Robust, sensitive, and quantitative single-cell proteomics based on ion mobility filtering

Jongmin Woo,[1,†] Geremy C. Clair,[2,†] Sarah M. Williams,[1] Song Feng,[2] Chia-Feng Tsai,[2] Ronald J. Moore,[2] William B. Chrisler,[2] Richard D. Smith,[2] Ryan T. Kelly,[1,3] Ljiljana Pasa-Tolic,[1] Charles Ansong,[2,*] Ying Zhu[1,*]

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

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

[3]Department of Chemistry and Biochemistry, Brigham Young University, Provo, Utah 84604, United States

[†] Jongmin Woo and Geremy C. Clair contributed equally to this work.

***Corresponding author:**

Dr. Charles Ansong (charles.ansong@pnnl.gov)

Dr. Ying Zhu (ying.zhu@pnnl.gov)

Abstract

Unbiased single-cell proteomics (scProteomics) promises to advance our understanding of the cellular composition of complex biological systems. However, a major challenge for current methods is their ability to identify and provide accurate quantitative information for low abundance proteins. Herein, we describe an ion mobility-enhanced mass spectrometry acquisition method, TIFF (Transferring Identification based on FAIMS Filtering), designed to improve the sensitivity and accuracy of label-free scProteomics. The TIFF method enabled unbiased proteome analysis to a depth of >1,700 proteins in single HeLa cells with >1,100 proteins consistently quantified, a significant improvement in overall performance. We applied the TIFF method to obtain temporal proteome profiles of >150 single murine macrophage cells during a lipopolysaccharide stimulation experiment and uncovered unanticipated temporal response trajectories. Further, we demonstrated, to our knowledge, the first application of scProteomics to classify cell populations of a human organ (the lung) without prior antibody labeling.

Introduction

Single-cell technologies have become the cornerstone of biomedical and cell biology research [1, 2]. The emergence of single-cell RNA sequencing (scRNA-seq) and related single-cell sequencing technologies has brought to light unappreciated cellular heterogeneity and reveal cell subpopulations obscured in bulk measurements [3]. However, many integrative studies have shown only low to moderate correlations between the abundance of RNA transcripts and their corresponding proteins [4, 5], as the translation of RNA into functional protein can be affected by diverse events such as alternative splicing and microRNA regulation [6]. Additionally, RNA measurements cannot infer post-translational modifications that modulate protein functions. Thus, there is an unmet need for broad proteome measurements at the single-cell level, which has lagged behind single-cell sequencing approaches.

Recent advances in sample preparation and mass spectrometry facilitate unbiased single-cell proteomics (scProteomics) [7-18]. Microfluidic sample processing devices and systems have improved protein digestion efficiency and sample recovery by minimizing adsorptive losses [14-17]. Tandem mass tag (TMT)-based isobaric labeling approaches have enabled multiplexed single-cell analysis in individual LC-MS runs [7, 9, 11, 18]. The miniaturization of capillary electrophoresis or liquid chromatography have improved separation resolution and enhanced electrospray ionization efficiency [19]. High-resolution MS analyzers combining with ion focusing devices such as the ion funnel, have increased detection sensitivity to the level where single molecules can be detected [20]. State-of-the-art methodologies in scProteomics now can detect from ~700 to ~1,000 proteins from cultured single mammalian cells (e.g., HeLa) using label-free approaches [8, 10, 12, 15] and from ~750 to ~1,500 proteins using TMT-labelling and signal boosting strategies [7, 9, 11, 18]. Despite these advances, scProteomics remains immature, and significant technical challenges remain, including not only limited proteome depth and poor quantification performance, but also low system robustness for large-scale single-cell studies.

Because of the lack of a global amplification method for proteins, the coverage and quantification performance of scProteomics largely rely on the capabilities of MS measurement (*e.g.* sensitivity, speed, dynamic range). Although targeted MS measurements enable the detection of low copy number proteins and even single molecules [20], these measurements are generally performed using narrow m/z windows [20] or tandem mass spectra [21] to minimize background signals. Background ions, originating from ambient air and solvent/reagent impurities, dominate MS spectra during full m/z range acquisition for scProteomics. These abundant ions quickly fill ion trapping devices (e.g., ion trap or ion routing multipole) and limit elongated ion trapping, which could otherwise accumulate more low-abundance ions of interest and improve detection sensitivity [22, 23]. The high background signals generated by these ions can also significantly reduce the dynamic range of MS analyzers and deteriorate feature detection during downstream data analysis.

