following shipment of islets in CMRL media (Thermo Fisher Scientific) overnight from Edmonton to Vancouver.

Islets were hand-picked and cultured in RPMI 1640 medium (Thermo Fisher Scientific, Cat#, 11879-020) supplemented with 5.5 mmol glucose, 10% FBS, and 100 units/mL penicillin/streptomycin for 24-72 hours prior to experiments to allow for recovery from shipment. Perfusion columns were loaded with 65 islets and perfused at 0.4 mL/min with 3 mM glucose Krebs-Ringer Modified Buffer (KRB) solution for 1-hour. Islets were perfused with 6 mM or 15 mM glucose, and subsequently 30 mM KCl. Parallel experiments examined responses to 5 mM leucine (Sigma, Cat. # L8912, dissolved in 1M HCl, then pH adjusted with 1M NaOH) and a 1:1 mixture of 0.75 mM oleic acid (Sigma, Cat. # 364525) and 0.75 mM palmitic acid (Sigma, Cat. #P5585) at either 3 mM or 6 mM glucose (see figure legends). Fatty acids stock solutions were prepared in 50% ethanol at 65°C for 30 minutes, and then added to a solution of fatty acid free bovine serum albumin (BSA) (Sigma, Cat. # A7030) in a 6:1 molar ratio in a 37 °C water bath for 60 min. Peak insulin secretion was defined as single point whereby the amount of insulin released during the first 15 minutes of a solution change was the highest. Samples were stored at -20 °C and insulin secretion was quantified using human (Millipore Cat. # HI-14K) insulin radioimmunoassay kits.

Oxygen consumption. Agilent's Seahorse XFE24 Analyzer with the islet capture microplates was used for oxygen consumption measurements. On the day of sample collection, islets were hand-picked to 70 islets per well in triplicate. Islets were placed in the center of the well, with islet capture screens used to keep islets in place, and incubated at 37°C for 1 hour without CO₂. Islets were exposed to a modified Mito Stress Test (Seahorse XF Cell Mito Stress Test Kit, Cat. No. 103015-100) where basal measurement was at 2.8 mM glucose and stimulation at 16.7 mM glucose, then with 5 μM oligomycin, 3 μM FCCP, and 5 μM of rotenone and antimycin A. All solutions were made using DMEM (from Agilent) supplemented with 1% FBS, sodium pyruvate, and L-glutamine. After the run was complete, islets were collected and a DNA assay (QuantiFluor® dsDNA System, Promega, Madison, WI, USA) was conducted for data normalization.

Electrophysiology was performed as described.^{54,69} Islets were dispersed to single cells, plated as single cells in 35-mm dishes, and cultured for 1-3 days in RPMI with 5 mM glucose (37°C, 5% CO₂). Patch-clamp was in the whole-cell configuration using a HEKA EPC-9 amplifier (Harvard Apparatus) with 1, 5 or 10 mM glucose. Cells were held at -70 mV and depolarizations were applied to cells to elicit activation of voltage-dependent Na⁺ and Ca²⁺ channels, and exocytosis responses were measured as increases in whole-cell capacitance. Responses are

normalized to initial cell size. Following electrophysiological measurements, cells were identified by post-hoc immunostaining for insulin with a rabbit anti-insulin primary antibody (Santa Cruz; #SC-9168; RRID: AB_2126540) and goat anti-rabbit Alexa Fluor488 secondary (ThermoFisher, #A-11076; RRID: AB_141930), and with a guinea pig anti-glucagon primary antibody (Sigma-Aldrich, #G2654; RRID: AB_259852) and goat anti-guinea pig Alexa Fluor 594 secondary (ThermoFisher, #A-11076; RRID: AB_141930); or following collection for scRNA-seq.

Omics data processing

Bulk RNA-seq 150 human islets were hand-picked on the day of distribution and placed in 1ml of Trizol (Invitrogen), vortexed and stored at -80°C until shipping to Oxford or Stanford for further processing. Ensembl gene IDs were mapped to Entrez IDs and counts from any duplicated Entrez IDs were summed. Transcripts with over 80% zeros were filtered out. The RNA-seq dataset contains seven batches, and significant batch effects were observed. Batch effects were corrected for using the Combat-Seq function in the *sva* R package (version 3.44.0)⁷⁰, which preserves the count nature of the data. Counts were converted to log-counts-per-million (logCPM) using the relative log expression (RLE) method in the *edgeR* R package (version 3.38.4).⁷¹ Transcripts with logCPM variance in the lowest 20th percentile were filtered out. Since the sequencing depth was different across different batches, a substantial number of transcripts had zero counts in some batches and low but robust counts in other batches. These batch-related drop-out patterns led to large numbers of significant differentially expressed features that disappeared if individual batches were held out. To address this problem, counts with a value of zero or one were replaced with NAs, excluding these individual values from all downstream statistical analyses and leading to more stable differential expression analysis results.

PatchSeq. Reads from FASTQ files (available in the NCBI Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) under accession numbers GSE124742,⁶⁹ GSE164875,⁵⁴ at PancDB,⁷² and unpublished data to be deposited) were trimmed using TrimGalore version 0.6.5 with auto-detection of adapter sequence, then aligned and counted using STAR version 2.7.9a to the GRCh38 human genome (Ensembl release 104). Cell count matrices were loaded into R⁷³ and a Seurat object was generated using the *Seurat* R package (version 4.3.0).⁷⁴ Only cells with patch-clamp data were included; data for additional cells are available from the listed accession numbers. Cell types were identified by projection, using the MapQuery command from the R package *Seurat*, onto a large set of human islet single cell RNA-seq data generated

from publicly available data (see below). Cells were filtered to remove any that were not α - or β -cells, had fewer than 500 total counts across all transcripts, had fewer than 300 unique transcripts with non-zero count values, or had greater than 20% mitochondrial DNA. Ensembl gene IDs were mapped to Entrez IDs and counts from any duplicated Entrez IDs were summed. Counts were separated by cell type (α - and β -) and glucose concentration (1, 5, and 10mM), resulting in six different matrices. Each matrix was filtered to remove transcripts with greater than 80% zeros. Counts were converted to logCPM values using the trimmed mean of M-values (TMM) method from the *edgeR* R package (version 3.38.4).⁷¹ TMM has been shown to perform better for single-cell RNA-seq data compared to the RLE method because it is robust to the high drop-out rate. Individual transcripts that had zero counts were replaced with NA. Electrophysiology parameters other than the half-inactivation sodium current were normalized to cell size. The sign of the calcium entry, sodium current, and calcium current (early and late) electrophysiology parameters was flipped such that greater values correspond to greater currents. Negative exocytosis values were replaced with zeros.

Pseudobulk RNA-seq. The single-cell counts described above were also summarized at the donor level using a pseudobulk approach. Counts were separated by cell type (α - and β -), creating two matrices. For each matrix, the number of cells per donor was counted, and only cells from donors with at least five cells were retained. Pseudobulk profiles were created by summing counts from all cells derived from the same donor. Ensembl gene IDs were mapped to Entrez IDs and counts from any duplicated Entrez IDs were summed. Each matrix was filtered to remove transcripts with greater than 80% zeros. Counts were converted to logCPM values using the TMM method from the *edgeR* R package (version 3.38.4).⁷¹ Individual transcripts that had zero or one count were replaced with NA values.

