bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

James M. Fulcher,[1] Lye Meng Markillie,[1] Hugh D. Mitchell, [1] Sarah M. Williams,[1] Kristin M. Engbrecht,[2] Ronald J. Moore,[3] Joshua Cantlon-Bruce,[4,5] Johannes W. Bagnoli,[5] Anjali Seth,[5] Ljiljana Paša-Tolić,[1] Ying Zhu[1, *]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99354, United States

[2]Chemical and Biological Signature Sciences, Pacific Northwest National Laboratory, Richland, WA 99352, United States

[3]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99354, United States

[4]Scienion AG, Volmerstraße 7, 12489 Berlin, Germany

[5]Cellenion SASU, 60 Avenue Rockefeller, Bâtiment BioSerra2, 69008 Lyon, France

*Correspondence**:** Dr. Ying Zhu, ying.zhu@pnnl.gov

ABSTRACT

Single-cell multiomics can provide comprehensive insights into gene regulatory networks, cellular diversity, and temporal dynamics. While tools for co-profiling the single-cell genome, transcriptome, and epigenome are available, accessing the proteome in parallel is more challenging. To overcome this limitation, we developed nanoSPLITS (nanodroplet SPlitting for Linked-multimodal Investigations of Trace Samples), a platform that enables unbiased measurement of the transcriptome and proteome from same single cells using RNA sequencing and mass spectrometry-based proteomics, respectively. We demonstrated the nanoSPLITS can robustly profile > 5000 genes and > 2000 proteins per single cell, and identify cell-type-specific markers from both modalities.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MAIN

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 lineages,[4] and rare cell types.[5] 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 genome, epigenome, transcriptome, proteome, and metabolome. Indeed, proteins are of particular interest in establishing cellular identities because they are the downstream effectors and their abundance cannot be easily inferred from other modalities, including mRNA[6]. While the simultaneous acquisition of multiple modalities such as transcriptome-genome[7] and transcriptome-epigenome[8] have demonstrated great depth and sensitivity, multimodal transcriptome-proteome[9-12] measurements are restricted to at most a few hundred protein targets. These measurements also require intermediate antibodies to recognize epitopes, which can be limited by availability and specificity[13].

A route for overcoming these limitations is through the adoption of a mass spectrometry-based proteomics approach. With the advance of microfluidic sample preparation[14] and isobaric labeling[15], single-cell proteomics (scProteomics) is now capable of measuring thousands of proteins from single cells in an unbiased manner.[16] Encouraged by these developments, we sought to acquire multimodal transcriptome-proteome measurements from the same single cell by integrating single-cell RNA sequencing (scRNAseq) with scProteomics. To enable efficient integration, we developed nanoSPLITS (nanodroplet SPlitting for Linked-multimodal Investigations of Trace Samples), a method capable of equally dividing nanoliter-scale cell lysates via two droplet microarrays and separately measuring them with RNA sequencing and mass spectrometry. NanoSPLITS builds on the nanoPOTS platform that allows for high-

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

efficiency proteomic preparation of single cells by miniaturizing the assay volumes to nanoliter scale volumes[16,17]. We have previously demonstrated reaction miniaturization not only reduces non-specific adsorption-related sample losses, but also enhances enzymatic digestion kinetics.[18] Similarly, we reason the use of nanoliter droplets can improve overall sample recovery of both mRNA transcripts and proteins for sensitive single-cell multiomics.

