BIOLOGY Methods & Protocols

Biology Methods and Protocols, 2019, 1–14

doi: 10.1093/biomethods/bpz019
Methods Manuscript

METHODS MANUSCRIPT

Single-cell transcriptomics from human pancreatic islets: sample preparation matters

Lori L. Bonnycastle [1], Derek E. Gildea[2], Tingfen Yan[1], Narisu Narisu[1], Amy J. Swift[1], Tyra G. Wolfsberg[2], Michael R. Erdos[1], and Francis S. Collins [1,*]

[1]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA and [2]Bioinformatics and Scientific Programming Core, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

*Correspondence address. Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. Tel: 301-496-2433; Fax: 301-402-2700; E-mail: collinsf@mail.nih.gov

Abstract

Single-cell RNA sequencing (scRNA-seq) of human primary tissues is a rapidly emerging tool for investigating human health and disease at the molecular level. However, optimal processing of solid tissues presents a number of technical and logistical challenges, especially for tissues that are only available at autopsy, which includes pancreatic islets, a tissue that is highly relevant to diabetes. To assess the possible effects of different sample preparation protocols on fresh islet samples, we performed a detailed comparison of scRNA-seq data generated with islets isolated from a human donor but processed according to four treatment strategies, including fixation and cryopreservation. We found significant and reproducible differences in the proportion of cell types identified, and more minor effects on cell-specific patterns of gene expression. Fresh islets from a second donor confirmed gene expression signatures of alpha and beta subclusters. These findings may well apply to other tissues, emphasizing the need for careful consideration when choosing processing methods, comparing results between different studies, and/or interpreting data in the context of multiple cell types from preserved tissue.

Keywords: scRNA-seq; islets; tissue processing; transcriptome

Introduction

Type 2 diabetes (T2D) is a growing worldwide problem characterized by insufficient insulin production in the face of peripheral insulin resistance. Genome-wide association studies have identified up to 243 T2D risk loci [1], and many of those point to the pancreatic islet as the relevant tissue. More than 90% of the T2D risk variants fall in noncoding regions, however, indicating that they act through regulatory mechanisms. Interrogating the transcriptome of the human islet, the source of insulin and other important metabolic hormones, is therefore expected to provide valuable insights into the pathophysiology of diabetes. Single-cell RNA sequencing (scRNA-seq) offers the ability to assess the transcriptomes of individual cells from pancreatic tissue, a heterogeneous population of exocrine cells (acinar and ductal) and hormone-secreting endocrine cells (beta, alpha, delta, gamma, and epsilon).

A number of scRNA-seq technologies with variable technical and cost efficiencies are currently available to investigate a wide range of organ/tissue systems (reviewed in [2–6]). But the biological validity of scRNA-seq data with regard to capturing the native transcriptional state of the cell is critically dependent on minimizing technical perturbations that may stress and/or

Received: 21 October 2019; **Revised:** 26 November 2019; **Editorial decision:** 28 November 2019; **Accepted:** 11 December 2019

damage cells. Such perturbations can occur during tissue harvesting, single-cell isolation, and/or preservation of cells prior to RNA-seq. These issues are expected to be less significant for tissues such as blood lymphocytes that can be obtained fresh, and that are already present in single-cell suspensions. But solid tissues obtained by biopsy or autopsy require more extensive treatments for harvesting and dissociation into single cells. The ability to use methods such as fixation or cryopreservation can offset the logistical complexities of matching the timing of tissue acquisition and sample preparation—but the effect of such methods on single-cell gene expression has received limited attention.

Human islets secrete insulin, glucagon, and other important hormones that regulate metabolism, and this a critically important tissue for research on the pathophysiology of diabetes. But human islets can currently only be obtained at autopsy, and only a subset of this limited resource is released for scientific research. Receiving fresh human pancreatic islets for experimental analyses is thus challenging, infrequent, and unpredictable. The timing of scRNA-seq processing of islet tissue relies on several factors including (i) tissue availability, (ii) tissue transport time to the processing and sequencing laboratory, and (iii) laboratory personnel and equipment availability at the time of receipt of tissue. Ideally, the ability to process multiple tissue samples from different donors in parallel could contribute to technical, time, and cost efficiencies, and reduce batch effects between scRNA-seq runs. Thus, it would be highly desirable to identify a sample preparation option that could preserve tissues and/or cells for scRNA-seq processing at a later time.

A few groups have preserved tissues/cells from various sources for scRNA-seq utilizing fixation [7–10] or cryopreservation [11, 12] strategies, and recovered RNA-seq data that correlate reasonably well with that obtained from fresh cells. But pancreatic islets may represent a particularly challenging case. In this work, we report a comparison of single-cell islet transcriptomic data from fresh material (considered as the "gold standard") to those obtained after methanol fixation or cryopreservation. We find significant cell type-specific differences with regard to cell abundance and gene expression that should inform future decisions about sample preparation for scRNA-seq.

Material and methods

Origin and processing of pancreatic islets

Purified human pancreatic islets were obtained from Prodo Laboratories (Aliso Viejo, CA) and were isolated from cadaverous donors whose organs were consented for research. As per Office for the Protection of Research Subjects policy, islets obtained from nonliving individuals do not fall under the guidelines of human subject research. All experimental protocols performed for this study are approved under the National Institutes of Health (NIH) guidelines.

The main source Donor 1 (whose islet cells were used in each of the preservation treatment groups in this manuscript) was a nondiabetic (HbA1c = 4.2%) 28-year-old Caucasian male, with BMI of 34.7. The isolated islets were 95% pure by dithizone staining and 95% viable by dye exclusion [13]. Islet size index indicates 93% of the purified islets were ≥100um in size. A second source Donor 2, referred to as "Fresh-Donor 2" and used to support findings of the Donor 1 "Fresh" samples, was a diabetic (HbA1c = 7.1%) 61-year-old Haitian male, with BMI of 27.31 and islet purity of 85%. The purified islets were cultured in Prodo

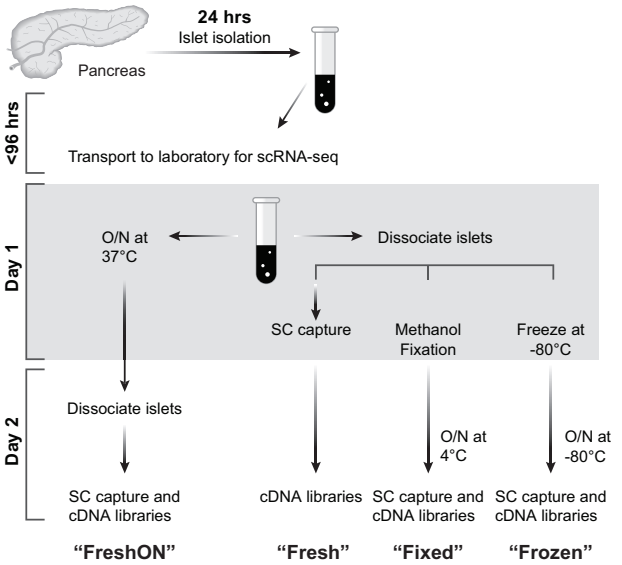

Figure 1: Experimental design and preservation treatment groups for Donor 1. For each of the four sample preparation methods, biological triplicates were prepared for scRNA-seq. Drawing by Julia Fekecs, NHGRI-NIH.

Islet Media [PIM(S)] complete media (Prodo Laboratories, Aliso Viejo, CA, USA) at a density of 10 000 Islet Equivalents (IEQ)/150 mm^2 for 72 h at 37°C. Islets were then packaged and transported to our laboratory at 4°C over a period of ~24 h. Upon receipt, islets were equilibrated to 37°C for 1 h prior to downstream processing for either bulk sequencing or scRNA-seq under four treatment conditions.

Main islet source (Donor 1) for single-cell preservation treatment groups

Twelve aliquots, each with 3000 IEQ from a single source of human pancreatic islets, were assembled to test, in triplicate four conditions of preserving islet cells for scRNA-seq (Figure 1). Treatment 1 islets ("Fresh") were dissociated to single cells immediately after 1-h equilibration at 37°C and then put through the scRNA-seq pipeline beginning on Day 1. Treatment 2 islets ("FreshON") were incubated overnight at 37°C prior to dissociation and scRNA-seq on Day 2. Treatment 3 islets ("Fixed") were dissociated to single cells immediately upon arrival, fixed with methanol and stored overnight at 4°C prior to scRNA-seq on Day 2. Treatment 4 islets ("Frozen") were dissociated to single cells immediately upon arrival, frozen at −80°C overnight prior to scRNA-seq on Day 2.

Additional islet source (Fresh-Donor 2) for Fresh treatment group

Two aliquots (to serve as replicates), each with 3000 IEQ, were tested under the "Fresh" condition as described for Donor 1. The two Fresh-Donor 2 samples were processed separately from 12 Donor 1 samples, including library preparation, sequencing, and data analysis.