NanoString. Fifty islets were hand-picked on the day of distribution and placed in 100 μ L RLT (Qiagen) with 1% beta mercaptoethanol (Sigma) and stored at -80°C until shipping to the University of British Columbia for processing. Briefly, 1.5 μ L cell lysate was used for hybridization with customized gene code sets for 16 hours. The hybridized complex was then loaded on one cartridge, and the NanoString nCounter Sprint Profiler was used to measure the target gene expression. Data are analyzed with nSolver4.0 software for quality control and normalization to a panel of housekeeping genes (NanoString, USA). NanoString data were collected from two different gene lists, the first with 132 transcripts and the second with 147 transcripts, and 156 unique transcripts across both. Significant batch effects were present, and so transcripts

measured in both NanoString arrays were batch corrected using the Combat function from the *sva* R package (version 3.44.0).⁷⁰ Given the much smaller number of transcripts compared to the RNA-seq data, no abundance or variance filters were applied.

Bulk proteomics. Islets were shipped to the University of British Columbia, hand-picked, and cultured in RPMI for 24-72 hours. After recovery, 300 islets were hand-picked again, washed with PBS, snap frozen as a pellet, and stored at -80°C . LC-MS/MS analysis was performed with $n = 3$ technical replicates, using a NanoElute UHPLC system (Bruker Daltonics) with Aurora Series Gen2 (CSI) analytical column coupled to a Trapped Ion Mobility-Time of Flight mass spectrometer (timsTOF Pro; Bruker Daltonics, Germany) operated in Data-Independent Acquisition-Parallel Accumulation-Serial Fragmentation (DIA-PASEF) mode. Protein abundances were filtered to remove any that had more than 50% missing values, median normalized, and log2 transformed. Missing values were imputed with the missForest method (random forest-based) using the *imp4p* R package (version 1.2.0).⁷⁵ Uniprot IDs were converted to Entrez IDs, and rows with duplicate IDs were aggregated by taking the mean. Proteins with variance in the lowest 20th percentile were filtered out.

For all omics datasets, both the minimally and fully processed matrices were saved so that they can be accessed by researchers on the Data Download page. Minimally processed datasets are the data in matrix form, but with no filters, normalization, missing value imputation, or batch correction applied. Raw data such as FASTQ files for the RNA-seq are not available for download through the tool. Instead, accession IDs indicating where they are available on public repositories are available in the donor metadata file ("RRID" column), or at the accessions listed in the data accessibility section below.

Statistical analysis

The following methods are called by the omics Analysis page to identify omics features significantly associated with selected metadata and pathways overrepresented or enriched in the association results.

Feature association analysis. Linear models are used to compute omics feature-metadata associations for the bulk RNA-seq, NanoString, pseudobulk RNA-seq, and bulk proteomics datasets. A linear model containing the selected metadata variable of interest and all covariates (if any) is fit to the expression or abundance levels of each feature in the selected omics dataset, using the *limma* R package (version 3.52.4).⁷⁶ If the user selects a categorical

metadata of interest, two dropdowns automatically appear so that a specific contrast can be specified (ie. type 2 diabetes versus no diabetes). Then, the coefficient of the metadata variable (continuous) or metadata contrast (categorical) of interest and its associated p-value are extracted from the model for each feature. P-values are adjusted using the false discovery rate method. The rationale for using the original *limma* method is that it has been shown to perform well for many omics types as long as the data are appropriately normalized and transformed. While more customized methods developed for specific omics types may perform slightly better when applied to the appropriate data type, the advantage of our approach is that it can be applied to many omics types without modification, making the results more comparable to each other and more consistent for users to interpret.

Due to the highly nonlinear nature of the relationships between single-cell gene expression and single-cell electrophysiology outcomes, Spearman correlation is used to compute feature-metadata relationships for the patchSeq data. Only single-cell electrophysiology outcomes are available for analysis with the single-cell expression data, and covariate adjustment is not supported. The single-cell expression profiles are also summarized as pseudobulk profiles for each donor; associations with donor-level clinical, technical, and functional outcomes can be investigated using this format.

Pathway analysis. Overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA) are supported for pathway analysis of the feature association results. Both methods are implemented using the *fgsea* R package (version 1.22.0).⁷⁷ The feature list used for ORA is determined by the p-value threshold specified in the association analysis. GSEA can be performed using either the model coefficient (equivalent to log2FC when the metadata is categorical) or the test statistic from the association analysis to rank the features. For the patchSeq data, the coefficient is the correlation estimate and there is no test statistic option as it is not compatible with the Spearman correlation results. There are six supported pathway libraries, including KEGG, Reactome, the full Gene Ontology term list (broken into Biological Process, Cellular Component, and Molecular Function), and the PANTHER summary of the Gene Ontology terms (also broken into the same three categories). Pathway analysis can sometimes return a high number of significant pathways or gene sets that consist of mostly the same features, leading to a redundant results list. Users have the option of collapsing redundant sets, which uses the “collapsePathways” function within the *fgsea* R package to filter the list of gene sets to only include those that have statistically significantly distinct feature lists. Pathway

analysis is not supported for the NanoString data, as there are too few features ($n = 156$) for adequate pathway coverage.

Pre-computed Results

Associations were computed between all omics feature-metadata variable pairs and between all metadata variable-metadata variable pairs. The linear model and Spearman correlation methods described above were used to compute the omics-metadata associations, with one modification. Some categorical variables have obvious specific contrasts, for example for diabetes status, one is usually interested in type 1 versus no diabetes and type 2 versus no diabetes. In these cases, we specify each contrast of interest using the methods described above. Other variables do not have obvious comparisons, for example donation type, so in this case an ANOVA style analysis is done and the statistics are extracted for the categorical variable coefficient rather than for a specific contrast. Sex, age, BMI, and culture time were included as covariates for analyses that used a linear model (all bulk omics data associations).

Three different methods were used to compute associations between pairs of metadata, depending on whether each metadata in the pair were continuous or categorical. Pearson correlation was used for continuous-continuous pairs, Cramer's V correlation was used for categorical-categorical pairs, and ANOVA was used for categorical-continuous pairs where the categorical variable is the independent variable, and the continuous variable is the dependent variable. Multiple hypothesis correction (FDR method) was performed across all results for the same statistical method. Some metadata variables such as percent purity and percent trapped are in between ordinal variables and a true continuous variable because they are estimates made by eye by a research technician, and the technician tends to estimate whole percentages (ie. 80%, 70%). For simplicity, we treat these as continuous variables.

