

Parallel measurement of transcriptomes and proteomes from same single cells using nanodroplet splitting

Received: 16 May 2024

Accepted: 1 November 2024

Published online: 05 December 2024

 Check for updates

James M. Fulcher¹✉, Lye Meng Markillie¹, Hugh D. Mitchell¹, Sarah M. Williams¹, Kristin M. Engbrecht², David J. Degnan³, Lisa M. Bramer³, Ronald J. Moore³, William B. Chrisler³, Joshua Cantlon-Bruce^{4,5}, Johannes W. Bagnoli⁵, Wei-Jun Qian³, Anjali Seth⁵, Ljiljana Paša-Tolić¹ & Ying Zhu^{1,6}✉

Single-cell multiomics provides comprehensive insights into gene regulatory networks, cellular diversity, and temporal dynamics. Here, we introduce nanoSPLITS (nanodroplet SPLitting for Linked-multimodal Investigations of Trace Samples), an integrated platform that enables global profiling of the transcriptome and proteome from same single cells via RNA sequencing and mass spectrometry-based proteomics, respectively. Benchmarking of nanoSPLITS demonstrates high measurement precision with deep proteomic and transcriptomic profiling of single-cells. We apply nanoSPLITS to cyclin-dependent kinase 1 inhibited cells and found phospho-signaling events could be quantified alongside global protein and mRNA measurements, providing insights into cell cycle regulation. We extend nanoSPLITS to primary cells isolated from human pancreatic islets, introducing an efficient approach for facile identification of unknown cell types and their protein markers by mapping transcriptomic data to existing large-scale single-cell RNA sequencing reference databases. Accordingly, we establish nanoSPLITS as a multiomic technology incorporating global proteomics and anticipate the approach will be critical to furthering our understanding of biological systems.

Multicellular organisms contain a variety of cell populations and subpopulations, which are well-organized in defined patterns to implement critical biological functions. The development and rapid dissemination of single-cell omic technologies have dramatically advanced our knowledge on cellular heterogeneity^{1–3}, cell lineages⁴, and rare cell types⁵. However, most existing technologies only capture single modalities of molecular information. Such measurement provides only a partial picture of a cell's phenotype, which is

determined by the interplay between the genome, epigenome, transcriptome, proteome, and metabolome. Moreover, proteins are of particular interest in establishing cellular identities because they are the downstream effectors and their abundance cannot be reliably inferred from other modalities, including transcripts⁶. Unfortunately, existing multimodal transcriptome-proteome measurements^{7–14} are restricted to at most a few hundred (and often significantly fewer) protein targets. These measurements also require intermediate

¹Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA. ²Nuclear, Chemistry, and Biology Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA. ³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA. ⁴Scienion AG, Volmerstraße 7, 12489 Berlin, Germany. ⁵Cellenion SASU, 60 Avenue Rockefeller, Bâtiment BioSerra2, 69008 Lyon, France. ⁶Present address: Department of Proteomic and Genomic Technologies, Genentech Inc., 1 DNA Way, South San Francisco 94080, USA.

✉ e-mail: james.fulcher@pnnl.gov; zhu.ying@gene.com

antibodies to recognize epitopes, which can be limited by availability and specificity^{15,16}.

A route for overcoming these limitations is through the adoption of a mass spectrometry (MS)-based approach. With the advance of microfluidic sample preparation¹⁷ and isobaric labeling¹⁸, single-cell proteomics (scProteomics) is now capable of measuring thousands of proteins from a single cell^{19–21}. Encouraged by these developments, we seek to acquire multimodal transcriptome-proteome measurements from the same single cells by integrating single-cell RNA sequencing (scRNAseq) with scProteomics. Towards this end, we present nanoSPLITS (nanodroplet SPLITting for Linked-multimodal Investigations of Trace Samples), a method capable of dividing single-cell lysates via two nanoliter droplet microarrays and separately measuring them with scRNAseq and MS-based scProteomics. NanoSPLITS is built on the nanoPOTS platform that allows for high-efficiency proteomic preparation of single-cells by miniaturizing the assay to nanoliter scale volumes^{19,22}. We have previously demonstrated reaction miniaturization not only reduces non-specific adsorption-related sample losses, but also enhances enzymatic digestion kinetics²². Similarly, we reason the use of nanoliter droplets can improve overall sample recovery of both mRNA transcripts and proteins for sensitive and comprehensive single-cell multiomics.

Herein, we demonstrate that nanoSPLITS-based multiomics can provide comprehensive molecular characterizations of both secondary and primary cells with insights into covarying functional protein/gene clusters, as well as unique phosphorylation events which are typically unavailable in single omics measurements. The capability to obtain the proteome profiles of individual cell types in complex tissue dissociations is highly desirable for therapeutic target discovery and biomarker identification. However, it is highly challenging to measure many single cells with current label free LC-MS-based proteomics workflows. To address this problem, we also develop an efficient approach for cell-type annotation and protein marker detection by mapping nanoSPLITS transcriptomics data to existing reference databases. This reference mapping approach significantly reduces the required cell numbers for clustering-based cell-type annotation.

Results

Optimization of nanoSPLITS provides deep transcriptome and proteome coverage in single cells

The overall workflow of the nanoSPLITS-based single-cell multiomics platform is illustrated in Fig. 1. Briefly, we employed an image-based single-cell isolation system to directly sort single cells (contained in ~400 pL volume droplets) into lysis buffer (200 nL), followed by a single freeze-thaw cycle to achieve cell lysis. Notably, the selective hydrophobic/hydrophilic patterning of nanoSPLITS chips ensures that droplets are constrained to the 1.2 mm wells on the chip. Next, the microchip containing single-cell lysate is manually aligned with a separate chip containing only cell lysis buffer. The droplet arrays on the two chips are merged for 15 s and then separated, before repeating twice (Supplementary Movie 1). As nanoSPLITS chip contains 48 wells of 1.2 mm diameter separated by 4.5 mm of space, chips can easily be aligned to accomplish droplet merging. The chip receiving lysate (“acceptor” chip) was then transferred into a 384-well plate for scRNAseq with Smart-seq 2 protocol¹, while the chip initially containing the lysate (“donor” chip) is digested using an n-dodecyl- β -D-maltoside (DDM)-based sample preparation protocol and directly analyzed with an ion-mobility-enhanced MS data acquisition method for scProteomics²³.

We first optimized cell lysis buffers to ensure compatibility with both proteomic and transcriptomic workflows. Typically, scProteomics utilizes a buffer containing DDM to reduce non-specific binding of proteins to surfaces while scRNAseq often includes recombinant protein-based RNase inhibitors to reduce mRNA degradation. To evaluate their impacts on both methods, we tested the inclusion of these additives in a moderately buffered hypotonic solution (10 mM Tris, pH 8) with 20 mouse alveolar epithelial cells (C10). The inclusion of 1 x RNase inhibitor modestly suppressed proteomic identifications, while 0.1% DDM had no significant impact on transcript identifications (Supplementary Fig. 1a,b). Furthermore, removal of RNase inhibitor from RNAseq analysis had minimal effect on gene identifications (Supplementary Fig. 1a). Therefore, we decided to use 0.1% DDM (w/v) in 10 mM Tris (pH 8) solution as the cell lysis buffer for our initial nanoSPLITS experiments.

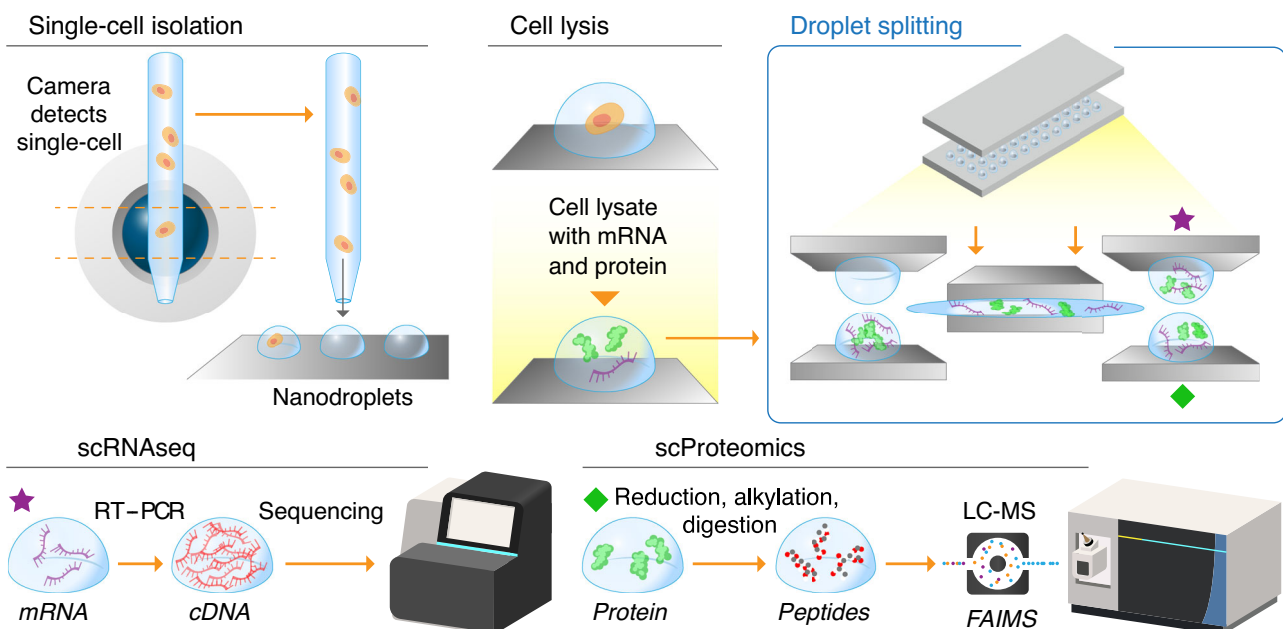


Fig. 1 | Overview of the nanoSPLITS-based single-cell multiomics platform. Schematic illustration showing the workflow including cell sorting, lysis, droplet merging/mixing, and droplet splitting for downstream scRNAseq and scProteomics measurement.