Dissociation of pancreatic islets

Each islet aliquot was dissociated as follows: 3000 IEQ in 1 mL Accutase solution (Innovation Cell Technologies, Inc) were incubated at 37°C for 10 min and then washed with 2 mL PIM(S)™ (Prodo Islet Media, Prodo Laboratory Inc, Irvine, CA, USA). This

was followed with a 10 min 37°C incubation in 3 mL PBS with 5 U Dispase I (Roche Diagnostics)/3 U DNase I (ThermoFisher Scientific), then washed once and resuspended in PIM(S). The final cell suspension was passed through a BD 40 μm cell strainer to remove aggregates, and then assessed for cell viability and cell numbers via staining with acridine orange and DAPI (Chemometec Nucleocounter NC-3000).

Fixation of dissociated islets (single cells)

Cells were fixed as per the 10X Genomics™ Sample Preparation Demonstrated Protocol Rev B. Briefly, dissociated islet cells were washed twice with chilled rehydration buffer Phosphate Buffer Saline (PBS), 1% Bovine Serum Albumin (BSA) and 500 U/mL Ribonuclease inhibitor), resuspended in 100 μL rehydration buffer, vortexed at low speed during addition of 900 μL chilled methanol, incubated on ice for 15 min, and stored overnight at 4°C. The following day, the fixed cell solutions were pelleted to remove methanol, washed once, and resuspended in rehydration buffer, passed through BD 40 μm cell strainer to remove debris, and assessed for cell counts.

Freezing dissociated islets (single cells)

Cells were resuspended in Recovery Cell Culture Freezing Medium (ThermoFisher Scientific), frozen overnight at −80°C, thawed in 37°C water bath (on Day 2), washed twice, and resuspended in PIM(S), and assessed for cell viability and cell numbers prior to processing for scRNA-seq.

Bulk RNA extraction, sequencing, and data alignment

One sample (2000 IEQ) of each of the two donor sources was processed for bulk RNA sequencing. Each sample was flash-frozen at −80°C for total RNA extraction (bulk islets). Total RNA was extracted and purified using Trizol reagent (Invitrogen) as previously described [14], yielding RNA Integrity Number (RIN) 9.5 and 9.4, for Donors 1 and 2, respectively. Strand-specific PolyA+-enriched messenger RNA (mRNA) libraries were generated with the TruSeq Stranded mRNA kit using an input of 1 ug mRNA and sequenced on an Illumina NovaSeq 6000. Total of 159 million and 193 million paired-end 151 bp reads were generated for Donors 1 and 2, respectively. RNA-seq reads passing the Illumina chastity filter were mapped to a reference sequence of hg19 (with chromosome M replaced with Cambridge Reference Sequence and chromosome Y pseudoautosomal region masked). Read alignment was performed with the Spliced Transcript Alignment to a Reference (STAR) software (version 2.5.1 [15]) using default parameters and a splice junction catalog based on the basic Gencode v19 annotations [16]. Duplicate read pairs were retained but read aligning to alternate haplotypes were excluded. Nonuniquely mapping reads and read pairs with unpaired alignments were also excluded. Read counts for each gene were determined with the Quality of RNA-Seq Toolset (QoRTs) v1.1.18 [17] and generated expression values as reads per kilobase transcripts per million mapped reads (RPKM). The quality and integrity of this bulk library were confirmed by assessing quality control (QC) metrics in QoRTs along with 10 other samples that were processed and sequence in the same batch.

scRNA-seq and read alignments

Single-cell mRNA isolation and sequencing library generation were performed with the 10X Genomics SC3'v2 chemistry according to manufacturer's instructions. This chemistry labels each input RNA molecule with a unique molecular identifier (UMI) that can be used during data analysis to account for PCR amplification bias. Barcoded sequencing libraries were quantified by Quant-IT PicoGreen dsDNA kit (P11496, Invitrogen), diluted at 3 nM and sequenced on Illumina HiSeq3000 using the following read length: 26 bp for Read1, 8 bp for I7 Index, and 98 bp for Read2. All 12 libraries from Donor 1 were pooled together and sequenced in 12 lanes, generating a total of more than 3.6 billion reads with an average of 301 million reads per sample. Fresh-Donor 2 replicate libraries were pooled together and sequenced separately from the 12 Donor 1 treatment group samples.

Demultiplexing of the Illumina base call files and generation of fastq files containing the scRNA-seq data were done using the "mkfastq" function in the CellRanger software package (10X Genomics, v. 1.2.1). Alignment of scRNA-seq reads to the human reference (GRCh37/hg19, Ensembl release 82) and transcript quantification was performed using the "count" function in CellRanger.

Single-cell RNA-seq data quality control filtering

Outlier cells potentially representing low-quality cells or multiple cell captures were removed by excluding cells with very low or high library sizes [18]. Thus, cells with either log-normalized UMI/cell or gene/cell values >2.5 standard deviations (SDs) or <−2.5 SDs from the median value were removed. Rarely expressed genes were also removed by excluding those expressed in <0.2% of cells in a sample.

Single-cell data normalization for cell type classification and differential gene expression

Gene expression data (represented by UMI count) were normalized by both library size and by $\log_e$ transformation. Gene-specific UMI counts in each cell were normalized by dividing the gene-specific UMI count by the total number of UMI in the cell and then multiplied by a scaling factor of 10 000 ("CP10K"). To adjust for 0 counts, 1 was added to all UMI values prior to $\log_e$ transformation (Seurat V2.0—"NormalizeData" function) (https://satijalab.org/seurat [19]). These normalized and log-transformed (ln[CP10K + 1]) expression values were used for downstream in cell type classifications and differential gene expression analyses.

Cell type classification

Gaussian mixture models (GMMs) were used to classify cell types according to the ln[CP10K + 1] expression values of known cell type-enriched "major" marker genes: *INS*—beta, *GCG*—alpha, *SST*—delta, *GHRL*—epsilon, *PPY*—gamma, *KRT19*—ductal, *PRSS1*—acinar, *COL1A1*—stellate, *VWF*—endothelial, and *SDS*—macrophage. GMMs were built using the R package mclust (v. 5.3) with two mixture components, each component representing cells expressing either low levels or high levels of each marker gene. Cell type identity was assigned to a cell if it expressed a major marker gene at a high level. Cells not expressing any major marker gene at the high threshold level were considered "unknown." Cells with high expression levels for more than one major marker gene were considered "mixed" cells. The mixed cells were removed from further downstream analysis.

Generation of heatmaps

The expression data shown in heatmaps are scaled z-scores that were calculated as follows: $z_{i,j} = (x_{i,j} - \mu_i)/\sigma_i$, where $x_{i,j}$ is

the normalized expression of gene i in cell j, μ_i is the mean expression of gene i for all cells of a sample, and σ_i is the SD of expression for gene i for all cells of a sample. Cells, shown as columns, were ordered and grouped according to cell type as determined from the expression levels of the "major marker" genes (shown in bold on the right side). Genes, shown as rows, were ordered and grouped according to previously known associations with the different islet cell types.

Differential gene expression for pathway (IPA) analyses

Pairwise differential gene expression comparisons were made across treatment groups or cell type (alpha or beta) subclusters. ln[CP10K] expression values were used in these analyses (Seurat V2.0—"FindMarkers" with default test method of Wilcoxon rank-sum). Genes with significant differential expression for each pairwise comparison ($P\text{-adj}_{\text{Bonfcorr}} < 0.05$ and ratio of expression change for the pairwise comparison ≤ 0.8 or ≥ 1.3) were used as input for Ingenuity Pathway Analysis (IPA—QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis).

Cell type-specific enriched genes

Triplicate samples were combined to generate four treatment-specific datasets. Within each treatment dataset, differential gene expression (Seurat V2.0—"FindMarkers" with default test method of Wilcoxon rank-sum) was used to determine the top genes enriched in each endocrine cell type (except for the rare epsilon cells). Specifically, one cell type was compared to all other types i.e. beta vs. the remainder of the islet cells ("other types"), alpha vs. other types, delta vs. other types, gamma vs. other types, etc. and this process was repeated for each treatment group independently. A "positive" fold change value (≥ 1.2) was taken as an enrichment of one cell over the "other types." The top 100 most significantly ($P\text{-adj}_{\text{Bonfcorr}} < 0.01$) and enriched genes observed in four major endocrine cell types (beta, alpha, delta, and gamma) from the Fresh treatment group was assessed for their presence in the corresponding lists of the top 100 significantly enriched genes from the FreshON, Fixed, and Frozen groups.

Correlation of gene expression across biological replicates

Gene expression in each cell was normalized for library size by dividing the gene-specific UMI count by the total number of UMI in the cell and multiplying by 10 000 ("CP10K"). A synthetic bulk value for each gene was then generated by averaging the gene CP10K expression value across all cells within each replicate sample, followed by $\log_e$ transformation. Only genes with nonzero values in all samples were used in the pairwise linear correlation (Pearson) analyses.

Correlation analyses of treatment groups and bulk sample

Triplicate samples were combined to generate four independent single-cell treatment-specific datasets. A synthetic bulk value for each gene was then generated by averaging the gene CP10K expression value across all cells within each treatment-specific dataset, followed by $\log_e$ transformation. For the bulk sample, RPKM values (as described above) were also $\log_e$ transformed for this analysis. Only genes with nonzero values in all datasets were used in the pairwise linear (Pearson) or rank-order (Spearman's) correlation analyses.