Single-cell database compilation

Reads from FASTQ files (available in the NCBI Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) under accession numbers GSE81076,⁷⁸ GSE81547,⁷⁹ GSE81608,⁸⁰ GSE83139,⁸¹ GSE84133,³⁷ GSE85241,⁸² GSE86469,⁸³ GSE154126;⁸⁴ from the European Molecular Biology Laboratory European Biology Institute (EMBL-EBI) under accession number E-MTAB-5061,⁸⁵ at PancDB up to donor HPAP-109⁷²) were trimmed using TrimGalore version 0.6.5 with auto-detection of adapter sequence, then aligned to the GRCh38 human genome, Ensembl release 104, and counted using STAR version 2.7.9a except for files from GSE84133, which were aligned and counted using the inDrops protocol; files from

GSE81076 and GSE85241, which were aligned and counted using the R package *scruff* (version 1.10.1)⁸⁶ with *Rsubread* (version 1.22.2);⁸⁷ and files to be deposited (unpublished), which were aligned and counted with Cell Ranger (10X Genomics Cell Ranger v6.1.2). Cell count matrices were loaded into R⁷³ and objects generated using the R package *Seurat* (version 4.3.0),⁷⁴ restricted to transcripts detected in more than 10 cells and cells with between 200 and 12,500 unique transcripts with non-zero count values; cells with percent mitochondrial transcripts in the top 25% for each dataset were also dropped. Transcript names were maintained as Ensembl identifiers to preserve the greatest number of unique transcripts. After combining all donors, transcripts expressed in fewer than 0.1% of cells were filtered. Donor HPAP-027 was dropped from the data set because there were very few cells (32 cells). The reciprocal PCA integration with references (HPAP-037, HPAP-040; HPAP-045, HPAP-050, HPAP-052, and HPAP-072) workflow from the R package *Seurat* (version 4.3.0)⁷⁴ was followed, with `k.anchor = 5`, then `FindNeighbors` and `FindCluster` (resolution = 1.2). Cell types were first manually annotated based the unambiguous, high level expression of marker genes (**Suppl Table 5**) in distinct clusters in UMAP space, except cytotoxic T cells, which did not form a distinct cluster but had restricted expression of *TRAC*. The remaining ambiguous cells were annotated using the `MapQuery` function by projecting unknown cells onto the manually annotated cells.

Deconvolution analysis

After normalization for tissue sample size, the intensity of a marker protein that is only abundant in one cell type and that has relatively constant levels is proportional to the abundance of that cell type. High-quality lists of marker genes have been published for islet endocrine cell types based on a meta-analysis of seven different human islet single-cell datasets.⁸⁸ Studies have also published the range of endocrine cell type proportions across human donors, as assessed by staining and microscopy - here, we assume an average of 0.55 β -cells, 0.30 α -cells, 0.10 δ -cells, and 0.05 PP cells.^{89,90} We used this prior knowledge to estimate the cell type composition of each donor using the proteomics data. First, islet endocrine marker gene lists from van Gurp et al.⁸⁸ were filtered to remove any that were classified as markers for more than one cell type and any that were not present in the proteomics data. This resulted in 72 genes for β -cells, 78 for α -cells, 15 for δ -cells, and 8 for PP cells. Next, we computed the proportion of each cell type for each donor, according to each marker gene ($prop_{c,m}$). This was done using the following formula:

$$prop_{c,m} = (abundance_m / median(abundance_m)) * Prop_c$$

For example, the proportion of β -cells in a given donor according to the marker gene *INS* would be computed as the abundance of insulin, divided by the median insulin abundance across all donors, and multiplied by the prior known median proportion of β -cells (0.55). The overall proportion of each cell type was calculated as the median of proportion estimates across all marker genes for that cell type. Finally, the proportion of non-endocrine cells was computed as the remainder after adding the estimated proportion of β -, α -, δ -, and PP cells and normalizing the endocrine proportions to satisfy an assumed maximum purity of 98%. When computing associations between individual omics features and cell type proportions, endocrine cell proportions were normalized as a proportion of the total endocrine signal and the non-endocrine proportion was included as a covariate to better distinguish cell type-specific signals.

Tool Implementation

The *HumanIslets* frontend is implemented using Angular version 14, with the PrimeNg components library. All statistical analyses are performed in R, are managed by a Java backend with Rserve, and are communicated from the frontend to the backend via REST API. Interactive visualizations are implemented with d3 in the frontend and static visualizations are generated in the backend using the *ggplot* R package (version 3.4.2).⁹¹ The R functions used to generate all statistical results and visualizations are available at the HumanIsletsR Github repository (<https://github.com/xia-lab/HumanIsletsR>).

Data accessibility

All raw data used in this study is available for download via the *HumanIslets* web-tool, with the following exceptions for raw sequencing and proteomics data (available elsewhere). Bulk RNA sequencing data and raw proteomics data are from a recent preprint³³ and has been deposited in the European Genome-phenome Archive (EGA; [EGAS00001007241](https://ega-archive.org/studies/EGAS00001007241)) and ProteomeXchange via MASSive (PXD045422), respectively. Patch-seq sequencing reads are from published papers,^{54,69} new data, and the Human Pancreas Analysis Program (HPAP) and is available in the NCBI Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA) under accession numbers [GSE124742](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124742), [GSE164875](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164875), and at the PANC-DB data portal of the HPAP (<https://hpap.pmacs.upenn.edu>),⁷² along with data to be deposited (unpublished). Data used to generate islet cell type-specific single-cell expression profiles is from HPAP (availability as above) and, from new data (GSE269204), and from published papers.^{37,78–85} The latter are

available via NCBI GEO and SRA via accession numbers GSE81076, GSE81547, GSE81608, GSE83139, GSE84133, GSE85241, GSE86469, GSE154126, and from the European Molecular Biology Laboratory European Biology Institute (EMBL-EBI) under accession number E-MTAB-5061.

ACKNOWLEDGEMENTS

The University of Alberta acknowledges that we are located on Treaty 6 territory, and respects the histories, languages, and cultures of First Nations, Métis, Inuit, and all First Peoples of Canada, whose presence continues to enrich our vibrant community.

We thank all families and donors for the generous gifts in support of diabetes and transplantation research, and the Human Organ Procurement and Exchange (HOPE) program, Southern Alberta Organ and Tissue Donation Program (SAOTDP), Trillium Gift of Life Network (TGLN), BC Transplant, Quebec Transplant, and other Canadian organ procurement organizations for their efforts procuring pancreas for research.

We appreciate the input of Viljem Pohorec (University of Maribor), Joan Camunas-Soler (Gothenberg), Jon Campbell (Duke University) and Jane Velghe (UBC).

FUNDING

Data collection was been supported by grants from the Canadian Institutes of Health Research (MacDonald - 186226, 148451; Johnson - 168857), JDRF (MacDonald - 2-SRA-2019-698-S-B), BCCHRI Child Health Integrative Partnership Strategy Funding (Verchere, Levings and Lynn), the National Institutes of Health (MacDonald - U01-DK-120447; MacDonald/Gloyn - U01-DK-123716; Gloyn - U01-DK105535, U01-DK085545, UM-1DK126185), and the Wellcome Trust (Gloyn - 095101, 200837, 106130, 203141). Proteomics infrastructure and analysis was supported by the UBC Life Sciences Institute, Canada Foundation for Innovation, BC Knowledge Development Fund, and Genome Canada/BC (264PRO, 374PRO).