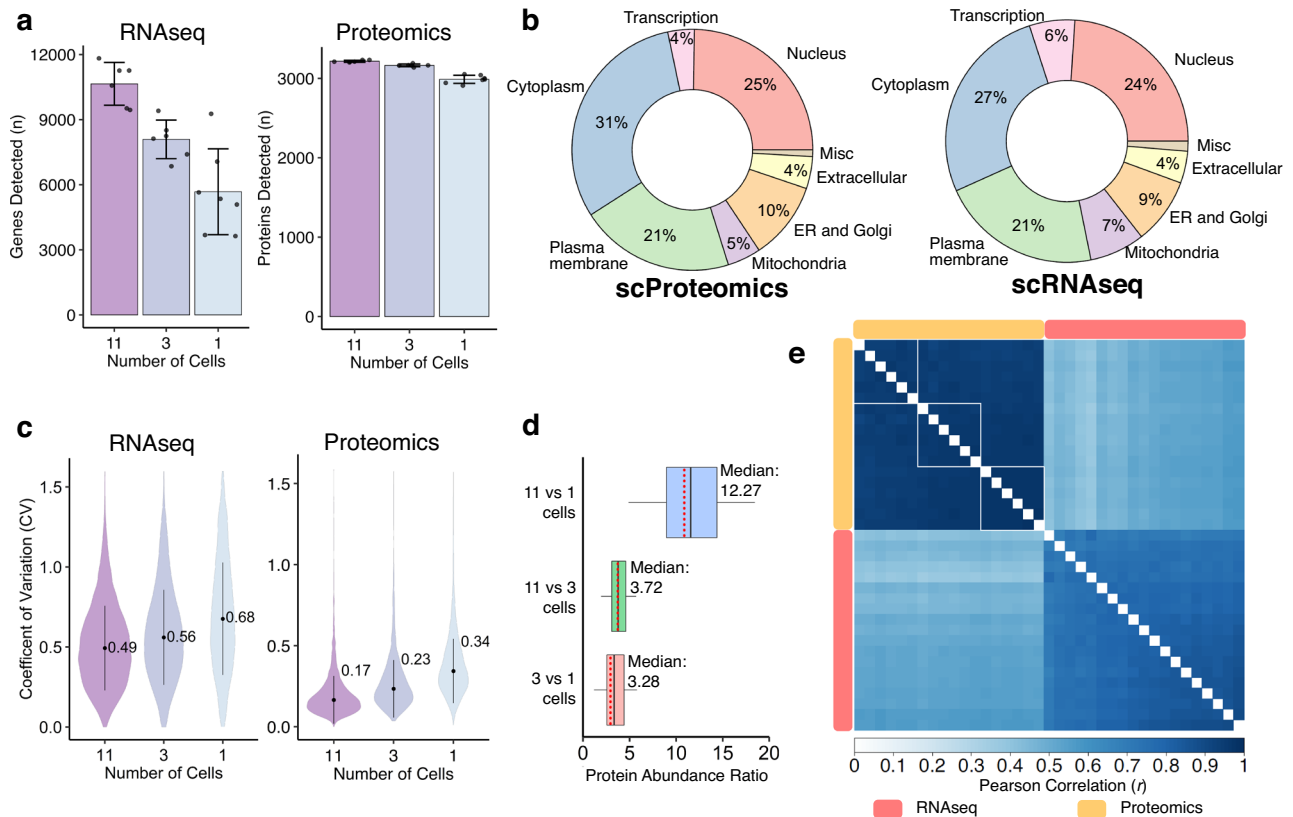


Fig. 2 | Qualitative and quantitative assessment of nanoSPLITS for transcriptome and proteome measurements. **a** Mean number of detected genes and proteins for each modality. Error bars indicate standard deviations (\pm s.d.). **b** CCO distributions of genes (scRNAseq) and proteins (scProteomics) identified in the single-cell data. **c** Distributions of the coefficients of variation (CV) for all genes and proteins with at least 2 observations across replicates. Indicated values represent median CVs, which are also shown at the center point within each distribution. Whiskers extending from the center point represent standard deviation (\pm s.d.). **d** The ratios of average protein abundances were calculated for comparisons between the different pooled cell samples. Experimental medians are indicated at

the black crossbar while the theoretical ratio for each comparison is shown at the red dotted line within each boxplot. Boxplot boundaries extend from the first (25%) and third (75%) quartile, representing the interquartile range (IQR). Boxplot whiskers extend to $0.75 \times \text{IQR}$ while outliers are not shown. **e** Pearson correlation heatmap with clustering of transcriptomics and proteomics results. Only the proteomic data showed complete clustering of 11, 3 and 1 cell samples, indicated by the inscribed white squares. Self-comparisons along the diagonal are excluded (white). For (**a–d**), and (**e**) sample size for all analyses were $n = 6$ for 11, 3, and 1 C10 cell, with the exception the 1 C10 cell group analyzed by scRNAseq which had an $n = 7$.

Next, we evaluated the split ratios between two 200 nL droplet arrays. Using fluorescein as an initial model, the nanoSPLITS procedure achieved splitting ratios between 46% to 47%, with 50% representing a perfectly equal split (Supplementary Fig. 2 and Supplementary Table 1). As the properties of fluorescein may not fully model proteins, we split 10 C10 cells ($n = 6$) and quantified protein abundances in both chips with LC-MS. Encouragingly, we found the split ratios to be consistent and precise across the chips, with a median CV of 0.12 for the 10 pooled cell samples (Supplementary Fig. 3a). Using the mean protein abundance for each protein identified from both chips, we determined the relative proportion of each protein found on either chip (Supplementary Fig. 3b). Surprisingly, the median proportion of protein retained on the donor chip was $\sim 75\%$, greater than what would be expected in a perfect split. One potential explanation is that the cell lysate may require more time for diffusion between the merged droplets, although the diffusion coefficient of GFP in water ($\sim 100 \mu\text{m}^2\text{s}^{-1}$) suggests merging on the order of seconds should be sufficient. In addition, a diffusion-based effect would be expected to have a size dependence, which is not apparent based on the proteins quantified (Supplementary Fig. 3c). Another explanation might be that the modest hydrophobic coating of the nanowells captured proteins on the surface, resulting in more proteins retained on the donor chip. Thus, we took advantage of this outcome and chose the donor chip for scProteomic measurements to obtain deeper proteome coverages (Fig. 1).

We next sorted several quantities of C10 cells (groups of 11, 3, and 1 cell into separate wells) and measured them using the multiomics workflow. Considering a minimum of 5 reads per gene for transcript identifications and 1% FDR cutoff for protein identifications (5% at ion level during match-between runs, MBR), robust coverages of genes and proteins were achieved across all tested conditions (Fig. 2a). As expected, coverage reduced with decreasing cell numbers. Single-cell transcriptome and proteome measurements provided 5848 and 2934 ($n = 7$ cells) identifications on average, respectively. To ensure nanoSPLITS did not introduce bias toward different cellular components due to the nanodroplet splitting process, we investigated the distribution of gene and protein identifications from single cells within cellular component ontology (CCO). We found scProteomics and scRNAseq had corresponding identifications within cellular components that encompassed all major organelles (Fig. 2b). Furthermore, 25% of the proteins from the scProteomics analyses had CCO localizations to the nucleus, 219 of which have known roles in transcription. This is notable as nuclear proteins are typically drivers in gene regulation and transcription, and current multimodal technologies have been limited in the ability to directly measure nuclear protein abundances. The distribution of nuclear proteins across cellular components closely mirrors that of the entire mouse proteome (Supplementary Fig. 3d), and the percentages are only 4% and 3% higher than the entire proteome for proteomic and transcriptomic datasets, respectively.

We next evaluated the quantitative reproducibility for each modality by calculating the coefficients of variation (CVs) of transcriptome and proteome abundances. Median transcriptome CVs ranged from 0.49 for 11 cells to 0.68 for single cells, while proteome median CVs ranged from 0.17 for 11 cells to 0.34 for single cells (Fig. 2c). The trend of increasing CVs from 11 to 1 cell was expected, as the mixed cell populations represent averages of underlying cell to cell variation. Notably, the median CV (0.68) for single-cells measured by Smart-seq2 were higher than the corresponding scProteomic median CV (0.34). This is also in agreement with recent reports that have shown median CVs across single-cells to be as high as 1.08^{19,21,24}. These higher CVs partially reflect the dynamic nature of mRNA transcript expression relative to their protein counterparts, which have longer half-lives on average, high copy numbers, and therefore can be more precisely measured²⁵. We also compared the ratios of the measured protein abundances between the different cell inputs. Encouragingly, the measured fold differences between the median intensities for 11, 3, and 1 C10 cell are very close to the expected theoretical values (Fig. 2d). For example, the median protein abundance ratio for 3 cells compared to single-cells was 3.28, within 10% of the theoretical 3-fold difference. These results provide strong evidence that nanoSPLITS-based single-cell multiomics platform can provide sensitive and reproducible quantitative measurements of both the transcriptome and proteome of the same single cells.

Finally, we determined the Pearson correlation coefficients (r) across and within modalities using conceptually similar normalized transformations for each modality (Fig. 2e and Supplementary Fig. 4; TPM, transcripts per million for transcriptomics, and riBAQ, relative intensity-based absolute quantification for proteomics). In line with the CV distributions (Fig. 2c), proteomics data had better agreement between samples compared with transcriptomics data, once again highlighting the relatively stable nature of the proteome and the dynamic nature of the transcriptome where many genes are often expressed in short transcriptional “bursts”²⁵. The cross-correlation between the transcriptome and proteome in single cells was moderate with most correlation coefficients falling within the range of 0.35–0.45, on par with previous reports^{19,21,25}.