Clustering of single cells and tSNE projections

ln[CP10K + 1] expression values were scaled prior to clustering (Seurat V2.0—"ScaleData" with regression by total UMI/per cell) to standardize the range of expression and to remove additional variation based on library size. The top 1000 most variable genes [43] from each sample were used to cluster cells (Seurat V2.0—"FindCluster"). The resulting clusters were visualized with t-Distributed Stochastic Neighbor Embedding (tSNE) (Seurat V2.0—"RunTSNE").

Results

Islet preservation treatment groups

We utilized a single shipment of fresh human pancreatic islets from a single donor (Donor 1) to assess the integrity of single-cell transcriptomic data from dissociated cells preserved under four treatment conditions (Figure 1): (i) "Fresh"—cells dissociated from islets upon arrival from the islet resource center and processed immediately for scRNA-seq, (ii) "FreshON"—islets incubated overnight at 37°C prior to dissociation to single cells and scRNA-seq on the second day, (iii) "Fixed"—cells dissociated from islets upon arrival and fixed in methanol and stored overnight at 4°C prior to scRNA-seq processing on the second day, and (iv) "Frozen"—cells dissociated from islets upon arrival and frozen overnight at −80°C prior to scRNA-seq processing on the second day. Treatment groups were performed in triplicate for a total of 12 samples, generating transcriptomic data from 19 521 cells of 12 libraries (Table 1).

The major hormone-secreting cells in pancreatic islets are alpha cells (glucagon), beta cells (insulin), delta (somatostatin), gamma or PPY-cells (pancreatic polypeptide), and epsilon (ghrelin). Cell type composition can be variable among islets within the same pancreas, suggesting a high level of heterogeneity in islet morphology [20]. As such, the RNA composition of a "sample" of islets is distinct from another islet sample from the same donor. Thus, we consider each treatment group to consist of three biological replicates—though we also acknowledge that the distinction between biological and technical replicates is somewhat blurry in this situation.

RNA-seq metrics and quality control exclusions

We performed single-cell mRNA isolation and library generation on the 10X Genomics Chromium System and sequenced the cDNA libraries on the Illumina HiSeq 3000 to an average depth of 302 million reads per sample (Table 1). Across the different samples, the mean reads per cell ranged from 114 456 to 295 332 depending on the number of cells captured and the total sequence reads for that sample. UMIs attached to each sequence read identified unique transcripts, in order to avoid inflated counts due to PCR amplification during library preparation. The mean UMI/cell for each sample varied from 9213 to 17 310, and the mean number of genes detected per cell varied from 2244 to 3555 genes. The mean number of genes per cell detected for each sample was consistent across the triplicate groups with the Frozen triplicates having the highest numbers due to the higher reads/cell.

We excluded outlier cells (with gene count and/or UMI count >2.5 SDs or <−2.5 SDs from the median value) as these are likely to represent poor quality cells or cells that are in fact a

Table 1: Donor 1 treatment sample groups and data metrics

Sample	Cells captured	# Reads	Mean reads/cell	Mean UMI/cell	Mean genes/cell	Cells after QC
1_Fresh	2264	391 388 035	172 874	11 884	2436	2171
2_Fresh	2183	341 262 069	156 327	11 974	2437	2097
3_Fresh	2309	428 796 247	185 706	11 793	2501	2214
Fresh average	2252	387 148 784	171 636	11 884	2458	2160
4_FreshON	1090	192 562 591	176 662	10 070	2383	1043
5_FreshON	1274	171 226 910	134 401	9882	2303	1236
6_FreshON	2056	235 322 213	114 456	9213	2247	1985
FreshON average	1473	199 703 905	141 840	9722	2311	1421
7_Fixed	322	60 472 739	187 803	9289	2244	313
8_Fixed	1778	289 568 463	162 861	13 069	2563	1721
9_Fixed	1412	361 658 801	256 132	13 856	2648	1356
Fixed average	1171	237 233 334	202 265	12 071	2485	1130
10_Frozen	882	260 483 384	295 332	17 310	3555	863
11_Frozen	2042	484 046 291	237 045	16 366	3432	1993
12_Frozen	1909	405 087 485	212 198	15 091	3240	1849
Frozen average	1611	383 205 720	248 192	16 256	3409	1568
All Samples	19 521	3 621 875 228				18 841

mix of two or more cells (Supplementary Table S1). However, we chose not to use indicators of ER stress, such as high expression of mitochondrial genes [21], as a criterion for exclusion, as we considered that metrics of cellular stress might be informative for the comparison of cell preservation methods. ER stress in pancreatic beta cell has been linked to cell proliferation and to the high demands for insulin synthesis [22, 23], and mitochondrial gene expression in islet beta cells occurs during cellular oxidative phosphorylation to trigger insulin secretion (reviewed in Ref. [24]). Although the contribution of technical artifacts is unclear, mouse islet scRNA-seq data are enriched for genes in mitochondrial dysfunction pathways [25] and exhibit beta-cell heterogeneity with respect to expression of ER stress markers [26]. Thus, we included cells regardless of mitochondrial gene expression levels. Ultimately, we performed downstream analyses on a total of 18 841 islet cells, averaging 1570 cells per sample (Table 1 and Supplementary Table S2). Median mitochondrial gene expression in each of the 12 samples ranged from 1.1% to 3.0% with the lowest levels in the Fixed group (median $\leq$1.5%; Supplementary Figure S1), consistent with previous reports comparing Fresh and Fixed datasets [8, 9].

Concordance of biological replicates and quality of scRNA-seq data

To assess the consistency and quality of the scRNA-seq data, we examined the (i) correlation of the average expression of a given gene across all cells within a sample to that of its replicates and (ii) clustering of matching replicate cell types by the multidimensionality reduction technique, tSNE.

We addressed the first measure by computing the average expression of a given gene across all cells in a sample to generate "synthetic bulk" values for that gene. We then used synthetic bulk values of each gene to perform pairwise correlation between the replicate samples. The average expression (synthetic bulk) profiles are highly correlated across the Fresh triplicate samples ($r \geq 0.988$; Figure 2A) as is the case with triplicate samples from each of the other three treatment groups ($r \geq 0.951$; Supplementary Figure S2).

Second, unsupervised clustering of cells confirmed the presence of multiple cell types in pancreatic islets. Combining results from triplicate samples generated several clusters in the tSNE analysis, with cells from each replicate dispersed within all clusters, suggesting high concordance among replicates (Fresh triplicates—Figure 2B; data not shown for other treatment groups). Furthermore, we were readily able to identify known islet cell types in the unsupervised clusters; enriched "major marker" genes, namely *GCG* (alpha cells), *INS* (beta cells), *SST* (delta cells), *GHRL* (epsilon cells), *PPY* (gamma cells), *PRSS1* (acinar cells), *KRT19* (ductal cells), and *COL1A1* (stellate cells), were generally each expressed in a single unique cluster (Fresh triplicates—Figure 2C). These observations attest to both the consistency and integrity of our scRNA-seq data to recapitulate cell type gene expression patterns and enable cell type-specific clustering.

Cell type classifications

For a more systematic determination of the identity of cells isolated from the pancreatic islets, we classified each endocrine or exocrine cell type based on the level of expression of their respective enriched "major marker" gene i.e. *GCG*, *INS*, *SST*, *PPY*, *GHRL*, *PRSS1*, and *KRT19*. We also included several other markers to identify associated cell types: *COL1A1* (stellate), *VWF* (endothelial), and *SDS* (macrophage). Since these major marker genes are not expressed in a single cell type as a precise on–off switch, but rather exhibit a bimodal distribution (Fresh triplicates endocrine cells—Figure 3A), we used GMMs to establish expression threshold values for each major marker. Only cells expressing a major marker gene in the "upper mode" of the distribution were classified as belonging to its respective cell type. Overlaying these cell type classifications onto tSNE clusters highlighted the identities of each cluster (Fresh triplicates—Figure 3B). Cells not expressing any of the major markers at threshold or above levels were classified as "unknown." Cells with more than one major marker genes at or exceeding their threshold values may represent multiple cell encapsulations during capture and thus classified as "mixed" cells. To note, this method of classifying mixed cells does not account for instances where the mixed cells are of the same type.