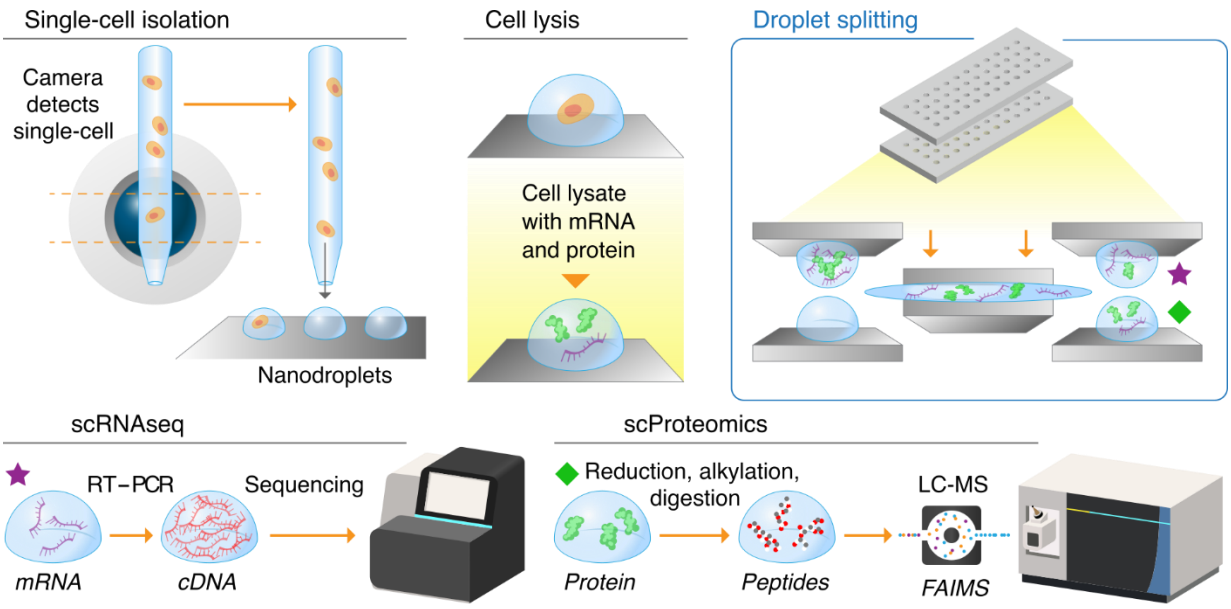

Fig.1: Overview of the nanoSPLITS-based single-cell multiomics platform. Schematic illustration showing the workflow including cell sorting, lysis, droplet merging/mixing, and droplet separation for downstream scRNAseq and scProteomics measurement.

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 into our optimized lysis buffer, followed by a freeze-thaw cycle to achieve cell lysis. Next, the microchip containing single-cell lysate was manually aligned with a separate chip containing only cell lysis buffer. The droplet arrays in the two chips were merged and separated for three rounds to achieve complete mixing (**Supplementary Movie 1**). One chip containing approximately half of the cell lysate can then be transferred into 384-well plate for scRNAseq based on Smart-seq 2[1]. For scProteomics, the remaining ~50% lysate is digested with a DDM-

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

based sample preparation protocol and directly analyzed with an ion-mobility-based MS data acquisition method[19]. Notably, when the same droplet volume (200 nL) was used in an evaluation experiment with a model fluorescent dye, the nanoSPLITS procedure can achieve splitting ratios between 46% to 47%, with 50% representing an equal split (**Supplementary Fig. S1** and **Supplementary Table S1**).

We first determined the optimal cell lysis buffer that is compatible with both scProteomics and scRNAseq workflows. Typically, scProteomics utilizes a buffer containing 0.1% n-dodecyl-β-D-maltoside (DDM) to reduce non-specific binding of proteins to surfaces[14], while scRNAseq includes recombinant protein-based RNase inhibitors to reduce mRNA degradation. To evaluate their impacts on both methods, we tested these additives in a moderately buffered hypotonic solution (10 mM Tris, pH 8) with 20 mouse alveolar epithelial cells (C10) (**Supplementary Fig. S2**). In short, we found the inclusion of 1 x RNase inhibitor suppressed proteomic identifications while 0.1% DDM had no significant impact on transcriptomic identifications. Furthermore, the removal of RNase inhibitor from RNAseq analysis had minimal effect on transcriptomic identifications. Therefore, we decided on a 10 mM Tris solution with 0.1% DDM as the cell lysis buffer for nanoSPLITS.

To evaluate the nanoSPLITS method, we sorted several quantities (11, 3, and 1) of C10 cells and measured them using the multiomics workflow (**Fig. 2**). Considering a 5 read minimum per gene for transcriptome identification and 1% FDR cutoff for protein identification, robust

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coverage of both genes and proteins could be achieved across all tested conditions (**Fig. 2a**).