Some data used in this web tool includes patch-seq data, and single-cell RNA-seq used for cell type expression analysis, from the Human Pancreas Analysis Program (HPAP-RRID:SCR_016202) Database (<https://hpap.pmacs.upenn.edu>), a Human Islet Research Network (RRID:SCR_014393) consortium (UC4-DK-112217, U01-DK-123594, UC4-DK-112232, and U01-DK-123716).

Consolidation of datasets and web tool development was supported by a research grant funded by the Canadian Institutes of Health Research, JDRF Canada, and Diabetes Canada (5-SRA-2021-1149-S-B/TG 179092) to MacDonald (Alberta), Xia (McGill), Johnson (UBC) and

Bruin (Carleton) with collaborators Gloyn (Stanford), Foster (UBC), Atlas (Ottawa), Mulvihill (Ottawa), Lynn (UBC), and Verchere (UBC).

Declaration of Interests

ALG's spouse is an employee of Genentech and holds stock options in Roche

AP serves on the Scientific Advisory Committee for Encellin Inc.

JX is founder of XiaLab Analytics, which provides omics data science training and support.

All other authors confirm no relevant interests to declare.

REFERENCES

1. Gray, D.W.R., McShane, P., Grant, A., and Morris, P.J. (1984). A method for isolation of islets of Langerhans from the human pancreas. *Diabetes* 33, 1055–1061. <https://doi.org/10.2337/diab.33.11.1055>.
2. Ricordi, C., Lacy, P.E., Finke, E.H., Olack, B.J., and Scharp, D.W. (1988). Automated method for isolation of human pancreatic islets. *Diabetes* 37, 413–420. <https://doi.org/10.2337/diab.37.4.413>.
3. Kneteman, N.M., and Rajotte, R.V. (1985). Isolation and cryopreservation of human pancreatic islets. *Life support Syst. : J. Eur. Soc. Artif. Organs* 3 *Suppl* 1, 712–718.
4. Gloyn, A.L., Ibberson, M., Marchetti, P., Powers, A.C., Rorsman, P., Sander, M., and Solimena, M. (2022). Every islet matters: Improving the impact of human islet research. *Nat. Metab.* 4, 970–977. <https://doi.org/10.1038/s42255-022-00607-8>.
5. Marchetti, P., Schulte, A.M., Marselli, L., Schoniger, E., Bugliani, M., Kramer, W., Overbergh, L., Ullrich, S., Gloyn, A.L., Ibberson, M., et al. (2019). Fostering improved human islet research: a European perspective. *Diabetologia* 62, 1514–1516. <https://doi.org/10.1007/s00125-019-4911-4>.
6. Hart, N.J., and Powers, A.C. (2019). Use of human islets to understand islet biology and diabetes: progress, challenges and suggestions. *Diabetologia* 62, 212–222. <https://doi.org/10.1007/s00125-018-4772-2>.
7. Shapira, S.N., Naji, A., Atkinson, M.A., Powers, A.C., and Kaestner, K.H. (2022). Understanding islet dysfunction in type 2 diabetes through multidimensional pancreatic phenotyping: The Human Pancreas Analysis Program. *Cell Metab.* 34, 1906–1913. <https://doi.org/10.1016/j.cmet.2022.09.013>.
8. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L.B. da S., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
9. Callaghan, S. (2020). On the importance of data transparency. *Patterns* 1, 100070. <https://doi.org/10.1016/j.patter.2020.100070>.
10. Boué, S., Byrne, M., Hayes, A.W., Hoeng, J., and Peitsch, M.C. (2018). Embracing transparency through data sharing. *Int. J. Toxicol.* 37, 466–471. <https://doi.org/10.1177/1091581818803880>.
11. Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat Biotechnol* 32, 1121–1133. <https://doi.org/10.1038/nbt.3033>.
12. Pagliuca, F.W., Millman, J.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic β cells in vitro. *Cell* 159, 428–439. <https://doi.org/10.1016/j.cell.2014.09.040>.
13. Ramzy, A., Thompson, D.M., Ward-Hartstonge, K.A., Ivison, S., Cook, L., Garcia, R.V., Loyal, J., Kim, P.T.W., Warnock, G.L., Levings, M.K., et al. (2021). Implanted pluripotent stem-cell-derived pancreatic endoderm cells secrete glucose-responsive C-peptide in patients with type 1 diabetes. *Cell Stem Cell* 28, 2047–2061.e5. <https://doi.org/10.1016/j.stem.2021.10.003>.

14. Viñuela, A., Varshney, A., Bunt, M. van de, Prasad, R.B., Asplund, O., Bennett, A., Boehnke, M., Brown, A.A., Erdos, M.R., Fadista, J., et al. (2020). Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun* 11, 4912. <https://doi.org/10.1038/s41467-020-18581-8>.
15. Greenwald, W.W., Chiou, J., Yan, J., Qiu, Y., Dai, N., Wang, A., Nariai, N., Aylward, A., Han, J.Y., Kadakia, N., et al. (2019). Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat Commun* 10, 2078. <https://doi.org/10.1038/s41467-019-09975-4>.
16. Wang, G., Chiou, J., Zeng, C., Miller, M., Matta, I., Han, J.Y., Kadakia, N., Okino, M.-L., Beebe, E., Mallick, M., et al. (2023). Integrating genetics with single-cell multiomic measurements across disease states identifies mechanisms of beta cell dysfunction in type 2 diabetes. *Nat. Genet.* 55, 984–994. <https://doi.org/10.1038/s41588-023-01397-9>.
17. Chiou, J., Zeng, C., Cheng, Z., Han, J.Y., Schlichting, M., Miller, M., Mendez, R., Huang, S., Wang, J., Sui, Y., et al. (2021). Single-cell chromatin accessibility identifies pancreatic islet cell type– and state-specific regulatory programs of diabetes risk. *Nat Genet* 53, 455–466. <https://doi.org/10.1038/s41588-021-00823-0>.
18. Kaestner, K.H., Powers, A.C., Naji, A., Consortium, H., and Atkinson, M.A. (2019). NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes*, db190058. <https://doi.org/10.2337/db19-0058>.
19. Brissova, M., Niland, J.C., Cravens, J., Olack, B., Sowinski, J., and Evans-Molina, C. (2019). The Integrated Islet Distribution Program answers the call for improved human islet phenotyping and reporting of human islet characteristics in research articles. *Diabetes* 68, 1363–1365. <https://doi.org/10.2337/dbi19-0019>.
20. Saunders, D.C., Messmer, J., Kusmartseva, I., Beery, M.L., Yang, M., Atkinson, M.A., Powers, A.C., Cartailier, J.-P., and Brissova, M. (2020). Pancreatlas: Applying an adaptable framework to map the human pancreas in health and disease. *Patterns* 1, 100120. <https://doi.org/10.1016/j.patter.2020.100120>.
21. Morgan, N.G., Richardson, S.J., Powers, A.C., Saunders, D.C., and Brissova, M. (2022). Images from the Exeter Archival Diabetes Biobank now accessible via Pancreatlas. *Diabetes Care* 45, e174–e175. <https://doi.org/10.2337/dc22-1613>.
22. Asplund, O., Storm, P., Chandra, V., Hatem, G., Ottosson-Laakso, E., Mansour-Aly, D., Krus, U., Ibrahim, H., Ahlqvist, E., Tuomi, T., et al. (2022). Islet Gene View—a tool to facilitate islet research. *Life Sci. Alliance* 5, e202201376. <https://doi.org/10.26508/lsa.202201376>.
23. Elgamal, R.M., Kudtarkar, P., Melton, R.L., Mummey, H.M., Benaglio, P., Okino, M.-L., and Gaulton, K.J. (2023). An integrated map of cell type-specific gene expression in pancreatic islets. *Diabetes* 72, 1719–1728. <https://doi.org/10.2337/db23-0130>.
24. Patil, A.R., Schug, J., Consortium, the H., Naji, A., Kaestner, K.H., Faryabi, R.B., and Vahedi, G. (2023). Computational workflow and interactive analysis of single-cell expression profiling of islets generated by the Human Pancreas Analysis Program. *bioRxiv*, 2023.01.03.522578. <https://doi.org/10.1101/2023.01.03.522578>.
25. Alonso, L., Piron, A., Morán, I., Guindo-Martínez, M., Bonàs-Guarch, S., Atla, G., Miguel-Escalada, I., Royo, R., Puiggròs, M., Garcia-Hurtado, X., et al. (2021). TIGER: The gene