We determined the accuracy of the single major marker classifications by examining the coexpression profiles of

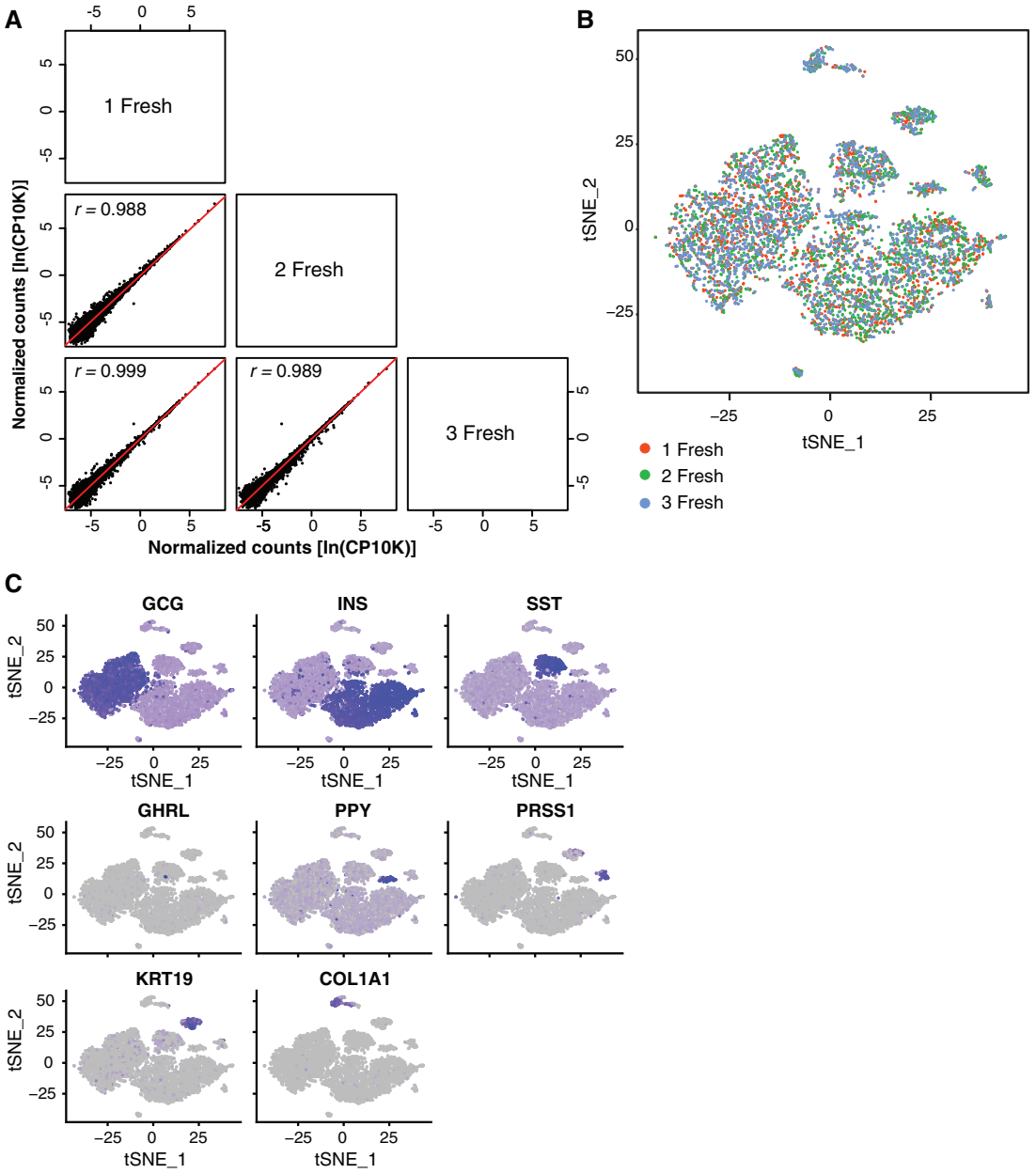

Figure 2: Fresh triplicate samples are highly correlated, and cell type-enriched marker genes are highly expressed in a single unique cluster. **(A)** Pairwise Pearson linear correlation of "synthetic" bulk expression values (generated as described in "Material and methods" section) for Fresh triplicate samples. Input values for these analyses were the natural log-normalized UMI counts, ln(CP10K), for 16 338 genes. **(B)** Unsupervised clustering and tSNE mapping (RunTSNE). While different cell type clusters are apparent, the three different Fresh samples give results that are completely interdigitated, suggesting a very high degree of similarity. **(C)** Revealing the basis of the clusters: each tSNE plot is highlighted by the location of cells showing high expression of major marker genes (increasing expression with darker shade of blue). Note that the overlapping location of the small *GHRL*-expressing cell cluster with the *SST*-expressing cell cluster is an artifact of the 2D display. The clusters do not overlap.

multiple genes known to be enriched in each cell type. The gene expression patterns for the Fresh triplicate samples with multiple known cell type genes previously reported in islet scRNA-seq datasets [25–30] confirmed our cell classifications (Figure 3C). Furthermore, putative mixed cell types show high gene expression patterns corresponding to more than one type of cell.

As expected, the major types are the beta and alpha cells, making up on average 39.2% and 34.6% of the Fresh triplicate samples, respectively (Table 2). The delta cells are the next most abundant population at 6–7%, with the remainder of the other cell types individually contributing <4% each. This composition of endocrine cells is within the range expected for human pancreatic islets, which has been observed to vary considerably between pancreata and also between islets within the same pancreas [31–33]. The proportion of mixed cells across the Fresh triplicates (average of 8.4%) is higher than seen previously with cell lines [34] especially since our value does not account for mixed cells of the same type. Most likely this is due to incomplete dissociation of cell aggregates and/or increased capture of single cells with genetic debris from damaged cells. A less likely factor is the presence of multihormonal-secreting endocrine

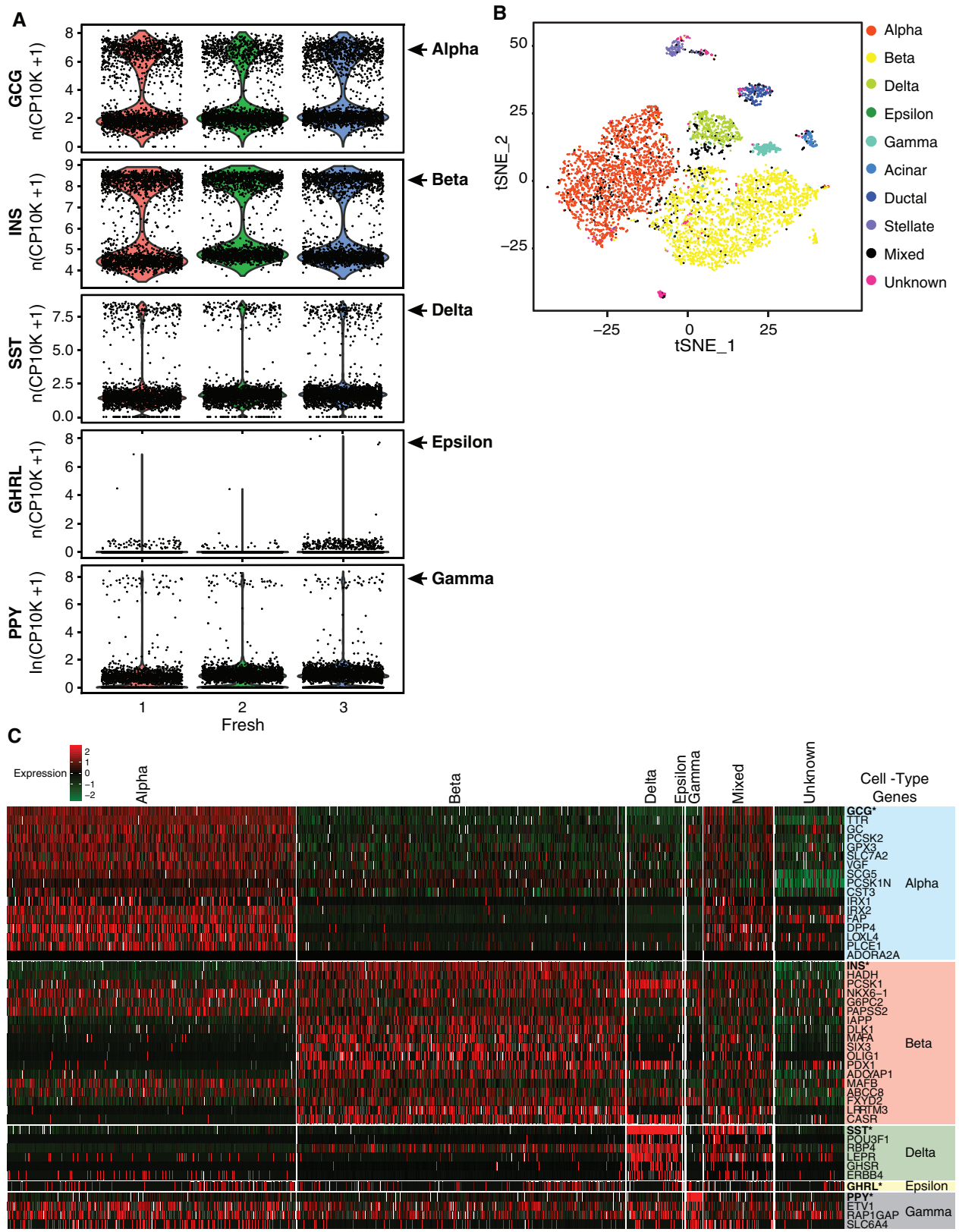

Figure 3: Fresh triplicate samples: cell type classification confirmed with cell type-specific clusters and coexpression patterns with multiple known cell type-specific marker genes. (**A**) Each panel displays the bimodal expression of a single endocrine cell type-enriched marker gene: *GCG* (alpha), *INS* (beta), *SST* (delta), *GHRL* (epsilon), and *PPY* (gamma). Each column represents one of three Fresh samples. GMMs were used to establish expression threshold for cell classification. (**B**) Cell type classifications are superimposed onto tSNE projected clusters (**C**) Heatmap of all cell types showing expression of 45 genes, each of which has been reported in the literature to be enriched in at least one of the five endocrine cell types. The 45 genes are listed on the far right grouped together according to their respective cell type associations i.e. alpha, beta genes, delta, epsilon, and gamma cell genes. Cells are in columns with color ribbon at top of heatmap representing level of expression.