We reasoned that the removal of background ions should dramatically enhance the sensitivity of MS detection and improve the proteome coverage and quantitation performance of scProteomics. A variety of approaches have been developed to minimize background signals including the use of a carbon filter in front of MS inlets to purify the ambient air [24], a picoliter-flow liquid chromatography (LC) system to reduce overall contaminates from air and solvent [19], a dynamic range enhancement applied to MS (DREAMS) data acquisition algorithm to reject high abundant ions before ion accumulation [23], and a high field asymmetric waveform ion mobility spectrometry (FAIMS) interface to remove singly charged ions [12]. FAIMS is an easily implemented approach that can further extend the MS dynamic range for scProteomics by filtering ions to form distinctive subpopulations based on ion mobility differences. Herein, we describe a FAIMS-enabled data acquisition method, TIFF (Transferring Identification based on FAIMS Filtering), that significantly improves proteome coverage and quantification accuracy for label-free scProteomics. We demonstrated the capability and scalability of the method by a large-scale study of macrophage stimulation with lipopolysaccharide (LPS) and by using it to classify dissociated human lung cells into distinct cellular populations.

Results

The TIFF concept

The TIFF method developed herein utilizes the accurate mass and time (AMT) tag approach [25], or other derivative approaches, such as "match between run" (MBR) implemented in MaxQuant [26], that rely on two measurements for peptide identification: the accurate mass-to-charge ratio (m/z) and the LC retention time (RT). We have previously demonstrated that MBR improves the proteome coverage and reduces missing values in nanoscale proteomics including scProteomics [14]. The recent integration of ion mobility devices, including FAIMS, at the interface between the LC system and mass spectrometer now provide an opportunity to use the additional ion-mobility separation dimension to reduce false discovery and improve coverage [27]. We take advantage of this advancement for the LC-FAIMS-MS-based TIFF method, which utilizes the FAIMS compensation voltage (CV) for a third matching feature (in addition to retention time and accurate mass) for peptide identification, as illustrated in Figure 1a. Briefly, a spectral library is constructed by repeatedly analyzing high-input samples on an LC-FAIMS-MS platform, with each LC-MS analysis utilizing a discrete FAIMS CV; in this case, CVs of -45V, -55V, -65V and -75V. Each peptide identified in the high-input analyses is associated with a unique 3D tag comprising LC retention time, accurate m/z, and FAIMS CV. Next, low-input samples (e.g., single-cells) are analyzed by cycling through the multiple FAIMS CVs (-45V, -55V, -65V, and -75V) within a single LC-MS analysis. A key aspect of the TIFF method is the mode of MS data acquisition, with most of the MS time spent on MS1 acquisition compared to a traditional standard MS or FAIMS MS (Supplementary Figure 1a) to enhance the accumulation of peptide ions for sensitive detection and precise quantitation. The fewer MS2 acquisitions generated within each cycle are sufficient to exploit the non-linear multi-sample alignment feature of MaxQuant. Subsequently, MS1 features in low-input samples (i.e. single-cells) are identified by matching to the spectral library generated with the high-input samples utilizing the unique 3D tag (i.e. LC retention time, accurate m/z, and FAIMS CV) employing the MBR algorithm within MaxQuant.

TIFF improves LC-MS sensitivity

We first verified the utility of FAIMS to remove singly charged ions ("chemical background" noise) and create more "room" for ion accumulation and enhance detection of low-abundance peptides. We analyzed single-cell equivalent amount (0.2 ng) of protein digests (CMK cells) with either a FAIMSpro interface or a standard interface. Representative mass spectra are shown in Supplementary Figure 1b. Without FAIMS, most dominating signals correspond to singly charged ions, some of which are known to originate from plasticizer (e.g. m/z 391.28) and air impurities (e.g. m/z 445.12, 462.29, and 519.14). Only a few multiply charged ions are assigned as peptide features. Because these highly abundant contaminants quickly fill ion accumulation (or trapping) regions, the median ion injection/accumulation time is only 30 ms across the whole LC-MS analysis (Figure 1b). In comparison, when FAIMS is used, most dominating ion signals are multiply charged (Supplementary Figure 1b), and the median ion injection times increase from ~30 ms to ~180 ms for a CV of -45 V, reaching the maximal time of 254 ms for the other three CVs (-55 V, -65 V, and -75 V). (Figure 1b). Benefiting from the low background and elongated ion accumulation, the median S/N of LC-MS features increased from 5.2 (STD) to 29.6 (FAIMS), representing a >5-fold increase for all the CV values (Figure 1c).