Classification of cell types and identification of markers with nanoSPLITS

Having established baseline characteristics of multimodal data, we then applied nanoSPLITS to analyze two cell lines: mouse epithelial (C10) and endothelial cells (SVEC). As nanoSPLITS uses only a portion of the total cellular content, we sought to ensure our multimodal measurements could distinguish cell types and detect gene/protein markers. As shown in Fig. 3a, both cell types and modalities were easily clustered based on correlations alone. Within-modality correlations were higher in scProteomics than scRNAseq for both cell types (Supplementary Fig. 5a). Cross-modality correlation analysis produced a slightly broader range of r (Supplementary Figs. 5a and 6a). We also compared the cross-modality correlations between same single-cells (“intracell”) and different single-cells (“intercell”), however there appeared to be no significant difference between them (Supplementary Fig. 5a). Overall, SVEC cells exhibited lower correlations and reduced measurement depth/precision (Supplementary Figs. 5b and 6a). The lower correlations and reduced measurement depth may be related to cell size²⁶, however we should note that factors related to metabolic state, proteome turnover or PTMs may also play an important role. On average, C10 cells had ~1800 overlapping identifications while SVEC cells had ~900 overlapping identifications between scRNAseq and scProteomic modalities. All proteins (3,609) detected in the scProteomic data could be detected as mRNA transcripts in the scRNAseq data; and, as expected, the majority of detected proteins were derived from mRNA transcripts of higher abundance (Supplementary Fig. 6b). Finally, we calculated mRNA-

protein correlations for each gene that was observed in both modalities. The distribution of mRNA-protein correlations was statistically different compared to a distribution of randomly sampled correlations, albeit modestly (Fig. 3b).

Next, we evaluated if the multiomics data could be used to identify cell-type-specific marker genes and proteins. Figure 3c shows the top 5 significantly enriched genes and proteins for each cell type. Interestingly, the overlap between scProteomics and scRNAseq of these significant markers was relatively low, indicating the widely used scRNAseq method may not be sufficient to provide reliable marker genes for protein-based functional assays. Encouragingly, the previously established SVEC-cell marker *H2-K1* was identified at both the protein and mRNA level (Fig. 3c). Dimensionality reduction with principal component analysis (PCA) delineated both cell types for scRNAseq and scProteomics despite the cell contents being divided (Supplementary Fig. 7). The integration of two modalities through a weighted nearest neighbor (WNN)²⁷ analysis provided robust clustering in the two-dimensional space and visual confirmation of SVEC marker *H2-K1* (Fig. 3d). Using canonical cell cycle markers²⁸, we could also identify individual C10 cells in different cell cycle phases, suggesting cell to cell variation was retained after the droplet splitting process (Fig. 3e and Supplementary Fig. 8). For example, cyclin-dependent kinase 1 (*Cdk1*) is upregulated at the transcriptional level in S and G2M phase C10 cells (Fig. 3e). Several other established cell cycle phase genes demonstrated similar differential abundance at the transcriptional level, including DNA topoisomerase II α (*Top2a*), cyclin B1 (*Ccnb1*), G2 and S phase expressed protein 1 (*Gtse1*), cytoskeleton-associated protein 5 (*Ckap5*), and anillin (*Anln*) (Supplementary Fig. 8). Furthermore, TOP2A protein appeared to be increased in G2/M cells as well.

Characterization of cell cycle features through the integration of scRNAseq, scProteomics, and scPhosphoproteomics

We next investigated more deeply and validated the cell cycle features identified in C10 cells after nanoSPLITS. Towards this end, C10 cells were arrested in G2/M phase with a CDK1 inhibitor (RO-3306) and compared to an untreated C10 cell population²⁹. NanoSPLITS facilitated comprehensive proteome and transcriptome measurements, with an average of 2942 proteins and 5559 genes identified in G2/M arrested cells as well as 2574 proteins and 4173 genes for untreated cells. The higher number of identifications for G2/M arrested cells can be attributed to their larger size (and therefore higher protein/mRNA content), which was noted during cell sorting (Supplementary Fig. 9). Although no phosphopeptide enrichment was performed³⁰, the FragPipe proteomics pipeline with FDR-controlled MBR also identified over 300 unique phosphopeptides, of which 138 were reproducibly observed with < 50% missing values. G2/M arrested and untreated cells could easily be clustered and separated by dimensional reduction (PCA and WNN-UMAP) for scProteomic and scRNAseq data (Supplementary Fig. 10). Overall, 3182 proteins and 4186 mRNA transcripts were tested for differential abundance, and 1930 protein/genes had paired data from nanoSPLITS. Quantitative analysis afforded 327 proteins, 1434 genes, and 29 phosphopeptides that were differentially abundant ($\log_2FC > 0.5$ or < -0.5 and $FDR < 0.01$). Importantly, covariant protein and mRNA clusters were identified by hierarchical clustering (Fig. 4a, b, Supplementary Fig. 11).

Of the proteomic clusters identified, clusters 2, 3, and 4 contained members with the strongest functional relationships. Cluster 2 (Fig. 4a, b) contained 7 proteins that were significantly more abundant in the untreated C10 cells. These were mainly extracellular matrix (ECM) proteins, including three collagens (COL1A1, COL3A1, and COL12A1), fibronectin (FN1), and procollagen C-endopeptidase enhancer 1 (PCOLCE). The remaining two proteins were endoplasmic reticulum junction formation protein lunapark (LNPK) and Synembryon-A (RIC8A), both of which play important roles within

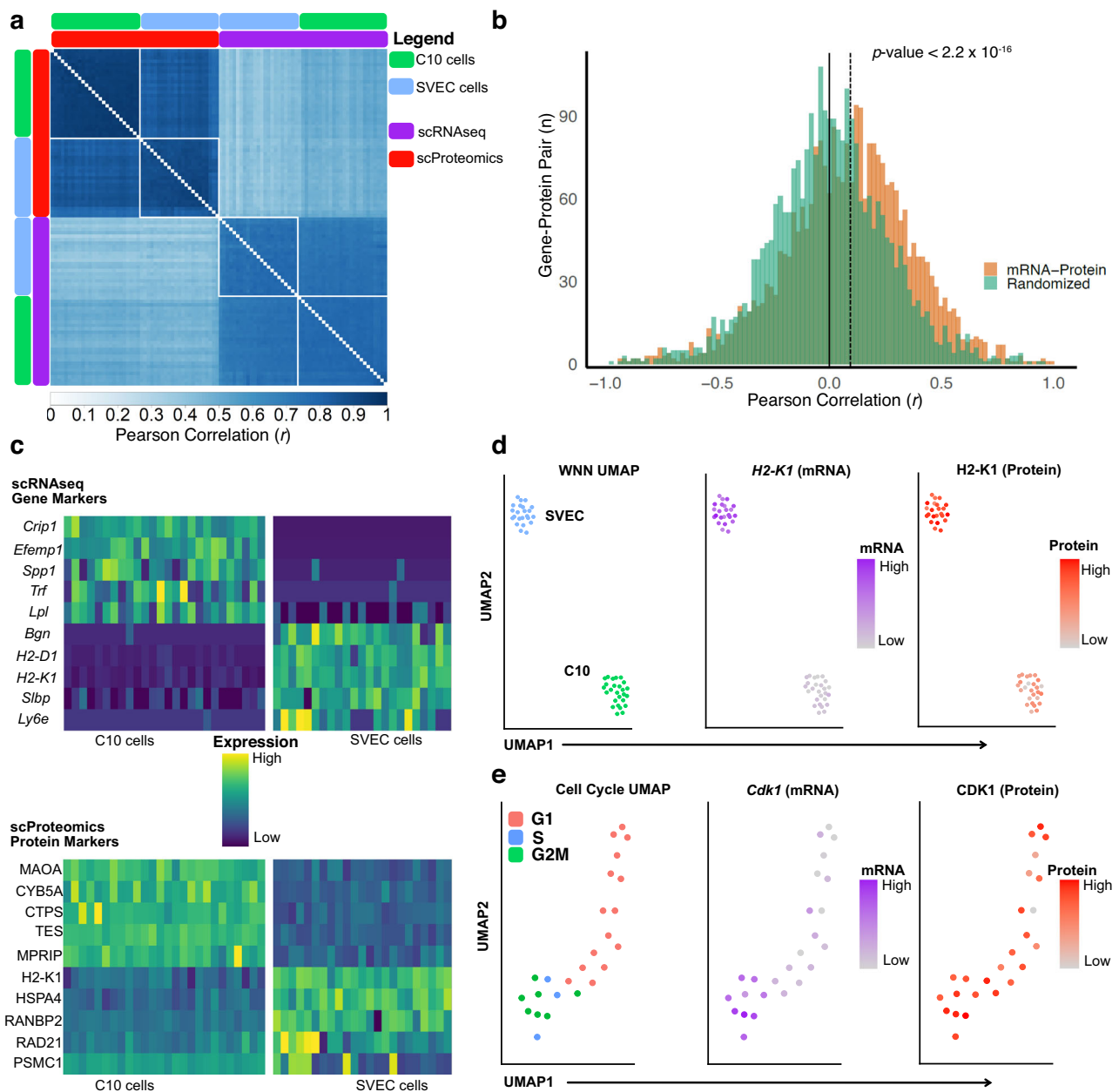


Fig. 3 | Underlying cell phenotype signatures are maintained after nanoSPLOTS.

a Pearson correlation heatmap with clustering of scRNAseq and scProteomics results for both single C10 ($n = 26$, paired) and SVEC ($n = 23$, paired) cells. **b** Histogram of mRNA-protein correlations for each gene quantified in both modalities with at least 4 observations (performed separately for C10 and SVEC cell types). Statistical testing was performed by a two-sided Wilcoxon rank-sum test. Solid and dashed line indicates median of randomized and experimental correlations, respectively. **c** Top 5 gene markers from scRNAseq data and protein markers from scProteomics data for each cell type. Candidate marker features were

determined using a two-sided Wilcoxon rank-sum test (adjusted p -values < 0.001). **d** Weighted-nearest neighbor (WNN) Uniform Manifold Approximation and Projection (UMAP) generated using Seurat to integrate scRNAseq and scProteomic data. Middle and right panels are colored based on H2-K1 gene (purple) and protein (red) abundance, respectively. **e** Feature-based UMAP generated for C10 cells using cell-cycle markers derived from scRNAseq data. Middle and right panels are colored based on *Cdk1* gene (purple) and protein (red) abundances, respectively. All abundance values shown in (c–e) are Z-scores after scaling and centering of data.

mitosis^{31,32}. Cluster 3 provided 13 proteins that were more abundant in G2/M arrested C10 cells, all of which were enriched for mitotic-relevant annotations (Supplementary Fig. 12). Interestingly, cluster 3 did not show concordance between the two modalities which we attribute to the temporal difference between peak translation and protein abundance in preparation for and during mitosis (Fig. 4b)³³. Cluster 4 contained genes with the largest \log_2 FC differences at both the mRNA and protein level, and was largely composed of glutathione S-transferases and cathepsins (Supplementary Fig. 13). For cluster 2

and cluster 4, strong concordance between the scProteomic and scRNAseq data was noted (Fig. 4b, see Supplementary Information for further discussion on covariant clusters identified in this study).

