bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license.

Droplet sample preparation for single-cell proteomics applied to the cell cycle

Andrew Leduc,[1,2] R. Gray Huffman,[1,2] & Nikolai Slavov[1,2,3,✉]

[1]Department of Bioengineering, Northeastern University, Boston, MA 02115, USA
[2]Barnett Institute, Northeastern University, Boston, MA 02115, USA
[3]Department of Biology, Northeastern University, Boston, MA 02115, USA
✉ Correspondence: nslavov@alum.mit.edu or nslavov@northeastern.edu
∈ Data, code & protocols: scope2.slavovlab.net/nPOP

Many biological functions, such as the cell division cycle, are intrinsically single-cell processes regulated in part by protein synthesis and degradation. Investigating such processes has motivated the development of single-cell mass spectrometry (MS) proteomics. To further advance single-cell MS proteomics, we developed a method for high throughput, automated nano-ProteOmic sample Preparation (nPOP). nPOP uses piezo acoustic dispensing to isolate individual cells in 300 picoliter volumes and performs all subsequent preparation steps in small droplets on a fluorocarbon coated slide. This design enables simultaneous sample preparation of thousands of single cells, including lysing, digesting, and labeling individual cells in volumes below 20 nl. Single-cell protein analysis using nPOP classified cells by cell type and by cell cycle phase. Furthermore, the data allowed us to quantify the covariation between cell cycle protein markers and thousands of proteins. Based on this covariation, we identify cell cycle associated proteins and functions that are shared across cell types and those that differ between cell types.

Introduction

Single-cell measurements are essential for understanding biological systems composed of different cell types[1,2]. Recent advances in single-cell RNA[3] and protein[4,5] analysis methods have facilitated the study of single-cell heterogeneity at unprecedented scale and depth. These emerging single-cell methods have the potential to go beyond classifying cell types, enabling the characterization of intrinsically single-cell processes[6], such as the cell division cycle (CDC). Indeed, single-cell analysis of the CDC obviates the need to synchronize cell populations and the associated perturbation artifacts[7–11]. Such analysis offers the possibility to explore, the coordination of the cell cycle

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

with metabolism and cell growth that had been observed in synchronized cell cultures[12–14]. Crucial aspects of the CDC are regulated post-transcriptionally by protein synthesis and degradation,[9,15,16] and their characterization demands single-cell protein analysis[6].

Traditionally, single-cell proteomic analyses have been performed by using fluorescent proteins or affinity reagents[17]. While these approaches are powerful[9,10], mass spectrometry (MS) has the potential to increase the specificity and depth of single-cell protein quantification[1,5,6,18]. For decades, MS has been a powerful tool for measuring the relative abundance of thousands of proteins in bulk samples consisting of thousands of cells or more[2,19–21]. Advances in sample preparation, mass-spectrometry data acquisition, and data analysis have enabled quantifying thousands of proteins in single cells as reviewed by ref.[4,5].

Bulk samples are often prepared for liquid chromatography tandem MS analysis by using relatively large volumes (hundreds of microliters) and chemicals (detergents or chaotropic agents like urea) that are incompatible with MS analysis and require removal by cleanup procedures. The large volumes and cleanup procedures entail sample losses that may be prohibitive for small samples, such as single mammalian cells[1,5,17]. Thus, numerous methods have been developed for preparing sub-microgram protein samples[22–25] and single-cell samples[26–35]. To enable some degree of parallel processing, some methods have been automated using multiwell plates[32,35,36]. Preparing small samples often uses sophisticated custom-made equipment[27–31]. While cleanup may result in significant sample losses, some methods, such as SP3[22,23] and iST[24], perform cleanup very efficiently even with submicrogram samples and might be used for single-cell analysis albeit such applications are not extensively tested.

We sought to develop a miniaturized and massively parallel sample preparation method to improve single-cell protein quantification and to use it for cell cycle analysis. Specifically, we aimed to develop a widely accessible, robust, and automated method that further reduces sample preparation volumes to a few nanoliters. Our goal was to carry out parallel sample preparation of thousands of single cells to increase the size of experimental batches and thus reduce batch effects[37–39]. To achieve high precision, we aimed to avoid any movement the samples during the sample preparation stage so that we could repeatedly dispense tiny volumes of reagents to each droplet containing a single cell. To achieve these goals, we used the the CellenONE cell sorting

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

and liquid handling system and developed nano-PrOteomic sample Preparation (nPOP), which allowed a 100-fold reduction of the sample volumes over the Minimal ProteOmic sample Preparation (mPOP) method[35,36,40,41]. Reducing sample volumes directly enabled more comprehensive CDC analysis than was previously possible with mPOP[35]. Integrating nPOP with the Single Cell ProtEomics by Mass Spectrometry (SCoPE2) workflow allowed us to classify cells based on the CDC phase. Furthermore, joint analysis of CDC proteins in U-937 and HeLa cells showed similarities and differences between the cell cycle progression of these cell types.

Results

A challenge for any small volume sample preparation method is choosing a suitable method for cell lysis. As demonstrated previously, we sought to obviate clean-up by avoiding MS-incompatible chemicals[35,41]. In carrying out the lysis on an open surface, we needed to control for several additional factors. First, the precision dispensing of the CellenONE relies on returning to the same XY coordinates. Since moving the surface reduces the accuracy of the dispensing reagents to the single-cell samples, we design nPOP without any slide movements. This required the cell lysis to be performed at room temperature on open surface. To satisfy these conditions, we sought to use 90 % dimethyl-sulfoxide (DMSO) for cell lysis. The low vapor pressure of DMSO at room temperature allows for precise control of evaporation. To test the efficacy of DMSO for cell lysis, we conducted bulk experiments to compare DMSO lysis to the more standard 6M urea lysis.

Using DMSO for Cell Lysis

We compared the efficiency of extracting proteins between DMSO and urea using SILAC quantification as shown in Fig. 1a. Equal amounts of light and heavy U-937 cells were lysed with Urea or DMSO. The samples were then diluted and combined for digestion. These results suggest that DMSO allows for efficient cell lysis without detectable biases against proteins originating from different cell compartments, Fig. 1a. .

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

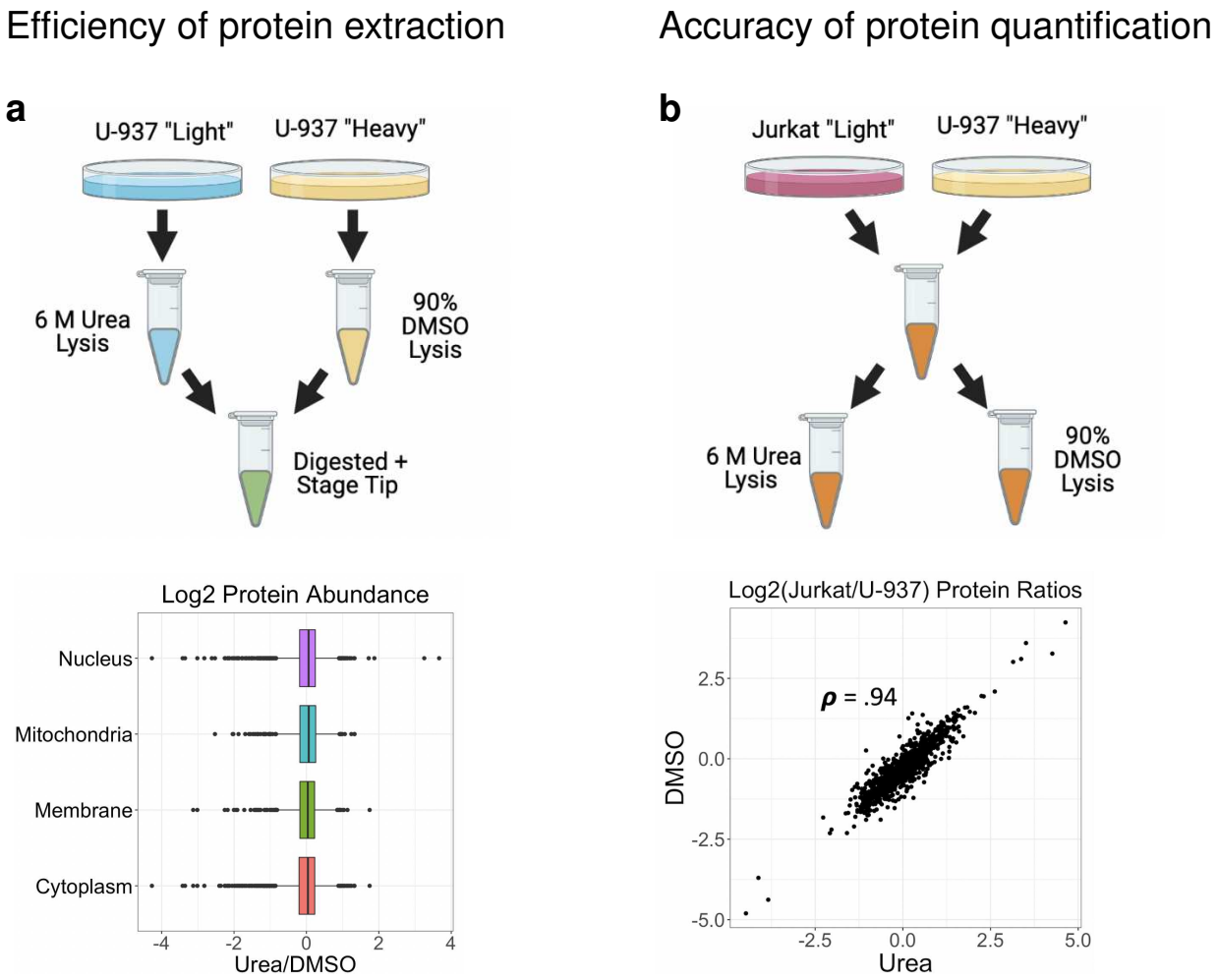

Figure 1 | Evaluating the efficiency of protein extraction by DMSO cell lysis. (**a**) Equal number of U-937 cells labeled with "Light" and "Heavy" isotopes via SILAC were lysed with urea or DMSO, diluted, and combined for digestion. The SILAC ratios for proteins from different cellular compartments show comparable protein recovery for DMSO and urea cell lysis. (**b**) Equal number of SILAC labeled "Light" Jurkat and "Heavy" U-937 cells were combined, and the mixed sample was then divided for cell lysis either by urea or or by DMSO. The agreement between the SILAC ratios from the two methods supports the use of DMSO lysis for quantitative protein analysis.

Next, we evaluated the accuracy of relative protein quantification with DMSO lysis. We lysed U-937 monocytes and Jurkat T-cells with both DMSO and urea and compared the protein ratios estimated from the cells lysed with DMSO and with urea, Fig. 1c. U-937 monocytes cultured in heavy SILAC media and Jurkat T-Cells cultured in standard media were combined in equal amounts and lysed with either 90 % DMSO or 6M urea as shown in Fig. 1c. The correlation between the protein ratios estimated for each condition suggests that DMSO lysis is compatible with accurate protein quantification, Fig. 1d. This gave us further confidence that DMSO lysis is well suited for miniaturizing sample preparation on an open surface without using MS-incompatible chemicals.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

nPOP Workflow

The workflow for nPOP sample preparation includes cell isolation, cell lysis, protein digestion, peptide labeling, and sample pooling as illustrated in Fig. 2a. Sample preparation starts with dispensing droplets of 4 nl DMSO for cell lysis. The droplets are organized as regular grids (Fig. 2b) to facilitate their pooling at the end of the experiment. The second step of nPOP is the isolation and dispensing of single cells into the DMSO droplets. Each single cell is isolated in a 0.3 nl droplet of PBS and added to a DMSO droplet for lysis, Fig. 2a. After 20 minutes for cell lysis, a perimeter of 12 nl droplets of water (for maintaining high local humidity) is deposited around the four sample arrays. The next nPOP step is protein digestion into peptides by the addition of trypsin with HEPES buffer. The addition brings the total volume to 13.5 nl, Fig. 2a. Samples are digested by 75 ng/μl trypsin for 5 hours. To further control evaporation, nPOP uses a humidifier to keep relative humidity inside the CellenONE at 75 %. During digestion, the temperature of the slide is dynamically controlled to 1 ^oC above the dew point, around 17 ^oC. After digestion, humidity is reduced, and the slide is brought to room temperature for labeling. The single cell droplets dry down on the slide to volumes of approximately 4 nl before labeling. TMT labels dissolved in DMSO are dispensed in volumes of 20 nl to the single cell droplets. Dissolving labels in DMSO is a distinctive aspect of nPOP that facilitates the manipulation of sub-nanoliter droplets of TMT solution. The most commonly used solvent for TMT, acetonitrile, is difficult to handle with CellenONE due to its density, volatility, and low surface tension. After samples are labeled for one hour at room temperature, labeling is quenched with two sequential 20 nl additions of 5 % hydroxylamine. Each addition is followed by a 20-minute incubation.

To pool all single-cell samples into a set, 1 μl of water is dispensed by a pipette onto each array of labelled samples. Samples are then pipetted directly into injection vial glass inserts containing isobaric carrier and reference previously prepared using the mPOP protocol[35]. To improve the recovery of labeled peptides, the footprint of each array is washed by 4 μl of acetonitrile, which is collected and added to the corresponding combined set. This wash is option and used to maximize the recovery of labeled peptides from the slide.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

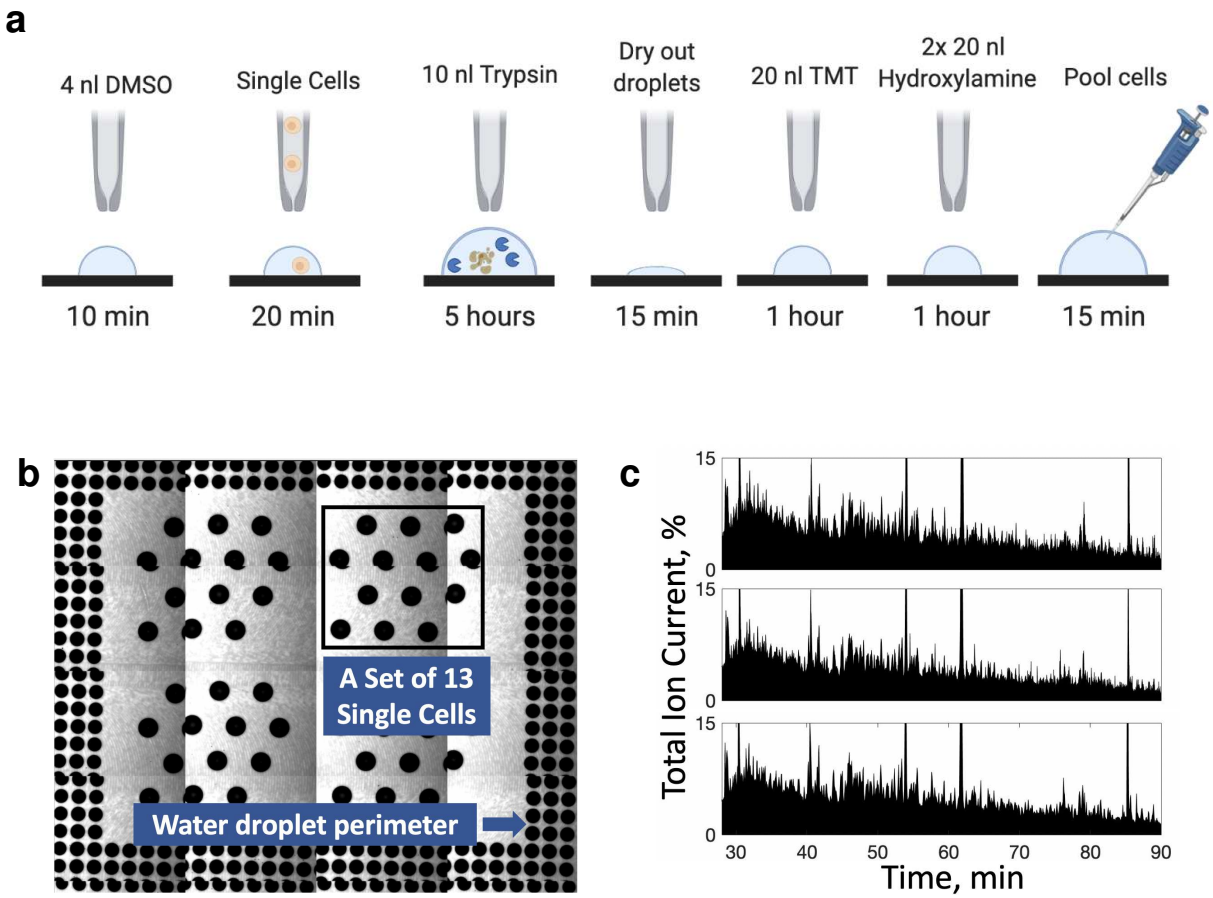

Figure 2 | Workflow of nano-PrOteomic sample Preperation (nPOP) (**a**) A schematic of the nPOP sample preparation method illustrates the steps of cell lysis, protein digestion, peptide labeling with TMT, and quenching with two additions of hydroxylamine. These steps are performed in parallel for all single cells and take place in small droplets. (**b**) A representative field of droplets post trypsin addition. Droplets with single cells are clustered in groups of 13, the number of cells labeled and combined into one SCoPE2 sets using TMTpro. The single-cell droplets are surrounded by a perimeter of water droplets for maintaining high local humidity. (**c**) Total ion current chromatograms from three runs demonstrate low contaminants and consistent chromatography.

Single-cell protein analysis with nPOP

nPOP is a general sample preparation method that can be used for either label-free MS analysis or multiplexed MS analysis as part of existing workflows reviewed by ref.[4,5]. Here, we demonstrate sample preparation by nPOP as part of the SCoPE2 protocol[36,40]. Specifically, we replaced Minimal ProteOmic sample Preparation (mPOP) module[35] with nPOP and used all other modules of the SCoPE2 workflow, including an isobaric carrier[42], Data-Driven Optimization of Mass Spectrometry (DO-MS)[43], Data-driven Alignment of Retention Times for IDentification (DART-ID)[44], and the SCoPE2 data analysis pipeline[40,45,46].

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

To evaluate the performance of nPOP for single-cell analysis, we applied it to two cell types, Hela cells and U-937 monocytes. The sample preparation was performed on two different days so that the data may reflect day-specific batch effects. The resulting SCoPE2 sets were run using roughly five days of instrument time, and the MS data were processed and evaluated via the SCoPE2 pipeline[36,45]. As a first benchmark for contaminant signal, we evaluated the RI intensity in negative controls. The negative controls correspond to droplets that did not receive single cells, and their RI intensities reflect cross-labeling and nonspecific background noise[36,40]. The RI intensities in two representative negative controls shown in Fig. 3a,b are mostly absent or very low, indicating that background noise is low for samples prepared with nPOP. The RI intensities for single cells also show that, as expected and previously observed[36], peptides from Hela cells are more abundant than peptides from U937 cells, likely reflecting the different cell sizes.

To further evaluate the single-cell data, the SCoPE2 pipeline calculates the coefficient of variation (CV) of relative levels for all peptides belonging to the same protein. The relatively low CV values indicate that protein quantification derived from different peptides is internally consistent, Fig. 3c. Furthermore, the small spread of the distribution for the median CVs indicates that each cell is treated consistently by the automated sample preparation technique.

Next, we performed principal component analysis (PCA) of the single-cell protein dataset using all quantified proteins, Fig. 3c. The PCA indicates two distinct clusters of cells. The clusters correspond the cell types and separate along the first principal component (PC1), which accounts for 66 % of the variance.To further validate the single cell quantification of proteins in single cell samples, we took the ratio of average proteins expression in the Hela cells and Monocytes and compared them to the ratios observed in bulk samples. Similar to previous analysis[36,40,47,48], the protein ratios in bulk samples agreed well with ratios in the single cells single cells. A Pearson correlation of 0.8 indicates that the single-cell protein quantification is consistent with the proteomic measurements of established bulk methods.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

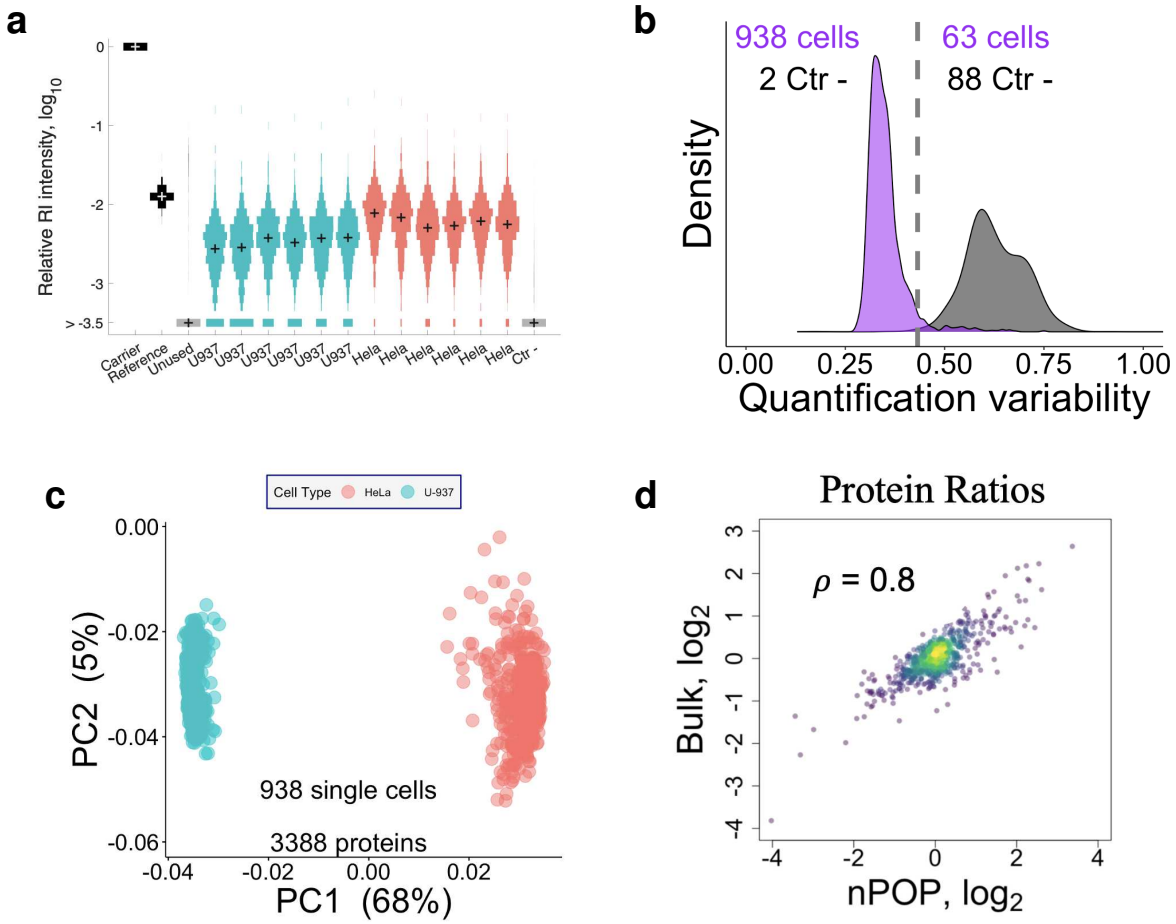

Figure 3 | Reporter ion intensities in single cells and in negative controls. The reporter ion intensities for a representative SCoPE2 sets prepared with nPOP are shown in (**a**). The panels show distributions of reporter ion intensities relative to the corresponding isobaric carrier for the set. RI intensities are mostly absent from negative control wells, which contains all reagents but not a single cell. (**b**) The consistency of protein quantification is estimated as the coefficient of variation (CV) of the relative levels of peptides originating from the same protein. The median CVs per cells are tightly distributed, suggesting high consistency of sample preparation. (**c**) Principal component analysis separates single-cell samples corresponding to HeLa cells or to U937 monocytes. (**d**) The peptides measurements in single cells within a cell type are averaged and correlated to samples of 200 cells prepared in bulk show that the relative protein quantification is consistent between single-cell and bulk samples.

Cell Cycle Analysis

We next sought to identify biological processes with concerted CDC dynamics that are common and those that are different between the cell division cycles of an epithelial cell line (HeLa) and a monocytic cell line (U-937). As a first step towards this analysis, we evaluated the potential to classify individual cells by their cell cycle phase. We intersected the list of confidently identified proteins in our data with a list of genes previously found to be CDC periodic[49]. We then pro-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

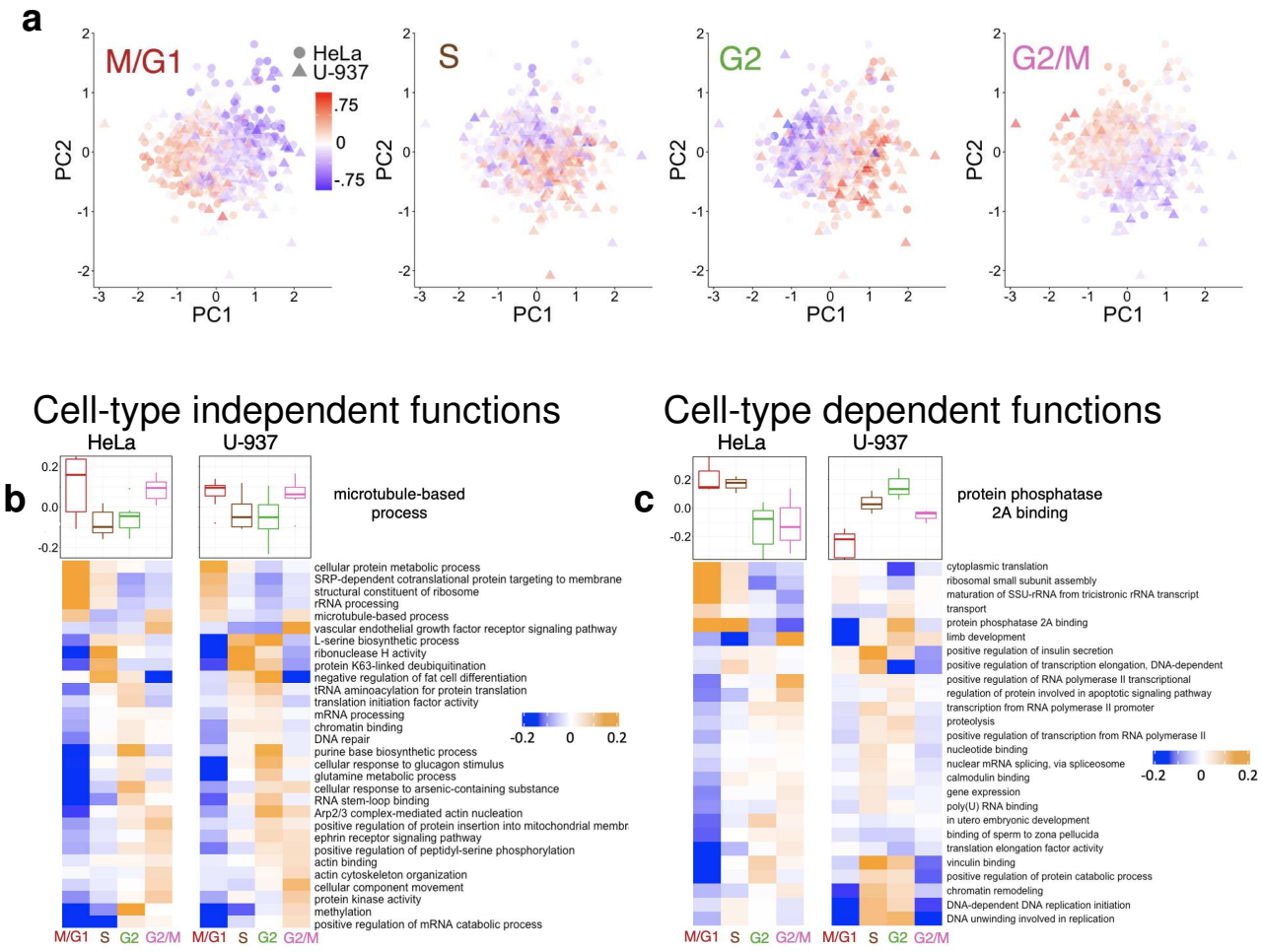

Figure 4 | Covariation of proteins and biological processes with the cell cycle (**a**) Common Principal Component Analysis (CPCA) of HeLa and U-937 cells in the space of CDC periodic genes. Cells in each CPCA plot are colored by the mean abundance of proteins annotated to M/G1, S, and G2 phases based on previous analysis[49]. (**b**) The boxplots display distributions for correlations between the phase markers and proteins from the large ribosomal subunit assembly GO term. The difference between these distributions was evaluated by 1-way ANOVA analysis to estimate statistical significance, $p < 0.0002$. The distributions for other GO terms that covary in a similar way between the two cell lines are summarized with their medians plotted as a heatmap. (**c**) Similar analysis and display as in panel a was used to visualize GO terms whose covariation with the CDC is cell type specific. Shown GO terms are at FDR $< 5\%$.

jected the proteomes of both Hela and U937 cells into a joint 2-dimensional space (Fig. 4) defined by common principal component analysis (CPCA); see methods. The CPCA was performed as previously described[40,50], and each cell was color-coded based on the mean abundance of protein markers for M/G1, S, and G2 phase. The cells from both cell types cluster by CDC phase (Fig. 4a), which suggests that the data capture CDC related protein dynamics.

Next we focused on identifying proteins that covary with the CDC periodic proteins used for defining CDC phases in Fig. 4a. To identify such covariation, we first estimated a marker vector

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

for each cell cycle phase as the median level of marker proteins for that phase. The phase marker vectors were correlated to the measured protein abundances of all proteins quantified across many single cells. Many proteins correlate strongly to the phase marker vectors, and for 34 of these proteins the correlations are statistically significant, sugesting that these proteins are CDC periodic.

To increase the statistical power of our covariation analysis, we next focused on the covariation of phase markers and proteins with similar functions as defined by the gene ontology (GO). Specifically, we compared the distributions of correlations between the 3 phase marker vectors and all quantified proteins from a GO term, as shown with the boxplots in Fig. 4b. For microtubule-based process, the distributions of correlations differ significantly between the CDC phases, and this phase-specific covariation is similar for the two cell types, Fig. 4b. Many other GO terms show covariation to the phase markers that is similar for the two cell types, and instead of displaying the boxplot distributions for all of them, we summarized the distributions of correlations with their medians displayed as a heatmap, Fig. 4b. Such functions with shared covariation include DNA repair, protein kinase activity, and cellular protein metabolic processes.

We performed similar protein set enrichment analysis on the correlations between proteins and phase markers to identify biological functions that covary with the cell cycle phases in a different way for the two cell types, Fig. 4c. Such functional groups of proteins that differed in CDC covariation include protein phosphate 2A binding, proteolysis, and protein transport, Fig. 4c.

Discussion

Existing single-cell omics methods excel at classifying cells by cell type. However, the regulatory dynamics resulting in cell to cell variability within a cell type are more challenging to analyze. To support such analysis, we introduce a highly parallel sample preparation that allows us to prepare hundreds to thousands of single cells in a given experiment. It allows for reduced volumes and increased consistency of single-cell proteomic sample preparation. Furthermore, it can enable processing thousands of single cells in parallel and thus empower high-throughput high-power biological analysis[51].

To maximize access and flexibility, nPOP used only commercially available equipment and

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

prepares single cells on an open surface that can be pragmatically reconfigured and adopted to different experimental designs. The open environment also obviates all sample movements and maximizing the consistency and precision of the sample preparation. The open layout using a hydrophobic slide can scale up to simultaneously preparing thousands of single cells. Furthermore, nPOP is amenable to different coatings or hydrophobic surfaces which have the potential to further improve recovery. While nPOP performed very well in our experiments, we have not tested it with samples contaminated by chemicals undermining MS analysis. Such samples are likely to be better handled by SP3[22,23] and iST[24].

nPOP allowed for deeper single cell proteomic analysis of the cell division cycle than our previous sample preparation method, mPOP[35]. The data allowed us to identify new proteins and functional groups of proteins associated with the cell cycle without the artifacts associated with synchronizing cell cultures[11]. These initial results demonstrate the feasibility of inferring co-regulation of biological processes from single-cell proteomics measurements.

Supplemental website Data and other resources can be found at: scope2.slavovlab.net/nPOP

Acknowledgments: We thank A. Chen for early experiments with the organic solvent lysis, A. Murphy for assistance with using CellenONE, and H. Specht for discussions and constructive comments. This work was funded by a New Innovator Award from the NIGMS from the National Institutes of Health to N.S. under Award Number DP2GM123497, an Allen Distinguished Investigator award through the Paul G. Allen Frontiers Group to N.S., a Seed Networks Award from CZI CZF2019-002424 to N.S., through a Merck Exploratory Science Center Fellowship, Merck Sharpe & Dohme Corp. to N.S. Funding bodies had no role in data collection, analysis, and interpretation.

Competing Interests: The authors declare that they have no competing financial interests.

Correspondence: Correspondence and materials requests should be addressed to nslavov@alum.mit.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

Author Contributions

Experimental design: A.L., G.H. and N.S.

Single-cell LC-MS/MS: A.L., G.H.

Sample preparation: A.L

Raising funding & supervision: N.S.

Data analysis: A.L. and N.S.

Writing & editing: A.L. and N.S.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

Methods

Cell Culture Hela cells were grown as adherent cultures in DMEM with high glucose (Sigma-Aldrich D5796), supplemented with 10 fetal bovine serum (FBS, Millipore Sigma F4135) and 1 penicillin-streptomycin (pen/strep, ThermoFisher 15140122). U-937 and Jurkat cells were grown as suspension cultures in RPMI medium (HyClone 16777-145) supplemented with 10 fetal bovine serum (FBS, Millipore Sigma F4135) and 1 penicillin-streptomycin (pen/strep, ThermoFisher 15140122). Cells were passaged when a density of 106 cells/ml was reached, approximately every two days.

Mass spectrometry analysis MS analysis was designed and performed according to the SCoPE2 guidelines and protocol[36,40,42]. Specifically, the single cells pooled into SCoPE2 sets were separated via online nLC on a Dionex UltiMate 3000 UHPLC; 1 μl out of 1.2 μl of sample was loaded onto a 25cm x 75 μl IonOpticks Aurora Series UHPLC column (AUR2-25075C18A). Buffer A was 0.1 % formic acid in water and buffer B was 0.1% formic acid in 80 acetonitrile / 20% water. A constant flow rate of 200nl/min was used throughout sample loading and separation. Samples were loaded onto the column for 20 minutes at 1% B buffer, then ramped to 5 B buffer over two minutes. The active gradient then ramped from 5% B buffer to 25% B buffer over 53 minutes. The gradient then ramped to 95% B buffer over 2 minutes and stayed at that level for 3 minutes. The gradient then dropped to 1% B buffer over 0.1 minutes and stayed at that level for 4.9 minutes. Loading and separating each sample took 95 minutes total. All samples were analyzed by a Thermo Scientific Q-Exactive mass spectrometer from minute 20 to 95 of the LC loading and separation process. Electrospray voltage was set to 1.8 V, applied at the end of the analytical column. To reduce atmospheric background ions and enhance the peptide signal-to-noise ratio, an Active Background Ion Reduction Device (ABIRD, by ESI Source Solutons, LLC, Woburn MA, USA) was used at the nanospray interface. The temperature of ion transfer tube was 250 ^oC and the S-lens RF level was set to 80. After a precursor scan from 450 to 1600 m/z at 70,000 resolving power, the top 7 most intense precursor ions with charges 2 to 4 and above the AGC min threshold of 20,000 were isolated for MS2 analysis via a 0.7 Th isolation window with a 0.3 Th offset. These ions were accumulated for at most 300ms before being fragmented via HCD at a normalized collision energy of 33 eV (normalized to m/z 500, z=1). The fragments were analyzed by an MS2 scan

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

with 70,000 resolution. Dynamic exclusion was used with a duration of 30 seconds with a mass tolerance of 10ppm.

DMSO and urea lysis sample preparation Jurkat cells and U-937 cells cultured in heavy SILAC media (containing $+10Da$ Arg and $+8Da$ Lys) were washed and re-suspended in PBS at 20,000 cells per μl. Two solutions of equal cell count containing Jurkat and U-937 cells were made mixed in 1:1 ratios. One sample was lysed by diluting cells in 90 % DMSO and the other was lysed in 6M urea. The DMSO cell lysate was diluted to a concentration of 33 % DMSO and urea lysate was diluted to 0.5 M. Both solutions were digested in 15 ng/μl of trypsin for 12 hours. Each sample was then desalted using C18 stage tips and run using data dependant acquisition.

Sample layout and Experimental Design nPOP reactions are carried out on the surface of a fluorocarbon coated glass slide. The layout is very flexible and adjustable to the experimental parameters. The droplets used for single-cell sample preparation are arranged in clusters, and the number of droplets per cluster equals the number of single cells per SCoPE2 set. Here we used TMTpro 16plex and 12 droplets per cluster, corresponding to the 12 isobaric labels used for single cells. Our design allowed fitting 36 clusters per slide and 4 glass slides, which results in simultaneously processing up to $12 \times 36 \times 4 = 1,728$ single cells. With TMTpro 18plex, this design can simultaneously process up to $14 \times 36 \times 4 = 2,016$ single cells. Reducing the space between the clusters can further increase the number of clusters per slide and thus the number of simultaneously prepared single cells. The droplet geometry was optimized to keep droplets from the same set close in proximity but prevent reaction volumes from merging. Negative control droplets in this context are defined as droplets that experience all sample preparation steps, except that no single cell is dispensed to these droplets. This layout of droplets is flexible and can be easily adopted to different configurations or larger array sizes.

Reagent Handling with CellenONE The CellenONE is set up with two nozzles. One nozzle handles cell suspensions. The other tip handles organic solvents and protein solutions. Reagents are loaded into a 384 well plate in volumes of 30 μl. When aspirating protein solutions, make sure to aspirate 20 μl to ensure the mixture is not diluted with system water. When dispensing DMSO, it is important to shut off humidifyer. This allows leftover DMSO on the tip to evaporate quickly so dispensing is not effected. After each sample preparation, tips are washed methanol and cleaned

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

under sonication to remove built up material from inside of tip and ensure optimal performance.

Carrier and reference channel preparation in bulk The isobaric carrier consisting of a 1:1 mixture of Hela and monocyte cells was prepared in bulk and aliquoted into carriers corresponding to 200 cells each. A single cell suspension of about 22,000 cells was transferred to a 200 μl PCR tube (USA Scientific 1402-3900) and then processed via the mPOP sample preparation method[35]. The reference channel was made from the same sample.

nPOP Sample Preparation To lyse cells, cells are dispensed into pools of DMSO. 4 nl droplets of DMSO are dispensed over the grid in the location of each reaction volume. Single cells are then dispensed in each each reaction volume. Cells are incubated in the 4 nL DMSO droplets for 20 minutes. After lysis, a solution containing trypsin and HEPES buffer is added to each reaction volume, for a final concentration of 100 ng/μl of trypsin and 1 mM HEPES and total volume of 14 nl.

The humidifier and cooling system is then turned on to prevent droplet evaporation. Relative humidity inside the CellenONE is set to 72 % and the water chiller temperature is set to dynamically chase one degree above the dew point. Mass spectrometry grade water is dispensed in a perimeter surrounding each grid to provide further control for the local humidity of the reaction volumes. The system is set to refresh the water droplet perimeter to control local humidity every 20 minutes for 5 hours as proteins digest.

After proteins have digested for 5 hours, the humidity and cooling controls are turned off. 20 nL of TMT labels suspended in DMSO and concentrated at 28mM are then dispensed to each reaction volume using the organic dispensing tip. When dispensing labels, humidifier was turned off to assist with dispensing. After single cells are left to label for 1 hour, 20 nL of 5 % hydroxylamine solution is added to each reaction volume to quench labeling reaction. Humidity and cooling controls are returned to previous settings for quenching labeling reaction. After 20 minutes, another addition of 30 nL of 5 % hydroxylamine is added.

After quenching proceeds for another 20 minutes, samples are pooled on plate in 2.5 μl of a 50 % 0.1 % formic/acetonitrile solution added via hand pipette. Samples are collected and added to a single glass HPLC insert (ThermoFisher C4010-630) and dried down to dryness in a speed-vacuum (Eppendorf, Germany) and either frozen at $-80\ ^oC$ for later analysis or immediately reconstituted

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

in 1.2 μl of 0.1 % formic acid (ThermoFisher 85178) for mass spectrometry analysis.

Analysis of raw MS data Raw data were searched by MaxQuant[20,52] 1.6.0.16 and 1.6.2.3 against a protein sequence database including all entries from the appropriate mouse or human SwissProt database (downloaded July 15, 2018 and July 30, 2018, respectively) and known contaminants such as human keratins and common lab contaminants. MaxQuant searches were performed using the standard work flow[53]. We specified trypsin specificity and allowed for up to two missed cleavages for peptides having from 5 to 26 amino acids. Methionine oxidation (+15.99492 Da) and protein N-terminal acetylation (+42.01056 Da) were set as a variable modifications. Carbamidomethylation was disabled as a fixed modification. All peptide-spectrum-matches (PSMs) and peptides found by MaxQuant were exported in the msms.txt and the evidence.txt files. SILAC data was searched in two batches (by date acquired) with match between runs enabled, using the default settings.

Single-cell filtering and normalization The single-cell data were processed and normalized by the SCoPE2 pipeline[36,40]. This pipeline is also implemented by the scp Bioconductor package[39,46]. Briefly, single cells with suboptimal quantification were removed prior to data normalization and analysis based on objective criteria: The internal consistency of protein quantification for each single cell was evaluated by calculating the coefficient of variation (CV) for proteins (leading razor proteins) identified with over 5 peptides for that cell. The coefficient of variation is defined as the standard deviation divided by the mean. The CVs were computed for the relative reporter ion intensities, i.e., the RI reporter ion intensities of each peptide were divided by their mean resulting in a vector of fold changes relative to the mean. Cells that fell outside the sharp CV distribution centered around 0.27 were removed from analysis with a threshold of 0.35. Data was normalized as by procedure outlined by Specht et al.[40,45].

Principal component analysis for single cell data sets From the protein x single cell matrix, all pairwise protein correlations (Pearson) were computed. Thus, for each protein, there was a computed vector of correlations with a length the same as the number of rows in the matrix (number of proteins). The dot product of this vector with itself was used to weight each protein prior to principal component analysis. The principal component analysis was performed on the correlation matrix of the weighted data.

SILAC data analysis When comparing relative protein levels in Jurkat and U-937 cells, SILAC

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

ratios for peptides were computed by taking dividing each channel by its median, and then taking the ratio of the light and heavy channels. When comparing absolute abundances between heavy and light U-937 cells to measure efficiency of extraction, label swap experiments were ran so that both lysis conditions were measured with both heavy and light labels. The raw intensities for corresponding lysis methods were averaged and the ratio between different lysis methods was plotted.

Common Principal Component Analysis To jointly analyze the cell cycle protein from HeLa and U-937 cells, we performed Common Principal Component Analysis (CPCA) in the space of 20 cell cycle dependant(CDC) proteins using the Krzanowski method[50]. Specifically, we computed the correlation matrices of CDC proteins in the U-937, $\mathbf{R}_u$, and in the HeLa cells, $\mathbf{R}_h$, and determined the eigenvector with the largest and second largest eigenvalue of the matrix $\mathbf{R}_u + \mathbf{R}_h$. Each matrix of proteins by cells was then multiplied by the common principal components to obtain two vectors of length number of cells. These vectors were then plotted against each other to project cells into a joint space as seen in Fig. 4a.

Identifying proteins that covary with CDC markers To identify proteins that covary with the phase marker vectors, we correlated the phase marker vectors to the measured protein levels. To minimize cell-type specific effects, we converted the levels of each proteins within a cell type to z-scores and then combined with the corresponding vector of z-scores from the other cell type. These combined vectors were correlated to the combined phase marker vectors using Spearman correlation.

Protein Set Enrichment Analysis To identify functionally coherent sets of proteins that covary with the CDC phase markers, we correlated each protein to the median abundance of CDC proteins that showed similarity between HeLa and U-937 cells as plotted in Fig. 4a. The resulting correlation vectors were analyzed by protein set enrichment analysis similar to previously reported analysis[54]. In the case of cell-type specific co-variation, we also used empirical bootstrapping to estimate the Z-score corresponding to each correlation, and then compared the distributions of Z-scores via ANOVA for estimating the statistical significance. Only GO Terms for which we had least 4 proteins were analyzed. We used ANOVA to estimate if the variance among the correlations of the proteins from the GO term and the CDC phase markers can be explained by the CDC.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

We then used the Benjamini-Hochberg method to estimate the corresponding q values (FDR; false discovery rare) for each GO term. Among the set of GO terms within 5 % FDR, we displayed in Fig. 4 the 20 GO terms whose correlations to the CDC phase markers was most similar or most different between the 2 cell lines.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

References

1. Specht, H. & Slavov, N. Transformative opportunities for single-cell proteomics. *Journal of Proteome Research* **17,** 2563–2916 (8 June 2018).
2. Rodriguez, H., Zenklusen, J. C., Staudt, L. M., Doroshow, J. H. & Lowy, D. R. The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. *Cell,* 1661–1670 (2021/04/03).
3. Ziegenhain, C. *et al.* Comparative analysis of single-cell RNA sequencing methods. *Molecular cell* **65,** 631–643 (2017).
4. Slavov, N. Single-cell protein analysis by mass spectrometry. *Current Opinion in Chemical Biology* **60,** 1–9. ISSN: 1367-5931 (2020).
5. Kelly, R. T. Single-Cell Proteomics: Progress and Prospects. *Molecular & Cellular Proteomics* **19,** 1739–1748 (2020).
6. Slavov, N. Unpicking the proteome in single cells. *Science* **367,** 512–513 (2020).
7. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews cancer* **9,** 153–166 (2009).
8. Silverman, S. *et al.* Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. *Proceedings of the National Academy of Sciences* **107,** 6946–6951 (2010).
9. Mahdessian, D. *et al.* Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* **590,** 649–654 (2021).
10. Stallaert, W. *et al.* The structure of the human cell cycle. *bioRxiv* (2021).
11. Cooper, S. The synchronization manifesto: a critique of whole-culture synchronization. *The FEBS Journal* **286,** 4650–4656 (2019).
12. Slavov, N. & Botstein, D. Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular Biology of the Cell* **22,** 1997–2009 (2011).
13. Slavov, N., Macinskas, J., Caudy, A. & Botstein, D. Metabolic cycling without cell division cycling in respiring yeast. *Proceedings of the National Academy of Sciences* **108,** 19090–19095. ISSN: 1091-6490 (Nov. 2011).
14. Aguilar, V. & Fajas, L. Cycling through metabolism. *EMBO molecular medicine* **2,** 338–348 (2010).
15. Murray, A. W. & Kirschner, M. W. Cyclin synthesis drives the early embryonic cell cycle. *Nature* **339,** 275–280 (1989).
16. Glotzer, M., Murray, A. W. & Kirschner, M. W. Cyclin is degraded by the ubiquitin pathway. *Nature* **349,** 132–138 (1991).
17. Levy, E. & Slavov, N. Single cell protein analysis for systems biology. *Essays In Biochemistry* **62,** 595–605 (4 2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

18. Slavov, N. Driving Single Cell Proteomics Forward with Innovation. *Journal of Proteome Research,* doi: 10.1021/acs.jproteome.1c00639m. https://doi.org/10.1021/acs.jproteome.1c00639m (2021).

19. Cravatt, B. F., Simon, G. M. & Yates Iii, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450,** 991 (2007).

20. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of proteome research* **10,** 1794–1805 (2011).

21. Scott, N. E. & Edgington-Mitchell, L. E. Systems biology and the rise and rise of omics approaches. *Current Opinion in Chemical Biology* **60.** Omics, A1–A3. ISSN: 1367-5931 (2021).

22. Hughes, C. S. *et al.* Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* **10,** 757 (2014).

23. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols,* 1 (2018).

24. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature methods* **11,** 319 (2014).

25. Kostas, J. C., Greguš, M., Schejbal, J., Ray, S. & Ivanov, A. R. Simple and Efficient Microsolid-Phase Extraction Tip-Based Sample Preparation Workflow to Enable Sensitive Proteomic Profiling of Limited Samples (200 to 10,000 Cells). *Journal of Proteome Research* **20,** 1676–1688 (2021).

26. Mellors, J. S., Jorabchi, K., Smith, L. M. & Ramsey, J. M. Integrated microfluidic device for automated single cell analysis using electrophoretic separation and electrospray ionization mass spectrometry. *Analytical chemistry* **82,** 967–973 (2010).

27. Lombard-Banek, C., Moody, S. A. & Nemes, P. Single-Cell Mass Spectrometry for Discovery Proteomics: Quantifying Translational Cell Heterogeneity in the 16-Cell Frog (Xenopus) Embryo. *Angewandte Chemie International Edition* **55,** 2454–2458 (2016).

28. Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature communications* **9,** 882 (2018).

29. Li, Z.-Y. *et al.* Nanoliter-scale oil-air-droplet chip-based single cell proteomic analysis. *Analytical chemistry* **90,** 5430–5438 (2018).

30. Woo, J. *et al.* High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *bioRxiv* (2021).

31. Lombard-Banek, C. *et al.* In Vivo Subcellular Mass Spectrometry Enables Proteo-Metabolomic Single-cell Systems Biology in a Chordate Embryo Developing to a Normally Behaving Tadpole (X. laevis). *Angewandte Chemie International Edition* (2021).

32. Liang, Y. *et al.* Fully Automated Sample Processing and Analysis Workflow for Low-Input Proteome Profiling. en. *Anal. Chem.* **93,** 1658–1666 (Jan. 2021).

33. Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv* (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

34. Hartlmayr, D. *et al.* An automated workflow for label-free and multiplexed single cell proteomics sample preparation at unprecedented sensitivity. *bioRxiv* (2021).

35. Specht, H. *et al.* Automated sample preparation for high-throughput single-cell proteomics. *bioRxiv 10.1101/399774*. https://doi.org/10.1101/399774 (2018).

36. Petelski, A. A. *et al.* Multiplexed single-cell proteomics using SCoPE2. *Nature Protocols,* doi: 10.1038/s41596-021-00616–z. https://doi.org/10.1038/s41596-021-00616-z (2021).

37. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161,** 1187–1201 (2015).

38. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).

39. Vanderaa, C. & Gatto, L. Utilizing Scp for the analysis and replication of single-cell proteomics data. *bioRxiv* (2021).

40. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biology* **22** (2021).

41. Marx, V. A dream of single-cell proteomics. *Nature Methods* **16,** 809–812 (2019).

42. Specht, H. & Slavov, N. Optimizing Accuracy and Depth of Protein Quantification in Experiments Using Isobaric Carriers. *Journal of Proteome Research* **20.** PMID: 33190502, 880–887 (2021).

43. Huffman, G., Chen, A. T., Specht, H. & Slavov, N. DO-MS: Data-Driven Optimization of Mass Spectrometry Methods. *J. of Proteome Res.* **18,** 2493–2500 (6 2019).

44. Chen, A. T., Franks, A. & Slavov, N. DART-ID increases single-cell proteome coverage. *PLOS Computational Biology* **15,** 1–30. https://doi.org/10.1371/journal.pcbi.1007082 (July 2019).

45. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Zenodo,* 10.5281/zenodo.4339954. 10.5281/zenodo.4339954 (2020).

46. Vanderaa, C. & Gatto, L. Mass Spectrometry-Based Single-Cell Proteomics Data Analysis. *Bioconductor,* 10.18129/B9.bioc.scp (2020).

47. Budnik, B., Levy, E., Harmange, G. & Slavov, N. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *bioRxiv 10.1101/102681*. https://doi.org/10.1101/102681 (2017).

48. 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 (2018).

49. Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular biology of the cell* **13,** 1977–2000 (2002).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted September 28, 2021. 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-NC-ND 4.0 International license.

50. Duras, T. A comparison of two estimation methods for common principal components. *Communications in Statistics* **5.** PMID: 33190502, 366–393 (2019).

51. Slavov, N. Increasing proteomics throughput. *Nature Biotechnology* **39,** 809–810. https://doi.org/10.1038/s41587-021-00881-z (2021).

52. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* **26,** 1367–1372 (2008).

53. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature protocols* **11,** 2301 (2016).

54. Franks, A., Airoldi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS computational biology* **13,** e1005535 (2017).