Table 2: Cell type recovery (%) is consistent across triplicates but varies across treatment groups

Sample	Cells Captured	Cells Analyzed	%Cell α	β	δ	ε	γ	Acinar	Ductal	Stellate	Endo	Macro	Mix	Unknown
1_Fresh	2264	2172	34.8	38.9	6.7	0.1	1.8	1.6	2.3	1.4	0.3	0.6	10.0	1.5
2_Fresh	2183	2098	31.3	44.1	6.1	0	1.9	2.0	3.6	1.9	0.3	1.0	5.3	2.5
3_Fresh	2309	2215	37.5	35.0	7.0	0.2	2.0	1.0	2.4	2.6	0.3	0.5	9.9	1.7
Fresh average	2252	2162	34.6	39.2	6.6	0.1	1.9	1.5	2.8	2.0	0.3	0.7	8.4	1.9
4_FreshON	1090	1044	27.4	43.6	**5.7**	0	2.6	1.1	1.4	**4.4**	1.2	1.7	7.5	3.4
5_FreshON	1274	1237	22.7	49.4	**5.6**	0	2.8	0.8	2.3	**3.7**	1.6	1.4	5.7	3.9
6_FreshON	2056	1986	29.9	44.5	**4.1**	0.1	1.4	1.1	2.0	**4.0**	0.8	1.5	8.3	2.4
FreshON aveage	1473	1422	27.2	45.7	4.9	0	2.1	1.0	1.9	4.0	1.1	1.5	7.3	3.1
7_Fixed	322	314	**27.7**	**48.1**	**3.8**	0	**0.3**	0.6	3.2	1.0	0	1.3	12.7	1.3
8_Fixed	1778	1722	**10.9**	**56.4**	**3.3**	0.1	**0**	2.0	5.2	1.3	0.3	0.1	19.2	1.2
9_Fixed	1412	1357	**11.9**	**54.7**	**3.4**	0	**0**	1.5	3.1	0.6	0	0.1	23.9	0.7
Fixed average	1171	1131	12.9	55.0	3.4	0.1	0.0	1.7	4.2	1.0	0.2	0.2	20.5	1.0
10_Frozen	882	864	31.7	43.2	**0.8**	0	**0.5**	0.2	1.7	2.2	0.3	0	17.8	1.5
11_Frozen	2042	1994	32.8	38.6	**0.5**	0	**0.5**	0.5	2.2	3.5	0.4	0	20.1	1.2
12_Frozen	1909	1850	32.7	46.1	**0.5**	0	**0.4**	0.8	3.2	4.3	0.7	0	10.2	1.1
Frozen average	1611	1569	32.5	42.3	0.6	0	0.5	0.6	2.5	3.6	0.5	0	15.8	1.2
All Samples	19 521	18 853												

Cells classified based on high expression of a single major marker: *INS*—beta, *GCG*—alpha, *SST*—delta, *GHRL*—epsilon, *PPY*—gamma, *KRT19*—ductal, *PRSS1*—acinar, *COL1A1*—stellate, *VWF*—endothelial, and *SDS*—macrophage. The proportion of each cell type recovered from the Fresh triplicate samples was compared (ANOVA, Tukey's HSD *post hoc* test) to those recovered in each of the other groups. Percent recovery values highlighted in pink and bolded fonts are significantly different ($P<.05$) compared to yields from the Fresh triplicates.

cells. Such subpopulations have been observed in mouse and human islet cells, and speculatively suggested as signatures of immature cells or dedifferentiation of mature cells into other cell types [35–38]. But in light of the known propensity for dissociation of solid tissues to lead to aggregates, we choose to be conservative and consider these as artifacts.

Preservation treatment and cell type recovery

Next, we turned to a comparison of results between the four different preservation methods. The proportions of the major endocrine cells (beta, alpha, delta, and gamma cells) captured are consistent within each treatment triplicate group except for two instances (Fixed sample 7_Fixed and Frozen sample 10_Frozen) where reagent clogging during the single-cell capture procedure resulted in significantly lower yields for both samples (Table 2). In contrast, the proportions of alpha, delta, and gamma cells recovered from the Fixed triplicate samples are significantly lower compared to those obtained from Fresh samples (ANOVA, Tukey's HSD *post hoc* test, $P < 0.05$), while the proportion of beta cells in Fixed samples is significantly higher than in Fresh samples. Similarly, the proportions of delta cells from FreshON or delta and gamma cells from Frozen samples were also significantly less than those obtained from the Fresh samples, suggesting cell type-specific loss after fixation or cryopreservation.

Although not statistically significant due to the high variance within the Fixed and Frozen triplicates, the average recovery of mixed cells is higher for these two groups. Methanol fixation dehydrates cells, causing protein to denature and precipitate, and can contribute to an increase in cells adhering together. Contamination of single cell isolates with genetic debris from damaged cells caused by freeze-thawing of Frozen samples may have contributed to increased mixed cells in this group. Thus, compared to Fresh islet samples, there is a significant difference in the recovery of the various cell types from the Fixed and Frozen samples, and the latter two sample groups also have a greater proportion of mixed cells.

Preservation treatment and gene expression

To assess the effects of cell preservation treatments on gene expression, we determined the (i) pairwise-correlation between treatment groups, (ii) cell type-specific expression patterns of all groups using multiple known markers for each cell type, (iii) pattern of clustering of all sample groups in tSNE plots, and (iv) overlap of the top 100 most significantly enriched genes for each cell type.

For pairwise correlation testing, we combined scRNA-seq data from triplicate samples and computed average gene expression values for all cells ("synthetic bulk") within a treatment group. Gene expression is highly correlated across treatment groups ($r \geq 0.959$; Figure 4A). The synthetic bulk samples are also well correlated with a frozen bulk islet sample from the same pancreas ($r \geq 0.789$; Figure 4A), suggesting that overall gene expression is neither drastically affected by islet processing to single cells nor by preservation treatments of single-cell preparations.

To refine the comparison of treatment groups, we tested the two major cell types, the beta and alpha cells, individually for correlation between groups. Beta-cell gene expression is highly correlated between treatment groups ($r \geq 0.962$; Figure 4B) as is the case with alpha cells ($r \geq 0.950$; Figure 4C). Additional testing with regard to the rank order (Spearman's correlation) of the most highly expressed to the lowest expressed genes confirms the high correlation of the various group datasets (beta $r \geq 0.975$, alpha $r \geq 0.959$; Supplementary Figure S3).

We also assessed all samples for their expression of multiple known cell type-specific marker genes. As was seen for the Fresh triplicate samples (Figure 3C), all samples show cell type-specific coexpression profiles regardless of treatment group (Supplementary Figure S4). This further confirms that technical