From the over 300 phosphopeptides from scProteomic data, 29 were dysregulated in G2/M arrested cells (Fig. 4b). We also compared the relative abundance of these phosphopeptides to their corresponding protein and mRNA abundances. Notably, vimentin (VIM) pSer55, a known phosphosite generated by CDK1 typically observed during the transition from prometaphase to metaphase^{34,35}, was

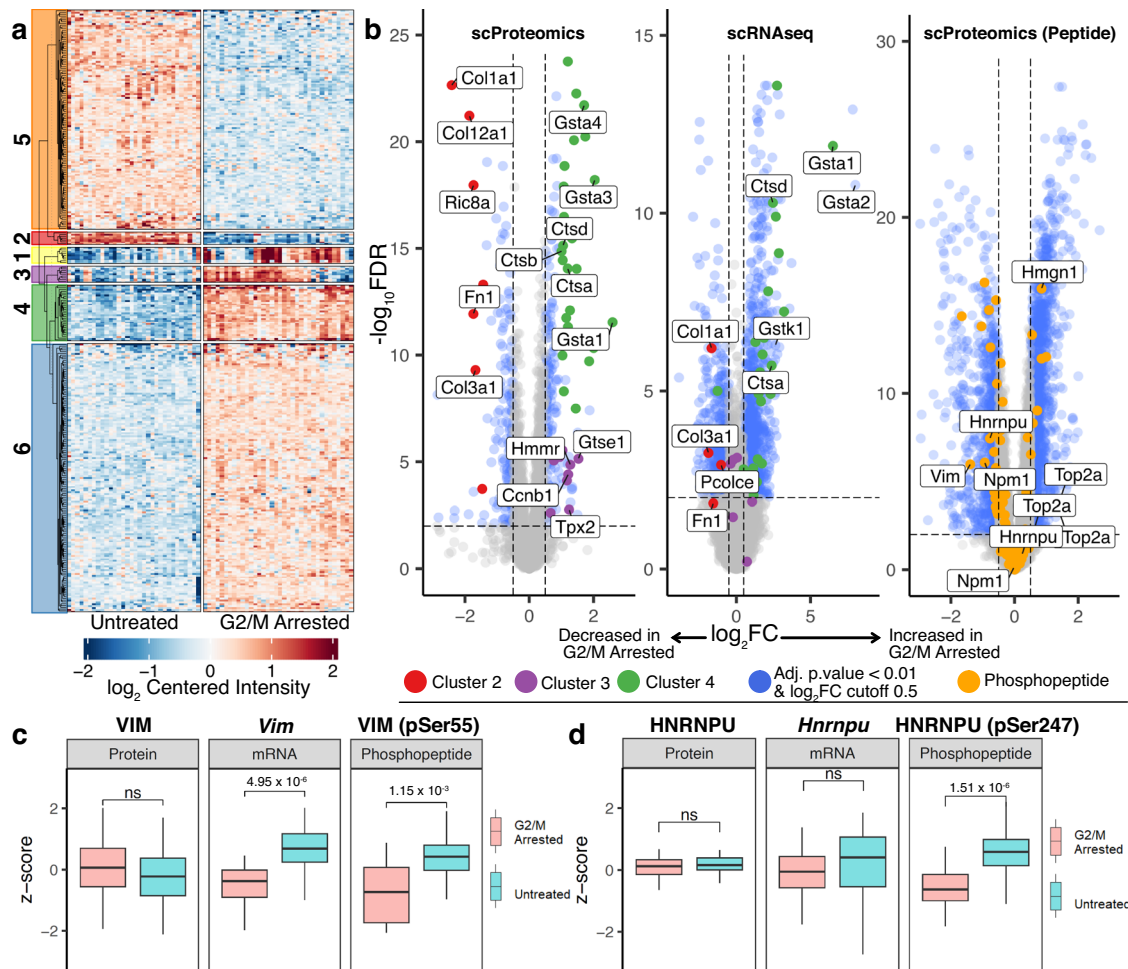


Fig. 4 | Protein, mRNA, and phosphopeptide cell cycle features from same single cells. **a** Clustergram of \log_2 centered intensities for differentially abundant proteins from scProteomic data with $FDR < 0.01$ and \log_2FC of ± 0.5 (327 proteins). Columns (cells) are clustered by K-means ($k = 2$), while rows cluster proteins ($k = 6$). Colored areas along the y-axis indicate protein clusters. **b** Volcano plot of G2/M arrested cells/untreated cells for scProteomic data (3182 proteins quantified), scRNAseq data (4186 genes with $< 50\%$ missing values), and peptide level (16,938 peptides). Differential abundance statistical tests were performed using a two-sided, limma/Empirical Bayes t-test with p -values adjusted for multiple hypothesis

tests. **c, d** Integration of scRNAseq, scProteomic, and phosphoproteomic data by comparing relative abundances from each modality for VIM and HNRNPU, respectively. Statistical differences were determined using a two-sided Wilcoxon rank-sum test. Adjusted p -values are presented above boxplots and “ns” indicates not significant. The center line in boxplots represents the median while the box boundaries extend from the first (25%) and third (75%) quartile, representing the interquartile range (IQR). Boxplot whiskers extend to $1.5 \times IQR$ while outliers are not shown. Sample sizes for each cell type were $n = 36$ for G2M arrested cells and $n = 32$ for untreated cells.

decreased in abundance significantly in G2/M arrested cells (Fig. 4c). This decrease in protein phosphorylation appeared to be independent of the global protein abundance (Fig. 4c), providing the most direct evidence for the loss of CDK1 kinase activity inhibited by RO-3306²⁹. Phosphosite pSer247 on heterogeneous nuclear ribonucleoprotein U (HNRNPU) was also found to be changing independent of overall protein abundance and followed a similar trend as VIM pSer55, suggesting it may be a target of CDK1 (Fig. 4d).

Leveraging large-scale single cell transcriptomics reference databases to confidently annotate cell types and identify functionally relevant protein markers

Having established the precision, depth, and scope of our nanoSPLITS measurements in secondary cell lines, we next applied nanoSPLITS to dissociated human pancreatic islet cells from two donors. The overall identification depth across the dissociated islet cells was comparable to C10 and SVEC cells, with an average of 1775 genes and 1949 proteins identified per cell. A notable advantage of having parallel scRNAseq data from nanoSPLITS measurements is the ability to annotate cells via

reference maps from published scRNAseq experiments. By integrating nanoSPLITS scRNAseq data into the reference-based mapping pipeline in Seurat (Azimuth)^{27,36}, we could assign nine different cell types across 106 cells (Fig. 5a). These cell types encompassed both pancreatic exocrine (acinar and ductal cells) and endocrine (alpha, beta, delta, and gamma cells) tissue. Notably, resident immune cells (representing only 0.6% of cells in the Azimuth reference atlas) in the pancreas could also be identified. Established protein markers showed strong agreement with the cell types annotated through Seurat (Supplementary Fig. 14). We next trained a sparse partial least squares discriminant analysis (splS-DA) model on the cell types annotated by Seurat. Applying the model to unannotated cells afforded 23 more annotated cells from the scProteomic data, with only 8 remaining “unclassified” (Supplementary Fig. 15). Therefore, our analysis encompassed 106 cells with scRNAseq data, 126 cells with scProteomic data, and 95 cells with both modalities (Supplementary Fig. 16).

With well-annotated cell types, we then focused on identifying unique protein markers for major endocrine cell populations and immune cells via pairwise abundance comparisons. Many proteins