To evaluate the anticipated improvements in MS sensitivity offered by the FAIMS and TIFF method, we investigated several metrics related to proteome coverage, including the number of peptide features (multiply charged MS feature), unique peptides, and proteins (Figure 1d-1f and Supplementary Figure 2a-2d). Concurrently, we investigated the benefits of modulating the number of CVs utilized in the TIFF method; comparing the utilization of 4 CVs versus 2 CVs. Briefly, we analyzed single-cell-level (0.2 ng) protein digests from three acute myeloid leukemia cell lines: CMK, K562, and MOLM14 with either a FAIMSpro interface or with a standard interface. Analyses with the FAIMSpro interface and TIFF method utilized either 4 CVs (-45 V, -55 V, -65 V, and -75 V) or 2 CVs (-45 V and -65 V). Compared to analyses performed with a standard interface, analyses performed with the FAIMSpro interface and the TIFF method increased the number of peptide features detected in the MS1 by >270% (Figure 1d). Most

of the increased peptide features appeared in the low-MS-intensity scale across all four FAIMS CVs (Supplementary Figure 2e). Compared to analyses performed with a standard interface, those employing the FAIMSpro interface and the TIFF method increased peptide identification by 160% (Figure 1e) and protein identification by 175% (Figure 1f), presumably due to the higher S/N and better detection of multiply charged ions. Modulation of CVs within the TIFF method had a modest effect on the number of peptide features, peptides, and proteins, with only a slight increase using 4 CVs as opposed to 2 CVs. However, utilizing 4 CVs in the TIFF method yielded >114% increases in summed peptide intensities compared to utilizing 2 CVs (Supplementary Figure 2f), which subsequently improved the quantification performance.

TIFF improves the quantification of scProteomics

Next, we evaluated whether the TIFF method improves quantification performance when compared with a standard approach. We compared the run-to-run reproducibility from triplicates using 0.2 ng of CMK cell digests with the standard, 2-CV TIFF, and 4-CV TIFF methods. While the distribution of the coefficients of variation was similar between the 2-CV TIFF and the standard methods (*p-value* = 0.31, Mann-whitney test), the median of the coefficients of variation for the 4-CV TIFF method was significantly reduced (*p-value* < 0.01, Mann-whitney test), from 15.6% to 12% (Figure 2a). Such an improvement could be attributed to the enhanced sensitivity of the 4-CV TIFF method, allowing more low-abundant peptides to be detected and identified. With the 4-CV TIFF method, over 80% of the proteins had no missing values and over 90% had no more than one missing value across the triplicates. Higher percentages of missing data were present with the 2-CV TIFF and standard methods (Figure 2b). To further assess the quantification accuracy of the 4-CV TIFF method, we performed a statistical analysis using samples from two cell types (CMK and K562). Proteins having at least 2 valid values in a given group were considered quantifiable. The 4-CV TIFF method exhibited a total of 2,345 quantifiable proteins that included ~98% of the proteins (1,052) using the standard method (Figure 2c). Because it was possible to quantify proteins more consistently with the TIFF method, we observed 1,053 differentially

abundant proteins (DAPs) (FDR < 0.05 and S_0=0.1) between the CMK and K562 cells, while only about half (i.e., 536 DAPs) were found using the standard method (Supplementary Figure 3a-3b). A total of 380 DAPs were common between the two methods. As shown in Figure 2d, the linear correlation coefficient of protein fold-changes between the two label-free methods is high (R = 0.95). The slope of linear regression is ~1 (*K*), demonstrating similar fold changes between the two methods. In the same way, the 4-CV TIFF method showed improved quantification results over the standard method in the comparison between MOLM14 and the other two cell types (Supplementary Figure 4a-4d and 5a-5d).

Recently, isobaric labeling-based methods have been introduced to increase the throughput of scProteomics. In these methods, single cells are labeled with individual tandem mass tags (TMT) and a high-input boosting sample, generated by labeling the protein digest of > 100 cells [11, 18], is added to the same TMT set. In previous studies, we showed that the boosting sample could reduce the dynamic range and compromise the quantification performance [9, 11]. We correlated the protein fold-changes obtained for CMK and K562 cells using the 4-CV TIFF method and a previous result employing the TMT method using 100× boosting [11] (Figure 2e). Although two very distinct proteomics strategies were used, the coefficient of correlation was relatively robust (R=0.68). However, the slope of linear regression was relatively low (*K*=0.24), indicating that higher fold changes and dynamic range can be obtained with the TIFF method based on label-free quantification. The improvement can be attributed to the absence of ratio compression of the label-free quantification method.

Overall, the results indicate that the TIFF method enables the identification and quantification of more proteins than standard label-free methods and enhances the quantification compared to both standard label-free and TMT-boosting scProteomics methodologies.

A streamlined label-free scProteomics platform

Having demonstrated that the TIFF method offers improvements in proteome coverage and quantification for mass-limited samples, we integrated it into our scProteomics pipeline that includes fluorescence-

activated cell sorting (FACS) for cell isolation [15], a robotically addressed nanowell chip for single-cell processing (nanoPOTS, Nanodroplet Processing in One pot for Trace Samples) [14], a nanoliter-scale LC autosampler for reliable sample injection [10], and a low-flow liquid chromatography system (LC column with 50 μm i.d.) [10]. Both single cells and pooled library cells can be isolated with FACS and processed with nanoPOTS. The integrated FACS-nanoPOTS-autosampler-TIFF-MS platform offers a complete solution from cell isolation to data acquisition for unbiased scProteomics, as well as other biological applications with mass-limited samples.

