Pacific Northwest
NATIONAL LABORATORY

PNNL-34730

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

August 2023

James M. Fulcher
Lye Meng Markillie
Hugh D. Mitchell
Sarah M. Williams
Kristin M. Engbrecht
Ronald J. Moore
William B. Chrisler
Ljiljana Paša-Tolić
Ying Zhu

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, **makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-34730

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

August 2023

James M. Fulcher
Lye Meng Markillie
Hugh D. Mitchell
Sarah M. Williams
Kristin M. Engbrecht
Ronald J. Moore
William B. Chrisler
Ljiljana Paša-Tolić
Ying Zhu

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

Single-cell multiomics provides comprehensive insights into gene regulatory networks, cellular diversity, and temporal dynamics. While tools for co-profiling single-cell genomes, transcriptomes, and epigenomes are available, accessing proteomes in parallel is more challenging. We developed 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 using RNA sequencing and mass spectrometry-based proteomics, respectively. nanoSPLITS can precisely quantify over 5000 genes, 2000 proteins, and 140 phosphopeptides per single cell and identify candidate cell markers from these modalities. By exploring Cdk1-mediated cell cycle arrest, we demonstrate how nanoSPLITS single-cell multiomics can provide comprehensive cellular characterization with insights into covarying protein/gene clusters, unique phosphorylation events, and mitotic pathways.

Acknowledgments

This research was supported by the Strategic Investments Initiative, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Contents

Figures

1.0 Introduction

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,(Picelli et al. 2014; Macosko et al. 2015; Klein et al. 2015) cell lineages,(Rust et al. 2020) and rare cell types.(Montoro et al. 2018) 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 mRNA(Liu, Beyer, and Aebersold 2016). Unfortunately, multimodal transcriptome-proteome(Stoeckius et al. 2017; Chung et al. 2021; Peterson et al. 2017; Frei et al. 2016; Darmanis et al. 2016) measurements are restricted to at most a few hundred (and often significantly fewer) protein targets. These measurements also require intermediate antibodies to recognize epitopes, which can be limited by availability and specificity(Baker 2015; Marcon et al. 2015).

A route for overcoming these limitations is through the adoption of a mass spectrometry-based approach. With the advance of microfluidic sample preparation(Zhu, Clair, et al. 2018) and isobaric labeling(Budnik et al. 2018), single-cell proteomics (scProteomics) is now capable of measuring thousands of proteins from single-cells globally.(Woo et al. 2021; Cong et al. 2021; Brunner et al. 2022) 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 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-efficiency proteomic preparation of single cells by miniaturizing the assay to nanoliter scale volumes(Zhu, Piehowski, et al. 2018; Woo et al. 2021). We have previously demonstrated reaction miniaturization not only reduces non-specific adsorption-related sample losses, but also enhances enzymatic digestion kinetics.(Zhu, Piehowski, et al. 2018) Similarly, we reason the use of nanoliter droplets can improve overall sample recovery of both mRNA transcripts and protein for sensitive single-cell multiomics.

2.0 Results

2.1 Optimization of nanoSPLITS provides precise and deep protein-mRNA coverage in single cells

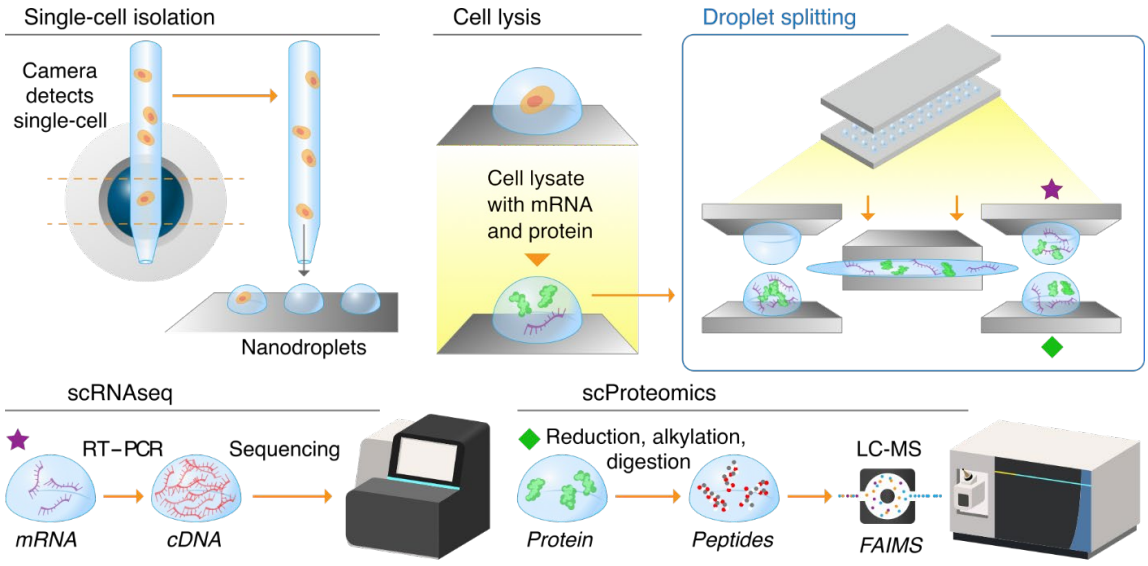

Figure 1. Overview of the nanoSPLITS-based single-cell multiomics platform.

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 lysis buffer, followed by a freeze-thaw cycle to achieve cell lysis. 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 seconds and then separated, before repeating twice more to mix completely. The chip receiving lysate (“acceptor” chip) was then transferred into a 384-well plate for scRNAseq with Smart-seq 2 protocol(Picelli et al. 2014), while the chip initially containing the lysate (“donor” chip) is digested with an n-dodecyl-β-D-maltoside (DDM)-based sample preparation protocol and directly analyzed with an ion-mobility-enhanced MS data acquisition method for scProteomics(Woo et al. 2022).

We first optimized cell lysis buffer conditions to ensure compatibility with 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 degradation of mRNA. 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). We found the inclusion of 1 x RNase inhibitor suppressed proteomic identifications while 0.1% DDM had no significant impact on transcriptomic identifications. Furthermore, removal of RNase inhibitor from RNAseq analysis had minimal effect on gene identifications. Therefore, we decided to use 0.1% DDM (w/v) in 10 mM Tris solution as the cell lysis buffer for nanoSPLITS workflow.

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. As the properties of fluorescein may not fully model proteins, we tested the protein splitting ratio by splitting 10 C10 cells (n = 6) and quantifying 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. Using the mean protein abundance for each protein identified from both chips, we determined the relative proportion of each protein found on either chip. Surprisingly, the median proportion of protein retained on the donor chip was greater than what would be expected in a perfect split (~75%). 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 (100 $\mu m^2 s^{-1}$) 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. The more convincing explanation might be that the chlorotrimethylsilane coating (to prevent hydrogen-bonding between mRNA molecules and the glass surface) of the nanowells captured

proteins on the surface, resulting in more proteins retained on the donor chip. Thus, we took advantage of the phenomena and chose the donor chip for scProteomic measurements to obtain deeper proteome coverage (Fig. 1).

As a proof-of-concept experiment to evaluate the nanoSPLITS, we next sorted several quantities (11, 3, and 1) of C10 cells and measured them using the multiomics workflow. Considering a minimum of 5 reads per gene for transcript identifications and 5% FDR cutoff for protein identifications, robust coverage of both, genes and proteins, could be achieved across all tested conditions (Fig. 2a).