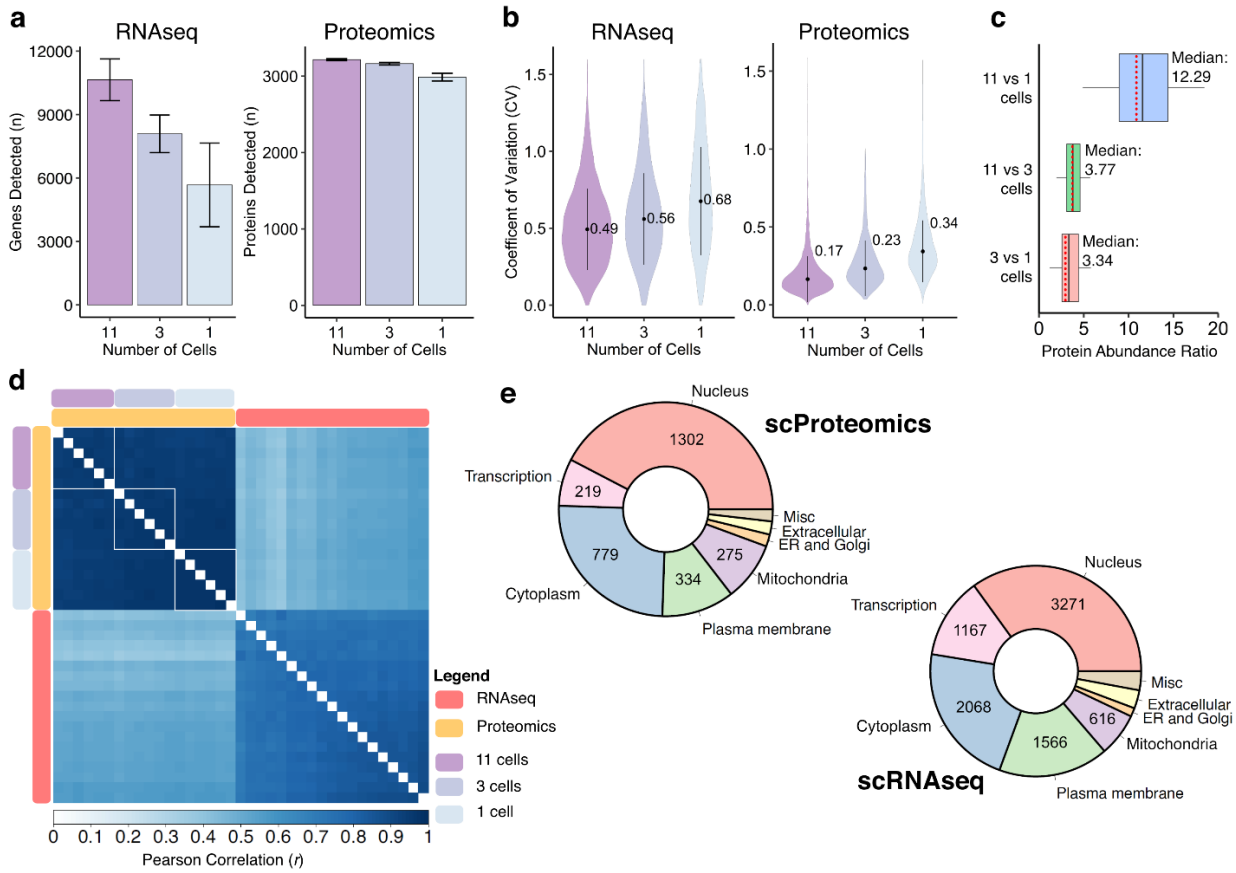

Fig. 2: Quantitative and qualitative assessment of transcriptome and proteome measurements after nanoSPLITS.

(**a**) Average numbers of detected genes and proteins. Error bars indicate standard deviations (±s.d.). (**b**) Distributions of the coefficients of variation (CV) for all proteins and genes with at least 2 observations. Indicated values represent median CV, which is also indicated at the center point within each distribution. (**c**) The ratios of protein abundance were calculated for comparisons between the different pooled cell samples (11 vs 1, 11 vs 3, and 3 vs 1). Experimental median is indicated at the black crossbar while the theoretical ratio for each comparison is shown at the red dotted line within each boxplot. (**d**) Pearson correlation heatmap with clustering of transcriptomics and proteomics results. (**e**) Cellular component gene ontologies were determined for each gene (scRNAseq) and protein (scProteomics) found in the single -cell data.