Proteome coverage of single HeLa cells

Several single-cell proteomics technologies have been previously evaluated using HeLa cells [8, 12, 15, 17]. We next used HeLa cells to benchmark the TIFF-based scProteomics workflow. Single cells were sorted into nanowells and fifty HeLa cells were pooled in microwells for library generation. Using a conventional tandem mass spectrometry approach (MS/MS), an average of 209 proteins were identified from single HeLa cells (Figure 3a). The number is comparable to our previously reported result (211 proteins) using a lower-flow LC-MS system (30 μm i.d. column) but without a FAIMS interface (Supplementary Table 1) [15], and 42% lower than that obtained using an ultra-low-flow LC system (20 μm i.d. column) and the newest generation (Eclipse) MS [8]. The utilization of the 4-CV TIFF method dramatically increased the coverage to an average of 1,212 (± 10%) identified protein across 10 analyses (Figure 3a). More than 75% of the proteins (725/961) identified using a similar LC-MS setup without FAIMS [15] were detected using the TIFF method (Figure 3b). The TIFF method more than doubled the total number of identifications made, reaching 1,771 unique proteins identified. The number of identifications obtained with the TIFF method is comparable to the one we obtained using a 20-μm-i.d. column, a FAIMS interface, an Eclipse MS, and a long LC gradient [12].

The quantification consistency was also evaluated. Using protein iBAQ intensities, 684 out of 1,771 proteins had no missing values across the 10 HeLa cells (Figure 3c). 1,103 proteins were identified in at

least 50% of the analyses. Pearson's correlation coefficients had a median value of 0.95 between any two HeLa cells, indicating high reproducibility of our integrated scProteomics pipeline (Figure 3d). Together, these results indicate that the integration of the TIFF method with high-efficiency single-cell preparation offers a sensitive and reliable scProteomics pipeline for label-free quantification.

Large-scale proteome profiling of single macrophage cells in response to lipopolysaccharide activation

To further evaluate our platform for large-scale scProteomics analysis, we profiled proteome changes of single murine macrophage cells (Raw 264.7) after 24 hr and 48 hr lipopolysaccharide (LPS) stimulation relative to unstimulated cells (control) (Figure 4a). We analyzed a total of 155 individual Raw 264.7 cells including 54 unstimulated cells, 52 24-hr stimulated cells, and 49 48-hr stimulated cells. Our analysis identified a total of 1,671 proteins across the 155 individual cells. The median number of proteins identified per cell was 451. While lower than the number of proteins identified from single HeLa cells described above, we note that Raw 264.7 cells have a median diameter of 10 μm [28] compared to ~17 μm for HeLa cells [13]; the 5-fold difference in cell volume likely accounts for the reduced coverage. We also observed control cells to have fewer identified proteins than LPS-stimulated cells. The median numbers of identified proteins were 307, 482, and 575 for control, 24-hr stimulation, and 48-hr stimulation, respectively (Supplementary Figure 6a). Previous reports have indicated that stimulated Raw 264.7 macrophages increased in size and changed morphology upon LPS stimulation, potentially accounting in part for the difference in identifications [28]. Of the 1,671 identified proteins, 519 were conservatively retained for quantitative analysis after filtering out proteins containing > 50% missing values in at least one experimental condition. Using a UMAP (the uniform manifold approximation and projection)-based dimensional reduction analysis, the 155 individual cells partitioned into three distinct clusters on a two-dimensional plot largely correspond to the three experimental conditions (Figure 4b). Five stimulated cells (3 from 24 hrs and 2 from 48 hrs) are clustered into the control group, indicating a small portion of Raw cells (~5%) are not sensitive to LPS stimulation.

To identify the differentially abundant proteins (DAPs) that drive the partitioning of the three clusters, we performed an ANOVA test analysis (permutation-based FDR <0.001, S_0=5). A total of 250 proteins were significantly modulated across the three groups (Supplementary Figure 6b). Gene ontology analysis results showed that proteins increased in abundance at 24 hr LPS stimulation (Cluster A in Supplementary Figure 6b) were primarily enriched in antigen processing and presentation processes (FE = 37.2 to 157.5, $p < 0.01$). Proteins increased at 24hr LPS stimulation and remained elevated through 48 hr LPS stimulation (Cluster B in Supplementary Figure 6b) were enriched in antigen processing and presentation (FE = 10.6, $p < 0.05$), response to LPS (FE = 2.5, $p < 0.05$), as well as oxidation-reduction (FE = 2.1, $p < 0.01$) processes, which are known to be a critical function of activated macrophage cells . Biological processes enriched in proteins increased after 48-hr LPS stimulation (Cluster C in Supplementary Figure 6b) included those related to protein exit from the endoplasmic reticulum (FE = 61.8, $p < 0.05$) and to foam cell differentiation (FE = 56.1, $p < 0.05$). The latter finding is in line with a previous report on the ability of LPS activated RAW 264.7 macrophages to differentiate into foam cells [29]. Proteins associated with cholesterol storage (FE = 47.5, $p < 0.05$) were also increased in abundance after 48 hr LPS stimulation. Storage of cholesterol ester or triglyceride has been suggested to lead to the formation of foam cells [30].