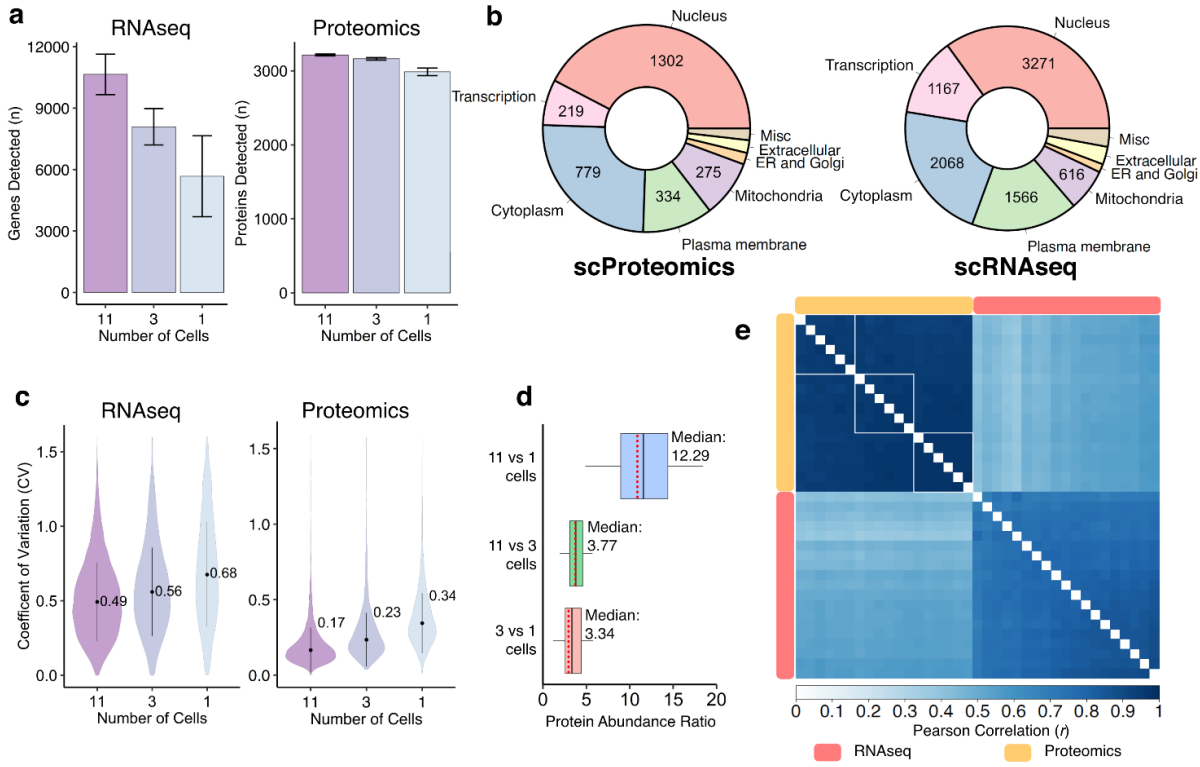

Figure 2. Qualitative and quantitative assessment of nanoSPLITS for transcriptome and proteome measurements

(**a**) Mean number of detected genes and proteins for each modality (n = 6, 6, and 7 for 11, 3, and 1 C10 cell, respectively). Error bars indicate standard deviations (±s.d.). (**b**) CCO of genes (scRNAseq) and proteins (scProteomics) identified in the single-cell data (n = 7). (**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 indicated at the center point within each distribution. (**d**) 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 medians are indicated at the black crossbar while the theoretical ratio for each comparison is shown at the red dotted line within each boxplot. (**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).

As expected, coverage reduced with decreasing cell numbers. Single-cell transcriptome and proteome measurements provided 5,848 and 2,934 identifications on average, respectively. To ensure nanoSPLITS did not introduce bias toward different cellular components due to the nanodroplet splitting process, we also investigated the distribution of gene and protein identifications from single cells within the cellular component ontology (CCO). We found scProteomics and scRNAseq had corresponding identifications within cellular components that encompassed all major organelles (Fig. 2b). Furthermore, 1,521 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.

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 modestly higher CVs for single cells were expected, as the mixed cell populations represent averages of the underlying cell to cell variation. Notably, we observed significantly higher CVs for transcript measurements relative to proteomic measurements, in agreement with recent reports(Woo et al. 2021; Brunner et al. 2022). Presumably, these higher CVs reflect the dynamic nature of mRNA relative to their protein counterparts, which have longer half-lives on average(Buccitelli and Selbach 2020). We also 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. 2d). 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. 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.

Finally, we determined the Pearson correlation coefficients (r) across and within modalities using conceptually similar normalized transformations for each modality (Fig. 2e; 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 dynamic nature of the transcriptome where many genes are often expressed in short transcriptional “bursts” (Buccitelli and Selbach 2020). The cross-correlation between the transcriptome and proteome in single cells was moderate with most correlation coefficients falling within the range of 0.35 to 0.45, on par with previous reports (Brunner et al. 2022; Woo et al. 2021; Buccitelli and Selbach 2020).

2.2 Classification of cell types and identification of markers with nanoSPLITS

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). As the nanoSPLITS approach uses approximately half the cellular content we sought to determine whether the multimodal measurements could precisely distinguish the two cell types and detect gene and/or protein markers, and we also desired to further characterize protein-transcript correlations at the single-cell level.

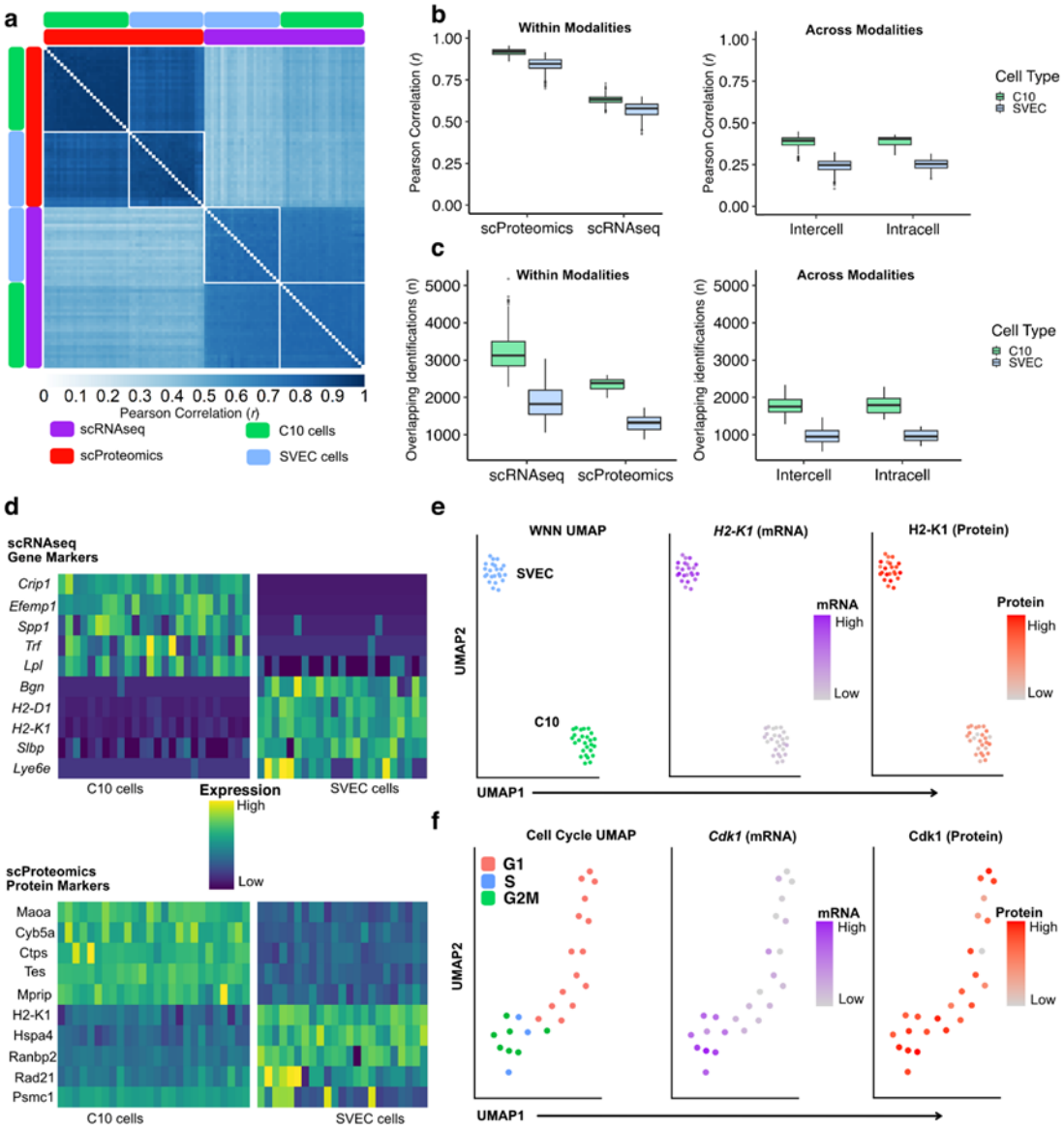

Figure 3. Underlying cell phenotype signatures are maintained after nanoSPLITS.