expression regulatory variation landscape of human pancreatic islets. *Cell Reports* 37, 109807. <https://doi.org/10.1016/j.celrep.2021.109807>.

26. Zimmerman, K.D., Espeland, M.A., and Langefeld, C.D. (2021). A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* 12, 738. <https://doi.org/10.1038/s41467-021-21038-1>.

27. Tung, P.-Y., Blischak, J.D., Hsiao, C.J., Knowles, D.A., Burnett, J.E., Pritchard, J.K., and Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7, 39921. <https://doi.org/10.1038/srep39921>.

28. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Kaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. *Nat. Commun.* 12, 5692. <https://doi.org/10.1038/s41467-021-25960-2>.

29. Fasolino, M., Schwartz, G.W., Patil, A.R., Mongia, A., Golson, M.L., Wang, Y.J., Morgan, A., Liu, C., Schug, J., Liu, J., et al. (2022). Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. *Nat. Metab.* 4, 284–299. <https://doi.org/10.1038/s42255-022-00531-x>.

30. Soufan, O., Ewald, J., Zhou, G., Hacariz, O., Boulanger, E., Alcaraz, A.J., Hickey, G., Maguire, S., Pain, G., Hogan, N., et al. (2022). EcoToxXplorer: Leveraging design thinking to develop a standardized web-based transcriptomics analytics platform for diverse users. *Environ. Toxicol. Chem.* 41, 21–29. <https://doi.org/10.1002/etc.5251>.

31. Costanzo, M.C., Grotthuss, M. von, Massung, J., Jang, D., Caulkins, L., Koesterer, R., Gilbert, C., Welch, R.P., Kudtarkar, P., Hoang, Q., et al. (2023). The Type 2 Diabetes Knowledge Portal: An open access genetic resource dedicated to type 2 diabetes and related traits. *Cell Metab.* 35, 695–710.e6. <https://doi.org/10.1016/j.cmet.2023.03.001>.

32. Benkahla, M.A., Sabouri, S., Kiosses, W.B., Rajendran, S., Quesada-Masachs, E., and Herrath, M.G. von (2021). HLA class I hyper-expression unmasks beta cells but not alpha cells to the immune system in pre-diabetes. *J. Autoimmun.* 119, 102628. <https://doi.org/10.1016/j.jaut.2021.102628>.

33. Kolic, J., Sun, W.G., Cen, H., Ewald, J., Beet, L., Moravcova, R., Rogalski, J.C., Sasaki, S., Sun, H., Rajesh, V., et al. (2023). Proteomic predictors of individualized nutrient-specific insulin secretion in health and disease. *medRxiv*, 2023.05.24.23290298. <https://doi.org/10.1101/2023.05.24.23290298>.

34. Lyon, J.G., Carr, A.L., Smith, N.P., Marfil-Garza, B., Spigelman, A.F., Bautista, A., O’Gorman, D., Kin, T., Shapiro, A.J., Peter, et al. Human research islet cell culture outcomes at the Alberta Diabetes Institute IsletCore. *bioRxiv*.

35. Lyon, J., Spigelman, A.F., Fox, J.E.M., and Macdonald, P.E. (2021). Sampling of human islets for quality control purposes v2. <https://doi.org/10.17504/protocols.io.bupbnvin>.

36. Lyon, J., Spigelman, A.F., Fox, J.E.M., and Macdonald, P.E. (2018). Human islet quantification and purity assessment v1. <https://doi.org/10.17504/protocols.io.si3ecgn>.

37. Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* 3, 346–360.e4. <https://doi.org/10.1016/j.cels.2016.08.011>.

38. Liang, R., Sun, P., Cai, X., Liu, T., Wang, L., Liu, N., Ding, X., Zhang, B., Liu, Z., Zou, J., et al. (2023). Characteristics of research-focused human islet preparations from organ donors with type 2 diabetes. *Islets* 15, 2219104. <https://doi.org/10.1080/19382014.2023.2219104>.
39. Lyon, J., Fox, J.E.M., Spigelman, A.F., Kim, R., Smith, N., O’Gorman, D., Kin, T., Shapiro, A.M.J., Rajotte, R.V., and MacDonald, P.E. (2016). Research-focused isolation of human islets from donors with and without diabetes at the Alberta Diabetes Institute IsletCore. *Endocrinology* 157, 560–569. <https://doi.org/10.1210/en.2015-1562>.
40. Ravier, M.A., Daro, D., Roma, L.P., Jonas, J.-C., Cheng-Xue, R., Schuit, F.C., and Gilon, P. (2011). Mechanisms of control of the free Ca^{2+} concentration in the endoplasmic reticulum of mouse pancreatic β -cells. *Diabetes* 60, 2533–2545. <https://doi.org/10.2337/db10-1543>.
41. Beauvois, M.C., Merezak, C., Jonas, J.-C., Ravier, M.A., Henquin, J.-C., and Gilon, P. (2006). Glucose-induced mixed $[\text{Ca}^{2+}]_c$ oscillations in mouse β -cells are controlled by the membrane potential and the SERCA3 Ca^{2+} -ATPase of the endoplasmic reticulum. *Am. J. Physiol.-Cell Physiol.* 290, C1503–C1511. <https://doi.org/10.1152/ajpcell.00400.2005>.
42. Xu, G.G., Gao, Z., Borge, P.D., Jegier, P.A., Young, R.A., and Wolf, B.A. (2000). Insulin regulation of β -cell function involves a feedback loop on SERCA gene expression, Ca^{2+} homeostasis, and insulin expression and secretion. *Biochemistry* 39, 14912–14919. <https://doi.org/10.1021/bi001260w>.
43. Dornbos, P., Singh, P., Jang, D.-K., Mahajan, A., Biddinger, S.B., Rotter, J.I., McCarthy, M.I., and Flannick, J. (2022). Evaluating human genetic support for hypothesized metabolic disease genes. *Cell Metab.* 34, 661–666. <https://doi.org/10.1016/j.cmet.2022.03.011>.
44. Varadi, A., Lebel, L., Hashim, Y., Mehta, Z., Ashcroft, S.J.H., and Turner, R. (1999). Sequence variants of the sarco(endo)plasmic reticulum Ca^{2+} -transport ATPase 3 gene (SERCA3) in Caucasian Type II diabetic patients (UK Prospective Diabetes Study 48). *Diabetologia* 42, 1240–1243. <https://doi.org/10.1007/s001250051298>.
45. Boer, P. de, Pirozzi, N.M., Wolters, A.H.G., Kuipers, J., Kusmartseva, I., Atkinson, M.A., Campbell-Thompson, M., and Giepmans, B.N.G. (2020). Large-scale electron microscopy database for human type 1 diabetes. *Nat. Commun.* 11, 2475. <https://doi.org/10.1038/s41467-020-16287-5>.
46. Perry, D.J., Shapiro, M.R., Chamberlain, S.W., Kusmartseva, I., Chamala, S., Balzano-Nogueira, L., Yang, M., Brant, J.O., Brusko, M., Williams, M.D., et al. (2023). A genomic data archive from the Network for Pancreatic Organ donors with Diabetes. *Sci. Data* 10, 323. <https://doi.org/10.1038/s41597-023-02244-6>.
47. Campbell-Thompson, M., Wasserfall, C., Kaddis, J., Albanese-O’Neill, A., Staeva, T., Nierras, C., Moraski, J., Rowe, P., Gianani, R., Eisenbarth, G., et al. (2012). Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. *DiabetesMetab. Res. Rev.* 28, 608–617. <https://doi.org/10.1002/dmrr.2316>.
48. Mularoni, L., Ramos-Rodríguez, M., and Pasquali, L. (2017). The pancreatic islet regulome browser. *Front. Genet.* 8, 13. <https://doi.org/10.3389/fgene.2017.00013>.
49. Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D., Orchard, P., Wolford, B.N., Kursawe, R., et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc. Natl. Acad. Sci.* 114, 2301–2306. <https://doi.org/10.1073/pnas.1621192114>.