As expected, coverage was reduced with the decreasing cell numbers. Single-cell transcriptome and proteome measurements provided 5,848 and 2,934 identifications on average, respectively. We next evaluated the quantitative reproducibility for each modality by calculating the coefficients of variations (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. 2b**). The modestly higher

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

CVs for single cells were expected, as the mixed cell populations represent averages of the underlying biological variations. Notably, we observed significantly higher CVs for the transcriptome compared to proteome, in agreement with recent reports[16,20]. Presumably, these higher CVs reflect the dynamic nature of mRNA relative to their protein counterparts, which have longer half-lives on average[21]. We next compared the ratios of the measured protein abundances between the different cell populations. Encouragingly, the experimental fold differences between the median intensities for 11, 3, and 1 C10 cell are very close to the expected theoretical values (**Fig. 2c**). For example, the median protein abundance ratio for 3 cells compared to single cells was 3.34, within 12% of the theoretical 3-fold difference. Taken together, these results provide strong evidence that nanoSPLITS-based single-cell multiomics platform can provide sensitive and reproducible measurement of both the transcriptome and proteome of the same single cells.

We next determined the Pearson correlation coefficients (*r*) across and within modalities using conceptually-similar normalized transformations for each modality (**Fig. 2d**; TPM, transcripts per million for transcriptomics, and riBAQ, relative intensity-based absolute quantification for proteomics). In line with the CV distributions (**Fig. 2b**), proteomics data had a better agreement between samples compared with transcriptomics data, once again highlighting the dynamic nature of transcriptome where many genes are often expressed in short transcriptional “bursts” [21]. To ensure nanoSPLITS did not introduce a bias toward different cellular components due to the nanodroplet splitting process, we also investigated the distribution of gene and protein identifications in single cells across several gene ontologies (GO).We found scProteomics and scRNAseq had corresponding identifications within cellular components that encompassed all major organelles (**Fig. 2e**). Furthermore, 1,521 proteins from the scProteomics analyses have GO localizations to the nucleus, 219 of which of have known roles in transcription. This is notable as nuclear proteins are typically drivers in gene regulation