(**a**). Pearson correlation heatmap with clustering of scRNAseq and scProteomic results for both single C10 (n = 26, paired) and SVEC (n = 23, paired) cells. (**b**) Box plots showing the distributions of Pearson correlations, separated by cell type (C10 and SVEC) and modality (scProteomics and scRNAseq). (**c**) The overlap in gene and protein identifications for each modality separately, as well as across modalities. (**d**) Top 5 gene markers from scRNAseq data and protein markers from scProteomics data 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) Uniform Manifold Approximation and Projection (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) abundance, respectively. (**f**) Feature-based UMAP generated for C10 cells using cell-cycle markers measured in the scRNAseq data. Middle and right panels are colored based on Cdk1 gene (purple) and protein (red) abundance, respectively. All abundance values shown in **d**, **e**, and **f** are derived from Z-scores after scaling and centering of data.

As shown in Fig. 3a, both cell types and modalities could be easily clustered based on correlations alone. In line with our first 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 a slightly broader range of r, ranging from 0.26 to as high as 0.52 with a median of 0.44. We also compared the cross-modality correlations between same single cells ("intracell") and different single cells ("intercell"), however there appeared to be no difference between them (Fig. 3b). Overall, SVEC cells exhibited lower correlations, presumably due to their smaller cell size and correspondingly reduced measurement depth/precision (Fig. 3c). The protein/gene overlap analysis shows 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. 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 transcripts of higher abundance. Finally, we calculated an mRNA-protein correlation for each gene that was observed in both modalities, provided there were at least 4 paired observations. Correlations were further separated by cell type to avoid Simpson's paradox(Franks, Airoldi, and Slavov 2017). Although there were 15 significant mRNA-protein correlations with an FDR < 0.05 and the distribution of mRNA-protein correlations was statistically different compared to a distribution of randomly sampled correlations (Mann-Whitney test, p-value < 2.2×10^{-16}), the mRNA-protein correlations still centered near an r of 0. These results highlight that even at the single cell level, challenge associated with the using mRNA measurements to predict protein counterparts remain.

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 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. Despite this, the previously established SVEC-cell marker H2-K1 was identified here at both, the protein and mRNA level (Fig. 3d). Dimensionality reduction with principal component analysis (PCA) delineated both cell types for scRNAseq and scProteomics despite the cell contents being divided. The integration of both modalities through an unsupervised weighted nearest neighbor (WNN)(Hao et al. 2021) analysis provided robust clustering in the two-dimensional space (Fig. 3e). This also provided us the ability to visualize both protein and mRNA abundances, offering visual confirmation of H2-K1 as a potential marker that is differentially abundant at the protein and gene level (Fig. 3e). Using canonical cell cycle markers,(Nestorowa et al. 2016) 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. 3f) . For example, cyclin-dependent kinase 1 (Cdk1) is upregulated at the transcriptional level in S and G2M phase C10 cells (Fig. 3f). Several other established cell cycle phase genes demonstrated similar differential abundance at the transcriptional level, including DNA topoisomerase IIα (Top2a), cyclin B1 (Ccnb1), G2 and S phase expressed protein 1 (Gtse1), cytoskeleton-associated protein 5 (Ckap5), and anillin (Anln). Furthermore, Top2a protein appeared to be increased in G2/M cells as well.

2.3 Characterization of cell cycle features through the integration of scRNAseq, scProteomics, and single-cell phosphoproteomics analyses of same single cells

Although C10 and SVEC cells could easily be discriminated, the large differences in protein/gene composition between the two distinct cell types precluded establishing whether the technology could be applied to study biological systems with more subtle changes. Therefore, we sought to investigate different cell cycle phase features for C10 cells by arresting them in G2/M phase using the Cdk1 cell cycle inhibitor RO-3306 and comparing them to an untreated

C10 cell population as control(Vassilev et al. 2006). The nanoSPLITS approach allowed us to obtain deep coverage of both the proteome and transcriptome, with an average of 2,942 proteins and 5,559 genes being identified in G2/M arrested cells, as well as 2,574 proteins and 4,173 genes for control 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. Although no phosphopeptide enrichment was performed(Orsburn, Yuan, and Bumpus 2022), the FragPipe proteomics pipeline with FDR-controlled MBR(Yu, Haynes, and Nesvizhskii 2021) also identified over 300 unique phosphopeptides, of which 138 of these were reproducibly observed (<50% missing value). Encouragingly, G2/M arrested and untreated cells could easily be clustered and separated by dimensionality reduction (PCA and WNN-UMAP) for scProteomic and scRNAseq data. Overall, 2,842 proteins and 4,186 mRNA transcripts had sufficient observations for testing differential abundance and 1,930 protein-gene pairs overlapped. Quantitative analysis afforded 327 proteins, 1,434 genes, and 29 phosphopeptides that were differentially abundant (log2FC > 0.5 or < -0.5 and FDR < 0.01). Importantly, covariant protein and mRNA clusters were identified (Fig. 4a and Fig. 4b).

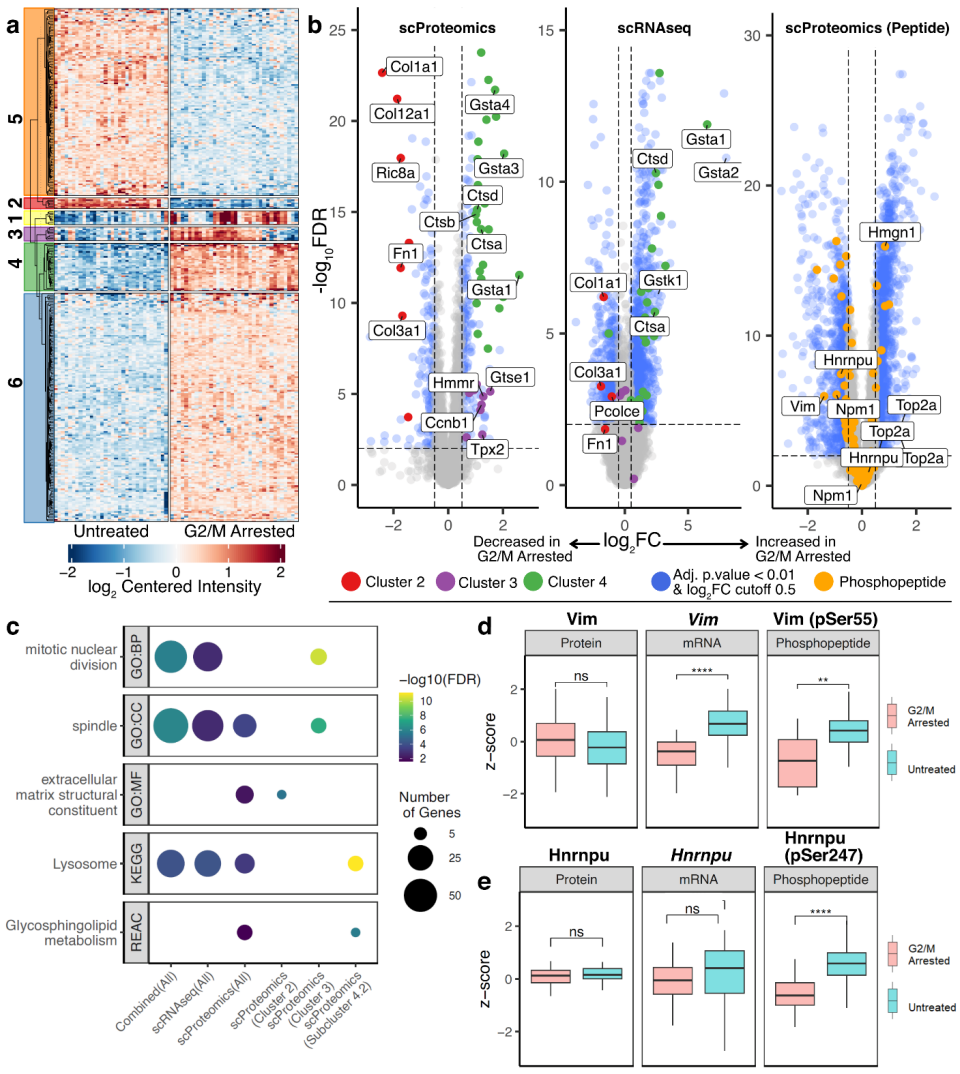

Figure 4. Defining the G2/M protein and mRNA landscape with nanoSPLITS.