50. Fox, J.E.M., Lyon, J., Dai, X.Q., Wright, R.C., Hayward, J., Bunt, M. van de, Kin, T., Shapiro, A.M.J., McCarthy, M.I., Gloyn, A.L., et al. (2015). Human islet function following 20 years of cryogenic biobanking. *Diabetologia* 58, 1503–1512. <https://doi.org/10.1007/s00125-015-3598-4>.
51. Bunt, M. van de, Fox, J.E.M., Dai, X., Barrett, A., Grey, C., Li, L., Bennett, A.J., Johnson, P.R., Rajotte, R.V., Gaulton, K.J., et al. (2015). Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *Plos Genet* 11, e1005694. <https://doi.org/10.1371/journal.pgen.1005694>.
52. Olsson, A.H., Yang, B.T., Hall, E., Taneera, J., Salehi, A., Nitert, M.D., and Ling, C. (2011). Decreased expression of genes involved in oxidative phosphorylation in human pancreatic islets from patients with type 2 diabetes. *Eur. J. Endocrinol.* 165, 589–595. <https://doi.org/10.1530/eje-11-0282>.
53. Wang, Z., Gurlo, T., Matveyenko, A.V., Elashoff, D., Wang, P., Rosenberger, M., Junge, J.A., Stevens, R.C., White, K.L., Fraser, S.E., et al. (2021). Live-cell imaging of glucose-induced metabolic coupling of β and α cell metabolism in health and type 2 diabetes. *Commun. Biol.* 4, 594. <https://doi.org/10.1038/s42003-021-02113-1>.
54. Dai, X.-Q., Camunas-Soler, J., Briant, L.J.B., Santos, T. dos, Spigelman, A.F., Walker, E.M., Drigo, R.A. e, Bautista, A., Jones, R.C., Avrahami, D., et al. (2022). Heterogenous impairment of α cell function in type 2 diabetes is linked to cell maturation state. *Cell Metab.* 34, 256-268.e5. <https://doi.org/10.1016/j.cmet.2021.12.021>.
55. Li, Y., Chen, G.-C., Moon, J.-Y., Arthur, R., Sotres-Alvarez, D., Daviglus, M.L., Pirzada, A., Mattei, J., Perreira, K.M., Rotter, J.I., et al. (2024). Genetic subtypes of prediabetes, healthy lifestyle, and risk of type 2 diabetes. *Diabetes*. <https://doi.org/10.2337/db23-0699>.
56. Ghatan, S., Rooij, J. van, Hoek, M. van, Boer, C.G., Felix, J.F., Kavousi, M., Jaddoe, V.W., Sijbrands, E.J.G., Medina-Gomez, C., Rivadeneira, F., et al. (2024). Defining type 2 diabetes polygenic risk scores through colocalization and network-based clustering of metabolic trait genetic associations. *Genome Med.* 16, 10. <https://doi.org/10.1186/s13073-023-01255-7>.
57. Chen, Y., Klimek-Abercrombie, A.M., Potter, K.J., Pallo, L.P., Soukhatcheva, G., Dai, L., Bellin, M.D., and Verchere, C.B. (2022). Elevated islet prohormone ratios as indicators of insulin dependency in auto-islet transplant recipients. *Am. J. Transplant.* 22, 1992–2005. <https://doi.org/10.1111/ajt.17076>.
58. Rodriguez-Calvo, T., Chen, Y.-C., Verchere, C.B., Haataja, L., Arvan, P., Leete, P., Richardson, S.J., Morgan, N.G., Qian, W.-J., Pugliese, A., et al. (2021). Altered β -cell prohormone processing and secretion in type 1 diabetes. *Diabetes* 70, 1038–1050. <https://doi.org/10.2337/dbi20-0034>.
59. Pang, Z., Lu, Y., Zhou, G., Hui, F., Xu, L., Viau, C., Spigelman, A.F., MacDonald, P.E., Wishart, D.S., Li, S., et al. (2024). MetaboAnalyst 6.0: Towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res.*, gkae253. <https://doi.org/10.1093/nar/gkae253>.
60. Gang, N., Allen, K.V., Villeneuve, P.J., MacDonald, H., and Bruin, J.E. (2022). Sex-specific associations between type 2 diabetes incidence and exposure to dioxin and dioxin-like pollutants: A meta-analysis. *Front. Toxicol.* 3, 685840. <https://doi.org/10.3389/ftox.2021.685840>.