Beyond functional enrichment analysis, our statistical analysis identified specific proteins previously described as being involved in the response process of macrophage cells to LPS stimulation. For example, immune responsive gene 1 (Irg1), known as a resistance-inducing protein against LPS [31], was up-regulated in macrophage cells exposed to LPS at both 24 and 48 hr (Supplementary Figure 6c). Irg1 is highly expressed during various infections or TLR ligand stimulation in macrophages, which have been reported to regulate macrophage innate immune responses by controlling proinflammatory cytokines [31]. Prostaglandin-endoperoxide synthase 2 (Ptgs2/Cox-2), an important precursor of prostacyclin enzyme which is expressed in macrophages exposed to LPS [32], was also significantly increased in LPS-stimulated macrophage cells (Supplementary Figure 6c). Transitional endoplasmic reticulum ATPase (Vcp, also

called p97) is involved in targeting and translocation of ubiquitinated proteins and the regulation of the inflammatory response in immune cells [33]. We observed increased abundances of Vcp at 24 hr LPS stimulation with a decrease to basal levels at 48 hr LPS stimulation (Supplementary Figure 6c). The perturbation of cellular ubiquitin homeostasis supports the concept that variations in protein ubiquitination may be key to infection by pathogens and are also involved in triggering the defense mechanism of macrophages. Heat shock cognate 71 kDa protein (Hspa8), known to be involved in the presentation of antigenic peptides by major histocompatibility complex (MHC) class II (MHCII) molecules for CD4 + T cells, was significantly increased in LPS stimulated cells (Supplementary Figure 6c) in line with previous studies that also showed this protein to be overexpressed in response to LPS stimulation [34].

Response trajectory analysis of macrophage cells under LPS stimulation

Although the unsupervised UMAP analysis largely discriminated the LPS stimulation phenotypes into three well-defined clusters (Figure 4b), the clusters were not homogeneous, with contaminant single cells cross-populating the clusters and yielding heterogeneous clusters/groupings (Figure 4b). We hypothesized that LPS stimulation may differentially affect individual single cells, leading to the heterogeneous temporal response to LPS stimulation observed in our experiment. To examine the response trajectory of macrophage cells under LPS stimulation, we applied CellTrails [13, 35], which facilitates the analysis of branching trajectories from single-cell expression data. Data from 155 macrophage cells were processed as described in Methods, and a total of 26 protein features were selected for CellTrails analysis to generate the response trajectory map of macrophage cells under LPS stimulation. Response trajectory mapping manifested four cellular subgroups, each of which was distinguished by distinct marker protein sets (Figure 4c). The distribution of single cells in a subgroup S1 comprised 96% of cells from control and 4% of cells from 48 hrs LPS; suggesting a subset of cells from 48 hrs LPS may either be resistant to LPS stimulation or become tolerized to LPS stimulation, returning to a molecular state resembling unstimulated cells, likely along a gradient. The distribution of single cells in S2 comprised 11% of cells

from control, 26% of cells from 24hr LPS, and 63% of cells from 48hr LPS; suggesting this to be an intermediate state in which both pro-inflammatory and anti-inflammatory features coexist [36]. The distribution of single cells in S3 comprised 100% of cells from 48 hrs LPS. The distribution of single cells in S4 comprised 94% of cells from 24hr LPS and 6% of cells from 48 hrs LPS; supporting the concept of tolerized cells transitioning to a molecular state resembling unstimulated control cells along a gradient. The trajectory plot order, beginning from S1, of S1 → S2 → S3 → S4 (Figure 4c) derives from the state similarity of the four subgroups. S2 closest to S1, likely driven by S2 being the only other subgroup with an appreciable proportion of control cells; S3 close to S2, likely due to both being characterized by high proportions 48 hrs LPS cells (63% distribution for S2 and 100% distribution for S3); S4 closer in proximity to S3, likely due to the modest overlap of 48 hrs LPS cells with S3 and the more substantive correlations between S1-S2 and S2-S3. A potential functional interpretation of the above is that initially (at 24 hrs), stimulated macrophages undergo a substantial trajectory change from their control unstimulated state; however, as the stimulation time extends (to 48 hrs), some fraction of cells return to a state closer to baseline (unstimulated control) (Figure 4d) in agreement with the 2D UMAP described above.