(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 (3,182 proteins quantified), scRNAseq data (4,186 genes with < 50% missing values), and peptide level (16,938 peptides). (**c**) Enrichment of GO terms (y-axis) for differentially abundant proteins and genes found in the scProteomic and scRNAseq data (x-axis) with FDR < 0.01 and log_2FC of +/- 0.5. Point size represents number of genes (proteins) and color represents -log10(FDR) (**d**) Integration of scRNAseq, scProteomic, and phosphoproteomic data comparing relative abundances from each modality for Vim. **** FDR < 0.00001, *** FDR < 0.0001, ** FDR < 0.001, and ns indicates not significant. (**e**) Same as (**d**) but with Hnrnpu

Of the proteomic clusters identified, clusters 2, 3, and 4 contained members with the strongest functional relationships. Cluster 2 (Fig. 4a and Fig. 4b) contained 7 proteins that were significantly more abundant in the untreated C10 cells. These were principally 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 Synembryn-A (Ric8A), both of which play important roles within mitosis(Woodard et al. 2010; Wang et al. 2016). Cluster 3 provided 13 proteins that were more abundant in G2/M arrested C10 cells, all of which had mitotic-relevant annotations (Fig.S11a). This included Anln, Gste1, Top2a, targeting protein for Xklp2 (Tpx2), and Ccnb1. Other notable mitotic proteins include aurora kinase B (Aurkb), the mitotic checkpoint serine/threonine-protein kinase BUB1 beta (Bub1b), and the mitotic spindle regulator ubiquitin-conjugating enzyme 2C (Ube2c). These proteins all displayed exceptionally strong correlations with each other, with an average r of 0.71 and some correlations as high as 0.9, such as between Top2a, importin subunit alpha-1 (Kpna2), and Tpx2.

Cluster 4 contained proteins with higher abundance in G2/M arrested cells like cluster 3, but notably all members had p-values of larger significance and no functional annotations related to cell cycle (Fig. 4b). Cluster 4 is also notable in that several of the scProteomic fold-changes are reproduced in the scRNAseq results, demonstrating concordance between mRNA and protein (Fig. 4b). Three subclusters could be identified within cluster 4 based on protein covariation. These subclusters, though small in membership, still provided statistically significant functional and pathway enrichment, with subcluster 4.1 showing clear enrichment for lysosomal processes and subcluster 4.2 being linked to hypoxia and metabolic pathways. Proteins with close functional relationships (for example, prosaponin, progranulin, cathepsin A, and cathepsin D) exhibited strong correlations with each other. Finally, the largest clusters (5 and 6) represented proteins with more modest differences in abundance for either condition.

The paired scRNAseq could cluster arrested and control C10 cells as well; however clustering of genes was less structured, which we attribute to the “bursty” nature of mRNA expression. Several of the proteins described above clustered together in the scRNAseq data. For example,

cluster 1 from the scRNAseq data contained genes that were less abundant in G2/M arrested C10 cells and included several of the proteins noted in cluster 2 from the scProteomic data (Col1a1, Col3a1, and Pcolce). We also found particularly strong concordance with mRNA and protein covariation for the extracellular modulator osteonectin (Sparc) and Col1a1 in G2/M arrested cells. Cluster 4 of the scRNAseq data contained genes with the greatest log2FC in G2/M arrested C10 cells, including several glutathione S-transferase genes and cathepsin genes, many of which were clustered together in the scProteomic results and showed concordance with protein abundance.

We next explored how the global functional enrichment compares between modalities, as well as to the clusters previously identified. Despite relatively poor overlap in differentially abundant genes and proteins, many enriched GO terms were shared between the two modalities (Fig. 4). For terms related to cell cycle and mitochondria, the combination of differentially expressed genes and proteins produced enrichments with greater statistical significance, suggesting the two modalities are more divergent in their identifications within these cases. Several of the protein clusters also produced enrichments with greater significance than what was found globally (Fig. 4c), and terms related to sphingolipid metabolism and ECM were exclusively found at the protein level.

From the over 300 phosphopeptides identified from the scProteomic data, 29 were differentially abundant in G2/M arrested cells (Fig. 4b). We also related 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(Yamaguchi et al. 2005; Chou et al. 1990), was significantly decreased in abundance in G2/M arrested cells (Fig. 4d), and this decrease in protein phosphorylation appears to be independent of the global protein abundance (Fig. 4d), providing the most direct evidence for the loss of Cdk1 kinase activity inhibited by RO-3306 (Vassilev et

al. 2006). Another phosphosite, pSer247, on heterogeneous nuclear ribonucleoprotein U (Hnrnpu) was also found to be independent of overall protein abundance, following a similar trend as Vim pSer55 and suggesting it may be a target of Cdk1 (Fig. 4e). Interestingly, the phosphorylation of Hnrpu (also known as scaffold attachment factor A, SAF-A) 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 identified(Douglas et al. 2015). Our data suggests this upstream priming by Cdk1 may be positioned at Ser247. Further evidence of Cdk1's direct action on Vim and Hnrnpu is supported by the observation that both serine residues are flanked by a proline, which is an established motif of proline-directed kinase substrates.

3.0 Discussion

Compared with previous technologies that utilize antibodies to infer protein abundances in multimodal experiments, the nanoSPLITS platform employs mass spectrometry to globally detect proteins and post-translational modifications. We have demonstrated how the nanoSPLITS approach can enable multimodal profiling of thousands of transcripts and proteins, as well as hundreds of phosphopeptides from same single cells. The multiomics data acquired on same single cells allowed us to precisely quantify the abundances of both transcripts and proteins, and identify marker genes and proteins from both modalities. Quantification of protein abundances from 11, 3, and 1 C10 cells demonstrated high-precision (median CV < 0.34 for single cells) and experimental abundance ratios were found to be close to theoretical. 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 studies(Buccitelli and Selbach 2020). We also demonstrated how nanoSPLITS could delineate different cell types and identify candidate marker genes or proteins. Notably, the overlap of marker genes/proteins between scRNAseq and scProteomics was low, suggesting identifying reliable markers could benefit from collecting data with both modalities.

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 mRNA-protein measurements derived from the same single cell. The mRNA-protein correlations measured could be distinguished from a distribution of randomly sampled measurements, and a trend towards more positive correlations was observed. Conversely, the overall distribution was found to center near an r of 0, providing further evidence that the lack of correlation between mRNA and protein levels is a general phenomenon(Taniguchi et al. 2010; Darmanis et al. 2016). Indeed, the inherent stochasticity of gene expression is well documented(Sanchez and Golding 2013). Overall, our results agree with the perspective that

correlations between mRNA and proteins at the single-cell level are less correlated relative to bulk measurements which “average out” sources of variance within a population of cells(Liu, Beyer, and Aebersold 2016; Darmanis et al. 2016). However, the modest shift in correlations towards positive r leaves open the possibility that at least some transcripts are more tightly linked to their translated products.