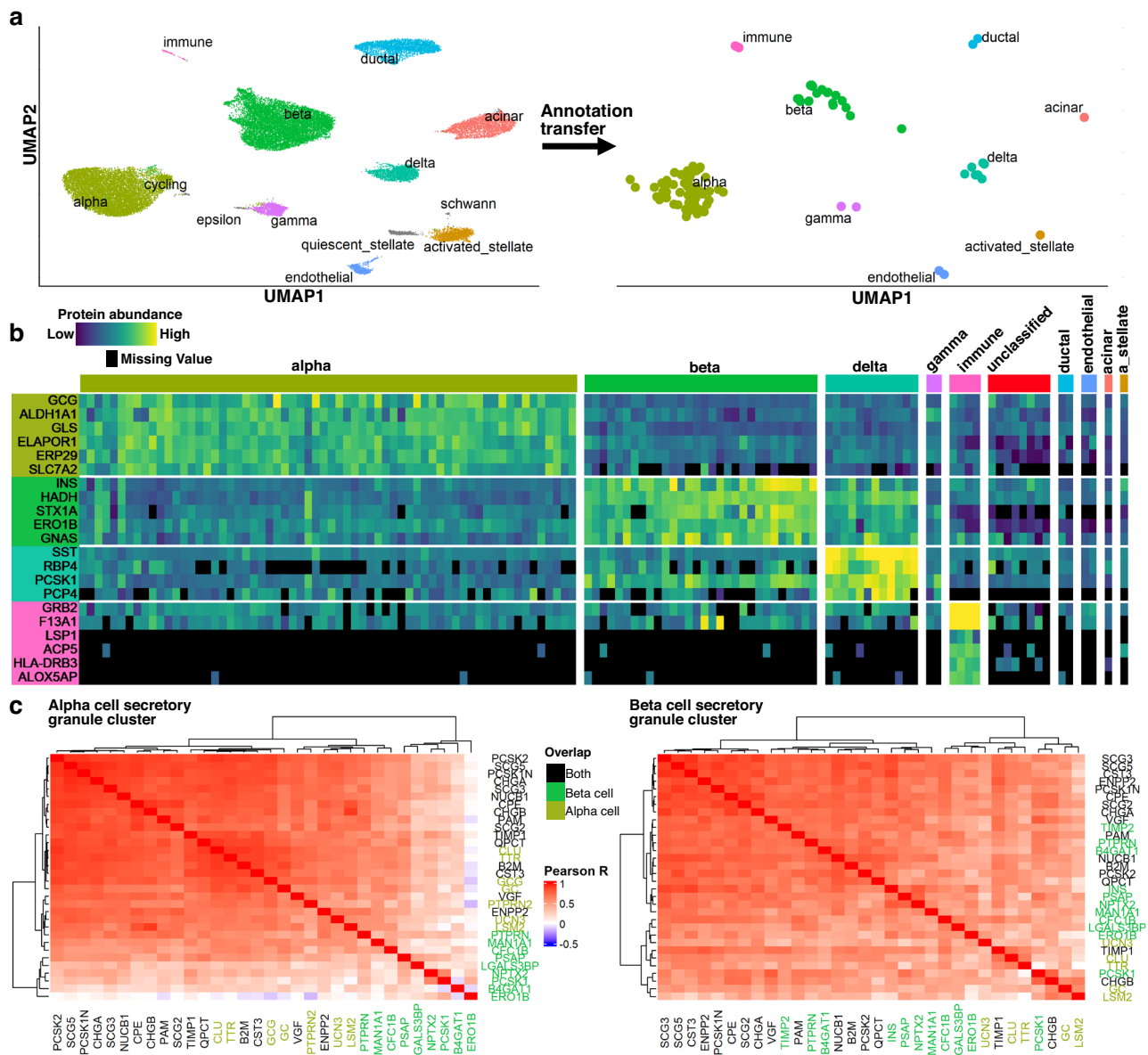


Fig. 5 | Multiomics analysis of dissociated pancreatic islet cells. **a** Left panel: UMAP of Azimuth human pancreatic reference map consisting of 35,289 cells. Cell types are indicated by color and inset text labels. Right panel: UMAP of 106 dissociated pancreatic islet cells from two human donors based on label transfer from Azimuth reference map to nanoSPLITS scRNAseq data. Cells with annotation scores of > 0.8 are shown. **b** Abundance heatmap of selected protein markers identified from nanoSPLITS scProteomics data for alpha cells ($n = 64$), beta cells ($n = 30$), delta cells ($n = 12$), immune cells ($n = 4$), and unclassified cells ($n = 8$), compared across all

annotated cell types. Statistically significant markers were determined using a two-sided, limma/Empirical Bayes t -test (adjusted p -values < 0.01) or a hypergeometric test to identify markers based on missing values across cell types (adjusted p -values < 0.01). **c** Clustergram of Pearson correlations for secretory granule proteins hierarchically clustered for alpha (left panel) and beta cells (right panel). “Overlap” indicates if proteins were observed in both alpha and beta cell secretory granule clusters or only one cluster (alpha or beta).

with cell-type specific profiles could be observed (Fig. 5b). This includes several of the well-established markers for pancreatic endocrine cells, such as GCG for alpha cells, INS beta cells, and SST for delta cells. Importantly, we also identified several potential marker proteins that, to the best of our knowledge, have only been established at the mRNA level in single-cells (ALDH1A1, SLC7A2, and GLS for alpha cells; HADH and ERO1B for beta cells; RBP4 and PCSK1 for delta cells)^{37,38}, as well as proteins that have not been identified as markers in large scale sequencing datasets (ERP29 and ELAPOR1 with alpha cells, STX1A with beta cells, and PCP4 with delta cells). ALDH1A1 is noteworthy as it plays a role in beta cell maturation, while SLC7A2 and GNAS have only recently been found to have critical roles in alpha and beta cell

function, respectively^{39–41}. Owing to the different lineage from which they were derived, immune cells presented with the strongest markers amongst the tested cell types. Like the pancreatic endocrine cells, established mRNA markers were observed (ACP5, HLA-DRB3, ALOX5AP, and LSP1) as well as several only observed at the protein level (GRB2 and F13A1)²⁷.

Finally, we investigated the functional protein covariation profiles of alpha and beta cells to identify differences in their hormone secretion pathways. Through Gaussian Mixture Model (GMM) clustering of the scProteomics results, a cluster of ~25 proteins was found in each cell type containing known secretory proteins as well as the major secretory hormones GCG and INS for alpha and beta cells, respectively

(Supplementary Figs. 17, 18). Both clusters were significantly enriched for the gene ontological term “secretory granule” (*adjusted p-value* 2.1×10^{-6} for beta cells and 1.6×10^{-6} for alpha cells). Proteins from both clusters were then used to construct hierarchical clustergrams specific to alpha cells or beta cells (Fig. 5c). From the 33 proteins in these secretory granule clusters, 16 core proteins were found to overlap in both cell types. Proteins uniquely correlated within the alpha cell secretory granule cluster included several of known functional relevance in alpha cells (GC, TTR, and UCN3)^{37,42,43}. Of the 11 proteins unique to the beta cell cluster, four were largely uncorrelated in alpha cells (NPTX2, B4GAT1, PCSK1, and ERO1B). NPTX2 has only recently been identified as a beta cell marker at the mRNA level, and ERO1B is an oxidoreductase necessary for proper folding of proinsulin^{37,44}. Finally, it is notable that metalloproteinase inhibitor TIMP2 deviates significantly from the related TIMP1, which appears to be a core protein in both secretory pathways. TIMP2 is frequently observed in beta cells (28 out of 30 beta cells) and unique to the beta cell secretory granule cluster, however only sparsely observed in alpha cells at much lower abundance (18 out of 64 alpha cells). The role of TIMP proteins in hormone secretion appears to be relatively unknown, however TIMP1 is at least known to localize to compartments resembling secretory vesicles in neutrophils⁴⁵.

Discussion

Compared with previous technologies that utilize antibodies to infer protein abundances in multimodal experiments, the nanoSPLITS platform employs MS to globally detect proteins and post-translational modifications (PTMs). We have established how the nanoSPLITS approach can enable multimodal profiling of thousands of transcripts and proteins, as well as hundreds of phosphopeptides from same single cells. Quantification of protein abundances from 1L, 3, and 1 C10 cells verified experimental abundance ratios were close to theoretical and highly precise. Despite splitting cellular contents, we found the molecular coverage to be broadly distributed across different cellular compartments for both modalities, and global across-modality correlations were in line with prior studies²⁵. Finally, applying nanoSPLITS to both secondary cell lines and human primary cells enabled delineation of different cell types and identification of candidate marker genes and/or proteins.

Although our experimental design was not targeted towards clarifying the relationship between mRNA and protein abundance, we could calculate the correlations for all genes with corresponding transcript-protein measurements derived from same single cells (C10 and SVEC in Fig. 3b, as well as dissociated pancreatic islet cells in Supplementary Fig. 19). The mRNA-protein correlations measured could be distinguished from distributions of randomly sampled measurements, and a trend towards more positive correlations was observed. However, the medians of the distributions were only modestly shifted from 0, providing further evidence that the lack of correlation between transcript and protein levels is a general phenomenon^{11,46}. Indeed, the inherent stochasticity of gene expression is well documented⁴⁷. Overall, our results agree with the perspective that correlations between transcripts and proteins at the single-cell level are less correlated relative to bulk measurements which “average out” different sources of variance within a population of cells^{6,11}. However, modest shifts in correlations towards positive r suggests that at least some transcripts are more tightly linked to their translated products. We anticipate that sampling larger sets of cells will provide unique insights into regulation of protein expression.

We also demonstrated nanoSPLITS can identify cell states, such as mitotic cell cycle phases. By comparing cells arrested in G2/M via the CDK1 inhibitor RO-3306 to an actively cycling C10 population, we could identify abundance changes for proteins and transcripts in line with the expected phenotype, as well as covarying protein/gene clusters relevant to mitotic processes. Over 300 phosphopeptides were

also identified without phosphopeptide enrichment, providing one of the deepest single-cell phosphoproteomic datasets to date^{30,48}. From this dataset, we found phosphosites pSer55 VIM and pSer247 HNRNPU were strongly impacted by CDK1 inhibition. Interestingly, the phosphorylation of HNRNPU by PLK1 at a separate site (pSer59) is known to be a key step in mitosis that requires an upstream priming phosphorylation event by CDK1, but the exact site that CDK1 acts on HNRNPU has yet to be identified⁴⁹. Our data suggests this upstream priming by CDK1 may be positioned at Ser247. These phosphoproteomic observations demonstrate how additional insight can be acquired from PTMs³⁰.

Finally, we developed a powerful approach to annotate unknown cell types by leveraging large-scale scRNAseq reference database with nanoSPLITS multiomic data. This highlights a unique capability of single-cell multimodal analyses: utilizing one modality (e.g., transcripts) as a bridge to inform cell types. This allows for discovery of cellular markers in other modalities (e.g., proteins). Such capability is particularly useful for identifying protein markers in complex cell populations with scProteomics, as only a small number of single cells need to be measured. Additionally, the approach avoids having to use the same data for both clustering and differential expression analysis (a process referred to as “double dipping” that inflates false-positives)⁵⁰. By performing nanoSPLITS on human pancreatic islet cells, we demonstrated how most islet cell types, including rare cell types such as immune cells, could be confidently annotated. We provide validation for previously identified mRNA-based markers and found several additional protein markers within these cell types as well. With the ongoing efforts in large-scale single-cell mapping such as Human Cell Atlas and HuBMAP^{51,52}, reference maps for many organs are readily available. Thus, we envision the integration of reference mapping with nanoSPLITS multiomics as a powerful approach for discovery proteomics in heterogeneous cell populations.