Figure 4e shows the proteomic abundance patterns of the 26 features underlying the CellTrails analysis according to each state aligned based on pseudotime classification that lend further credence to this concept. For example, Hnrnpa1, Sec22b, Tmed2, Rpl22, and Rpl35 proteins related to the release of inflammatory mediators have been shown to gradually increase in abundance levels toward the late state[37-39]. Anti-inflammatory related Cfl1, Gng12, Manf, and Ptpn6 proteins show high expression in the S3 state and then decrease again. Notably, S100a10, Rps28, Irg1, and Ptgs2 proteins increase the amount of abundance from S4 to S3, which is the most active period of macrophages, showing that cells at that time have a proinflammatory function[31, 40, 41]. Through trajectory analysis, we found that Arglu1, Nip7, Nudcd2, and Mtmr2 proteins, whose functions are not yet known for the activation of macrophages, elevate in S4 and then gradually decrease with time. The expressions of Cfh and Lmo1, known as

immune suppressors [42], decreased at the beginning of LPS stimulation (S2 and S4), and proteins with similar patterns such as Lgalsl, Atp5i, Cox7a2, and Pde8b proteins were also detected. Overall, CellTrails analysis of scProteomics data predicted the phenotypic change according to the response time of LPS stimulated macrophages, and showed long-term effects of LPS on single macrophages, which are masked in bulk proteomics.

Application to dissociated primary cells from human lung

A powerful capability offered by single-cell omics technologies is their ability to distinguish cell types (including sub-types) without prior knowledge, including classifying cell populations directly from tissue samples obtained from animal models and human clinical samples. While single-cell RNA sequencing is now a well-established method to distinguish and classify cell populations based on transcript abundances [1, 2], a similar capability is lacking for scProteomics. A key benefit of classifying cell populations based on protein abundances via unbiased scProteomics is that it directly provides a protein-level readout of markers for each cell population identified, facilitating isolation of cell populations of interest using antibody-based enrichment approaches for functional characterization. To initially explore the scProteomics platform for cell-type classification from tissue single-cell suspension samples, we analyzed non-depleted and non-labeled primary cells from the lung of a 2-year-old donor (Figure 5a). 19 single cells were processed and analyzed using the TIFF-based scProteomics workflow, resulting in a total of 986 identified proteins with an average of 390 identified proteins per single cell (Supplementary Table 2). We retained proteins identified in at least 8 of the 19 single cells (40% presence) for quantitative analysis, resulting in 402 quantifiable proteins (Supplementary Table 3). PCA analysis of the 402 proteins suggested the presence of at least three cell populations in the lung tissue single-cell suspension (Figure 5b).

To identify proteins distinguishing these populations, we performed the ANOVA test (permutation-based FDR < 0.05, $S_0 = 0$), revealing 99 proteins (~20% of quantifiable proteins) were differentially abundant

across the three cell population groups/clusters (Supplementary Table 4) as visually represented in Figure 5c. Cell type identity was assigned to each of the three cell population groups by comparing markers from the scProteomics data to lung cell type markers previously enumerated by bulk proteomics of sorted cell populations of human lung endothelial, epithelial, immune, and mesenchymal cells [5]. Correspondence across the scProteomic and bulk proteomic markers revealed Cluster 1 represented a lung endothelial cell population, Cluster 2 represented a lung immune cell population, and Cluster 3 represented a lung epithelial cell population (Supplementary Figure 7). For example, Caveolin-1 (CAV1) and Polymerase I and transcript release factors (PTRF), which were highly abundant in single-cell cluster 1 (Figure 5d and Supplementary Figure 7), are known to structurally maintain the specialized lipid raft of caveola in lung endothelial cells [43]. L-Plastin (LCP1) protein, important for alveolar macrophage development and antipneumococcic response [44], was highly abundant in bulk sorted immune cells as well as Cluster 2 (Figure 5d and Supplementary Figure 7). Pulmonary surfactant associated protein B (SFTPB) which facilitates alveolar stability by modulating surface tension [45] is known to be preferentially enriched in lung epithelial cells. SFTPB was highly abundant in bulk sorted epithelial cells as well as Cluster 3 (Figure 5d and Supplementary Figure 7). The above results demonstrate the utility of the scProteomics platform for cell-type classification from non-depleted whole tissue single-cell suspension samples.