We also investigated how nanoSPLITS can assign more subtle 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 numerous changes in protein and mRNA abundance in line with the expected phenotype. Increases in the abundance of cyclin-dependent kinase 2 (Cdk2) and cyclin-B1 (Ccnb1) were noted, as well as other cyclin-dependent kinases such as Cdk16, Cdk11b, and Cdk6., This revealed broad reorganization of cyclin-CDK complexes in response to the inhibition of Cdk1,in agreement with prior work investigating CDK1 loss(Lau et al. 2021). Over 300 phosphopeptides were also identified without phosphopeptide enrichment, providing one of the deepest single-cell phosphoproteomic datasets to date(Orsburn, Yuan, and Bumpus 2022; Tsai et al. 2023). We identified the previously described Cdk1 phosphosite pSer55 Vim and the lesser characterized pSer247 Hnrnpu as being directly impacted by Cdk1 inhibition, with the latter phosphosite being implicated as a priming site for Cdk1(Yamaguchi et al. 2005; Chou et al. 1990; Douglas et al. 2015). These phosphoproteomic findings provided additional insight into post-translational regulation(Orsburn, Yuan, and Bumpus 2022), exemplifying the benefits of global mass-spectrometry based proteomics in single-cell analysis.

One advantage of single-cell approaches is the identification of covarying protein or gene clusters in specific cell types or biological contexts(Budnik et al. 2018; Leduc et al. 2022). In the context of cell cycle arrest, we identified several covarying clusters of proteins and genes including a cluster composed exclusively of canonical cell cycle proteins. Many of these proteins had exceptionally strong correlations, such as Top2a and Kpna2 with r of 0.92 in G2/M arrested

cells. The strength of this correlation suggests the relative abundances of these proteins are tightly regulated. Tpx2, a highly correlated protein within this cluster that is involved in spindle assembly, is known to be sequestered by importins-α/β with 1:1 stoichiometry through its nuclear localization sequence (NLS) before being released during G2 phase by Hmmr (strongly correlated in this cluster as well) at the nuclear envelope (Safari et al. 2020; Chu et al. 2018). Considering Top2a contains a similar NLS(Mirski, Gerlach, and Cole 1999), our data raises the possibility that Top2a is regulated in a similar manner. Interestingly, many of the mitotic proteins described above did not show concordance with the scRNAseq data. One interpretation is that once the critical concentrations of these mitotic proteins are met, only lower levels of transcription are necessary to maintain proteins at that level. Furthermore, abundance (as measured by mRNA) of Top2a, Gtse1, and Ccnb1 has been shown to peak at G2 or in early mitosis(Hwang, McKenna, and Muschel 1998; Nielsen et al. 2020) before rapid degradation at the protein level during mitotic exit(Abdelbaki et al. 2020). Indeed, the G2/M arrested cells appear to be in the later stages of mitosis based on the mitotic protein cluster we identified(Kelly et al. 2022). Therefore, it can be inferred that under cell cycle arrest the peak transcriptional states of many mitotic proteins are not continuously maintained.

We also observed a protein cluster containing ECM proteins (Pcolce, Fn1, Col1a1, Col3a1, and Col12a1), Lnpk, and Ric8a decreasing in abundance in G2/M arrested cells. Lnpk is known to be an endoplasmic reticulum(ER)-shaping protein, and reduction of Lnpk facilitates the transition from tubular to sheet ER morphology during mitosis(Wang et al. 2016). Ric8a is a key regulator during metaphase for mitotic spindle orientation, and reduction of Ric8a has been demonstrated to lead to prolonged mitosis and spindle defects(Woodard et al. 2010). As Ric8a was considerably less abundant in G2/M arrested cells, it may play a direct role in Cdk1-inhibited mechanism of arrest. The remaining ECM proteins also shed light on cell morphological changes that occur during mitosis. It has been established that Cdk1 regulates

the remodeling of cell adhesion complexes during the cell cycle, and endogenous inactivation of Cdk1 triggers this remodeling event(Jones et al. 2018; Jones, Zha, and Humphries 2019). Typically, this remodeling is thought to involve rapid recycling of cytoskeletal components. Our results demonstrate these ECM proteins are being reduced at both the transcriptional and translational level during G2/M arrest, suggesting ECM degradation and transcriptional repression are also key to the adhesion complex remodeling process during mitosis. Hence, nanoSPLITS-based single-cell multiomics can not only reconstruct known mitotic processes but also identify new processes underlying biology.

The nanoSPLITS platform holds promise to become a powerful discovery tool for biomedical applications, such as characterizing tissue, peripheral blood cell, and circulating tumor cell heterogeneity. Notably, nanoSPLITS is not restricted to the two modalities applied here (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 (and/or nanoSPLITS-derived technologies) will also enable greater insight into how different modalities interact with each other to orchestrate single-cell phenotypes in health and disease.

Although a low throughput approach was employed in this study, high-throughput multiplexing approaches such as CEL-Seq(Hashimshony et al. 2012) for transcriptomics and SCoPE-MS(Budnik et al. 2018) for proteomics can be readily integrated into the nanoSPLITS workflow. The integration of multiplexing approaches to nanoSPLITS would enable effective analysis of thousands of single cells with reasonable throughput and cost(Dephoure and Gygi 2012). Regardless, nanoSPLITS enables scProteomic data to be directly connected to much higher-throughput scRNAseq results, providing the opportunity for proteomic data to be bridged to large single-cell atlases(Tabula Sapiens et al. 2022; Hu 2019). Additionally, the scRNAseq component of nanoSPLITS can serve as a preselection tool to identify rare cell phenotypes for

further analysis by scProteomics, obviating the need for collecting data on thousands of cells with scProteomics.

4.0 Methods

4.1 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), 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) ,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 ThermoFisher Scientific (Waltham, MA, USA). SMART-Seq V4 Plus kit (Cat# R400753) was purchased from Takara Bio USA.

4.2 Design, fabrication, and assembly of the nanoSPLITS chips

The nanoSPLITS chips were fabricated using standard photolithography, wet etching, and silanization as described previously(Zhu, Piehowski, et al. 2018) . Two different chips were designed and used in this 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.

4.3 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). 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 1,000 × 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 x 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.

4.4 CellenONE cell sorting and nanoSPLITS workflow

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 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(Woo et al. 2022), 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 settings, single-cell images, and metadata can be accessed at the GitHub repository provided (https://github.com/Cajun-data/nanoSPLITS).

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 (Movie S1). 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 seconds 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 (with the exception of the data generated for Fig. S3). 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.

4.5 Sample preparation and LC-MS/MS analysis

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 50ºC 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 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(Zhu, Piehowski, et al. 2018). 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 (ThermoFisher Scientific) with FAIMS, 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 TIFF method(Woo et al. 2022) samples, 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 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 the TIFF library 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. Precursor ions with intensities > 1E4 were fragmented by 30% HCD and scanned with an AGC of 2E4 and an IT of 254 ms.

4.6 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 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 (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) using STAR (https://github.com/alexdobin/STAR). BAM file outputs were mapped to genes using htseq-count with default settings. TPM counts were derived using an R script based on TPM procedure.

4.7 Database searching and data analysis

All proteomic data raw files were processed by FragPipe(Kong et al. 2017) 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 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 methionine, N-terminal acetylation, and S/Y/T phosphorylation. Protein and peptide identifications were filtered to a false discovery rate of less than 0.01 within FragPipe. For the TIFF methodology, IonQuant 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). With regards to quality control filtering of samples in scProteomics and scRNAseq, thresholds were set based on the degree of protein or gene missingness. For scRNAseq data, a minimum of 5 read counts were used to filter genes with

low abundance. When mRNA and protein abundances were directly correlated or compared between each other, TPM and riBAQ values were applied, otherwise only intensity or log2(intensity) were used with scProteomic data. With scRNAseq data, raw counts were used for analysis in Seurat and CPM was utilized for within-modality correlations. For cases where dimensionality reduction was applied to proteomic data, imputation was performed with k-nearest neighbors imputation on proteins with <50% missing values. When K-means clustering was applied, appropriate range for clusters was identified using the “elbow method,” defined by plotting within-cluster sum of squares versus number of clusters. GO analysis was performed with the gprofiler2 R package and web application. All of the figures generated, and associated code are included in R markdown or script files at the nanoSPLITS GitHub repository (https://github.com/Cajun-data/nanoSPLITS).