The nanoSPLITS platform holds promise to become a discovery tool for biomedical applications through characterization of cellular heterogeneity and by facilitating the integration of RNAseq data with protein abundance and PTM measurements at the single-cell level. In addition to phosphorylation, PTMs such as N-terminal acetylation and pyro-glutamate at N-terminal Glu/Gln were found (Supplementary Data 1), exemplifying the benefits of MS-based proteomics in single-cell analysis. Although a low throughput approach was employed in this study, high-throughput multiplexing approaches such as CEL-Seq⁵³ for transcriptomics and SCoPE-MS¹⁸ for proteomics can be readily integrated into the nanoSPLITS workflow. Our initial experiments indicated nanoSPLITS workflow can be adjusted to be compatible with downstream multiplexing labelling step. For example, the Tris buffer can be replaced with HEPES or TEAB buffer to be compatible with Tandem Mass Tags (TMT) labeling. The technique is also not restricted to the two modalities demonstrated here (transcriptomics and proteomics/phosphoproteomics); other modalities such as metabolomics, genomics, and epigenomics can conceptually be integrated into the workflow. Biomolecule enrichment technologies such as oligo (dT) beads for mRNA and immobilized metal affinity chromatography (IMAC) for phosphopeptide enrichment can be integrated to improve both recovery and reproducibility. As more analytical frameworks for integrating multimodal data are created, we anticipate nanoSPLITS (and/or nanoSPLITS-derived technologies) will also enable greater insight into how different modalities interact with each other to orchestrate cell phenotypes in health and disease.

Methods

Reagents and chemicals

Deionized water (18.2 M Ω) was purified using a Barnstead Nanopure Infinity system (Los Angeles, CA, USA). n-dodecyl- β -D-maltoside (DDM), CDK1 inhibitor RO-3306, iodoacetamide (IAA), ammonium bicarbonate (ABC), and formic acid (FA) were obtained from Sigma (St.

Louis, MO, USA). Nuclease-free water (not DEPC-treated), mass spectrometry grade Trypsin Platinum (Promega, Madison, WI, USA) and Lys-C (Wako, Japan) were reconstituted in 50 mM ABC before usage. Dithiothreitol (DTT, No-Weigh format), acetonitrile (ACN) with 0.1% FA, and water with 0.1% FA (MS grade) were purchased from Thermo Fisher Scientific (Waltham, MA, USA). SMART-Seq V4 Plus kit (Cat# R400753) was purchased from Takara Bio USA.

Design, fabrication, and assembly of the nanoSPLITS chips

The nanoSPLITS chips were fabricated using standard photolithography, wet etching, and silanization as described previously²². Two different chips were designed and used in this study. Both contained 48 (4 × 12) nanowells with a well diameter of 1.2 mm. The inter-well distance for the first chip was 2.5 mm while the second was 4.5 mm. Chip fabrication utilized a 25 mm × 75 mm glass slide pre-coated with chromium and photoresist (Telic Company, Valencia, USA). After photoresist exposure, development, and chromium etching (Transene), select areas of the chip were protected using Kapton tape before etching to a depth of -5 μm with buffered hydrofluoric acid. The freshly etched slide was dried by heating it at 120 °C for 1 h and then treated with oxygen plasma for 3 min (AP-300, Nordson March, Concord, USA). 2% (v/v) heptadecafluoro-1,1,2,2-tetrahydrodecyl-dimethylchlorosilane (PFDS, Gelest, Germany) in 2,2,4-trimethylpentane was applied onto the chip surface and incubated for 30 min to allow for silanization. The remaining chromium covering the wells was removed with etchant, leaving elevated hydrophilic nanowells surrounded by a hydrophobic background. To prevent retention of mRNA via interaction with free silanols on the hydrophilic surface of the nanowells, freshly etched chips were exposed to chlorotrimethylsilane under vacuum overnight to passivate the glass surface. A glass frame was epoxied to a standard glass cover slide so that it could be easily removed from the 2.5 mm inter-well distance chips for droplet splitting. For the 4.5 mm inter-well distance chips, PEEK chip covers were machined to fit the chip. Chips were wrapped in parafilm and aluminum foil for long-term storage and intermediate steps during sample preparation.

Cell culture

Two murine cell lines (NALIA clone C10 is referred to as C10 and is a non-transformed alveolar type II epithelial cell line derived from normal BALB/c mouse lungs; SVEC4-10, an endothelial cell line derived from axillary lymph node vessels) were cultured at 37 °C and 5% CO₂ in Dulbecco's Modified Eagle's Medium supplemented with 10% fetal bovine serum and 1× penicillin-streptomycin (Sigma, St. Louis, MO, USA). For the RO-3306 treatment, C10 cells were seeded at 200,000 cells per dish and incubated overnight. Treated cells were cultured with 10 μM RO-3306 for 36 h before harvesting. Control C10 cells were cultured similarly with vehicle (DMSO). The cultured cell lines were collected in a 15 ml tube and centrifuged at 1000 × g for 3 min to remove the medium. Cell pellets were washed three times by PBS, then counted to obtain cell concentration. PBS was then added to achieve a concentration of -200 × 10⁶ cells/mL. Immediately before cell sorting, the cell-containing PBS solution was passed through a 40 μm cell strainer (Falcon™ Round-Bottom Polystyrene Test Tubes with Cell Strainer Snap Cap, FisherScientific) to remove aggregated cells.

Human islets (1000–3000 islet equivalents) were obtained from Prodo Laboratories. Islets were immediately cultured in PIM(S) media in a humidified incubator at 37 °C and 5% CO₂ after being received. After 24 h in PIM(S) media, islets cells were collected and centrifuged at 100 × g for 5 min and resuspended in TrypLE™ for cell dissociation by incubating in a 37 °C water bath for 15 to 30 min along with gentle pipetting until islets were visibly dispersed and no clumps of cells remained. The dissociated islet cells were centrifuge at 300 × g for 5 min; the supernatant was removed, and cells were resuspended in PBS. The dissociated cells were stained with 0.1 μM Calcein AM and

15 μM propidium iodide for cell viability prior to cell sorting with the cellenONE.

nanoSPLITS buffer optimization experiment

Using nuclease-free water (Thermo Fisher Scientific, cat# 4387936), several 10 mM Tris pH 8 test buffers were created containing 0.1% DDM and/or 1 × RNase inhibitor. 10 μL of each buffer was added to four wells within a 96 well plate. 20 C10 cells were then sorted into each well before snap freezing with liquid nitrogen. Immediately after thawing and centrifugation at 2500 g, 5 μL from each well was transferred to a separate 96 well PCR plate containing 7.5 μL of 3' SMART-Seq CDS primer II before heating 70 °C for 3 min. 7.5 μL of RT mix (4 μL 5x ultra low first-strand buffer, 1 μL 48 μM SMART-Seq V4 oligonucleotide, 0.5 μL 40 units/μL RNase inhibitor, 2 μL SMARTScribe II reverse transcriptase) was then added before incubation at 42 °C for 90 min and 70 °C for 10 min. 30 μL of PCR master mix (25 μL SeqAmp PCR buffer, 1 μL PCR primer II A, 3 μL water, 1 μL SeqAmp DNA polymerase) was then added to each tube before performing 18 cycles of PCR (98 °C for 10 s, 65 °C for 30 s, 68 °C for 3 min). Isolation of cDNA was performed with Ampure XP beads with 80% ethanol washes. cDNA concentration and quality were determined with a Qubit fluorometer and Agilent fragment analyzer before next generation-sequencing, respectively.

The remaining 5 μL was retained and processed for label free proteomic analysis. Briefly, 5 μL of extraction buffer containing DTT and DDM was added to cell lysate to bring each sample to a final concentration of 1 mM DTT and 0.1% DDM before incubation at 60 °C for 1 h. 2 μL of 12 mM IAA was then added for a final concentration of 2 mM IAA before a 30 min incubation at 37 °C. 2 μL of 2.5 ng/μL Lys-C and 10 ng/μL of trypsin was added before incubation at 37 °C for 10 h. Enzymatic digestion was quenched by adding formic acid to a concentration of 1% before drying samples under vacuum. Peptides were reconstituted in 3 μL 5% acetonitrile 0.1% FA and transferred to a polypropylene microPOTS for proteomic analysis.

Fluorescein nanoSPLITS experiment

To the fluorescein-containing nanoSPLITS chip, 200 nL 0.01% 5,6-carboxyfluorescein solution containing 0.1% DDM was dispensed onto each well. For the PBS-containing chip, 250 nL PBS solution containing 0.1% DDM was dispensed onto each well. A slide cover was placed on both chips before wrapping them tightly in aluminum foil and placing them on ice to prevent evaporation until imaging. For imaging, both chips were placed on a chilled aluminum slide-holder and immediately imaged with brightfield light, followed by a Cy2 spectral filter using an AlphaMager FluorChemQ. Following imaging of the unsplit chips, two 2 cm² of 1/32" thick polyethylene foam was placed on one chip. The upper-chip was slowly lowered onto the bottom-chip, while carefully aligning the wells on both chips. Once the upper-chip was sitting on the separating foam, equal pressure was applied on the sides of the chip so that the droplets from both chips merged. Pressure was held for 15 seconds before releasing. The droplets were merged twice more following this process. The post-split chips were immediately placed back in the imager and final images were acquired. Quantification of droplet splitting was performed with the "Fiji" distribution of ImageJ (version 1.5). Briefly, Cy2 emission images were converted to grayscale. Regions of interest were selected (chip wells) and analyzed using the standard particle analysis in ImageJ. Each region of interest produced an average pixel intensity that was normalized by droplet area before using for quantification.