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

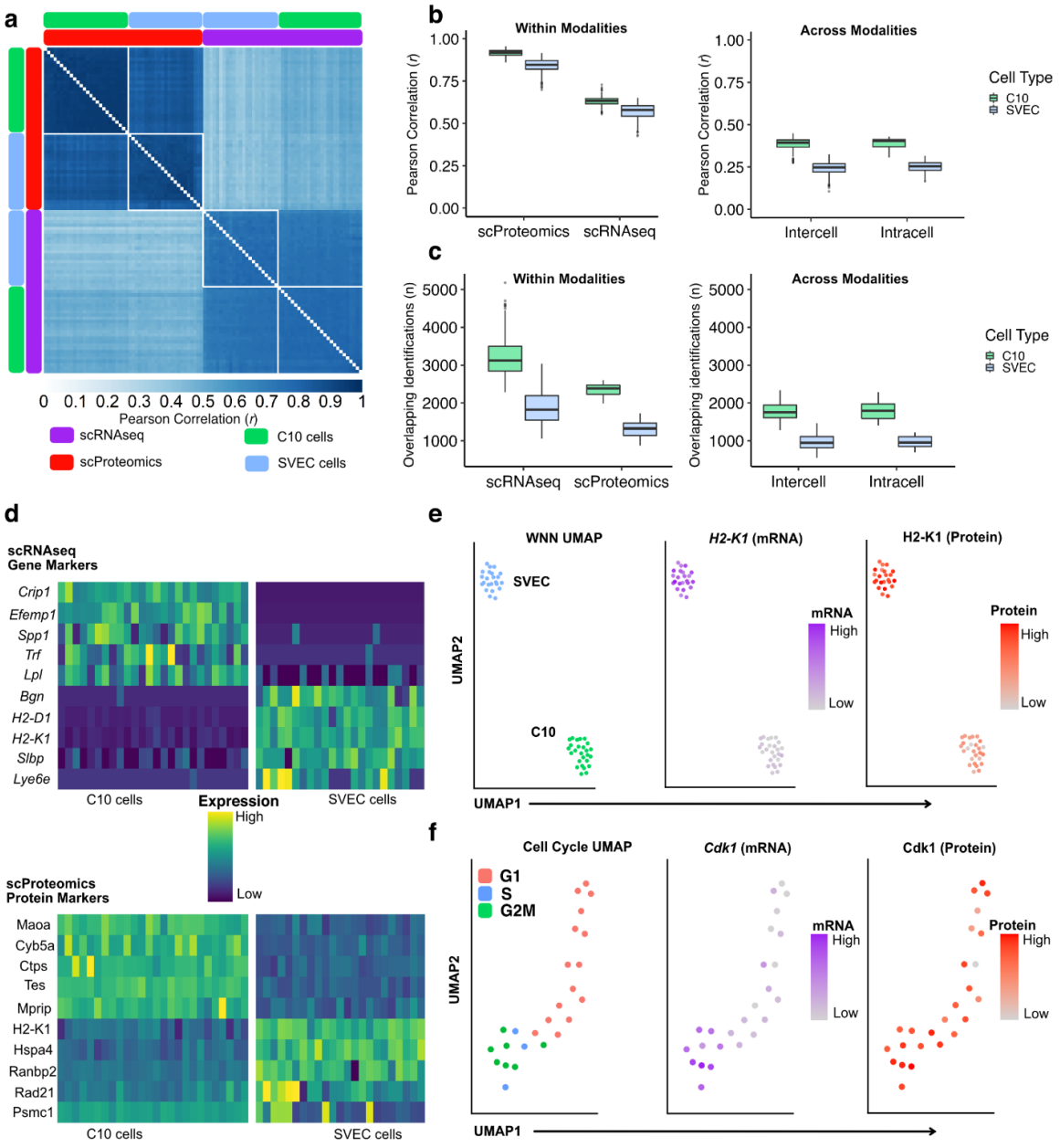

Fig.3: Underlying cell phenotype signatures are maintained after nanoSPLITS. (a). Pearson correlation heatmap with clustering of transcriptomics and proteomics results for both single C10 and SVEC cells. (**b**) Distributions of Pearson correlations, separated by cell type and modality (scProteomics and scRNAseq). (**c**) The overlap in gene and protein identifications were determined for each modality separately, as well as across the modalities. (**d**) Top 5 gene markers from scRNAseq data and protein markers from scProteomics data were determined for each cell type. Candidate marker features were determined using a Wilcoxon Rank Sum test (FDR corrected *p*-values <0.001). (**e**) Weighted-nearest neighbor (WNN) UMAP generated using Seurat in order to integrate the scRNAseq and scProteomic data. Middle and right panels are colored based on *H2-K1* gene (purple) and protein (red) expression, respectively. (**f**) UMAP generated for C10 cells based on cell-cycle features measured in the scRNAseq data. Middle and right panels are colored based on *Cdk1* gene (purple) and protein (red) expression, respectively. All expression values shown in d, e, and f are derived from *Z*-scores after scaling and centering of data.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and transcription, and current multimodal technologies have been limited in the ability to directly measure nuclear protein abundances.

Having established baseline characteristics of multimodal data, we then applied nanoSPLITS to a larger single-cell multimodal analysis encompassing two cell types, mouse epithelial (C10) and endothelial cells (SVEC). Because the nanoSPLITS approach uses only half the mRNA or protein contents, we sought to determine whether the multimodal measurements could precisely distinguish the two cell types and detect gene or protein markers. As shown in **Fig. 3a**, both cell types and modalities could easily be clustered based on correlations alone. In line with our pilot experiment, within-modality correlations were higher in scProteomics than scRNAseq for both cell types (**Fig. 3b**). Cross-modality correlation analysis between scRNAseq and scProteomics produced *r* ranging from 0.31 to as high as 0.56, which fell in the range of previously reported mRNA-protein correlations[21]. We also compared the cross-modality correlations between the same single cells (intracell) and the correlations between different single cells (intercell). As shown in Fig. 3b, no significant difference was observed. This is not entirely unsurprising, considering most of the variation between different cells can be attributed to only a small number of genes driving cell cycle progression. These low numbers of genes would not have a significant impact on global correlations. Overall, SVEC cells had slightly lower correlations across the board, presumably due to their smaller cell size and corresponding reduced measurement depth and precision (**Supplementary Fig. S3**). The protein/ gene overlap analysis demonstrates how measurement depth is strongly linked to cell size (**Fig. 3c**). On average, C10 cells had ~1,800 overlapping identifications while SVEC cells had ~900 overlapping identifications across modalities. Next, we evaluated if the multiomics data could be used to identify cell-type-specific marker genes and proteins. Fig. 3d shows the top-5 significant enriched genes and proteins for each cell type. Interestingly, the overlap of

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

these significant markers was relatively low. Despite this, the previously established SVEC-cell marker *H2-K1* was identified here at both the protein and mRNA level (**Fig. 3d**)[22].