5.0 References

Abdelbaki, A., H. B. Akman, M. Poteau, R. Grant, O. Gavet, G. Guarguaglini, and C. Lindon. 2020. "AURKA destruction is decoupled from its activity at mitotic exit but is essential to suppress interphase activity." *J Cell Sci* 133 (12). https://doi.org/10.1242/jcs.243071. https://www.ncbi.nlm.nih.gov/pubmed/32393600.

Baker, M. 2015. "Reproducibility crisis: Blame it on the antibodies." *Nature* 521 (7552): 274-6. https://doi.org/10.1038/521274a. https://www.ncbi.nlm.nih.gov/pubmed/25993940.

Brunner, A. D., M. Thielert, C. Vasilopoulou, C. Ammar, F. Coscia, A. Mund, O. B. Hoerning, N. Bache, A. Apalategui, M. Lubeck, S. Richter, D. S. Fischer, O. Raether, M. A. Park, F. Meier, F. J. Theis, and M. Mann. 2022. "Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation." *Mol Syst Biol* 18 (3): e10798. https://doi.org/10.15252/msb.202110798. https://www.ncbi.nlm.nih.gov/pubmed/35226415.

Buccitelli, C., and M. Selbach. 2020. "mRNAs, proteins and the emerging principles of gene expression control." *Nat Rev Genet* 21 (10): 630-644. https://doi.org/10.1038/s41576-020-0258-4. https://www.ncbi.nlm.nih.gov/pubmed/32709985.

Budnik, B., E. Levy, G. Harmange, and N. Slavov. 2018. "SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation." *Genome Biol* 19 (1): 161. https://doi.org/10.1186/s13059-018-1547-5. https://www.ncbi.nlm.nih.gov/pubmed/30343672.

Chou, Y. H., J. R. Bischoff, D. Beach, and R. D. Goldman. 1990. "Intermediate filament reorganization during mitosis is mediated by p34cdc2 phosphorylation of vimentin." *Cell* 62 (6): 1063-71. https://doi.org/10.1016/0092-8674(90)90384-q. https://www.ncbi.nlm.nih.gov/pubmed/2169348.

Chu, Tony L.H., Marisa Connell, Lixin Zhou, Zhengcheng He, Jennifer Won, Helen Chen, Seyed M.R. Rahavi, Pooja Mohan, Oksana Nemirovsky, Abbas Fotovati, Miguel Angel Pujana, Gregor S.D. Reid, Torsten O. Nielsen, Nelly Pante, and Christopher A. Maxwell. 2018. "Cell Cycle–Dependent Tumor Engraftment and Migration Are Enabled by Aurora-A." *Molecular Cancer Research* 16 (1): 16-31. https://doi.org/10.1158/1541-7786.Mcr-17-0417. https://doi.org/10.1158/1541-7786.MCR-17-0417.

Chung, H., C. N. Parkhurst, E. M. Magee, D. Phillips, E. Habibi, F. Chen, B. Z. Yeung, J. Waldman, D. Artis, and A. Regev. 2021. "Joint single-cell measurements of nuclear proteins and RNA in vivo." *Nat Methods* 18 (10): 1204-1212. https://doi.org/10.1038/s41592-021-01278-1. https://www.ncbi.nlm.nih.gov/pubmed/34608310.

Cong, Yongzheng, Khatereh Motamedchaboki, Santosh A. Misal, Yiran Liang, Amanda J. Guise, Thy Truong, Romain Huguet, Edward D. Plowey, Ying Zhu, Daniel Lopez-Ferrer, and Ryan T. Kelly. 2021. "Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell." *Chemical Science* 12 (3): 1001-1006. https://doi.org/10.1039/D0SC03636F. http://dx.doi.org/10.1039/D0SC03636F.

Darmanis, S., C. J. Gallant, V. D. Marinescu, M. Niklasson, A. Segerman, G. Flamourakis, S. Fredriksson, E. Assarsson, M. Lundberg, S. Nelander, B. Westermark, and U. Landegren. 2016. "Simultaneous Multiplexed Measurement of RNA and Proteins in Single Cells." *Cell Rep* 14 (2): 380-9. https://doi.org/10.1016/j.celrep.2015.12.021. https://www.ncbi.nlm.nih.gov/pubmed/26748716.

Dephoure, N., and S. P. Gygi. 2012. "Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast." *Sci Signal* 5 (217): rs2. https://doi.org/10.1126/scisignal.2002548. https://www.ncbi.nlm.nih.gov/pubmed/22457332.

Douglas, P., R. Ye, N. Morrice, S. Britton, L. Trinkle-Mulcahy, and S. P. Lees-Miller. 2015. "Phosphorylation of SAF-A/hnRNP-U Serine 59 by Polo-Like Kinase 1 Is Required for Mitosis." *Mol Cell Biol* 35 (15): 2699-713. https://doi.org/10.1128/mcb.01312-14.

Franks, Alexander, Edoardo Airoldi, and Nikolai Slavov. 2017. "Post-transcriptional regulation across human tissues." *PLOS Computational Biology* 13 (5): e1005535. https://doi.org/10.1371/journal.pcbi.1005535. https://doi.org/10.1371/journal.pcbi.1005535.

Frei, A. P., F. A. Bava, E. R. Zunder, E. W. Hsieh, S. Y. Chen, G. P. Nolan, and P. F. Gherardini. 2016. "Highly multiplexed simultaneous detection of RNAs and proteins in single cells." *Nat Methods* 13 (3): 269-75. https://doi.org/10.1038/nmeth.3742. https://www.ncbi.nlm.nih.gov/pubmed/26808670.

Hao, Y., S. Hao, E. Andersen-Nissen, W. M. Mauck, 3rd, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, and R. Satija. 2021. "Integrated analysis of multimodal single-cell data." *Cell* 184 (13): 3573-3587 e29. https://doi.org/10.1016/j.cell.2021.04.048. https://www.ncbi.nlm.nih.gov/pubmed/34062119.

Hashimshony, T., F. Wagner, N. Sher, and I. Yanai. 2012. "CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification." *Cell Rep* 2 (3): 666-73. https://doi.org/10.1016/j.celrep.2012.08.003. https://www.ncbi.nlm.nih.gov/pubmed/22939981.

Hu, Bmap Consortium. 2019. "The human body at cellular resolution: the NIH Human Biomolecular Atlas Program." *Nature* 574 (7777): 187-192. https://doi.org/10.1038/s41586-019-1629-x. https://www.ncbi.nlm.nih.gov/pubmed/31597973.

Hwang, A., W. G. McKenna, and R. J. Muschel. 1998. "Cell cycle-dependent usage of transcriptional start sites. A novel mechanism for regulation of cyclin B1." *J Biol Chem* 273 (47): 31505-9. https://doi.org/10.1074/jbc.273.47.31505. https://www.ncbi.nlm.nih.gov/pubmed/9813064.

Jones, M. C., J. A. Askari, J. D. Humphries, and M. J. Humphries. 2018. "Cell adhesion is regulated by CDK1 during the cell cycle." *J Cell Biol* 217 (9): 3203-3218. https://doi.org/10.1083/jcb.201802088. https://www.ncbi.nlm.nih.gov/pubmed/29930204.

Jones, M. C., J. Zha, and M. J. Humphries. 2019. "Connections between the cell cycle, cell adhesion and the cytoskeleton." *Philos Trans R Soc Lond B Biol Sci* 374 (1779): 20180227. https://doi.org/10.1098/rstb.2018.0227. https://www.ncbi.nlm.nih.gov/pubmed/31431178.

Kelly, Van, Aymen al-Rawi, David Lewis, Georg Kustatscher, and Tony Ly. 2022. "Low Cell Number Proteomic Analysis Using In-Cell Protease Digests Reveals a Robust Signature for Cell Cycle State Classification." *Molecular & Cellular Proteomics* 21 (1): 100169. https://doi.org/https://doi.org/10.1016/j.mcpro.2021.100169. https://www.sciencedirect.com/science/article/pii/S1535947621001419.

Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, and M. W. Kirschner. 2015. "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." *Cell* 161 (5): 1187-1201. https://doi.org/10.1016/j.cell.2015.04.044. https://www.ncbi.nlm.nih.gov/pubmed/26000487.