61. Hoyeck, M.P., Matteo, G., MacFarlane, E.M., Perera, I., and Bruin, J.E. (2022). Persistent organic pollutants and β -cell toxicity: a comprehensive review. *Am. J. Physiol.-Endocrinol. Metab.* 322, E383–E413. <https://doi.org/10.1152/ajpendo.00358.2021>.
62. Wagner, R., Eckstein, S.S., Yamazaki, H., Gerst, F., Machann, J., Jaghutriz, B.A., Schürmann, A., Solimena, M., Singer, S., Königsrainer, A., et al. (2022). Metabolic implications of pancreatic fat accumulation. *Nat. Rev. Endocrinol.* 18, 43–54. <https://doi.org/10.1038/s41574-021-00573-3>.
63. Liu, P., Ewald, J., Pang, Z., Legrand, E., Jeon, Y.S., Sangiovanni, J., Hacariz, O., Zhou, G., Head, J.A., Basu, N., et al. (2023). ExpressAnalyst: A unified platform for RNA-sequencing analysis in non-model species. *Nat. Commun.* 14, 2995. <https://doi.org/10.1038/s41467-023-38785-y>.
64. Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., and Xia, J. (2022). Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat. Protoc.* 17, 1735–1761. <https://doi.org/10.1038/s41596-022-00710-w>.
65. Kin, T., and Shapiro, A.M.J. (2010). Surgical aspects of human islet isolation. *Islets* 2, 265–273. <https://doi.org/10.4161/isl.2.5.13019>.
66. Lyon, J., Kin, T., Bautista, A., Smith, N., O, D., MacDonald, P., and Manning, J. (2019). Isolation of human pancreatic islets of Langerhans for research v2. protocols.io. <https://doi.org/10.17504/protocols.io.xgsfjwe>.
67. Spigelman, A. (2019). Static glucose-stimulated insulin secretion (GSIS) protocol - human islets v2 (protocols.io.wy4ffyw). protocols.io. <https://doi.org/10.17504/protocols.io.wy4ffyw>.
68. Kolic, J., Beet, L., Overby, P., Cen, H.H., Panzhinskiy, E., Ure, D.R., Cross, J.L., Huizinga, R.B., and Johnson, J.D. (2020). Differential effects of voclosporin and tacrolimus on insulin secretion from human islets. *Endocrinology* 161, bqaa162-. <https://doi.org/10.1210/endo/bqaa162>.
69. Camunas-Soler, J., Dai, X.-Q., Hang, Y., Bautista, A., Lyon, J., Suzuki, K., Kim, S.K., Quake, S.R., and MacDonald, P.E. (2020). Patch-seq links single-cell transcriptomes to human islet dysfunction in diabetes. *Cell Metab.* 31, 1017–1031.e4. <https://doi.org/10.1016/j.cmet.2020.04.005>.
70. Leek, J.T., Johnson, W.E., Parker, H.S., Fertig, E.J., Jaffe, A.E., Zhang, Y., Storey, J.D., and Torres, L.C. (2024). sva: Surrogate Variable Analysis. Preprint at R.
71. Chen, Y., Lun, A.T.L., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* 5, 1438. <https://doi.org/10.12688/f1000research.8987.2>.
72. Kaestner, K.H., Powers, A.C., Naji, A., Consortium, H., and Atkinson, M.A. (2019). NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes* 68, 1394–1402. <https://doi.org/10.2337/db19-0058>.
73. Team, R.C. (2024). R: A language and environment for statistical computing. Preprint at R Foundation for Statistical Computing.

74. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573-3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
75. Gianetto, Q.G. (2021). imp4p: Imputation for Proteomics. Preprint.
76. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 43, e47–e47. <https://doi.org/10.1093/nar/gkv007>.
77. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012–060012. <https://doi.org/10.1101/060012>.
78. Grün, D., Muraro, M.J., Boisset, J.-C., Wiebrands, K., Lyubimova, A., Dharmadhikari, G., van den Born, M., van Es, J., Jansen, E., Clevers, H., et al. (2016). De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* 19, 266–277. <https://doi.org/10.1016/j.stem.2016.05.010>.
79. Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K., and Quake, S.R. (2017). Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* 171, 321-330.e14. <https://doi.org/10.1016/j.cell.2017.09.004>.
80. Xin, Y., Kim, J., Okamoto, H., Ni, M., Wei, Y., Adler, C., Murphy, A.J., Yancopoulos, G.D., Lin, C., and Gromada, J. (2016). RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab* 24, 608–615. <https://doi.org/10.1016/j.cmet.2016.08.018>.
81. Wang, Y.J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., Golson, M.L., and Kaestner, K.H. (2016). Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* 65, 3028–3038. <https://doi.org/10.2337/db16-0405>.
82. Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., et al. (2016). A single-cell transcriptome atlas of the human pancreas. *Cell Syst* 3, 385-394.e3. <https://doi.org/10.1016/j.cels.2016.09.002>.
83. Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., Kycia, I., Robson, P., and Stitzel, M.L. (2017). Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res* 27, 208–222. <https://doi.org/10.1101/gr.212720.116>.
84. Avrahami, D., Wang, Y.J., Schug, J., Feleke, E., Gao, L., Liu, C., Consortium, H., Naji, A., Glaser, B., and Kaestner, K.H. (2020). Single-cell transcriptomics of human islet ontogeny defines the molecular basis of β -cell dedifferentiation in T2D. *Mol Metab* 42, 101057. <https://doi.org/10.1016/j.molmet.2020.101057>.
85. Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.-M., Andréasson, A.-C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M.K., et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab* 24, 593–607. <https://doi.org/10.1016/j.cmet.2016.08.020>.
86. Wang, Z., Hu, J., Johnson, W.E., and Campbell, J.D. (2019). scruff: an R/Bioconductor package for preprocessing single-cell RNA-sequencing data. *BMC Bioinformatics* 20, 222–222. <https://doi.org/10.1186/s12859-019-2797-2>.

87. Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* 47, e47–e47. <https://doi.org/10.1093/nar/gkz114>.
88. Gurp, L. van, Fodoulia, L., Oropeza, D., Furuyama, K., Bru-Tari, E., Vu, A.N., Kaddis, J.S., Rodríguez, I., Thorel, F., and Herrera, P.L. (2022). Generation of human islet cell type-specific identity genesets. *Nat. Commun.* 13, 2020. <https://doi.org/10.1038/s41467-022-29588-8>.
89. Olehnik, S.K., Fowler, J.L., Avramovich, G., and Hara, M. (2017). Quantitative analysis of intra- and inter-individual variability of human beta-cell mass. *Sci. Rep.* 7, 16398. <https://doi.org/10.1038/s41598-017-16300-w>.
90. Ashcroft, F.M., and Rorsman, P. (2012). Diabetes mellitus and the β cell: The last ten years. *Cell* 148, 1160–1171. <https://doi.org/10.1016/j.cell.2012.02.010>.
91. Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., and Woo, K. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York).

Figure Legends

Figure 1. A human research islet phenotyping, sample and data collection, and data analysis and visualization pipeline.

A) Human research pancreas received by the Alberta Diabetes Institute (ADI) IsletCore is processed for sample collection (grey) and islet isolation (FFPE – formalin fixed paraffin embedded).

B) Metadata and data are collected on the organ donor, organ and tissue processing, islet and cellular function, and molecular profiling via several omics platforms.

C) The *HumanIslets* tool comprises both a backend server for data storage and computation, and a user-friendly frontend for interactive data analysis and visualization.

Figure 2. Profiles of donors, islets and data types within the database.

A-E) Distribution of donor age (A), BMI (B), HbA1c (C), sex (D) and corrected diabetes status (E) across the 540 donors currently in the database.

F) Distribution of islet isolation outcomes across the 540 donors currently in the database.