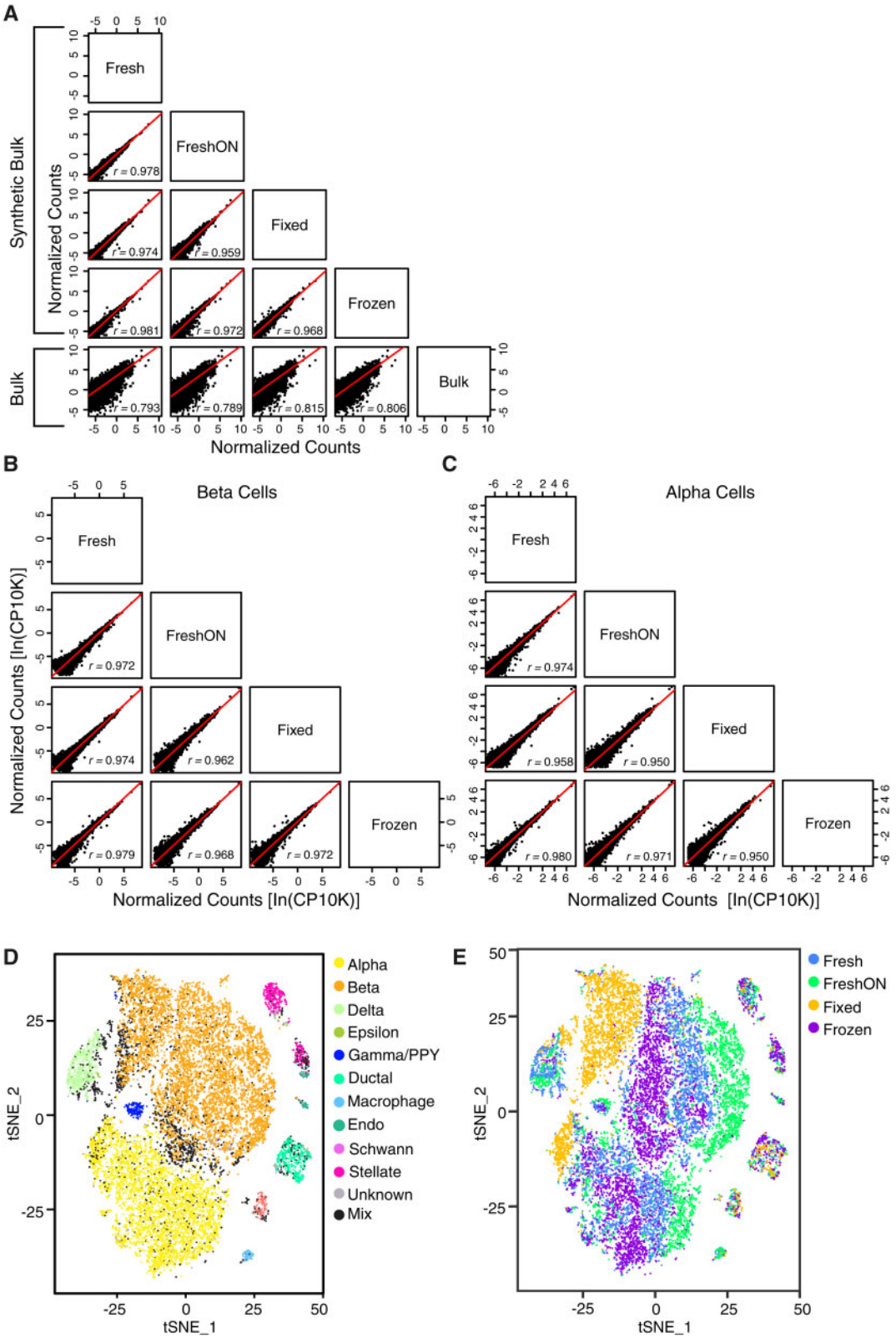

Figure 4: All treatment groups—12 samples: gene expression is highly correlated across treatment groups and generally exhibits cell type-specific clustering. (**A**) Pairwise Pearson linear correlation of "synthetic" bulk expression values from three replicate samples (generated as described in "Material and methods" section) between treatment-specific datasets. Input values for these analyses were the natural log of normalized UMI counts, ln(CP10K), for 16 486 genes. Pairwise correlation was also performed with one bulk sample [natural log of normalized counts, ln(FPKM)] and the synthetic bulk values from the four treatment groups. (**B**) Pairwise Pearson linear correlation of "synthetic" bulk expression values for beta cells between treatment-specific datasets. (**C**) Same as in (B) but only for alpha cells. (**D**) tSNE of all 12 samples identified by cell type assignments. (**E**) tSNE of all 12 samples identified by treatment group.

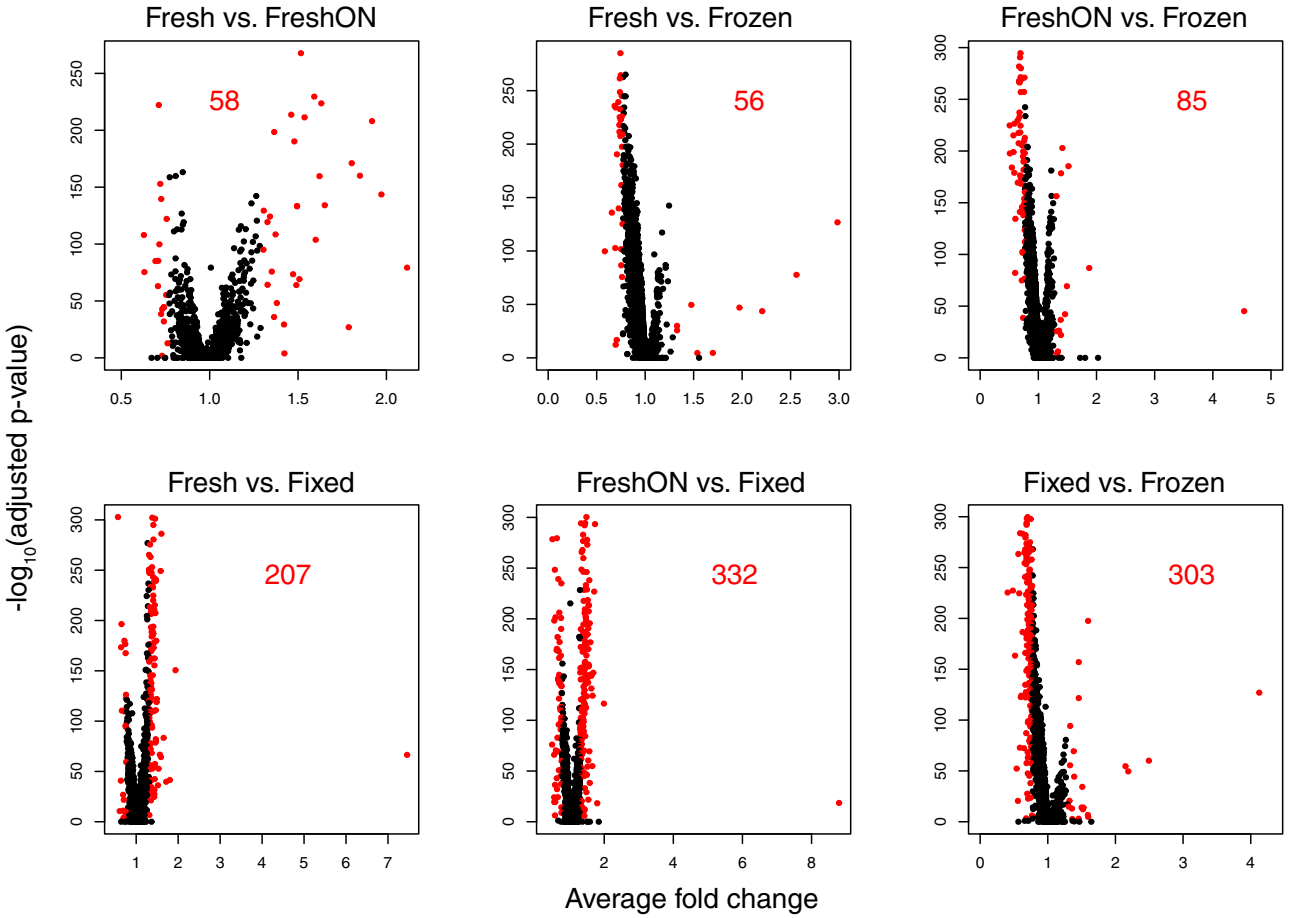

Figure 5: Volcano plots showing results of pairwise-differential gene expression across four treatment groups: genes designated to have significant differential expression for each pairwise comparison (red dots) have (1) adj $P < 0.05$ and (2) ratio of expression change ≤ 0.8 or ≥ 1.3. The number of significant differentially expressed genes is listed in red font in each panel.

variations introduced by the preservation treatments are relatively minor compared to the biological similarities of the cell-specific datasets.

Further support for the integrity of the data from all treatment groups comes from the coalescing of similar cell types as visualized by tSNE mapping with all 12 samples of the four groups. Cell type identities are maintained for cells from all sample groups as they formed cell type-specific clusters, aligning well with their cell type assignments based on expression levels of their respective major marker gene (Figure 4D). However, within each of the cell type clusters, samples separated by treatment group, suggesting significant variations between the groups (Figure 4E). We were concerned that this might be an artifact of having a sparse gene dataset leading to a high rate of dropout events ("zero counts") for gene transcripts that are not captured (reviewed in Ref. [39]), but the separation persisted when clustering was performed with the top 5% (597) of the most highly expressed genes.