Dimensionality reduction with principal component analysis (PCA) showed delineation of both cell types for scRNAseq and scProteomics despite only having half the cell contents (**Supplementary Fig. S4**). The integration of both modalities through an unsupervised weighted nearest neighbor (WNN)[23] analysis provided robust clustering in the two-dimensional space (**Fig. 3e**). This also provided us the ability to visualize both protein and mRNA abundances, confirming *H2-K1* to be a marker that is differentially expressed at the protein and gene level (**Fig. 3c**). Using canonical cell cycle markers[24] we could also identify sub-populations constituting specific cell cycle phases, demonstrating that even subtle cell to cell variation was retained after the droplet splitting process (**Fig. 3d**) . For example, the well-established marker cyclin-dependent kinase 1 (*Cdk1*) is upregulated at the transcriptional and translational level in S and G2M phase C10 cells (**Fig. 3f**, **Supplementary Fig. S5**, and **Supplementary Fig. S6**).

Taken together, we demonstrate how the nanoSPLITS approach can enable multimodal profiling of thousands of mRNA transcripts and proteins from the same single cells. The multiomics data allowed us to precisely quantify the abundances of both mRNA transcripts and proteins and identify marker genes and proteins from both modalities. Compared with previous technologies that utilize antibodies to infer protein abundances, the nanoSPLITS platform employs mass spectrometry to unbiasedly detect proteins, which is highly valuable for uncovering rare cell populations that lack reliable protein markers. We expect nanoSPLITS could become a powerful discovery tool for biomedical applications, such as characterizing tissue heterogeneity and circulating tumor cells. Notably, nanoSPLITS is not restricted to the two modalities (transcriptomics and proteomics); other modalities such as metabolomics, genomics, and epigenomics can conceptually be integrated into the workflow. As more analytical frameworks for integrating multimodal data are created, we anticipate nanoSPLITS

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

will enable greater insight into how different modalities interact with each other to control single-cell phenotypes in various contexts such as perturbations, mitosis/meiosis, and differentiation.

Although a low throughput approach was employed in this study, high-throughput multiplexing approaches such as CEL-Seq[25] for transcriptomics and SCoPE-MS[15] for proteomics can be readily integrated into the nanoSPLITS workflow. The integration of multiplexing approaches to nanoSPLITS would enable analysis of thousands of single cells with reasonable instrument time and overall cost[26]. Finally, recent advances in multimodal single-cell data analysis have enabled new avenues for harmonization across modalities by means of multi-omic datasets as molecular bridges[27]. The generation of proteome and transcriptome bridge datasets could readily be accomplished using nanoSPLITS, opening the proteome to reference mapping.

METHODS

Reagents and chemicals

Deionized water (18.2 MΩ) was purified using a Barnstead Nanopure Infinity system (Los Angeles, CA, USA). n-dodecyl-β-D-maltoside (DDM), iodoacetamide (IAA), ammonium bicarbonate (ABC), and formic acid (FA) were obtained from Sigma (St. Louis, MO, USA). Nuclease-free water (not DEPC-treated) ,Trypsin (Promega, Madison, WI, USA) and Lys-C (Wako, Japan) were dissolved 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 previously[18,28]. Two different chips were designed and used in this

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

study. Both contained 48 (4 x12) 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 x 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 (NAL1A clone C1C10 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). The cultured cell lines were collected in a 15 ml tube and centrifuged at 1,000 × g for 3 min to remove the medium. Cell pellets were

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

washed three times by PBS, then counted to obtain cell concentration. PBS was then added to achieve a concentration of 200 x 10^6 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) in order to remove aggregated cells.

CellenONE cell sorting

Before cell sorting, nanoSPLITS chips were prepared by the addition of 200-nL hypotonic solution consisting of 0.1% DDM in 10 mM Tris 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 µs, a nozzle voltage of 80 V, a frequency of 500 Hz, a LED delay of 200 µs, and a LED pulse of 3 µs. 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 µm in diameter, maximum circularity of 1.15, and maximum elongation of 2, while SVEC cells were 24 to 32 µm in diameter, maximum circularity of 1.15, and maximum elongation of 2. All cells were sorted based on brightfield 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 scProteomics[19], 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 -80ºC, which partially served to induce cell lysis via freeze-thaw. All associated

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

settings, single-cell images, and metadata can be accessed at the GitHub repository provided (https://github.com/Cajun-data/nanoSPLITS).