CellenONE cell sorting and nanoSPLITS workflow

Before cell sorting, all nanoSPLITS chips were prepared by the addition of 200 nL hypotonic solution consisting of 0.1% DDM in 10 mM Tris, pH 8 to each nanowell. A cellenONE instrument equipped with a glass piezo capillary (P-20-CM) for dispensing and aspiration was utilized for single-cell isolation. Sorting parameters included a pulse length of 50

μ s, a nozzle voltage of 80 V, a frequency of 500 Hz, a LED delay of 200 μ s, and a LED pulse of 3 μ s. The slide stage was operated at dew-point control mode to reduce droplet evaporation. Cells were isolated based on their size, circularity, and elongation in order to exclude apoptotic cells, doublets, or cell debris. For C10 cells, this corresponded to 25 to 40 μ m in diameter, maximum circularity of 1.15, and maximum elongation of 2, while SVEC cells were 24 to 32 μ m in diameter, maximum circularity of 1.15, and maximum elongation of 2. All cells were sorted based on bright field images in real time. The pooled C10 experiment had 11, 3, and 1 C10 cells sorted into each nanowell on a single 2.5 mm inter-well distance chip. For the SVEC and C10 comparison experiment, a single 48 well chip with 4.5 mm inter-well distance was used for each cell type and had a single cell sorted into each well. To perform the transferring identifications based on FAIMS filtering (TIFF) methodology for scProteomics²³, a library chip was also prepared containing 20 cells per nanowell, with each cell type sorted separately on the same chip to reduce technical variation. After sorting, all chips were wrapped in parafilm and aluminum foil before being snap-frozen and stored at -80°C , which may partially serve to induce cell lysis with a single freeze-thaw. All associated settings, single-cell images, and metadata can be accessed at nanoSPLITS GitHub repository [<https://github.com/Cajun-data/nanoSPLITS>]⁵⁴.

To accomplish splitting of the cell lysate, chips were first allowed to thaw briefly on ice. For each split, a complementary chip was prepared that contained the same 200 nL of 0.1% DDM in 10 mM Tris, pH 8 on each nanowell. The bottom chip containing the cell lysate was placed on an aluminum chip holder that was pre-cooled to 4°C within a PCR workstation (AirClean Systems AC600). Precut 1/32" thick polyurethane foam was placed around wells on the exterior of this bottom chip while the top chip was slowly lowered onto the polyurethane foam (Supplementary Movie 1). Wells were manually aligned for each chip before manual pressure was applied equally across the chip to merge the droplets for each chip. Pressure was held for 15 s before releasing. The droplets were merged twice more following this process. For consistency, the top chip was used for scRNAseq in all experiments while the bottom chip that initially contained the cell lysate was utilized in scProteomics (apart from the data generated for Supplementary Fig. 3). After merging, the top chip was immediately transferred into a 96-well or 384-well UV-treated plate containing RT-PCR reagents. For the pooled C10 (11, 3, and 1 cell) experiment, the transfer was performed by adding 1 μ L of RT-PCR buffer to each nanowell before withdrawing the entire volume and adding it to a 96-well plate. For all other nanoSPLITS experiments, the transfer was accomplished by laying the 4.5 mm inter-well distance chip onto a 384-well plate containing wells with the RT-PCR mix, sealed with a PCR plate seal, and then centrifuged at $3500 \times g$ for 1 min.

Sample preparation and LC-MS/MS analysis for scProteomics

All post-split chips were first allowed to dry out before placing them into the humidified nanoPOTS platform for sample processing. Protein extraction was accomplished by dispensing 150 nL of extraction buffer containing 50 mM ABC, 0.1% DDM, 0.3 x diluted PBS, and 2 mM DTT and incubating the chip at 50°C for 90 min. Denatured and reduced proteins were alkylated through the addition of 50 nL 15 mM IAA before incubation for 30 min in darkness at room temperature. Alkylated proteins were then digested by adding 50 nL 50 mM ABC with 0.1 ng/nL of Lys-C and 0.4 ng/nL of trypsin and incubating at 37°C overnight. The digestion reaction was then quenched by adding 50 nL of 5% formic acid before drying the chip under vacuum at room temperature. All chips were stored in a -20°C until LC-MS analysis.

We employed the in-house assembled nanoPOTS autosampler for LC-MS analysis. The autosampler contains a custom packed SPE column (100 μ m i.d., 4 cm, 5 μ m particle size, 300 \AA pore size C18 material, Phenomenex) and analytical LC column (50 μ m i.d., 25 cm

long, 1.7 μ m particle size, 190 \AA pore size C18 material, Waters) with a self-pack picofrit (cat. no. PF360-50-10-N-5, New Objective, Littleton, MA). The analytical column was heated to 50°C using AgileSleeve column heater (Analytical Sales and services, Inc., Flanders, NJ). Briefly, samples were dissolved with Buffer A (0.1% formic acid in water) on the chip, then trapped on the SPE column for 5 min. After washing the peptides, samples were eluted at 100 nL/min and separated using a 60 min gradient from 8% to 35% Buffer B (0.1% formic acid in acetonitrile).

An Orbitrap Eclipse Tribrid MS (ThermoFisher Scientific) with FAIMS, operated in data-dependent acquisition mode, was used for all analyses. Source settings included a spray voltage of 2400 V, ion transfer tube temperature of 200°C , and carrier gas flow of 4.6 L/min. For TIFF method²³ samples, ionized peptides were fractionated by the FAIMS interface using internal compensation voltage stepping (-45 , -60 , and -75 V) with a total cycle time of 0.8 s per compensation voltage. Fractionated ions within a mass range 350–1600 m/z were acquired at 120,000 resolution with a max injection time of 500 ms, AGC target of 1E6, RF lens of 30%. Tandem mass spectra were collected in the ion trap with an AGC target of 20,000, a "rapid" ion trap scan rate, an isolation window of 1.4 m/z , a maximum injection time of 120 ms, and a HCD collision energy of 30%. For TIFF, a single library sample was collected at each compensation voltage (-45 , -60 , and -75 V) using independent LC-MS runs with slight modifications to the above method where cycle time was increased to 2 s and maximum injection time was set to 118 ms. Precursor ions with a minimum intensity of 1E4 were selected for fragmentation by 30% HCD and scanned in an ion trap with an AGC of 2E4 and an IT of 150 ms. Precursor ions with intensities $>1\text{E}4$ were fragmented by 30% HCD and scanned with an AGC of 2E4 and an IT of 254 ms.

RT-PCR, sequencing, and read mapping for scRNAseq

Following the transfer of samples into a 384-well plate containing RT-PCR buffer with 3' SMART-Seq CDS Primer IIA (SMART-Seq[®] v4 PLUS Kit, cat# R400753), the samples were immediately denatured at 72°C for 3 min and chilled on ice for at least 2 min. Full length cDNA was generated by adding RT mix to each tube and incubating at 42°C for 90 min; followed by heat inactivation at 70°C for 10 min. 18 cycles of cDNA amplification were done to generate enough cDNA for template library according to SMART-Seq[®] v4 PLUS Kit instruction. The SMART-Seq Library Prep Kit and Unique Dual Index Kit (cat# R400745) were used to generate barcoded template library for sequencing. Single-read sequencing of the cDNA libraries with a read length of 150 was performed on NextSeq 550 Sequencing System using NextSeq 500/550 High Output v2 kit (150 cycles, cat#20024907). Data quality was assessed with FastQC and read-trimming was conducted using BBduk. Reads were aligned to the mouse genome (Genome Reference Consortium Mouse Build 39) or human genome (Human genome assembly GRCh38) using STAR. BAM file outputs were mapped to genes using 'featureCounts' as part of the Subread package with default settings or using the -O, -M and --fraction settings to allow genes with overlapping features to be counted.

Database searching and data analysis

All proteomic data raw files were processed by FragPipe⁵⁵ (version 17.1 for C10 and SVEC cells, version 18.0 for G2/M arrested C10 cells, and version 20.0 for human islet cells) and searched against the *Mus musculus* or *Homo sapiens* UniProt protein sequence database with decoy sequences (UP000000589 containing 17,201 forward entries, accessed 12/21 and UP000005640 containing 20,596 forward entries, accessed 05/23). Search settings included a precursor mass tolerance of ± 20 ppm, fragment mass tolerance of ± 0.5 Da, deisotoping, strict trypsin as the enzyme (with allowance for N-terminal semi-tryptic peptides for the human pancreatic islet cell experiment), carbamidomethylation as a fixed modification, and several variable modifications,

including oxidation of methionine, N-terminal acetylation, S/Y/T phosphorylation (for CDK1-inhibition experiment), and N-terminal pyro-glutamate formation at Q/E (for human pancreatic islet cell experiment). Protein and peptide identifications were filtered to a false discovery rate of less than 0.01 within FragPipe. For the TIFF methodology, IonQuant MBR and MaxLFQ were set to “TRUE” and library MS datasets were assigned as such during the data import step. An MBR FDR of 0.05 at ion level was used to reduce false matching and normalization of samples was not applied. FragPipe result files were then imported into RStudio (Build 461) for downstream analysis in the R environment (version 4.1.3). For the pooled C10 samples, median normalization was performed with the “median_normalization” function from the proDA R package separately for each level of input (11, 3, and 1 cells) in order to avoid distorting the expected feature intensities. With all other experiments, median normalization was performed across all single-cells. With regards to quality control filtering of scProteomics and scRNAseq data, cells with total identifications of more than 2 standard deviations reduced from the average within their respective modality were removed from analysis. For scRNAseq data, a minimum of 5 read counts were used to filter genes with low abundance for all experiments with C10 and SVEC cells. In cases where complete data was needed, imputation was performed with k-nearest neighbors imputation ($n=10$ neighbors) with < 60% missing values.