G) Heatmap showing the availability of metadata and data types within the database and for download.

Figure 3. Effect of culture time on transcript and protein expression, and the impact of including ‘culture time’ as covariate in analysis.

A) The time in culture following research islet isolation and prior to sample collection or shipment. Times are clustered due to collection and shipment during business hours (usually in the morning).

B-C) Effect of culture time on transcript (B) and protein (C) expression in human islets. Adjusted p-value <0.05 indicated in red.

D-F) Impact of culture time on expression of *FGF* measured by RNA-seq (D), NanoString (E), or *FGF* measured by proteomics (F).

G) Gene set enrichment analysis using T-statistic as weighting against KEGG (black), Reactome (blue), and GO: Biological Process (red) genesets. Selected significant pathways (FDR<0.05) are shown. See also **Suppl Table X**.

H) Impact of adjusting for culture time on gene-level p-values for the association of bulk RNA-seq data and diabetes status (T2D vs. None). Transcripts were either significant (FDR<0.05) both with and without adjusting for culture time (dark grey), only when adjusting for culture time (red), only when not adjusting for culture time (blue), or not significant in either scenario (light grey).

Figure 4. Effect of preparation purity on transcript and protein expression, and the use of cell type deconvolution analysis in correction for exocrine contamination.

A) Estimate purity of preparations following isolation, with donors ranked in order by purity.

B-C) Although islets are hand-picked prior to these analyses, post-isolation purity has an impact on both transcript (B) and protein (C) expression in human islets. Adjusted p-value <0.05 indicated in red.

D) Cell type distributions computed from proteomic data across donors, ranked by 'non-endocrine' cell proportions.

E-F) Correlation of GALK1 protein with the proportion of non-endocrine cells (E) and confirmation in single-cell RNA-seq data that *GALK1* is expressed primarily in acinar and ductal cells (F).

G) Impact of adjusting for the proportion of non-endocrine cells on protein-level p-values for the association of bulk proteomics data and diabetes status (T2D vs. None). Proteins were either significant (FDR<0.05) both with and without adjusting for non-endocrine cells (dark grey), only when adjusting for non-endocrine cells (red), only when not adjusting for non-endocrine cells (blue), or not significant in either scenario (light grey).

H) Of α -, β - and non-endocrine-associated RNA-seq and proteomics features found to be significantly different (adjusted p-value <0.05) when comparing T2D versus no diabetes with and without adjustment for the computed proportion of non-endocrine cell types (non-adjusted RNA-seq, n=117; adjusted RNA-seq, n=90; non-adjusted proteomics, n=134; adjusted proteomics, n=134).

Figure 5. Mini-case study: Linking SERCA3b expression with diabetes and islet function.

A) Analysis of proteomic data comparing the corrected diabetes status of 'no diabetes' versus 'type 2 diabetes'. Analyses were run with no correction, with addition of age, sex and BMI as covariates, and then with addition of 'proportion non-endocrine' as a covariate. Adjusted p-value <0.05 indicated in red.

B) Web-tool Results Table, with top hits sorted by adjusted p-value.

C) Clicking on 'SC' (single-cell) raises a visualization of single-cell RNA-seq data for a particular hit (in this case *ATP2A3*) to confirm cell-type specific expression.

D) Clicking on a row of the Results Table raises the protein expression of a hit (in this case *ATP2A3*) separated by the initial query. This example shows *ATP2A3* expression in donors with no diabetes (ND), pre-type 2 diabetes (pre-T2D) and T2D (Type2). A potential outlier donor is indicated in the red circle.

E) Clicking on the potential outlier raises the Donor View page for that donor for inspection. Shown here are the donor metadata, HbA1c and pancreas weight in relation to the total database donor population, and cell type composition for this potential outlier (R241). Composition calculations are from the proteomics deconvolution, with exocrine cells shown as a proportion of all cells and endocrine cell types shown as a proportion of all endocrine cells. Inset, the presence of both insulin (green) and glucagon (red) positive cells in R241 is confirmed by fluorescence immunostaining in banked FFPE biopsy from this donor.

F) Proteomic data downloaded from the Data Download page, can be analyzed by the user. In this case the potential outlier was removed from analysis of *ATP2A3* expression (****-p<0.0001; one-way ANOVA followed by Tukey post-test to compare groups).

G) Gene set enrichment analysis, via the 'pathway analysis' tab highlights pathways up-regulated (blue) and down-regulated (red) in T2D. Results were exported in table format and

plotted using local software. Within the tool, clicking on a pathway result raises a heatmap showing genes/proteins contributing to the pathway. Significant hits are indicated by ***.

H) In the results table (panel B), clicking on 'NCBI' takes users to the NCBI gene page for a given hit, while clicking on 'T2DKP' takes users to the T2D Knowledge Portal page. Data extracted from the T2DKP page for top endocrine-enriched hits demonstrates Human Genetic Evidence (HuGE) scores suggesting potential links to human metabolism and diabetes. Red shows HuGE scores for *ATP2A3*.

I) In the results table (panel B), clicking on the hit name (in this case *ATP2A3*) takes users to the Feature View results for that particular hit. Here, *ATP2A3* is seen to be negatively (red) associated with T2D at both the transcript and protein levels, and positively (blue) associated with some measures of beta-cell function and insulin secretion.

Tables

Table 1. Number of omics features significantly associated with key metadata (adj. P-value < 0.05, FDR method). Table cells with an “N/A” indicate that the analysis was not performed because there were fewer than 10 donors with omics profiles per group for categorical variables, or fewer than 10 donors overall for continuous variables. Note that HbA1c data are not available for all donors and there were a few donors with T1D, thus the sample size is slightly less than the number of omics profiles for some of the T2D-ND and HbA1c analyses.

	Bulk RNA-seq	Nanostring	Proteomics	Pseudobulk RNA-seq (α)	Pseudobulk RNA-seq (β)
Features	14 314	156	6138	16 874	14 950
Donor profiles	117	190	134	52	30
Significant batch effects	Yes	Yes	No	No	No
Key diabetes outcomes					
T2D-ND	418	51	1519	0	N/A
HbA1c (%)	5	94	781	0	5
Selected clinical covariates					
Age (years)	105	39	46	0	0
Sex	27	0	20	12	8
BMI	0	1	0	0	0
Selected technical covariates					
Cold ischemic time (h)	2	0	2	0	0
Purity (%)	2245	98	2414	1	0
Culture time (h)	7547	30	1	0	0

Table 2. Features positively associated (FDR < 0.05) with each estimated cell type proportion. Beta, alpha, delta, and PP proportions are the % of all endocrine cells, and the association analysis for these cell types included non-endocrine % as a covariate. The percentage of features that were used as marker genes is shown in brackets.

Cell type	Proteomics	RNA-seq	Nanostring
Non-endocrine %	1491 (0%)	849 (0%)	20 (0%)
Beta cell %	352 (5%)	312 (11%)	3 (0%)
Alpha cell %	652 (10%)	168 (13%)	7 (57%)
Delta cell %	1036 (1%)	0 (0%)	0 (0%)
PP cell %	721 (1%)	0 (0%)	0 (0%)