NanoSPLITS process

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 on each nanowell. The bottom chip containing the cell lysate was placed on an aluminum chip holder that was pre-cooled to 4ºC 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 in order to merge the droplets for each chip. Pressure was held for 15 seconds before releasing. The droplets were merged twice more following this process. For consistency, the top chip which received 50% of the lysate was used for scRNAseq in all experiments while the bottom chip that initially contained the cell lysate was utilized in scProteomics. 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 1µL of RT-PCR buffer to each nanowell before withdrawing the entire volume and adding it to a 96-well plate. For the C10 and SVEC comparison experiment, 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 3,500 x g for 1 minute.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

150 nL of extraction buffer containing 50 mM ABC, 0.1% DDM, 0.3X diluted PBS, and 2 mM DTT, and incubating the chip at 60ºC for 60 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 37ºC 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 -20ºC 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 µm i.d., 4 cm, 5 µm particle size, 300 Å pore size C18 material, Phenomenex) and analytical LC column (50 µm i.d., 25 cm long, 1.7 µm particle size, 190 Å 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 (Thermo Scientific) with FAIMSpro, operated in data-dependent acquisition mode, was used for all analyses. Source settings included a spray voltage of 2,400 V, ion transfer tube temperature of 200ºC, and carrier gas flow of 4.6 L/min. For the TIFF test samples[19], ionized peptides were fractionated by the FAIMS interface using internal CV stepping (-45, -60, and -75 V) with a total cycle time of 0.8 s per CV. 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 from 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 the TIFF library

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

samples, a single CV was used for each LC-MS run 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.

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, TaKaRa, cat# R400753); the samples were immediately denatured at 72ºC 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 42ºC for 90 min; followed by heat inactivation at 70ºC 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 (TaKaRa, 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, Illumina, 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) using STAR (https://github.com/alexdobin/STAR). BAM file outputs were mapped to genes using htseq-count[29] with default settings. TPM counts were derived using an R script based on TPM procedure[30].

Database searching and data analysis

All proteomic data raw files were processed by FragPipe[31] version 17.1 and searched against the *Mus musculus* UniProt protein sequence database with decoy sequences (Proteome ID: UP000000589 containing 17,201 forward entries, accessed 12/02/21). Search settings included

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a precursor mass tolerance of +/- 20 ppm, fragment mass tolerance of +/- 0.5 Da, deisotoping, strict trypsin as the enzyme, carbamidomethylation as a fixed modification, and several variable modifications, including oxidation of methione, and N-terminal acetylation. Protein and peptide identifications were filtered to a false discovery rate of less than 0.01 within FragPipe. For the TIFF method, IonQuant[32] match-between-runs (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. FragPipe result files were then imported into RStudio (Build 461) for downstream analysis in the R environment (version 4.1.3). All of the figures generated and associated code are included in R markdown files at the nanoSPLITS GitHub repository (https://github.com/Cajun-data/nanoSPLITS).

DATA AVAILABILITY

The mass spectrometry raw data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with dataset identifier MSV000089280 (FTP Password: Nano4108). The raw RNA-seq data has been deposited to the Gene Expression Omnibus (GEO) under the identifier GSE201575.

CODE AVAILABILITY

All scripts, tables, single-cell images, and metadata used for figure generation are available at the nanoSPLITS GitHub repository (https://github.com/Cajun-data/nanoSPLITS).

ACKNOWLEDGMENTS

We thank Matthew Monroe for his assistance in depositing the raw proteomic data into MassIVE, as well as Nathan Johnson for the illustration utilized in Figure 1. This work was supported by a Laboratory Directed Research and Development award (I3T) from Pacific

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Northwest National Laboratory (L.M.M.) and the NIH Common Fund under award UG3CA256959 (L.P.T.). A portion of this research was performed on a project award (10.46936/prtn.proj.2021.51834/60000340) from the Environmental Molecular Sciences Laboratory (C.B.J.), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830.

AUTHOR CONTRIBUTIONS

Y.Z., L.P-T., L.M.M., and J.M.F. conceptualized and designed the research. J.M.F. and S.M.W. fabricated nanoSPLITS microchips. K.M.E., J.M.F., J.C.B., and Y.Z. performed the cell culture, cell sorting, and proteomic sample preparation. S.M.W., J.M.F., and R.J.M. performed 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., J.W.B., and S.M.W. 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. Battelle Memorial Institute has submitted a U.S. provisional patent application for the design of nanoSPLITS devices and the associated operation methods (Application number: 63/250,011 filed 09/29/2021; Inventors: Ying Zhu, James M Fulcher, Lye Meng Markillie, and Ljiljana Paša-Tolić; Status of application: Pending). Other authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

REFERENCES

1 Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols* **9**, 171-181, doi:10.1038/nprot.2014.006 (2014).