For the pancreatic islet experiment, batch correction was performed with the ComBat.NA function from the MSnSet.utils R package. Sparse partial least squares discriminant analysis (spl-DA) was utilized based on previously described algorithm⁵⁶. From the scProteomic data, common contaminants were removed, and cells were filtered to have at least 3000 peptide identifications. Three cell types were used to create different classes: alpha, beta, and delta (the remaining cells were considered as an alternative class). 10-fold cross-validation was used to evaluate the model’s predictive ability and balanced accuracy was used to assess sensitivity/specificity. GO analysis was performed with the gprofiler2 R package and web application. All raw p -values determined through statistical testing were corrected for multiple hypothesis testing (where relevant) using the Benjamini-Hochberg procedure. All experiments were assessed with standardized quality control metrics (e.g., sensitivity and data completeness) as recommended by Vanderaa and Gatto (Supplementary Fig. 20)⁵⁷.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The mass spectrometry raw data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with dataset identifiers [MSV000089280](https://doi.org/10.26434/chemrxiv-2024-00000), [MSV000090828](https://doi.org/10.26434/chemrxiv-2024-00000), and [MSV000093330](https://doi.org/10.26434/chemrxiv-2024-00000). The raw RNAseq data has been deposited to the Gene Expression Omnibus (GEO) under the identifiers [GSE201575](https://doi.org/10.1101/2024.01.15.575755), [GSE219047](https://doi.org/10.1101/2024.01.15.575755), and [GSE247519](https://doi.org/10.1101/2024.01.15.575755). All scripts, result tables, single-cell images, and metadata used for figure generation are available at the nanoSPLITS GitHub repository [<https://github.com/Cajun-data/nanoSPLITS>]⁵⁴. Source data are provided with this paper.

Code availability

Associated code for the data analysis is included in R markdown or R script files at the nanoSPLITS GitHub repository [<https://github.com/Cajun-data/nanoSPLITS>]⁵⁴.

References

- Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

- Macosko, E. Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
- Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
- Rust, K. et al. A single-cell atlas and lineage analysis of the adult *Drosophila* ovary. *Nat. Commun.* **11**, 5628 (2020).
- Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018).
- Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
- Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
- Chung, H. et al. Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nat. Methods* **18**, 1204–1212 (2021).
- Peterson, V. M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
- Frei, A. P. et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016).
- Darmanis, S. et al. Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells. *Cell Rep.* **14**, 380–389 (2016).
- Albayrak, C. et al. Digital Quantification of Proteins and mRNA in Single Mammalian Cells. *Mol. Cell* **61**, 914–924 (2016).
- Lin, J. & Tay, S. Ultra-Sensitive Quantification of Protein and mRNA in Single Mammalian Cells with Digital PLA. *Methods Mol. Biol.* **2386**, 157–169 (2022).
- Genshaft, A. S. et al. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* **17**, 188 (2016).
- Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274–276 (2015).
- Marcon, E. et al. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat. Methods* **12**, 725–731 (2015).
- Zhu, Y. et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew. Chem. Int. Ed. Engl.* **57**, 12370–12374 (2018).
- Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
- Woo, J. et al. High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nat. Commun.* **12**, 6246 (2021).
- Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chem. Sci.* **12**, 1001–1006 (2020).
- Brunner, A. D. et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **18**, e10798 (2022).
- Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **9**, 882 (2018).
- Woo, J. et al. Three-dimensional feature matching improves coverage for single-cell proteomics based on ion mobility filtering. *Cell Syst.* **13**, 426–434.e424 (2022).
- Ziegenhain, C. et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol. Cell* **65**, 631–643.e634 (2017).
- Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
- Rosenberger, F. A., Thielert, M. & Mann, M. Making single-cell proteomics biologically relevant. *Nat. Methods* **20**, 320–323 (2023).
- Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).

28. Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31 (2016).
29. Vassilev, L. T. et al. Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. *Proc. Natl Acad. Sci. USA* **103**, 10660–10665 (2006).
30. Orsburn, B. C., Yuan, Y. & Bumpus, N. N. Insights into protein post-translational modification landscapes of individual human cells by trapped ion mobility time-of-flight mass spectrometry. *Nat. Commun.* **13**, 7246 (2022).
31. Woodard, G. E. et al. Ric-8A and Gi alpha recruit LGN, NuMA, and dynein to the cell cortex to help orient the mitotic spindle. *Mol. Cell Biol.* **30**, 3519–3530 (2010).
32. Wang, S., Tukachinsky, H., Romano, F. B. & Rapoport, T. A. Cooperation of the ER-shaping proteins atlastin, lunapark, and reticulons to generate a tubular membrane network. *Elife* **5** <https://doi.org/10.7554/eLife.18605> (2016).
33. Tanenbaum, M. E., Stern-Ginossar, N., Weissman, J. S. & Vale, R. D. Regulation of mRNA translation during mitosis. *eLife* **4**, e07957 (2015).
34. Yamaguchi, T. et al. Phosphorylation by Cdk1 induces Plk1-mediated vimentin phosphorylation during mitosis. *J. Cell Biol.* **171**, 431–436 (2005).
35. Chou, Y. H., Bischoff, J. R., Beach, D. & Goldman, R. D. Intermediate filament reorganization during mitosis is mediated by p34cdc2 phosphorylation of vimentin. *Cell* **62**, 1063–1071 (1990).
36. Segerstolpe, A. et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab.* **24**, 593–607 (2016).
37. van Gurp, L. et al. Generation of human islet cell type-specific identity genesets. *Nat. Commun.* **13**, 2020 (2022).
38. Muraro, M. J. et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* **3**, 385–394 e383 (2016).
39. Augsornworawat, P., Maxwell, K. G., Velazco-Cruz, L. & Millman, J. R. Single-Cell Transcriptome Profiling Reveals beta Cell Maturation in Stem Cell-Derived Islets after Transplantation. *Cell Rep.* **32**, 108067 (2020).
40. Taneera, J. et al. GNAS gene is an important regulator of insulin secretory capacity in pancreatic beta-cells. *Gene* **715**, 144028 (2019).
41. Spears, E. et al. Pancreatic islet α cell function and proliferation requires the arginine transporter SLC7A2. *bioRxiv*, 2023.2008.2010.552656 (2023).
42. Viloria, K. et al. GC-Globulin/Vitamin D-Binding Protein Is Required for Pancreatic alpha-Cell Adaptation to Metabolic Stress. *Diabetes* **72**, 275–289 (2023).
43. van der Meulen, T. et al. Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. *Nat. Med.* **21**, 769–776 (2015).
44. Zito, E., Chin, K. T., Blais, J., Harding, H. P. & Ron, D. ERO1-beta, a pancreas-specific disulfide oxidase, promotes insulin biogenesis and glucose homeostasis. *J. Cell Biol.* **188**, 821–832 (2010).
45. Price, B., Dennison, C., Tschesche, H. & Elliott, E. Neutrophil tissue inhibitor of matrix metalloproteinases-1 occurs in novel vesicles that do not fuse with the phagosome. *J. Biol. Chem.* **275**, 28308–28315 (2000).
46. Taniguchi, Y. et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
47. Sanchez, A. & Golding, I. Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**, 1188–1193 (2013).
48. Tsai, C. F. et al. A streamlined tandem tip-based workflow for sensitive nanoscale phosphoproteomics. *Commun. Biol.* **6**, 70 (2023).
49. Douglas, P. et al. Phosphorylation of SAF-A/hnRNP-U Serine 59 by Polo-Like Kinase 1 Is Required for Mitosis. *Mol. Cell Biol.* **35**, 2699–2713 (2015).
50. Song, D., Li, K., Ge, X. & Li, J.J. ClusterDE: a post-clustering differential expression (DE) method robust to false-positive inflation caused by double dipping. *Res. Sq.* <https://doi.org/10.21203/rs.3.rs-3211191/v1> (2023).
51. Lindeboom, R. G. H., Regev, A. & Teichmann, S. A. Towards a Human Cell Atlas: Taking Notes from the Past. *Trends Genet.* **37**, 625–630 (2021).
52. Snyder, M. P. et al. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
53. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
54. Fulcher, J. M. Parallel measurement of transcriptomes and proteomes from same single cells using nanodroplet splitting. *GitHub/Cajun-data/nanoSPLITS* (2024). <https://doi.org/10.5281/zenodo.13922100>
55. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 (2017).
56. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Comput Biol.* **13**, e1005752 (2017).
57. Vanderaa, C. & Gatto, L. Revisiting the Thorny Issue of Missing Values in Single-Cell Proteomics. *J. Proteome Res.* **22**, 2775–2784 (2023).

Acknowledgements

We thank Drs. Clayton Mathews and Jing Chen at the University of Florida for providing a detailed protocol for culturing human islets, and Dr. Matthew Monroe for his assistance in depositing the raw proteomic data into MassIVE, as well as Nathan Johnson for the illustration utilized in Fig. 1. This work was supported by a Laboratory Directed Research and Development award (I3T) from Pacific Northwest National Laboratory (Y.Z.). A portion of this research was performed on a National Institutes of Health (NIH) Common Fund, Human Biomolecular Atlas Program (HuB-MAP) grant UH3CA256959 (L.P.-T.). A portion of this research was also performed on a project award (10.46936/prtn.proj.2021.51834/60000340) from the Environmental Molecular Sciences Laboratory (J.C.B. and L.P.-T.), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RLO1830.

Author contributions

Y.Z., J.M.F., L.P.-T., and L.M.M. conceptualized and designed the research. J.M.F. and S.M.W. fabricated nanoSPLITS microchips. K.M.E., L.M.M., J.M.F., J.C.B., W.B.C. and Y.Z. performed the cell culture, cell sorting, and proteomic sample preparation. J.M.F., D.J.D., and L.M.B. performed data analysis. S.M.W., J.M.F., and R.J.M. performed the LC-MS analysis. L.M.M. and H.D.M. performed the RNAseq and associated read mapping. J.M.F., Y.Z., and L.M.M. wrote the manuscript. J.M.F., Y.Z., L.P.-T., J.C.B., A.S., W.J.Q., J.W.B., and L.M.M. edited the manuscript. The final manuscript was read and approved by all authors.

Competing interests

J.C.B., J.W.B. and A.S. are employees of Scienion/Cellenion. Y.Z. is an employee of Genentech Inc. and shareholder of Roche Group. Battelle Memorial Institute has submitted a U.S. patent application for the design of nanoSPLITS devices and the associated operation methods (Application number: 17/954,834, filed 09/28/2022; Inventors: Y.Z., J.M.F.,

L.M.M., and L.P.T.; Status of application: Pending). Other authors declare no other competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54099-z>.

Correspondence and requests for materials should be addressed to James M. Fulcher or Ying Zhu.

Peer review information *Nature Communications* thanks Benjamin Orsburn, Amanda Smythers, and Ákos Végvári for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© Battelle Memorial Institute, Scienion US Inc. and Cellenion SASU 2024