To further assess the source of these variations, we performed pairwise-differential gene expression analyses across the four groups and identified differentially expressed genes (Figure 5 and Supplementary Table S3A). In the Fixed vs. Fresh comparison, 207 genes were differentially expressed ($P\text{-adj}_{\text{Bonfcorr}} < 0.05$ and ratio of expression change ≤ 0.8 or ≥ 1.3) with 181/207 downregulated in Fixed samples. Pathway analysis (IPA) with all 207 differentially expressed genes revealed significant enrichment of genes in the eukaryotic initiation factor (EIF2) signaling pathway ($P < 10^{-85}$; Supplementary Table S3B). We also observed this trend of downregulation for the majority of the differentially expressed genes in Fixed samples compared to FreshON (266/332) or Frozen (278/303; Supplementary Table S3A). IPA analysis once again showed significant enrichment for EIF2 signaling pathway genes in both the comparisons of Fixed vs. FreshON ($P < 10^{-96}$) and Fixed vs. Frozen ($P < 10^{-59}$; Supplementary Table S3C and S3D). The next three most significantly enriched pathways were eIF4 and p70S6K signaling, mTOR signaling, and oxidative phosphorylation which exhibited a similar trend of lower enrichment in Fixed samples compared to the other three treatment groups (ranging from $P < 10^{-17}$ to $P < 10^{-32}$). In contrast, pairwise comparisons of Fresh, FreshON, and Frozen highlighted pathways with much lower significance (best $P < 10^{-11}$) and include those associated with SPINK1 pancreatic cancer and oxidative phosphorylation (Supplementary Table S3E–S3G). EIF2 is required for the initiation of cellular translation and phosphorylation of EIF2 during the stress response reduces protein synthesis to conserve resources for cell recovery [40]. eIF4 and p70S6K signaling, as well as mTOR signaling genes, are also important regulators of protein translation and their enrichment here correlates with that observed for EIF2 genes. Furthermore, EIF2 signaling and oxidative phosphorylation were the top functional pathways identified for the most abundantly expressed genes in human

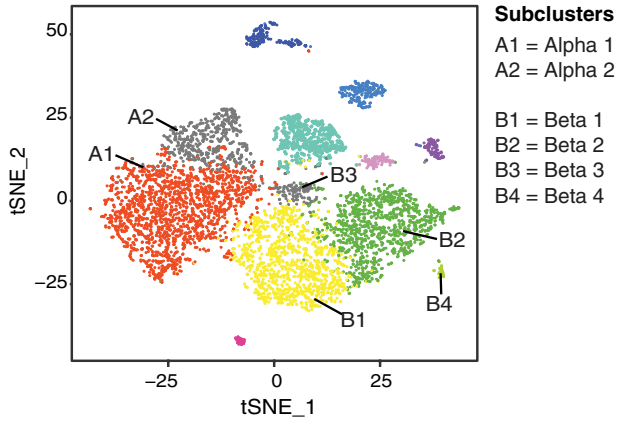

Figure 6: Fresh triplicate samples from Donor 1: tSNE plot highlighting two alpha subclusters (A1 and A2) and four beta subclusters (B1, B2, B3, and B4).

islet endocrine cells [25]. Our findings here support a major role of these genes in pancreatic islets and that these compelling differences between the Fixed compared to Fresh, FreshON, and Frozen are likely contributing to the Fixed beta-cell cluster being slightly more separated from the beta cells of the other three treatments groups. A reasonable explanation is that the effect of fixation is to downregulate multiple transcripts mediated through EIF2 and related pathways. However, there are no doubt other contributions to the variation in gene expression observed across treatment groups.

Lastly, within each treatment group, we identified the top 100 significantly enriched genes in each of the four endocrine cell types (beta, alpha, delta, and gamma) and determined the overlap of these gene lists across the four treatment groups. We first performed differential gene expression to identify genes enriched ($P\text{-adj}_{\text{Bonfcorr}} < 0.01$ and fold change ≥ 1.2) in one cell type compared to the remainder of the islet cell types. Since we consider Fresh samples as the "gold standard" for these analyses, we compared the gene list from this group (Supplementary Table S4) to the FreshON, Fixed and Frozen groups. Totally, 75–78% of the top 100 enriched genes in Fresh beta cells are also found in the top 100 beta-cell-enriched genes in the other three treatment groups (Supplementary Figure S5). Similarly, 79–84% of the most enriched genes in Fresh alpha cells are also present in the top genes of other treatment groups. In contrast, the minor cell types (delta and gamma) show a trend toward reduced overlap with the top 100 cell type-enriched genes identified from fresh samples. More specifically, the Frozen delta and gamma cell top gene lists include only 26% and 22% of the genes found in the Fresh top 100, respectively. We did not detect this lower correlation in gene expression across treatment groups when testing all cell types together, as the higher variation of the minor cell populations is likely masked by the more consistent expression of the much larger beta and alpha populations. These differences are not surprising in light of the biased cell type recovery observed for the minor endocrine cell types in the Fixed and Frozen groups.

Cell type heterogeneity

We assessed our dataset for evidence of alpha- and beta-cell heterogeneity, findings previously suggested by several groups [26, 28, 41–44]. Graph-based clustering followed by tSNE projections of the Fresh triplicate samples suggests two alphas (A1 and A2) and four betas (B1, B2, B3, and B4) subclusters (Figure 6). As was seen for the across treatment groups comparison with all cell types, pathway analysis (IPA) of differentially expressed genes (Supplementary Table S5A) revealed highly significant differences in enrichment for genes involved in EIF2 signaling between the subclusters of both alpha and beta cells (Supplementary Table S5B–S5G). Genes differentially expressed between alpha A1 vs. A2 subclusters are enriched in pathways related to EIF2 signaling ($P < 10^{-124}$), eIF4 and p70S6K signaling ($P < 10^{-37}$), and mTOR ($P < 10^{-35}$). Similarly, genes differentially expressed between beta B1 and either the B2 or B3 subclusters are enriched in the same pathways: EIF2 signaling ($P < 10^{-125}$), eIF4 and p70S6K signaling ($P < 10^{-37}$), and mTOR ($P < 10^{-35}$). The next most significant pathway for beta subclusters is oxidative phosphorylation, where B1 differs from B3 ($P < 10^{-3}$), B4 ($P < 10^{-9}$), and B2 ($P < 10^{-11}$). Alpha and beta subclusters are also present in other treatment groups and have very similar characteristics, whereby subclusters differ by expression of genes enriched in pathways as identified in the Fresh samples.

Confirmation of findings in islets from a second donor

To address the possibility of islet source-related variation, we analyzed islets from a second donor processed in our laboratory under the identical conditions as the "Fresh" islet samples in the current treatment comparison dataset. Replicate samples from this additional donor are hereafter referred to as "Fresh-Donor 2." Once again, the average expression (synthetic bulk) profile for Fresh-Donor 2 is well correlated with a frozen bulk islet sample from the same pancreas ($r \geq 0.822$; Supplementary Figure S6). We then confirmed the concordance of overall islet gene expression by performing pairwise correlation testing of the "Fresh" samples with the "Fresh-Donor 2" samples. We combined scRNA-seq data from the replicate samples for each donor source to generate a synthetic bulk gene expression value for all cells from either "Fresh" or "Fresh-Donor 2." Gene expression is highly correlated across the two donor sources ($r \geq 0.969$; Supplementary Figure S7). Next, we performed unsupervised clustering of cells from Fresh-Donor 2, identified two alpha and two beta subclusters (Supplementary Figure S8) and performed pairwise differential gene expression analysis between the respective cell type subclusters (Supplementary Table S6A). IPA analysis of the differentially expressed genes between the alpha or beta subclusters also showed enrichment for genes in the EIF2 signaling pathway ($P < 10^{-57}$), mTOR ($P < 10^{-17}$), and eIF4 and p70S6K signaling ($P < 10^{-18}$; Supplementary Table S6B and S6C).

Thus, the presence of alpha and beta cell subclusters is reproducible across two donors and based on the variable

abundance of transcripts involved in stress-related responses, including those related to insulin secretion. Such potentially stress-related heterogeneity has also been noted in other islet studies [26, 41, 45]. Whether these represent "natural" *in vivo* stress-induced beta [45] and alpha subpopulations, or are triggered by tissue processing/dissociation as has been reported for mouse muscle stem cells [46] and human tumor tissues [47], or both, is unclear. Lastly, differences between some of the subclusters may be rather minor, and potentially exaggerated by the limitations of tSNE mapping algorithm.

Discussion

The ability to interrogate human pancreatic islets at the single-cell level has provided new insights into the insulin-producing beta cell, the glucagon-producing alpha cell, and their interactions with the other surrounding islet cell types to maintain glucose homeostasis. We show that a variety of processing protocols can potentially be used to dissociate islet cells for subsequent scRNA-seq. Upon receipt of islets into the laboratory, whole islets can be processed immediately or kept overnight at 37°C prior to dissociation and single-cell capture. Alternatively, methanol fixation or cryopreservation can be employed after dissociation but prior to capture, allowing for some flexibility for islet processing.

There are, however, significant consequences of the specific protocol chosen. We observed a bias in cell type recovery for three of the five endocrine cell types for preserved samples compared to Fresh samples. Specifically, we found a reduced yield of alpha, delta, and gamma cells in fixed samples, and a reduced level of delta and gamma cells in cryopreserved samples. This suggests that the viability and/or recovery efficiency of different endocrine cell types can be significantly affected by the choice of sample preparation. Such a bias was suggested in a prior report for the lower recovery of beta cells from fluorescence-activated cell sorted islet cells compared to the proportion observed from histochemical analysis of the same islets [27].

With respect to gene expression patterns of the recovered individual cells, the separation by treatment on tSNE plots documents the presence of a measurable effect on gene expression. But this effect is small in comparison to the biological similarities, as cell type identity is retained for cells in all treatments. The magnitude of the similarities is also supported by the very high pairwise correlation of gene expression between treatment groups.