Kong, A. T., F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, and A. I. Nesvizhskii. 2017. "MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics." *Nat Methods* 14 (5): 513-520. https://doi.org/10.1038/nmeth.4256. https://www.ncbi.nlm.nih.gov/pubmed/28394336.

Lau, H. W., H. T. Ma, T. K. Yeung, M. Y. Tam, D. Zheng, S. K. Chu, and R. Y. C. Poon. 2021. "Quantitative differences between cyclin-dependent kinases underlie the unique functions of CDK1 in human cells." *Cell Rep* 37 (2): 109808. https://doi.org/10.1016/j.celrep.2021.109808. https://www.ncbi.nlm.nih.gov/pubmed/34644583.

Leduc, A., R. G. Huffman, J. Cantlon, S. Khan, and N. Slavov. 2022. "Exploring functional protein covariation across single cells using nPOP." *Genome Biol* 23 (1): 261. https://doi.org/10.1186/s13059-022-02817-5.

Liu, Y., A. Beyer, and R. Aebersold. 2016. "On the Dependency of Cellular Protein Levels on mRNA Abundance." *Cell* 165 (3): 535-50. https://doi.org/10.1016/j.cell.2016.03.014. https://www.ncbi.nlm.nih.gov/pubmed/27104977.

Macosko, E. Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, and S. A. McCarroll. 2015. "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell* 161 (5): 1202-1214. https://doi.org/10.1016/j.cell.2015.05.002. https://www.ncbi.nlm.nih.gov/pubmed/26000488.

Marcon, E., H. Jain, A. Bhattacharya, H. Guo, S. Phanse, S. Pu, G. Byram, B. C. Collins, E. Dowdell, M. Fenner, X. Guo, A. Hutchinson, J. J. Kennedy, B. Krastins, B. Larsen, Z. Y. Lin, M. F. Lopez, P. Loppnau, S. Miersch, T. Nguyen, J. B. Olsen, M. Paduch, M. Ravichandran, A. Seitova, G. Vadali, M. S. Vogelsang, J. R. Whiteaker, G. Zhong, N. Zhong, L. Zhao, R. Aebersold, C. H. Arrowsmith, A. Emili, L. Frappier, A. C. Gingras, M. Gstaiger, A. G. Paulovich, S. Koide, A. A. Kossiakoff, S. S. Sidhu, S. J. Wodak, S. Graslund, J. F. Greenblatt, and A. M. Edwards. 2015. "Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation." *Nat Methods* 12 (8): 725-31. https://doi.org/10.1038/nmeth.3472. https://www.ncbi.nlm.nih.gov/pubmed/26121405.

Mirski, S. E., J. H. Gerlach, and S. P. Cole. 1999. "Sequence determinants of nuclear localization in the alpha and beta isoforms of human topoisomerase II." *Exp Cell Res* 251 (2): 329-39. https://doi.org/10.1006/excr.1999.4587.

Montoro, D. T., A. L. Haber, M. Biton, V. Vinarsky, B. Lin, S. E. Birket, F. Yuan, S. Chen, H. M. Leung, J. Villoria, N. Rogel, G. Burgin, A. M. Tsankov, A. Waghray, M. Slyper, J. Waldman, L.

Nguyen, D. Dionne, O. Rozenblatt-Rosen, P. R. Tata, H. Mou, M. Shivaraju, H. Bihler, M. Mense, G. J. Tearney, S. M. Rowe, J. F. Engelhardt, A. Regev, and J. Rajagopal. 2018. "A revised airway epithelial hierarchy includes CFTR-expressing ionocytes." *Nature* 560 (7718): 319-324. https://doi.org/10.1038/s41586-018-0393-7. https://www.ncbi.nlm.nih.gov/pubmed/30069044.

Nestorowa, S., F. K. Hamey, B. Pijuan Sala, E. Diamanti, M. Shepherd, E. Laurenti, N. K. Wilson, D. G. Kent, and B. Gottgens. 2016. "A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation." *Blood* 128 (8): e20-31. https://doi.org/10.1182/blood-2016-05-716480. https://www.ncbi.nlm.nih.gov/pubmed/27365425.

Nielsen, C. F., T. Zhang, M. Barisic, P. Kalitsis, and D. F. Hudson. 2020. "Topoisomerase IIalpha is essential for maintenance of mitotic chromosome structure." *Proc Natl Acad Sci U S A* 117 (22): 12131-12142. https://doi.org/10.1073/pnas.2001760117. https://www.ncbi.nlm.nih.gov/pubmed/32414923.

Orsburn, Benjamin C., Yuting Yuan, and Namandjé N. Bumpus. 2022. "Insights into protein post-translational modification landscapes of individual human cells by trapped ion mobility time-of-flight mass spectrometry." *Nature Communications* 13 (1): 7246. https://doi.org/10.1038/s41467-022-34919-w. https://doi.org/10.1038/s41467-022-34919-w.

Peterson, V. M., K. X. Zhang, N. Kumar, J. Wong, L. Li, D. C. Wilson, R. Moore, T. K. McClanahan, S. Sadekova, and J. A. Klappenbach. 2017. "Multiplexed quantification of proteins and transcripts in single cells." *Nat Biotechnol* 35 (10): 936-939. https://doi.org/10.1038/nbt.3973. https://www.ncbi.nlm.nih.gov/pubmed/28854175.

Picelli, S., O. R. Faridani, A. K. Bjorklund, G. Winberg, S. Sagasser, and R. Sandberg. 2014. "Full-length RNA-seq from single cells using Smart-seq2." *Nat Protoc* 9 (1): 171-81. https://doi.org/10.1038/nprot.2014.006. https://www.ncbi.nlm.nih.gov/pubmed/24385147.

Rust, K., L. E. Byrnes, K. S. Yu, J. S. Park, J. B. Sneddon, A. D. Tward, and T. G. Nystul. 2020. "A single-cell atlas and lineage analysis of the adult Drosophila ovary." *Nat Commun* 11 (1): 5628. https://doi.org/10.1038/s41467-020-19361-0. https://www.ncbi.nlm.nih.gov/pubmed/33159074.

Safari, Mohammad S., Matthew R. King, Clifford P. Brangwynne, and Sabine Petry. 2020. "Branching microtubule nucleation is controlled by importin-mediated inhibition of TPX2 phase separation." *bioRxiv*: 2020.09.01.276469. https://doi.org/10.1101/2020.09.01.276469. https://www.biorxiv.org/content/biorxiv/early/2020/09/01/2020.09.01.276469.full.pdf.

Sanchez, A., and I. Golding. 2013. "Genetic determinants and cellular constraints in noisy gene expression." *Science* 342 (6163): 1188-93. https://doi.org/10.1126/science.1242975. https://www.ncbi.nlm.nih.gov/pubmed/24311680.

Stoeckius, M., C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, and P. Smibert. 2017. "Simultaneous epitope and transcriptome measurement in single cells." *Nat Methods* 14 (9): 865-868. https://doi.org/10.1038/nmeth.4380. https://www.ncbi.nlm.nih.gov/pubmed/28759029.