2 Macosko, Evan Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).

3 Klein, Allon M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187-1201, doi:https://doi.org/10.1016/j.cell.2015.04.044 (2015).

4 Rust, K. *et al.* A single-cell atlas and lineage analysis of the adult Drosophila ovary. *Nature Communications* **11**, 5628, doi:10.1038/s41467-020-19361-0 (2020).

5 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324, doi:10.1038/s41586-018-0393-7 (2018).

6 Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535-550, doi:10.1016/j.cell.2016.03.014 (2016).

7 Macaulay, I. C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nature Methods* **12**, 519-522, doi:10.1038/nmeth.3370 (2015).

8 Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380-1385, doi:doi:10.1126/science.aau0730 (2018).

9 Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods* **14**, 865-868, doi:10.1038/nmeth.4380 (2017).

10 Chung, H. *et al.* Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nature Methods* **18**, 1204-1212, doi:10.1038/s41592-021-01278-1 (2021).

11 Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nature Biotechnology* **35**, 936-939, doi:10.1038/nbt.3973 (2017).

12 Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nature Methods* **13**, 269-275, doi:10.1038/nmeth.3742 (2016).

13 Baker, M. Reproducibility crisis: Blame it on the antibodies. *Nature* **521**, 274-276, doi:10.1038/521274a (2015).

14 Zhu, Y. *et al.* Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angewandte Chemie International Edition* **57**, 12370-12374, doi:https://doi.org/10.1002/anie.201802843 (2018).

15 Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biology* **19**, 161, doi:10.1186/s13059-018-1547-5 (2018).

16 Woo, J. *et al.* High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nature Communications* **12**, 6246, doi:10.1038/s41467-021-26514-2 (2021).

17 Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature Communications* **9**, 882, doi:10.1038/s41467-018-03367-w (2018).

18 Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat Commun* **9**, 882, doi:10.1038/s41467-018-03367-w (2018).

19 Woo, J. *et al.* Three-dimensional feature matching improves coverage for single-cell proteomics based on ion mobility filtering. *Cell Systems*, doi:https://doi.org/10.1016/j.cels.2022.02.003 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2022.05.17.492137; this version posted May 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

20 Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Molecular Systems Biology* **18**, e10798, doi:https://doi.org/10.15252/msb.202110798 (2022).

21 Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nature Reviews Genetics* **21**, 630-644, doi:10.1038/s41576-020-0258-4 (2020).

22 Connell, K. A. & Edidin, M. A mouse lymphoid endothelial cell line immortalized by simian virus 40 binds lymphocytes and retains functional characteristics of normal endothelial cells. *The Journal of Immunology* **144**, 521 (1990).

23 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529, doi:https://doi.org/10.1016/j.cell.2021.04.048 (2021).

24 Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20-e31, doi:10.1182/blood-2016-05-716480 (2016).

25 Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Reports* **2**, 666-673, doi:https://doi.org/10.1016/j.celrep.2012.08.003 (2012).

26 Dephoure, N. & Gygi, S. P. Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Sci Signal* **5**, rs2-rs2, doi:10.1126/scisignal.2002548 (2012).

27 Hao, Y. *et al.* Dictionary learning for integrative, multimodal, and scalable single-cell analysis. *bioRxiv*, 2022.2002.2024.481684, doi:10.1101/2022.02.24.481684 (2022).

28 Liu, W. W., Zhu, Y., Feng, Y. M., Fang, J. & Fang, Q. Droplet-Based Multivolume Digital Polymerase Chain Reaction by a Surface-Assisted Multifactor Fluid Segmentation Approach. *Anal Chem* **89**, 822-829, doi:10.1021/acs.analchem.6b03687 (2017).

29 Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

30 Zhao, Y. *et al.* TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *Journal of Translational Medicine* **19**, 269, doi:10.1186/s12967-021-02936-w (2021).

31 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. *Nature Methods* **14**, 513-520, doi:10.1038/nmeth.4256 (2017).

32 Yu, F., Haynes, S. E. & Nesvizhskii, A. I. IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs. *Molecular & Cellular Proteomics* **20**, doi:10.1016/j.mcpro.2021.100077 (2021).