To our knowledge, this is the first in depth and side-by-side comparison of scRNA-seq data generated from different sample preps of a single source of human pancreatic islets. The high concordance of biological replicate samples within each treatment group reaffirms that robust single-cell transcriptomic data from solid tissues is achievable. Methanol fixation and cryopreservation are both shown to be effective for preserving the integrity of islet beta cells. However, there is a notable cell type recovery bias against the cell types other than beta cells, and gene expression changes in delta and gamma cells. Thus, there is a need for caution when interpreting results from the affected cell populations.

These results have important implications for future studies of pancreatic islet transcriptomes. An important question that these studies will hope to address is the difference in endocrine cell abundance and expression patterns between normal and diabetic islets. Researchers attempting to compare data across studies will need to exercise great caution, especially if different protocols were utilized for sample preparation. Ideally, to limit the inclusion of experimental steps that may introduce bias, the protocol referred to here as "Fresh" should be utilized. In cases where analysis of Fresh samples is not practical, using the same preservation method across experiments would be strongly advised, especially when cell abundance is being compared.

An important variable that is unmeasured in this work relates to the variability in the quality of donated islet samples and steps prior to receipt of the islet material in the research laboratory. Although suppliers of islets for research have relatively similar protocols for harvest and shipping, there is still notable variability in islet center-specific protocols and timing of delivery. Establishing standardized protocols will be critical if comparisons between different donors are to have experimental validity [48].

The cell type-specific effects of preservation treatments observed here for pancreatic islets are likely to hold true for other solid tissues. Investigators seeking to undertake a comparison of normal and diseased target tissues would be well advised to conduct similar analyses of reproducibility (Supplementary Figure S9) prior to embarking on more detailed and costly investigations.

Supplementary data

Supplementary data are available at *Biology Methods and Protocols* online.

Data availability

Sequence data are available from dbGAP with the accession phs001188.v2.p1: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001188.v2.p1

Code availability

No custom codes were used in this study.

ACKNOWLEDGEMENTS

We would like to thank the NIAMS Sequencing Technology Core for sequence generation from scRNA-seq libraries, the NISC Comparative Sequencing Program for sequence generation from the bulk RNA-seq library; Julia Fekecs for assistance with figure and table preparations; Efsun Arda for her helpful advice and two-step islet dissociation protocol; and Daniel Bar, Michael Stitzel, and Steve Parker for their helpful discussions.

Funding

This work was supported by the National Institutes of Health Intramural Research Program of the National Human Genome Research Institute [ZIA HG000024 to F.S.C.].

Conflict of Interest Statement. None declared.

REFERENCES

1. Mahajan A, Taliun D, Thurner M *et al*. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;**50**:1505–13.
2. Picelli S. Single-cell RNA-sequencing: the future of genome biology is now. *RNA Biol* 2017;**14**:637–50.

3. Hedlund E, Deng Q. Single-cell RNA sequencing: technical advancements and biological applications. *Mol Aspects Med* 2018;**59**:36–46.
4. Guillaumet-Adkins A, Heyn H. Single-cell genomics unravels brain cell-type complexity. *Adv Exp Med Biol* 2017;**978**:393–407.
5. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* 2018;**18**:35–45.
6. Birnbaum KD. Power in numbers: single-cell RNA-seq strategies to dissect complex tissues. *Annu Rev Genet* 2018;**52**: 203–21.
7. Thomsen ER, Mich JK, Yao Z *et al.* Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat Methods* 2016;**13**:87–93.
8. Alles J, Karaiskos N, Praktiknjo SD *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol* 2017;**15**:44.
9. Chen J, Cheung F, Shi R *et al.* PBMC fixation and processing for chromium single-cell RNA sequencing. *J Transl Med* 2018;**16**: 198.
10. Attar M, Sharma E, Li S *et al.* A practical solution for preserving single cells for RNA sequencing. *Sci Rep* 2018;**8**:2151.
11. Guillaumet-Adkins A, Rodriguez-Esteban G, Mereu E *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol* 2017;**18**:45.
12. Nguyen QH, Pervolarakis N, Blake K *et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* 2018;**9**:2028.
13. Scharp DW, Arulmoli J, MorganK *et al.* Advances in human islet processing: manufacturing steps to achieve predictable islet outcomes from research pancreases. *OBM Transplant* 2019; **3**:1–73.
14. Lawlor N, Marquez EJ, Orchard P *et al.* Multiomic profiling identifies cis-regulatory networks underlying human pancreatic beta cell identity and function. *Cell Rep* 2019;**26**: 788–801.e786.
15. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.
16. Harrow J, Frankish A, Gonzalez JM *et al.* GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res* 2012;**22**:1760–74.
17. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* 2015;**16**:224.
18. Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor [version 2; peer review: 3 approved, 2 approved with reservations]. *F1000Res* 2016;**5**:2122.
19. Butler A, Hoffman P, Smibert P *et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–20.
20. Steiner DJ, Kim A, Miller K *et al.* Pancreatic islet plasticity: interspecies comparison of islet architecture and composition. *Islets* 2010;**2**:135–45.
21. Ilicic T, Kim JK, Kolodziejczyk AA *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 2016; **17**:29.
22. Marchetti P, Bugliani M, Lupi R *et al.* The endoplasmic reticulum in pancreatic beta cells of type 2 diabetes patients. *Diabetologia* 2007;**50**:2486–94.
23. Sharma RB, O'Donnell AC, Stamateris RE *et al.* Insulin demand regulates beta cell number via the unfolded protein response. *J Clin Invest* 2015;**125**:3831–46.
24. Mulder H. Transcribing beta-cell mitochondria in health and disease. *Mol Metab* 2017;**6**:1040–51.
25. Xin Y, Kim J, Okamoto H *et al.* RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab* 2016; **24**:608–15.
26. Baron M, Veres A, Wolock SL *et al.* A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst* 2016;**3**:346–60.e344.
27. Segerstolpe A, Palasantza A, Eliasson P *et al.* Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab* 2016;**24**:593–607.
28. Wang YJ, Schug J, Won KJ *et al.* Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* 2016;**65**:3028–38.
29. Lawlor N, George J, Bolisetty M *et al.* Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res* 2017;**27**:208–22.
30. Dorajoo R, Ali Y, Tay VSY *et al.* Single-cell transcriptomics of East-Asian pancreatic islets cells. *Sci Rep* 2017;**7**:5024.
31. Brissova M, Fowler MJ, Nicholson WE *et al.* Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *J Histochem Cytochem* 2005;**53**: 1087–97.
32. Cabrera O, Berman DM, Kenyon NS *et al.* The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci USA* 2006;**103**:2334–39.
33. Blodgett DM, Nowosielska A, Afik S *et al.* Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. *Diabetes* 2015;**64**: 3172–81.
34. Zheng GX, Terry JM, Belgrader P *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;**8**: 14049.
35. Chiang MK, Melton DA. Single-cell transcript analysis of pancreas development. *Dev Cell* 2003;**4**:383–93.
36. Katsuta H, Akashi T, Katsuta R *et al.* Single pancreatic beta cells co-express multiple islet hormone genes in mice. *Diabetologia* 2010;**53**:128–38.
37. Butler AE, Campbell-Thompson M, Gurlo T *et al.* Marked expansion of exocrine and endocrine pancreas with incretin therapy in humans with increased exocrine pancreas dysplasia and the potential for glucagon-producing neuroendocrine tumors. *Diabetes* 2013;**62**:2595–604.
38. Teo AKK, Lim CS, Cheow LF *et al.* Single-cell analyses of human islet cells reveal de-differentiation signatures. *Cell Death Discov* 2018;**4**:14.
39. Haque A, Engel J, Teichmann SA *et al.* A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 2017;**9**:75.
40. Wek RC, Jiang HY, Anthony TG. Coping with stress: eIF2 kinases and translational control. *Biochem Soc Trans* 2006;**34**:7–11.
41. Muraro MJ, Dharmadhikari G, Grun D *et al.* A single-cell transcriptome atlas of the human pancreas. *Cell Syst* 2016;**3**: 385–94.e383.
42. Bader E, Migliorini A, Gegg M *et al.* Identification of proliferative and mature beta-cells in the islets of Langerhans. *Nature* 2016;**535**:430–34.
43. Dorrell C, Schug J, Canaday PS *et al.* Human islets contain four distinct subtypes of beta cells. *Nat Commun* 2016;**7**:11756.
44. Qiu WL, Zhang YW, Feng Y *et al.* Deciphering pancreatic islet beta cell and alpha cell maturation pathways and characteristic features at the single-cell level. *Cell Metab* 2018;**27**:702.
45. Xin Y, Dominguez Gutierrez G, Okamoto H *et al.* Pseudotime ordering of single human beta-cells reveals states of insulin production and unfolded protein response. *Diabetes* 2018;**67**: 1783–94.

46. van den Brink SC, Sage F, Vertesy A *et al*. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods* 2017;**14**:935–36.
47. O'Flanagan CH, Campbell KR, Zhang AW *et al*. Dissociation of solid tumour tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses. *bioRxiv*. doi: 10.1101/683227
48. Hart NJ, Powers AC. Use of human islets to understand islet biology and diabetes: progress, challenges and suggestions. *Diabetologia* 2019;**62**:212–22.