Tabula Sapiens, Consortium, R. C. Jones, J. Karkanias, M. A. Krasnow, A. O. Pisco, S. R. Quake, J. Salzman, N. Yosef, B. Bulthaup, P. Brown, W. Harper, M. Hemenez, R. Ponnusamy,

A. Salehi, B. A. Sanagavarapu, E. Spallino, K. A. Aaron, W. Concepcion, J. M. Gardner, B. Kelly, N. Neidlinger, Z. Wang, S. Crasta, S. Kolluru, M. Morri, A. O. Pisco, S. Y. Tan, K. J. Travaglini, C. Xu, M. Alcantara-Hernandez, N. Almanzar, J. Antony, B. Beyersdorf, D. Burhan, K. Calcuttawala, M. M. Carter, C. K. F. Chan, C. A. Chang, S. Chang, A. Colville, S. Crasta, R. N. Culver, I. Cvijovic, G. D'Amato, C. Ezran, F. X. Galdos, A. Gillich, W. R. Goodyer, Y. Hang, A. Hayashi, S. Houshdaran, X. Huang, J. C. Irwin, S. Jang, J. V. Juanico, A. M. Kershner, S. Kim, B. Kiss, S. Kolluru, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, B. Li, G. B. Loeb, W. J. Lu, S. Mantri, M. Markovic, P. L. McAlpine, A. de Morree, M. Morri, K. Mrouj, S. Mukherjee, T. Muser, P. Neuhofer, T. D. Nguyen, K. Perez, R. Phansalkar, A. O. Pisco, N. Puluca, Z. Qi, P. Rao, H. Raquer-McKay, N. Schaum, B. Scott, B. Seddighzadeh, J. Segal, S. Sen, S. Sikandar, S. P. Spencer, L. C. Steffes, V. R. Subramaniam, A. Swarup, M. Swift, K. J. Travaglini, W. Van Treuren, E. Trimm, S. Veizades, S. Vijayakumar, K. C. Vo, S. K. Vorperian, W. Wang, H. N. W. Weinstein, J. Winkler, T. T. H. Wu, J. Xie, A. R. Yung, Y. Zhang, A. M. Detweiler, H. Mekonen, N. F. Neff, R. V. Sit, M. Tan, J. Yan, G. R. Bean, V. Charu, E. Forgo, B. A. Martin, M. G. Ozawa, O. Silva, S. Y. Tan, A. Toland, V. N. P. Vemuri, S. Afik, K. Awayan, O. B. Botvinnik, A. Byrne, M. Chen, R. Dehghannasiri, A. M. Detweiler, A. Gayoso, A. A. Granados, Q. Li, G. Mahmoudabadi, A. McGeever, A. de Morree, J. E. Olivieri, M. Park, A. O. Pisco, N. Ravikumar, J. Salzman, G. Stanley, M. Swift, M. Tan, W. Tan, A. J. Tarashansky, R. Vanheusden, S. K. Vorperian, P. Wang, S. Wang, G. Xing, C. Xu, N. Yosef, M. Alcantara-Hernandez, J. Antony, C. K. F. Chan, C. A. Chang, A. Colville, S. Crasta, R. Culver, L. Dethlefsen, C. Ezran, A. Gillich, Y. Hang, P. Y. Ho, J. C. Irwin, S. Jang, A. M. Kershner, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, S. Liu, G. B. Loeb, W. J. Lu, J. S. Maltzman, R. J. Metzger, A. de Morree, P. Neuhofer, K. Perez, R. Phansalkar, Z. Qi, P. Rao, H. Raquer-McKay, K. Sasagawa, B. Scott, R. Sinha, H. Song, S. P. Spencer, A. Swarup, M. Swift, K. J. Travaglini, E. Trimm, S. Veizades, S. Vijayakumar, B. Wang, W. Wang, J. Winkler, J. Xie, A. R. Yung, S. E. Artandi, P. A. Beachy, M. F. Clarke, L. C. Giudice, F. W. Huang, K. C. Huang, J. Idoyaga, S. K. Kim, M. Krasnow, C. S. Kuo, P. Nguyen, S. R. Quake, T. A. Rando, K. Red-Horse, J. Reiter, D. A. Relman, J. L. Sonnenburg, B. Wang, A. Wu, S. M. Wu, and T. Wyss-Coray. 2022. "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans." *Science* 376 (6594): eabl4896. https://doi.org/10.1126/science.abl4896. https://www.ncbi.nlm.nih.gov/pubmed/35549404.

Taniguchi, Y., P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. 2010. "Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells." *Science* 329 (5991): 533-8. https://doi.org/10.1126/science.1188308. https://www.ncbi.nlm.nih.gov/pubmed/20671182.

Tsai, Chia-Feng, Yi-Ting Wang, Chuan-Chih Hsu, Reta Birhanu Kitata, Rosalie K. Chu, Marija Velickovic, Rui Zhao, Sarah M. Williams, William B. Chrisler, Marda L. Jorgensen, Ronald J. Moore, Ying Zhu, Karin D. Rodland, Richard D. Smith, Clive H. Wasserfall, Tujin Shi, and Tao Liu. 2023. "A streamlined tandem tip-based workflow for sensitive nanoscale phosphoproteomics." *Communications Biology* 6 (1): 70. https://doi.org/10.1038/s42003-022-04400-x. https://doi.org/10.1038/s42003-022-04400-x.

Vassilev, L. T., C. Tovar, S. Chen, D. Knezevic, X. Zhao, H. Sun, D. C. Heimbrook, and L. Chen. 2006. "Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1." *Proc Natl Acad Sci U S A* 103 (28): 10660-5. https://doi.org/10.1073/pnas.0600447103. https://www.ncbi.nlm.nih.gov/pubmed/16818887.

Wang, S., H. Tukachinsky, F. B. Romano, and T. A. Rapoport. 2016. "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. https://www.ncbi.nlm.nih.gov/pubmed/27619977.

Woo, J., G. C. Clair, S. M. Williams, S. Feng, C. F. Tsai, R. J. Moore, W. B. Chrisler, R. D. Smith, R. T. Kelly, L. Pasa-Tolic, C. Ansong, and Y. Zhu. 2022. "Three-dimensional feature matching improves coverage for single-cell proteomics based on ion mobility filtering." *Cell Syst* 13 (5): 426-434 e4. https://doi.org/10.1016/j.cels.2022.02.003. https://www.ncbi.nlm.nih.gov/pubmed/35298923.

Woo, J., S. M. Williams, L. M. Markillie, S. Feng, C. F. Tsai, V. Aguilera-Vazquez, R. L. Sontag, R. J. Moore, D. Hu, H. S. Mehta, J. Cantlon-Bruce, T. Liu, J. N. Adkins, R. D. Smith, G. C. Clair, L. Pasa-Tolic, and Y. Zhu. 2021. "High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip." *Nat Commun* 12 (1): 6246. https://doi.org/10.1038/s41467-021-26514-2. https://www.ncbi.nlm.nih.gov/pubmed/34716329.

Woodard, G. E., N. N. Huang, H. Cho, T. Miki, G. G. Tall, and J. H. Kehrl. 2010. "Ric-8A and Gi alpha recruit LGN, NuMA, and dynein to the cell cortex to help orient the mitotic spindle." *Mol Cell Biol* 30 (14): 3519-30. https://doi.org/10.1128/MCB.00394-10. https://www.ncbi.nlm.nih.gov/pubmed/20479129.

Yamaguchi, T., H. Goto, T. Yokoyama, H. Sillje, A. Hanisch, A. Uldschmid, Y. Takai, T. Oguri, E. A. Nigg, and M. Inagaki. 2005. "Phosphorylation by Cdk1 induces Plk1-mediated vimentin phosphorylation during mitosis." *J Cell Biol* 171 (3): 431-6. https://doi.org/10.1083/jcb.200504091. https://www.ncbi.nlm.nih.gov/pubmed/16260496.

Yu, Fengchao, Sarah E. Haynes, and Alexey I. Nesvizhskii. 2021. "IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs." *Molecular & Cellular Proteomics* 20. https://doi.org/10.1016/j.mcpro.2021.100077. https://doi.org/10.1016/j.mcpro.2021.100077.

Zhu, Y., G. Clair, W. B. Chrisler, Y. Shen, R. Zhao, A. K. Shukla, R. J. Moore, R. S. Misra, G. S. Pryhuber, R. D. Smith, C. Ansong, and R. T. Kelly. 2018. "Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS." *Angew Chem Int Ed Engl* 57 (38): 12370-12374. https://doi.org/10.1002/anie.201802843. https://www.ncbi.nlm.nih.gov/pubmed/29797682.

Zhu, Y., P. D. Piehowski, R. Zhao, J. Chen, Y. Shen, R. J. Moore, A. K. Shukla, V. A. Petyuk, M. Campbell-Thompson, C. E. Mathews, R. D. Smith, W. J. Qian, and R. T. Kelly. 2018. "Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells." *Nat Commun* 9 (1): 882. https://doi.org/10.1038/s41467-018-03367-w. https://www.ncbi.nlm.nih.gov/pubmed/29491378

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov