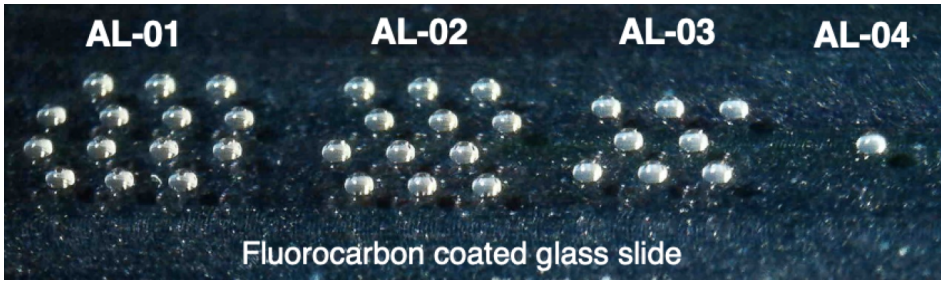
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Exploring functional protein covariation across single cells using nPOP

Andrew Leduc,[1] R. Gray Huffman,[1] Joshua Cantlon,[2] Saad Khan,[1] & Nikolai Slavov[1,✉]

[1]Departments of Bioengineering, Biology, Chemistry and Chemical Biology, Single Cell Proteomics Center, and Barnett Institute, Northeastern University, Boston, MA 02115, USA
[2]Scienion AG, Phoenix, AZ 85042, USA
✉ Correspondence: nslavov@alum.mit.edu or nslavov@northeastern.edu
∈ Data, code & protocols: scp.slavovlab.net/nPOP

Many biological processes, such as the cell division cycle, are reflected in protein covariation across single cells. This covariation can be quantified and interpreted by single-cell proteomics with sufficiently high throughput and accuracy. Toward this goal, we developed the nano-ProteOmic sample Preparation (nPOP) method for single-cell proteomics. 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. We used nPOP to prepare 1,888 single cells and 128 negative controls in a single batch. Their analysis enabled quantifying the covariation between thousands of proteins and cell-cycle protein markers. Many protein sets covaried with the cell cycle similarly across all cell types and states, reflecting cell-type independent cell cycle functions. However, the cell cycle covariation of other protein sets differed markedly between cell types, even within subpopulation of melanoma cells expressing markers for drug-resistance priming. The cells expressing these markers accumulated in the G1 phase of the cell cycle and exhibited different covariation of enzymes catabolizing glucose. These results demonstrate that protein covariation across single cells may reveal functionally concerted biological differences between closely related cell states.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Single-cell measurements are commonly used to identify different cell types from tissues composed of diverse cells[1,2]. This analysis is powering the construction of cell atlases, which can pinpoint cell types affected by various physiological processes. This cell classification requires analysing large number of cells and may tolerate measurement errors[1,3,4].

In addition to classifying cells by type, single-cell measurements may reveal regulatory processes within a cell type and even associate them with different functional outcomes[5–7]. For example, the covariation among proteins across single cells from the same type may reflect cell intrinsic dynamics, such as the cell division cycle[5,8]. Furthermore, protein covariation may reflect protein interactions within complexes or cellular states, such as senescence[5]. However, estimating and interpreting protein covariation within a cell type requires high quantitative accuracy and high throughput[5,9]. Indeed, protein differences within a cell type are smaller than across cell types and can be easily swamped by batch effects and measurement noise.

Therefore, we sought to minimize measurement noise to levels consistent with estimating and interpreting protein covariation across single cells from the same cell type. Towards this goal, we aimed to reduce batch effects and background noise, since these factors undermine the accuracy of single-cell proteomics by mass spectrometry (MS)[10–14]. Specifically, we aimed to develop a widely accessible, robust, and automated sample preparation method that reduces volumes to a few nanoliters. Our goal was to perform parallel sample preparation of thousands of single cells to increase the size of experimental batches and thus reduce batch effects[11,15,16]. To achieve high precision, we aimed to avoid any movement of the samples during the sample preparation stage so that we could repeatedly dispense 1-10nl volumes of reagents to each droplet containing a single cell. To this end, we used the the CellenONE cell sorting 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[13,17–19].

Here we demonstrate that nPOP can prepare over 2,000 single-cell samples in one batch and enables analysis of protein covariation within a cell type, and even within subpopulations of melanoma cells. The protein covariation identified biological processes that covary with the cell

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

division cycle (CDC) in a cell type dependent and independent manner. Furthermore, this analysis indicated differences between the CDC of melanoma cells expressing priming markers for drug resistance.

Results

Sample preparation

To reduce batch effects and background signal, our goal was to maximize the number of single cells prepared in parallel while minimizing the reaction volumes used in sample preparation. To this end, we explored the idea of performing all sample preparation steps in droplets on the surface of a uniform glass slide, Fig. 1. This allows the freedom to arrange single cells in any geometry that best fits the experimental design, Fig. 1c,d. To facilitate this idea, we needed clean reagents, compatible both with analysis by LC-MS and our open surface design. To this end, we introduced the use of 100 % dimethyl sulfoxide (DMSO) as a reagent for cell lysis and protein extraction. Its low vapor pressure enables nanoliter droplets to persist on the surface of the open glass slide. Furthermore, its compatibility with MS analysis obviates sample cleanup and the associated losses and workflow complications. Control experiments indicate that DMSO efficiently delivers proteins to MS analysis without detectable bias for cellular compartments (Fig. S1a-c) and supports accurate relative protein quantification (Fig. S1d).

Having benchmarked DMSO's lytic efficacy, we integrated it into the nPOP workflow by printing an experimenter-defined array of DMSO droplets into which single cells are subsequently dispensed for lysis, Fig. 1a. After lysis, proteins are digested for 4 hours by adding the protease trypsin dissolved in aqueous buffer. To control evaporation throughout digestion, slide temperature and chamber humidity are automatically maintained around set points. Furthermore, we dispense a perimeter of water droplets around the samples, Fig. 1d. See methods for details.

For labeling peptides post-digest, we found that the commonly used approach of dissolving labels in acetonitrile was incompatible with our platform's dispensing method due to the reagent's low density and surface tension. By substituting DMSO for acetonitrile as the label solvent, we

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observed robust dispensing and labeling performance for picoliter dispense volumes over hundreds of samples. We validated this approach by measuring labeling efficiency in pooled samples and found that over 99 % of all detected peptides were TMT labeled, see methods.

Finally, nPOP pools clusters of labeled single cells that comprise a set into a single droplet, aspirates it, and dispenses it into a 384 well plate. All of these steps are fully automated. The 384-well plate can then be sealed with foil and placed into a compatible autosampler for sample injection and MS analysis, Fig. 1a,b.

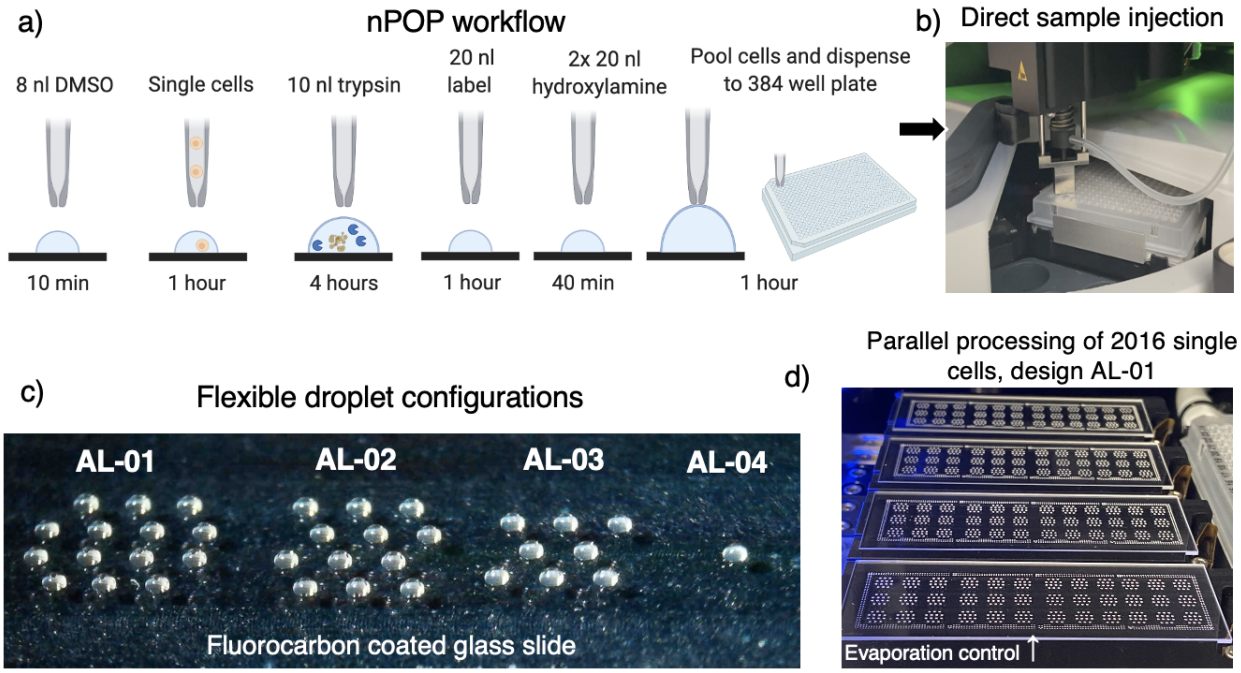


Figure 1 | Parallel preparation of thousands of single cells by mPOP (**a**) A schematic of the nano-PrOteomic sample Preparation (nPOP) method illustrates the steps of cell lysis, protein digestion, peptide labeling with TMT, quenching of labeling reaction, and sample collection. These steps are performed for each single cell (corresponding to a single droplet). (**b**) After barcoding, single-cell samples are automatically pooled into sets and transferred into a 384-well plate, which is then placed into the autosampler for automated injection for LC-MS/MS. Any system that support 384-well plate injection (such as Dionex 3000) can implement this workflow. (**c**) The slide based approach allows programming different droplet layouts, such as the 4 examples shown in the picture. (**d**) Four slides with 2,016 single cells from an nPOP experiment using droplet configuration AL-01. Samples are surrounded by a perimeter of water for local humidity control. Slides are placed on a cooling surface to further prevent evaporation.

To evaluate nPOP's ability to analyze protein covariation within and across cell types, we analyzed two cell lines, WM989 melanoma and U-937 monocyte cells. Single-cell samples were prepared using the AL-01 sample layout design, which can prepare up to 2,016 single cells in

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

one day, Fig. 1d. Using this design, we successfully analyzed 1,556 single cells (Fig. 2a) as part of a single batch. This is lower than the 2,016 capacity due to: 1) including 128 negative controls (droplets that did not receive single cells), 2) having 175 single cells excluded from analysis (Fig. 2a) and 3) 15 sets lost because of LC malfunctions. To increase the depth and consistency of proteome coverage, the single-cell samples were analyzed by prioritized Single Cell ProtEomics (pSCoPE) introduced by Huffman *et al.*[20], following the guidelines of the SCoPE2 protocol[13,18].

Single-cell data quality controls

The MS analysis quantified on average 997 proteins and 2,630 peptides per single cell, with 2,844 proteins quantified across the 1,543 single cells prepared by nPOP, Fig. 2a. To quantify the level of background noise in these measurements, we evaluated the reporter ion intensities in negative controls, which reflect cross-labeling and nonspecific background noise[13,18]. These intensities are mostly absent or very low (Fig. 2b), indicating that background noise is low for singe-cell samples prepared with nPOP. The intensities for single cells also show that peptides from melanoma cells are more abundant than peptides from monocyte cells, reflecting the different cell sizes Fig. 2b. To further test the extent to which higher reporter ion signal in the melanoma cells reflects larger cell size, we plotted the measured diameter for single cells against the average reporter ion signal, Fig. 2c. The correlation between cell diameter and average intensity, $\rho = 0.81$, indicates that nPOP consistently delivers proteins to MS analysis without significant variations from cell to cell.

As an additional QC metric, we evaluated the agreement in relative quantification derived from different peptides originating from the same protein. The agreement is significantly higher in the single cells than the negative controls, Fig. 2d. Furthermore, the small spread of the distribution for the quantitative variability suggests high consistency of the automated sample preparation technique.

By increasing the sample throughput, nPOP reduces the number of sample preparation batches needed and thus the associated batch effects. Indeed, because all single cells of this study were prepared in one batch on the same day, no sample preparation batch corrections needed to be applied to the data.

Next, we performed principal component analysis (PCA) of the single-cell protein dataset us-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

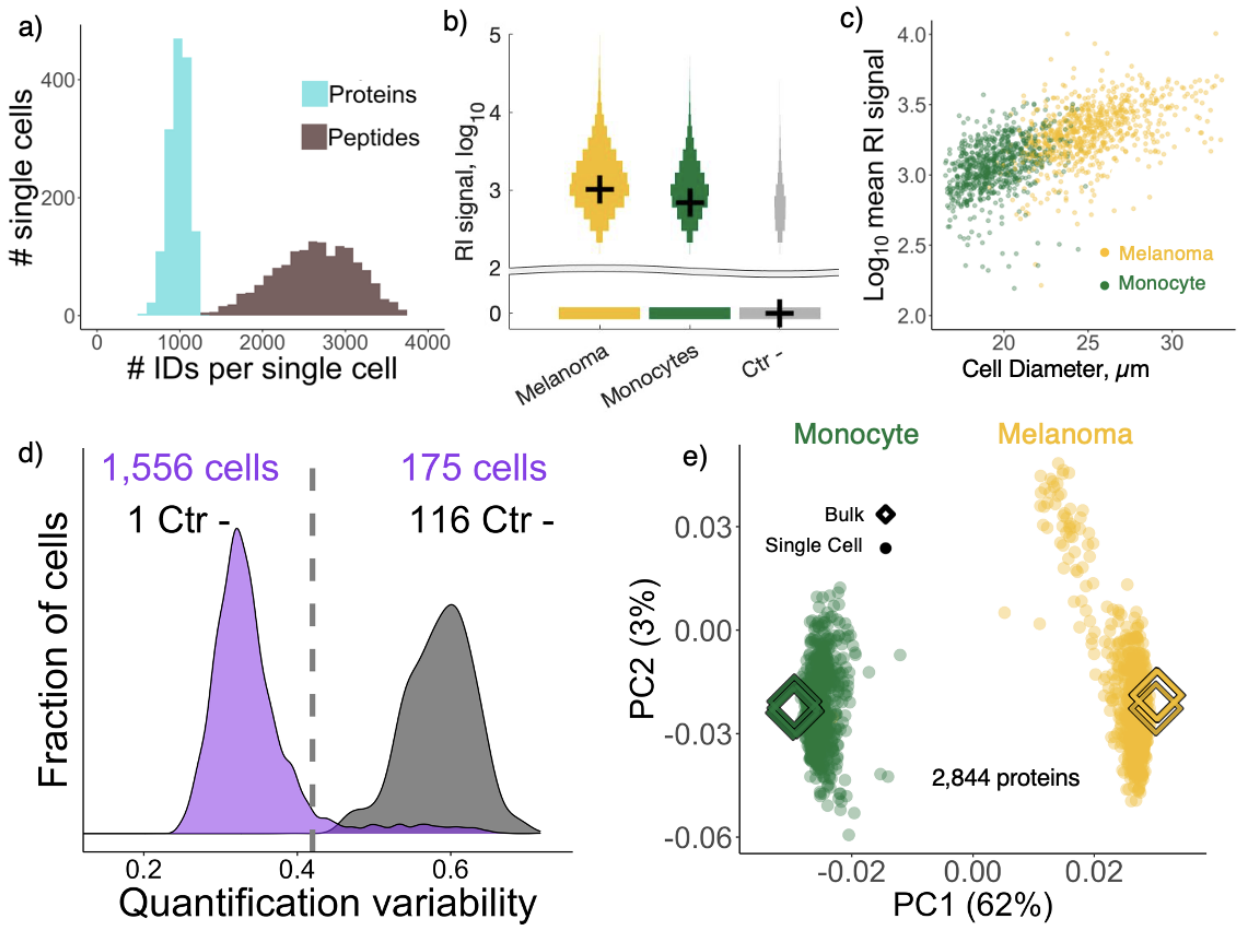


Figure 2 | Proteome coverage and quality controls. (**a**) Number of proteins and peptides quantified per single cell from multiplexed samples prepared by nPOP and analyzed using 60 min active gradients on Q-exactive classic. (**b**) Distributions of reporter ion (RI) intensities for all melanoma, monocyte, and negative controls. Intensities are mostly absent from negative control wells, which contain all reagents but not a single cell. (**c**) The average reporter ion intensity is plotted against the measured diameter of cells. The correlation suggests consistent delivery of proteins. (**d**) The consistency of protein quantification is estimated as the coefficient of variation (CV) of the relative levels of peptides originating from the same protein. (**e**) Principal component analysis separates single-cell samples corresponding to melanoma cells or to U937 monocytes. 200 cell bulk samples are projected onto PCA to demonstrate agreement between bulk and single cell measurements.

ing all quantified proteins, Fig. 2e. The PCA indicates three distinct clusters of cells. The clusters correspond to the cell types, with two subpopulations of melanoma cells. The cell types separate along the first principal component (PC1), which accounts for 59 % of the total variance. To evaluate whether this separation reflects technical artifacts, such as differences in cell size or missing data, or biological differences between cell types, we projected the proteomes of 200-cell samples of melanoma and monocyte cells analyzed by established bulk methods, Fig. 2d. The results

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

indicate that the clustering of single cells is fully consistent with the underlying biology.

Protein covariation with the cell cycle

Next, we moved on to the more challenging problem of quantifying CDC-related protein covariation within a cell type. As a first step towards this analysis, we evaluated the potential to classify individual cells by their cell cycle phase. To obtain a list of proteins whose abundance varies periodically with the cell division cycle, we first sorted populations of each cell type based off their DNA content, Fig. 3a. We quantified the proteomes of the sorted cells and identified proteins whose abundance differs in G1, S, and G2/M phase for both cell types.

To construct robust markers for each phase, we averaged the abundances of groups of proteins corresponding to each phase of the cell cycle. For each CDC phase, we constructed two markers from non-overlapping sets of proteins. Positive correlation between markers from the same phase served as internal validation based on the expectation that proteins peaking in the same phase positively covary. Conversely, we expect markers for different phases to negatively correlate to each other, Fig. 3b. Markers were first constructed in the space of monocyte cells and correlations between markers were cross-validated in melanoma cells, Fig. 3a. Having validated the protein markers, we averaged protein markers within phase for downstream analysis.

To further explore our ability to capture CDC-related protein dynamics, we projected the proteomes of both melanoma and monocyte cells into a joint 2-dimensional space of the CDC marker proteins defined by principal component analysis (Fig. 3b). Each cell was then color-coded based on the mean abundance of a protein marker for the indicated CDC phase, Fig. 3c. The cells from both cell types cluster by CDC phase, which further suggests that the data capture CDC-related protein dynamics.

Next we focused on identifying proteins that covary with the CDC periodic markers. To identify such covariation, the phase marker vectors were correlated to the measured protein abundances of all proteins quantified across many single cells. Many proteins (for 121 for melanoma and 113 for monocyte cells) are significantly correlated (FDR < 0.01), suggesting that these proteins are CDC periodic (supp. table 2). The list includes well characterized CDC proteins and new ones, such as NPM1 which facilitates ribosome biogenesis is positively correlated with G1 phase in both

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

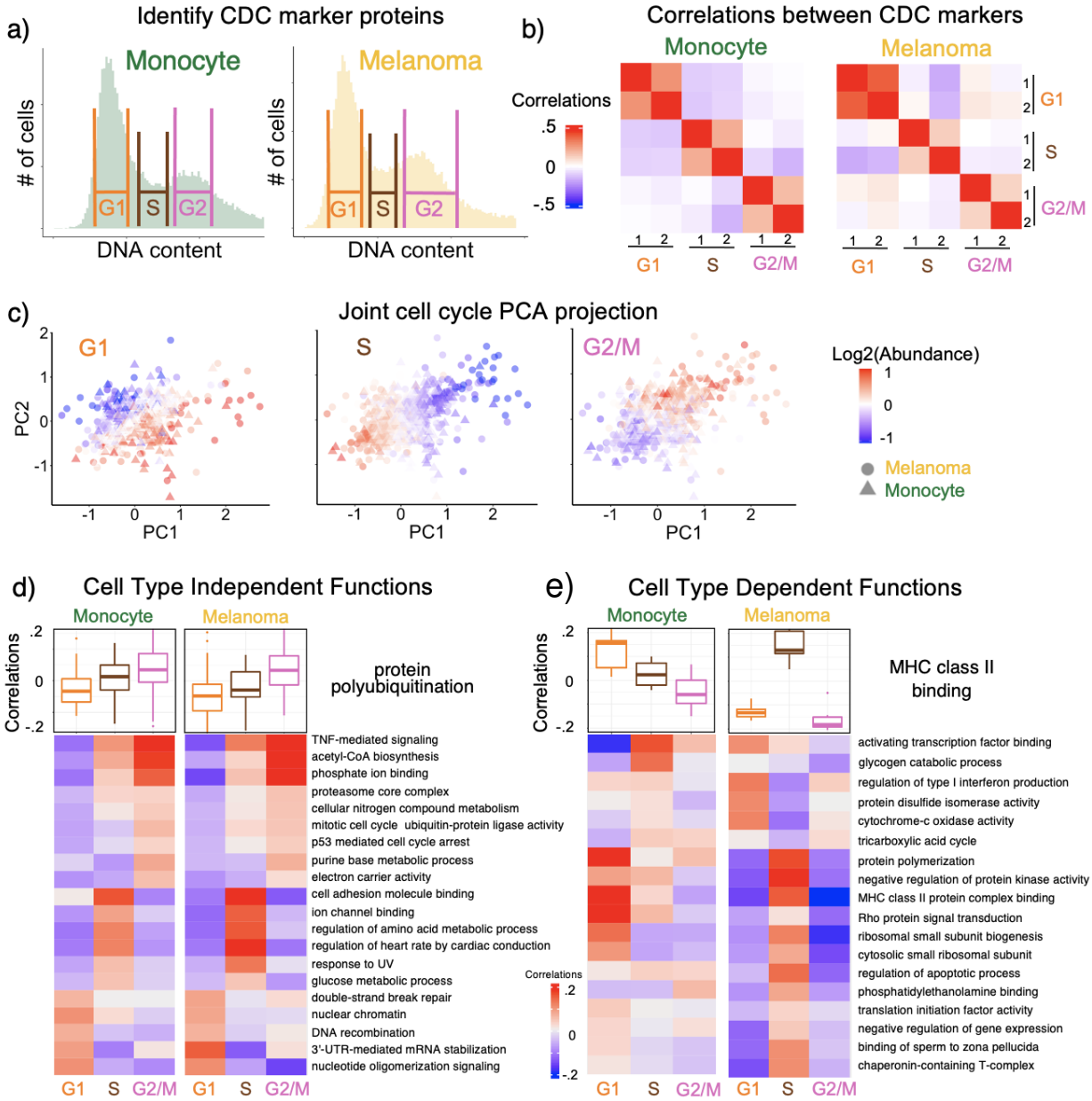


Figure 3 | Identifying functional sets of proteins that covary with CDC-markers. (**a**) Distributions of DNA content for FACS-sorted cells used to identify proteins whose abundance varies with CDC phase. (**b**) Correlations between CDC protein markers computed within the single cells from each type. (**c**) Principal Component Analysis (PCA) of melanoma and monocyte cells in the space of CDC periodic genes. Cells in each PCA plot are colored by the mean abundance of proteins annotated to the indicated phase. (**d**) The boxplots display distributions for correlations between the CDC-phase markers and proteins from the polyubiquitination GO term. The difference between these distributions was evaluated by 1-way ANOVA analysis to estimate statistical significance, FDR $< 5\%$. 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. (**e**) Similar analysis and display as in panel d was used to visualize GO terms whose covariation with the CDC differs between the two cell types.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

melanoma and monocyte populations, $p < 10^{-15}$, $p < 10^{-8}$, respectively.

To increase the statistical power and identify functional covariation with the CDC, we next focused on the covariation of phase markers and proteins with similar functions as defined by the gene ontology (GO). For each GO term, we computed the correlations between its proteins and each CDC-phase-associated marker vector. The distributions of these correlations (as shown with the boxplots in Fig. 3d) were then compared to evaluate statistical significance. For protein polyubiquitination, the distributions of correlations differ significantly between the CDC phases, and this phase-specific covariation is similar for the two cell types. 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. 3d. Such functions with shared covariation include proteolysis in G2/M phase which is consistent with the role of protein degradation in cell cycle progression[21]. Additionally, terms related to DNA repair and translation are correlated with G1 markers confirming the role of cell growth and DNA repair post mitosis. The majority of the 117 significant GO terms (supp. table 3) showed concerted trends between the two cell types, highlighting the conservation CDC related processes.

In addition to finding groups of proteins that show similar cell cycle covariation between cell types, several GO terms also co-varied differentially between the cell types, Fig. 3e. They include terms related to cell signalling, metabolic, and immunological processes, which may reflect the role of the monocyte as an immune cell.

Protein covariation within melanoma subpopulations

We next turned our attention to the two distinct clusters of melanoma cells observed in Fig. 2d. Recent studies of these melanoma cells identified a transcriptomic signature associated with a cell state that is more likely to resist treatment by the cancer drug vemurafenib[7,22]. To test if our distinct clusters corresponded with these signatures, we color coded the cells by the abundance the markers corresponding to each state[7].

The results indicate that the large subpopulation (Cluster A) have higher abundance of the non-primed markers ($p < 10^{-15}$) while the smaller subpopulation (Cluster B) have higher abundance of

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

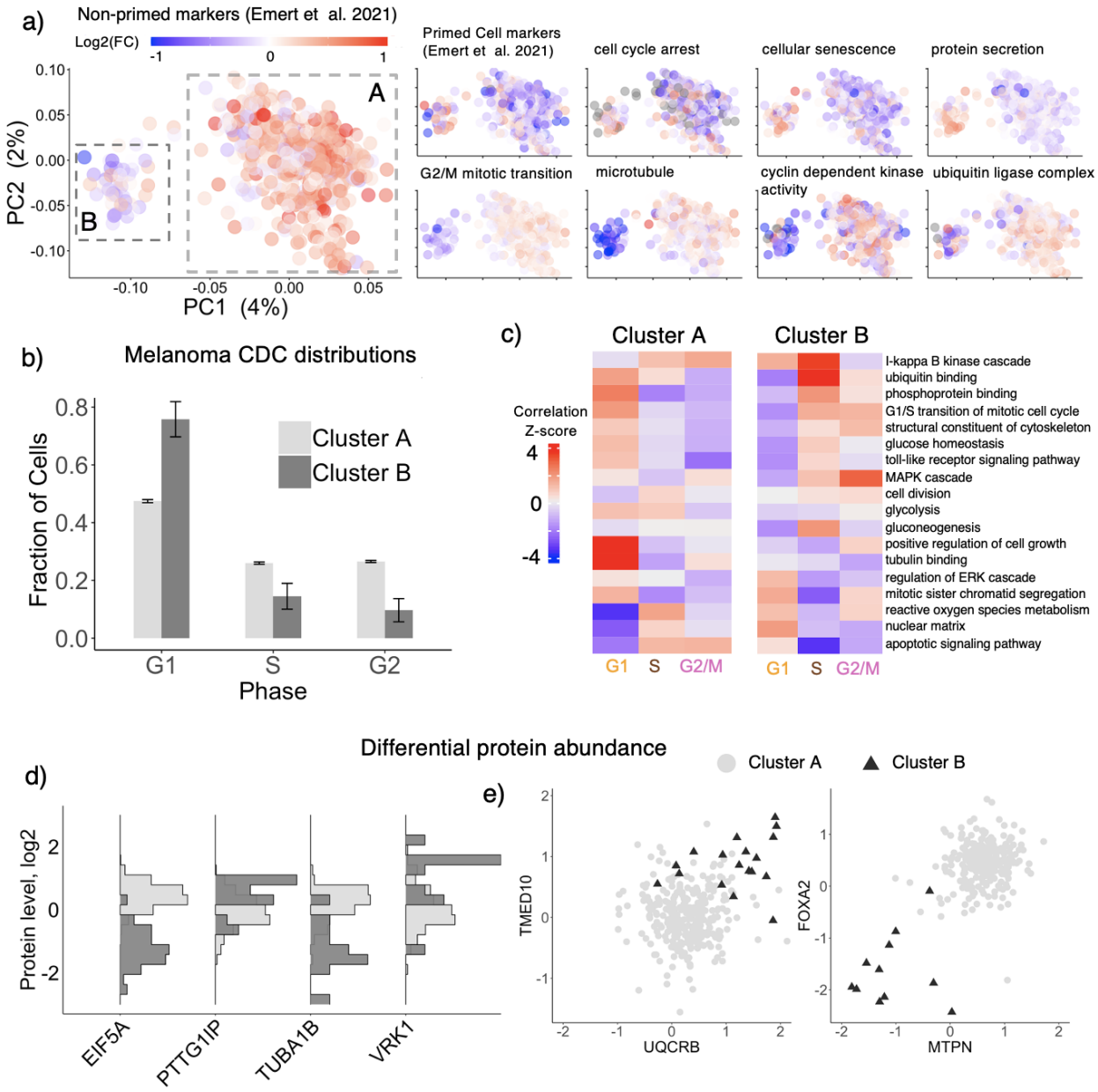


Figure 4 | Protein differences between melanoma subpopulations. (**a**) PCA of melanoma cells indicates two distinct clusters. Single cells are colored based in the protein abundances corresponding to transcripts previously identified as markers of primed cells[7]. The single cells were also colored by the average abundance of protein sets exhibiting significant enrichment clusters A and B. (**b**) Distributions of cells by CDC-phase for cells from cluster A and B. CDC-phases were determined from marker proteins from Fig. 3. (**c**) Protein sets showing distinct covariation in subpopulation A and B. The analysis and display are as in Fig. 3e. Marginal (**d**) and joint (**e**) distributions of protein abundances differentiating clusters A and B.

the primed markers ($p = 0.0002$), Fig. 4a. Having established correspondence between the populations, we sought to identify additional protein differences between the two clusters by performing

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Protein Set Enrichment Analysis (PSEA). It resulted in 200 sets of functionally related proteins exhibiting differential abundance at FDR $< 1\%$, (supp. table 4). A selection of these results are displayed in Fig. 4a, in which cells are color-coded by the median abundance of the proteins associated with the protein set. Protein sets related to G2/M transition of mitosis, cyclin-dependant kinase activity and protein degradation are more abundant in cluster A. In contrast, protein sets with increased abundance in cluster B are related to senescence and cell-cycle arrest. These results suggest that cells in Cluster A are more proliferative than those in Cluster B., consistent with prior report[22].

To further explore CDC differences, we quantified the distribution of cells in each CDC phase across the two sub-populations. We found a substantially larger fraction of cells in cluster B in G1 phase, 78 %, while only 4 % of cells were assigned to G2 phase Fig. 4b. This result further bolsters the conclusion that cells in Cluster B divide slower than those in Cluster A. We next sought to identify additional groups of proteins that co-vary with CDC phase between the two populations. Upon repeating the analysis from Fig. 3e on both melanoma populations, we found several sets of proteins that correlated significantly to the CDC markers but differed between clusters A and B. Specifically, positive regulation of cell growth is strongly correlated to G1 markers in population A but not in cluster B, consistent with slower proliferation of population B. This trend is also reflected in proteins functioning in glycolysis and glucose homeostasis, Fig. 3e.

To further probe the differences between clusters A and B, we identified differentially abundant proteins, 234 at FDR $< 1\%$. The marginal distribution of abundances of these proteins showed minimal overlap (Fig. 4d). These proteins include the surface protein Transmembrane emp24 domain-containing protein 10 (increased abundance in cluster B) and transcription factor Hepatocyte nuclear factor 3-beta (decreased abundance in cluster B), Fig. 4e. All differentially abundant proteins are listed in supplementary table 5.

The joint distributions of proteins with higher abundance in cluster B covary positively (Fig. 4e), which is consistent with the observed co-occurrence of priming transcript markers[6,7]. However, the limited number of analyzed cells from cluster B constrains our ability to confidently quantify this co-variation, which is likely to be strongest in the rare "jackpot" cells defined by Shaffer *et al.*[6].

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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[5]. To support such analysis, we introduce a highly parallel sample preparation method that allowed us to prepare over a thousand single cells in a single batch. It also 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[9,23].

nPOP allowed for deeper single-cell proteomic analysis of the cell division cycle than possible using the minimal sample preparation method (mPOP)[17]. The improvements to throughput and reduction in background signal allowed us to identify new proteins and functional groups of proteins associated with the cell cycle without the artifacts associated with synchronizing cell cultures[24].

Additionally, we were able to identify CDC protein covariation distinct for a subpopulation of melanoma cells expressing markers of priming for drug resistance. This subpopulation is proliferative, with most cells being in G1 phase. Furthermore, we identified differentially abundant proteins between the two melanoma clusters and functional groups of proteins that covary differentially with the respect to the cell cycle markers. 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: scp.slavovlab.net/nPOP and slavovlab.net/Leduc_et_al_2022

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Author Contributions

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

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

Sample preparation: A.L, S.K, J.C

Raising funding & supervision: N.S.

Data analysis: A.L. , S.K. 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 March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Cell culture U-937 and Jurkat cells were grown as suspension cultures in RPMI medium (HyClone 16777-145) supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Cells were passaged when a density of $10e6$ cells/ml was reached, approximately every two days. The Melanoma cells (WM989-A6-G3, a kind gift from Arjun Raj, University of Pennsylvania) were grown as adherent cultures in TU2% media which is composed of 80% MCDB 153 (Sigma-Aldrich M7403), 10% Leibovitz L-15 (ThermoFisher 11415064), 2 % fetal bovine serum, 0.5% penicillin-streptomycin and 1.68mM Calcium Chloride (Sigma-Aldrich 499609). Cells were passaged at 80% confluence (approximately every 3-4 days) in T75 flasks (Millipore-Sigma, Z707546) using 0.25% Trypsin-EDTA (ThermoFisher 25200072) and re-plated at 30% confluence.

HPAF-II cells (ATCC CRL-1997) were cultured in EMEM (ATCC 30-2003), CFPAC-I cells (ATCC CRL-1918) were cultured in IMDM (ATCC 30-2005), and BxPC-3 cells (ATCC CRL-1687) were cultured in RPMI 1640 (ATCC 30-2001). All media were supplemented with 10% fetal bovine serum (FBS, Millipore Sigma F4135) and 1% penicillin-streptomycin (pen/strep, Thermo Fisher 15140122). Cells were passaged at 70% confluence.

Lysis validation experiment 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.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

method[17]. The reference channel was made from the same sample.

Bulk melanoma and monocyte samples Additional bulk samples of melanoma and monocyte cells were prepared for validating quantification of single cells. Cell pellets of 100,000 monocyte and melanoma cells were suspended in 50 μl of mass spectrometry grade water and lysed and digested via mPOP sample preparation[17]. Samples were then labeled with 16plex TMTpro, combined, and diluted down to a concentration of 400 cells/μl for analysis by LC-MS.

Reagent handling with CellenONE The CellenONE is equipped with two piezo dispensing capillaries (PDC). One PDC is dedicated to handle cell suspensions. The other PDC is dedicated for all other reagent handling including 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 deactivate the humidifier. This allows residual DMSO left on the tip of the PDC to evaporate quickly so dispensing is not effected. After each sample preparation, PDCs are washed with ethanol and cleaned under sonication to remove any built up material from inside of the PDC and ensure optimal performance.

Sample preparation and experimental design nPOP reactions are carried out on the surface of a fluorocarbon coated glass slide. The array 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 18-plex and 14 droplets per cluster, corresponding to the 14 isobaric labels used for single cells. We used design AL-01, which allowed fitting 36 clusters per slide and 4 glass slides on the temperature controlled target holder for simultaneous processing of up to $14 \times 36 \times 4 = 2,016$ single cells. Reducing the space between clusters can further increase the number of clusters per slide and thus the number of simultaneously prepared single cells. The design can be modified easily to accommodate higher and lower plex labels, as demonstrated with plexDIA[25]. The array layout was optimized to keep droplets from the same set close in proximity but prevent reaction

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

volumes from merging.

Once an array layout is selected, 8nL of DMSO is dispensed to each location of the array forming the initial reaction volume for each single cell reaction. Lysis begins when cells are dispensed inside a droplet of about 300pL of PBS into these reaction volumes of DMSO. After lysis, 10 nl of solution containing trypsin and HEPES buffer is added to each reaction volume, for a final concentration of 120 ng/μl of trypsin and 5 mM HEPES and total volume of 18 nl.

The humidifier and cooling system is then turned on to prevent droplet evaporation. Relative humidity inside the CellenONE is set to 75 % and the 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 40 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, sample clusters are pooled by aspirating them off the slide surface in 10 μl of a 100 % acetonitrile solution via CellenONE PDC and syringe pump controls. Pooled samples are then transferred into a 384-well plate (ThermoFisher AB1384) and dried down to dryness in a speed-vacuum (Eppendorf, Germany) and either frozen at $-80\ ^oC$ for later analysis or immediately reconstituted in 1.1 μl of 0.1 % formic acid (ThermoFisher 85178) for mass spectrometry analysis.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA sorting for Bulk CDC Analysis Melanoma and monocyte cells were incubated using Vybrand Dye Cycle (Thermo V35003) following manufacturers instructions. Cells were sorted via the Beckman CytoFLEX SRT. Post sorting, cells were pelleted and washed with Mass Spectrometry grade water and resuspended in water at a concentration of 2000 cells/μl. Cells were then frozen at -80 degrees celsius for 10 minutes and then heated to 90 C for 10 minutes for lysis. Proteins were then digested over night in a solution of 15 ng/μl of trypsin. Samples were analyzed via data independent acquisition.

LC-MS platform MS analysis was designed and performed according to the SCoPE2 guidelines and protocol[13,18,26]. Specifically, the single cells pooled into SCoPE2 sets were separated via online nLC on a Dionex UltiMate 3000 UHPLC; 1 μl out of 1.1 μl of sample was picked up out of a 384 well plate (Thermo AB1384) placed on an auto sampler height adjuster for PCR plates (Thermo 6820.4089) and 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.

Single-Cell MS Data Acquisition A prioritized analysis workflow[20] was used to increase con-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sistency of identification and depth of coverage for the nPOP-prepared single-cell data shown in Fig. 2, Fig. 3, and Fig. 4. A spectral library was built from two injections of a 10x concentrated aliquot of combined carrier and reference sample analyzed by DIA instrument methods 1 and 2, as well as an injection of a 5x concentrated aliquot of combined carrier and reference sample analyzed by DIA method 1. Both of these instrument methods are detailed in the methods section of Huffman, *et al.*[20]. A subsequent injection of a 1x concentrated aliquot of carrier and reference sample was analyzed by DIA instrument method 1 to serve as a retention-time-calibration run. The results from this retention-time-calibration run were searched with Spectronaut to generate a prioritized inclusion list for subsequent scout runs and prioritized single-cell analyses. The prioritized inclusion lists were then imported into MaxQuant.Live (v. 2.0.3 with priority tiers) and used to analyze 1x concentrated carrier and reference samples or nPOP prepared single-cell samples, with settings detailed below.

Inclusion-list generation for scout experiments Spectronaut search results of the retention-time-calibration run were filtered to EG.PEP ≤ 0.02 and EG.Qvalue ≤ 0.05. Additionally, precursors without TMTPro modifications (+ 304.2071 Da) on the peptide n-terminus or lysine residue were filtered out. The distribution of precursor intensities for the remaining precursors was then subset into tertiles for use in priority tier assignment. These precursors were then filtered such that a maximum of four peptides per protein were selected, with the most intense peptides per protein being selected. Filtered peptides with precursor intensities in the top intensity tertile were placed on the top priority tier, peptides with intensities in the middle intensity tertile were placed on the middle priority tier, and peptides with intensities in the bottom intensity tertile were placed on the bottom priority tier. All species matching the original EG.PEP and EG.Qvalue filtration characteristics that were not previously selected for a priority tier were assigned a priority below the previous bottom tier. These priority-tier-assigned peptides were then enabled for participation in MaxQuant.Live's realtime-retention-time-alignment algorithm, as well as MS2 upon detection. Any remaining PSMs outside of the original filtration criteria (EG.PEP ≤ 0.02 and the EG.Qvalue ≤ 0.05) were enabled for participation in MaxQuant.Live's realtime-retention-time-alignment algorithm, but not sent for MS2 upon detection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Scout experiment instrument method and raw data analysis $1\mu l$ injections of a 1x concentrated aliquot of mixed carrier-reference material were analyzed using the instrument method detailed in the prioritized acquisition parameters section and MaxQuant.Live parameters indicated in the associated table.

The two raw files associated with these experiments were then searched using MaxQuant (v. 1.6.17.0) using a FASTA containing all entries from the human SwissProt database (`swissprot_human_20211005.fasta`, 20,386 proteins). TMTPro 16plex was enabled as a fixed modification on peptide n-termini and lysines via the reporter ion MS2 submenu. Methionine oxidation (+ 15.99492 Da) and protein n-terminal acetylation (+ 42.01056 Da) were enabled as variable modifications, and trypsin was selected for in silico digestion with enzyme mode set to specific. Up to 2 missed cleavages were allowed per peptide with a minimum length of 7 amino acids. Second peptide identifications were disabled, calculate peak properties was enabled, and msScans was enabled as an output file. PSM FDR and protein FDR were set to 1.

Pre-prioritization shotgun experiment instrument method and raw data analysis One $1\mu l$ injection of a 1x concentrated aliquot of mixed carrier-reference material was analyzed using the LC settings indicated above. The following MS1 settings were used: 70k resolution, 1e6 AGC target, 100ms maximum injection time, and a scan range of 450Th to 1600Th. MS2 scans were acquired with the following settings: 70k resolution, 1e6 AGC target, 300ms maximum injection time, loop count (i.e. top-n) of 7, Isolation window of 0.7Th with a 0.3Th offset, fixed first mass of 100 m/z, NCE of 33, and a centroid spectrum data type. The minimum AGC target was 2e4, apex triggering was disabled, and charge exclusion was enabled for unassigned charge states, as well as charge states greater than 6. The peptide match setting was disabled, exclude isotopes was enabled, and dynamic exclusion was set to 30 seconds. Voltage was set to 0 for the first 25 minutes, sweep gas was applied from minute 24.6 to 25 to dislodge any accumulated droplets from the capillary tip. From minute 25 to 80, voltage was set to 1.7kV, capillary temp to 250 °C, and the S-lens RF level to 80. From minute 94.20 to 94.60, sweep gas was applied to dislodge any accumulated droplets

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from the capillary tip.

The raw file generated by this analysis was searched using the same maxquant settings as indicated in the Scout experiment instrument method and raw data analysis section.

Prioritized inclusion list generation The PSMs generated from the scout runs using intensity-dependent-tiers (wAL00191 and wAL00192) were partitioned into three categories: PSMs at PEP ≤ 0.02 (set α), PSMs with $0.02 <$ PEP ≤ 0.05 (set β), and PSMs with PEPs > 0.05 (set γ). Then the same set of PEP filters defined above for wAL00191 and wAL00192 were applied to the results of a DDA analysis conducted on an injection of a 1x concentrated aliquot of carrier and reference material to generate sets δ, ϵ, and ζ. Furthermore, these last three precursor sets were assembled such that they each contained a unique set of precursors with respect to one another and the preceding set of precursors.

Sets α and δ were combined and filtered such that a maximum of 4 peptides per protein were selected, choosing those precursors with the highest precursor intensities, to form the top priority tier candidates. The excluded precursors from this filtration were then combined with sets β and ϵ to make up the middle priority tier candidates. Peptides from sets γ and ζ were then combined to form the bottom priority tier candidates.

The results from the retention-time-calibration experiment were then intersected with the priority tier sets, and the PSMs matching each set were given a corresponding priority index for use by MaxQuant.Live. Up to 8,600 of the most abundant remaining retention-time-calibration-experiment-associated PSMs were then added to the bottom priority tier to provide additional identifiable precursors when higher priority precursors were not detected. All selected precursors were then enabled for participation in the MaxQuant.live real-time-retention-time-alignment algorithm, and for MS2 upon detection. All remaining PSMs that were not part of the priority tiers were then selected for participation in the MaxQuant.live real-time-retention-time-alignment algorithm, but not for MS2 upon detection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Limited FASTA file generation for raw data analysis corresponding to prioritized samples The `swissprot_human_20211005.fasta` was read into the R environment using the seqinr[27] package, and only those proteins with peptides present on the inclusion list were retained to generate the `AndrewsnPOP_FASTA_v2.fasta` file, containing 3535 proteins, used to search the resulting prioritized single-cell experiments

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

DIA MS Acquisition of Bulk CDC populations Samples were run using the V1 method from[25]. This method contains 5 MS1 scans at 140k resolution and 28 MS2 scans per duty cycle. This method offers improved MS1 level quantification.

Analysis of DDA MS data Raw data were searched by MaxQuant[28,29] 1.6.17.0 against a protein sequence database including entries from the appropriate human SwissProt database (downloaded July 30, 2018) and known contaminants such as human keratins and common lab contaminants. Fasta was limited to proteins which were included on prioritization list. MaxQuant searches were performed using the standard work flow[30]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Analysis of Data Independent Acquisition MS data Data Independent Acquisition runs were searched with DIA-NN v1.8.0[31] using an in silico fasta generated library enabled by deep learning. Raw files were searched together with match between runs (MBR). DIA-NN search settings: Library Generation was set to "IDs, RT, & IM Profiling", Quantification Strategy was set to "Peak height", scan window = 1, Mass accuracy = 10 ppm, and MS1 accuracy = 5 ppm, "Remove likely interferences", "Use isotopologues", and "MBR" were enabled.

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

Estimating labeling efficiency for TMT dissolved in DMSO

To estimate the labeling efficiency of for TMT dissolved in DMSO, we conducted two experiments. In the first experiment, we dispensed 50 ng of monocyte digest dissolved in HPLC grade water and 100 mM Triethyl ammoniumbicarbonate buffer, pH 8.5. We also lysed, digested, and labeled 200 monocytes in 10 drops of 20 cells per drop following nPOP protocol.

Both samples (Raw files AL078 and AL080 respectively) were searched by maxquant with TMTpro specified as a variable modification on lysines and n-termini. We then computed the labeling efficiency as the ratio of number possible lysine or n-termini that could be labeled containing TMT-16 labels divided by number of possible labeling sites. (See supplementary code)

Single-cell filtering and normalization The single-cell data were processed and normalized by the SCoPE2 pipeline[13,18]. This pipeline is also implemented by the scp Bioconductor package[32,33]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 distribution were removed from analysis with a threshold of 0.41. Data was normalized as by procedure outlined by Specht et al.[13,34].

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.

Melanoma sub population protein set enrichment analysis Protein set enrichment analysis was performed by t-test between Cluster A and B on the un-imputed data. It was required that a given gene set had at least 4 proteins measured in the single cells and that each population had at least 80 % of cells with protein observations. The distribution of p-values was corrected for multiple hypothesis testing with the BH method. Only GO terms were reported with Q value less than 0.0001 were reported.

Differential protein abundance testing

Differential protein abundance testing was performed using precursor-level quantitation. To account for variation in sample loading amounts, precursors from each sample were normalized to their sample-median. Then, each precursor was normalized by its mean across samples to convert it to relative levels. The normalized relative precursor intensities from different replicates were grouped by their corresponding protein groups and compared by ANOVA to estimate the significance of differential protein abundance across samples/conditions. This comparison captures both the variability between different replicates and different peptides originating from the same pro-

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tein. To correct for multiple hypotheses testing, we used the Benjamini-Hochberg (BH) method to estimate q-values for differential abundance of proteins and protein sets.

Constructing CDC phase markers Phase markers were constructed from proteins identify with differential abundant each CDC phase in both monocyte and melanoma cells. These proteins were first identified on the bulk level. To further narrow the list of proteins used to create phase markers, we used proteins that contained multiple, positively correlated peptides in the single cell samples.

Phase markers were then constructed by averaging the abundances of all possible combinations of 2 or 3 proteins corresponding to each phase of the cell cycle. We selected groups two markers for each CDC phase that were positively correlated. This served as validation as we expect proteins that are highly abundant in same phase will positively covary. We then further filtered groups of protein markers

Markers were first constructed in the space of monocyte cells and correlations between markers were validated in melanoma cells Fig. 3 b. Having validated the protein markers, we averaged protein markers within phase for downstream analysis.

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 of each protein using Spearman correlation. We adjusted the distribution of p-values obtained from the Spearman correlation test using the BH method and filtered results at FDR $< 1\%$.

Cell cycle 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 melanoma and monocyte cells as plotted in Fig. 3a. The resulting correlation vectors were analyzed by protein set enrichment analysis similar to previously reported analysis[35]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. 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. 3 the 20 GO terms whose correlations to the CDC phase markers was most similar or most different between the 2 cell lines.

Assigning cells to CDC phase We took a greedy approach to assign cells to a CDC phase. First, we created a vector comprised of length 3X the number of cells, where each value was the average abundance of the G1, S, or G2 marker proteins in a given cell. We then sorted this vector from highest to lowest. Subsequently iterated down the list and sorted cells into the G1, S, or G2 bin based off the phase of each value. We sorted 50 % of cells into the G1 bin, 25 % of cells into the S and G2 bins based off the distribution observed from the bulk FACS CDC sorting.

MaxQuant.live settings for prioritized analysis

Global settings: Survey scan	**Scout exp.**	**nPOP samples**
ScanDataAsProfile	True	True
PositiveMode	True	True
MaxIT	100ms	100ms
Resolution	70,000	70,000
AgcTarget	1,000,000	1,000,000
MzRange	(450,1258)	(450,1258)
BoxCarScans	0	0
Global settings: TopN	**Scout exp.**	**nPOP samples**
NumOfMS2Scans	0	0

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RealtimeCorrection	**Scout exp.**	**nPOP samples**
MzTolerances	(4.5,5)	(4.5,5)
RetentionTimeTolerances	(0.01,2)	(0.01,2)
SigmaScaleFactorRt	3	3
PeptideHistoryLength	2	2
MinUsedCorrectionPeptides	15	15
IntensityPeakRatioThreshold	.01	.01
PeptideDetectionIsoPeaks	2	2
IsotopeTolerance	9	9
Ms2DetectionNeeded	False	False
Ms2ExcludeDetectedPeptides	False	False
Ms2MinNormIntensity	0.1	0.1
Ms2MzTolerance	20	20

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TargetedMs2	**Scout exp.**	**nPOP samples**
BatMode	False	False
AutoPriority	True	True
DefaultPriority	0	0
MaxNumOfScans	1	1
WindowAndOffsetInDalton	False	False
ScanDataAsProfile	False	False
WindowSize	0.5	0.5
MzOffset	0	0
LowerMzBound	100	100
CollisionEnergy	33	33
LifeTime	2,100ms	2,100ms
Resolution	70,000	70,000
MaxIT	300ms	300ms
AgcTarget	1,000,000	1,000,000
PositiveMode	True	True

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figures

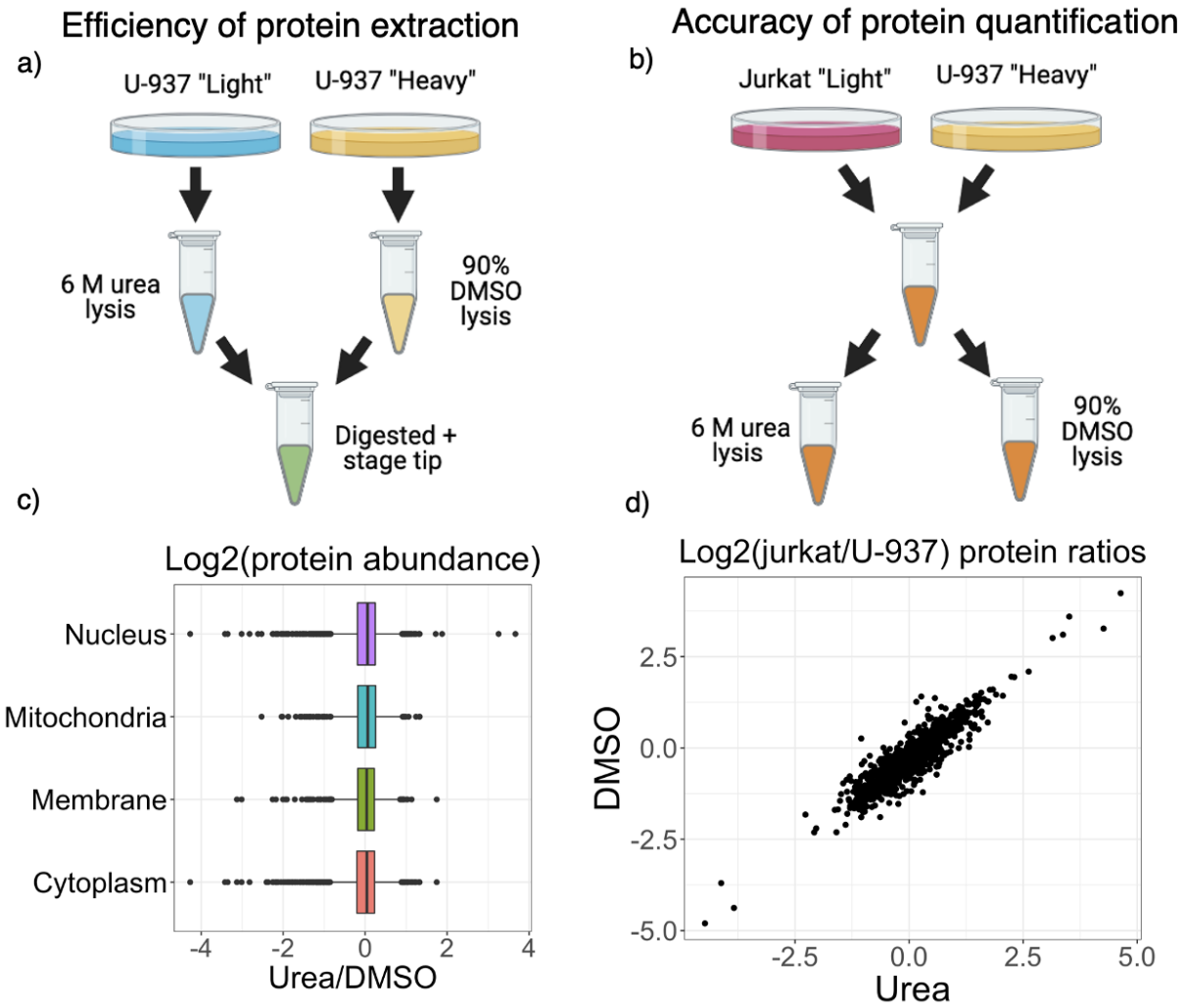


Figure S1 | Evaluating the efficiency of protein extraction by DMSO cell lysis. (**a**) Equal number of U-937 cells labeled with "Light" and "Heavy" isotopes via SILAC were lysed with urea or DMSO, diluted, and combined for digestion. The SILAC ratios for proteins from different cellular compartments 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Regev, A. *et al.* Science forum: the human cell atlas. *elife* **6,** e27041 (2017).
2. Specht, H. & Slavov, N. Transformative opportunities for single-cell proteomics. *Journal of Proteome Research* **17,** 2563–2916 (8 June 2018).
3. Ziegenhain, C. *et al.* Comparative analysis of single-cell RNA sequencing methods. *Molecular cell* **65,** 631–643 (2017).
4. Slavov, N. Unpicking the proteome in single cells. *Science* **367,** 512–513 (2020).
5. Slavov, N. Learning from natural variation across the proteomes of single cells. *PLOS Biology* **20,** 1–4. https://doi.org/10.1371/journal.pbio.3001512 (Jan. 2022).
6. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. en. *Nature* **546,** 431–435 (June 2017).
7. Emert, B. L. *et al.* Variability within rare cell states enables multiple paths toward drug resistance. *Nature biotechnology* **39,** 865–876 (2021).
8. Mahdessian, D. *et al.* Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* **590,** 649–654 (2021).
9. Slavov, N. Scaling Up Single-Cell Proteomics. *Molecular & Cellular Proteomics* **21,** 100179. ISSN: 1535-9476 (2022).
10. Slavov, N. Single-cell protein analysis by mass spectrometry. *Current Opinion in Chemical Biology* **60,** 1–9. ISSN: 1367-5931 (2020).
11. Vanderaa, C. & Gatto, L. Replication of single-cell proteomics data reveals important computational challenges. *Expert Review of Proteomics* **18,** 835–843 (2021).
12. Kelly, R. T. Single-Cell Proteomics: Progress and Prospects. *Molecular & Cellular Proteomics* **19,** 1739–1748 (2020).
13. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biology* **22** (2021).
14. Singh, A. Towards resolving proteomes in single cells. en. *Nat. Methods* **18,** 856 (Aug. 2021).
15. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161,** 1187–1201 (2015).
16. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).
17. Specht, H. *et al.* Automated sample preparation for high-throughput single-cell proteomics. *bioRxiv 10.1101/399774.* https://doi.org/10.1101/399774 (2018).
18. Petelski, A. A. *et al.* Multiplexed single-cell proteomics using SCoPE2. *Nature Protocols* **16,** 5398–5425 (2021).
19. Marx, V. A dream of single-cell proteomics. *Nature Methods* **16,** 809–812 (2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Huffman, R. G. *et al.* Prioritized single-cell proteomics reveals molecular and functional polarization across primary macrophages. *bioRxiv 2022.03.16.484655.* https://doi.org/10.1101/2022.03.16.484655 (2022).

21. Glotzer, M., Murray, A. W. & Kirschner, M. W. Cyclin is degraded by the ubiquitin pathway. *Nature* **349,** 132–138 (1991).

22. Fallahi-Sichani, M. *et al.* Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. *Molecular systems biology* **13,** 905 (2017).

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

24. Cooper, S. The synchronization manifesto: a critique of whole-culture synchronization. *The FEBS Journal* **286,** 4650–4656 (2019).

25. Derks, J. *et al.* Increasing the throughput of sensitive proteomics by plexDIA. *bioRxiv*. https://doi.org/10.1101/2021.11.03.467007 (2021).

26. 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).

27. Charif, D. & Lobry, J. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *Biological and Medical Physics, Biomedical Engineering* (eds Bastolla, U., Porto, M., Roman, H. & Vendruscolo, M.) ISBN : 978-3-540-35305-8, 207–232 (2007).

28. 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).

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

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

31. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature methods* **17,** 41–44 (2020).

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

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

34. 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).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.24.441211; this version posted March 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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