Discussion

In this study, we developed an ion mobility-enhanced MS acquisition method, TIFF (Transferring Identification based on the FAIMS Filtering), and coupled with our previously described nanoPOTS scProteomics workflow [10, 14] to improve the sensitivity and accuracy of unbiased scProteomics. MS acquisition efficiency was significantly improved by filtering out singly charged background ions and allowing elongated ion accumulation to enable ultrasensitive detection. Compared with our previous FAIMS-based scProteomics workflow using an ultra low-flow LC column (20-μm-i.d.) and long gradient,[12] the TIFF method dramatically improved both system robustness and analysis throughput to enable large-scale single-cell studies. Our TIFF-based workflow enabled the identification of >1,700

proteins and quantification of ~1,100 proteins from single HeLa cells in a label-free analysis. We demonstrated the robustness and scalability of the scProteomics workflow via a large-scale analysis of 155 single macrophage cells under different LPS stimulation conditions to reveal the biological processes at single-cell level. Finally, we demonstrated the capability for the TIFF-based scProteomics workflow to classify cell populations of a human organ (the lung).

While our label-free analysis of single cultured cells (e.g. HeLa) yielded >1000 proteins identified and similar numbers of proteins quantified, a similar analysis of single primary cells (e.g. human lung cells) resulted in the identification of significantly fewer proteins, presumably due to the fact that culture conditions enhance protein productions. This again highlights the need to further improve the overall sensitivity of current scProteomics platforms to enable routine deep single-cell proteome analyses of primary cells derived from tissues of animal models and human donors. One strategy for improving overall sensitivity is by further improving protein/peptide recovery. Protein/peptide recovery during sample processing procedures could be increased using smaller nanowells or low-binding surfaces to reduce adsorptive loss. Another strategy for improving overall sensitivity is through enhancing peptide separation resolution and ionization efficiency. With the advance of nanoLC pump technologies, the LC flow rates could be reduced to low nanoliter and to even picoliter-scale to further enhance peptide separation resolution and ionization efficiency. MS instrumentation with high ion-transmission optics and sensitive detectors could provide further enhancements in proteome coverage for single cells. In addition to FAIMS, other ion mobility-based technologies including trapped ion mobility spectrometry (TIMS)[46, 47], and particularly, structures for lossless ion manipulation (SLIM)[48, 49] can offer improved ion separation and overall ion utilization efficiencies. With all these developments, we believe the proteome depths of scProteomics will reach the level of single-cell RNA sequencing and eventually become an indispensable tool in biological and medical researches.

Acknowledgments

We thank the LungMAP HTC for providing dissociated primary human lung cells (U01HL148860 to G.P.). We thank the insightful discussions from Aman Makaju at ThermoFisher Scientific (San Jose, CA) on the FAIMS interface. This work was supported by a Laboratory Directed Research and Development award (I3T) from Pacific Northwest National Laboratory (Y.Z.) and the NIH grants U01 HL122703 (C.A.) and P41 GM103493 (R.D.S.). The research was also supported by the Intramural program at EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

Contributions

J. W., G. C., L. P. T., C. A., and Y. Z. proposed the method and designed the research. J.W., G. C., C.F.T., S.M.W, R. J. M., W. B. C, and Y. Z. performed cell culture, FACS sorting, proteomic sample preparation, and LC-MS experiments. J. W. and Y. Z. optimized the TIFF method. J.W., G. C., S. F., C. A., and Y. Z. analyzed the data. J. W., G. C., S. F., R. D. S., R. T. K., L. P. T., C. A., and Y. Z. wrote the manuscript.

Competing interests The authors declare they have no competing interests.

Data Availability The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifier MSV000085937.

References

1. Wilson, N.K. et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16**, 712-724 (2015).
2. Hou, Y. et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* **26**, 304-319 (2016).

3. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat Rev Genet* **20**, 257-272 (2019).
4. Qian, X. et al. Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* **165**, 1238-1254 (2016).
5. Du, Y. et al. Integration of transcriptomic and proteomic data identifies biological functions in cell populations from human infant lung. *Am J Physiol Lung Cell Mol Physiol* **317**, L347-L360 (2019).
6. Buettner, F. et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* **33**, 155-160 (2015).
7. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**, 161 (2018).
8. Cong, Y. et al. Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry. *Anal Chem* **92**, 2665-2671 (2020).
9. Dou, M. et al. High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. *Anal Chem* **91**, 13119-13127 (2019).
10. Williams, S.M. et al. Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatography-Mass Spectrometry for High-Throughput Single-Cell Proteomics. *Anal Chem* (2020).
11. Tsai, C.F. et al. An improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) strategy for precise quantitative single-cell proteomics. *Mol Cell Proteomics* **19**, 828-838 (2020).
12. Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *bioRxiv*, 132449 (2020).
13. Zhu, Y. et al. Single-cell proteomics reveals changes in expression during hair-cell development. *Elife* **8**, e50777 (2019).
14. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat Commun* **9**, 882 (2018).
15. Zhu, Y. et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angew Chem Int Ed Engl* **57**, 12370-12374 (2018).
16. Shao, X. et al. Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. *Anal Chem* **90**, 14003-14010 (2018).
17. Li, Z.Y. et al. Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. *Anal Chem* **90**, 5430-5438 (2018).
18. Specht, H. et al. Single-cell mass-spectrometry quantifies the emergence of macrophage heterogeneity. *bioRxiv*, 665307 (2019).
19. Xiang, P. et al. Picoflow Liquid Chromatography-Mass Spectrometry for Ultrasensitive Bottom-Up Proteomics Using 2-mum-i.d. Open Tubular Columns. *Anal Chem* **92**, 4711-4715 (2020).
20. Makarov, A. & Denisov, E. Dynamics of ions of intact proteins in the Orbitrap mass analyzer. *J Am Soc Mass Spectrom* **20**, 1486-1495 (2009).
21. Amenson-Lamar, E.A., Sun, L., Zhang, Z., Bohn, P.W. & Dovichi, N.J. Detection of 1zmol injection of angiotensin using capillary zone electrophoresis coupled to a Q-Exactive HF mass spectrometer with an electrokinetically pumped sheath-flow electrospray interface. *Talanta* **204**, 70-73 (2019).
22. Meier, F., Geyer, P.E., Virreira Winter, S., Cox, J. & Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat Methods* **15**, 440-448 (2018).
23. Pasa-Tolic, L. et al. Increased proteome coverage for quantitative peptide abundance measurements based upon high performance separations and DREAMS FTICR mass spectrometry. *J Am Soc Mass Spectrom* **13**, 954-963 (2002).
24. Luber, C.A. et al. Quantitative proteomics reveals subset-specific viral recognition in dendritic cells. *Immunity* **32**, 279-289 (2010).

25. Pasa-Tolic, L., Masselon, C., Barry, R.C., Shen, Y. & Smith, R.D. Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* **37**, 621-624, 626-633, 636 passim (2004).
26. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).
27. Prianichnikov, N. et al. MaxQuant Software for Ion Mobility Enhanced Shotgun Proteomics. *Molecular & Cellular Proteomics* **19**, 1058-1069 (2020).
28. Saxena, R.K., Vallyathan, V. & Lewis, D.M. Evidence for lipopolysaccharide-induced differentiation of RAW264.7 murine macrophage cell line into dendritic like cells. *J Biosci* **28**, 129-134 (2003).
29. Funk, J.L., Feingold, K.R., Moser, A.H. & Grunfeld, C. Lipopolysaccharide stimulation of RAW 264.7 macrophages induces lipid accumulation and foam cell formation. *Atherosclerosis* **98**, 67-82 (1993).
30. Feingold, K.R. et al. Mechanisms of triglyceride accumulation in activated macrophages. *J Leukoc Biol* **92**, 829-839 (2012).
31. Li, Y. et al. Immune responsive gene 1 (IRG1) promotes endotoxin tolerance by increasing A20 expression in macrophages through reactive oxygen species. *J Biol Chem* **288**, 16225-16234 (2013).
32. Tang, T. et al. Macrophage responses to lipopolysaccharide are modulated by a feedback loop involving prostaglandin E2, dual specificity phosphatase 1 and tristetraprolin. *Sci Rep* **7**, 4350 (2017).
33. Fenech, E.J. et al. Interaction mapping of endoplasmic reticulum ubiquitin ligases identifies modulators of innate immune signalling. *Elife* **9**, e57306 (2020).
34. Zhang, A., Zhou, X., Wang, X. & Zhou, H. Characterization of two heat shock proteins (Hsp70/Hsc70) from grass carp (Ctenopharyngodon idella): evidence for their differential gene expression, protein synthesis and secretion in LPS-challenged peripheral blood lymphocytes. *Comp Biochem Physiol B Biochem Mol Biol* **159**, 109-114 (2011).
35. Ellwanger, D.C., Scheibinger, M., Dumont, R.A., Barr-Gillespie, P.G. & Heller, S. Transcriptional Dynamics of Hair-Bundle Morphogenesis Revealed with CellTrails. *Cell Rep* **23**, 2901-2914 e2913 (2018).
36. O'Carroll, C., Fagan, A., Shanahan, F. & Carmody, R.J. Identification of a unique hybrid macrophage-polarization state following recovery from lipopolysaccharide tolerance. *J Immunol* **192**, 427-436 (2014).
37. Das, A.S. et al. Post-transcriptional regulation of C-C motif chemokine ligand 2 expression by ribosomal protein L22 during LPS-mediated inflammation. *FEBS J* (2020).
38. Descoteaux, A. et al. The secretory pathway-resident SNARE Sec22b regulates nitric oxide and cytokine production in dendritic cells. *The Journal of Immunology* **204**, 229.224-229.224 (2020).
39. Maurel, M. et al. Control of anterior GRadient 2 (AGR2) dimerization links endoplasmic reticulum proteostasis to inflammation. *EMBO Mol Med* **11**, e10120 (2019).
40. Lou, Y. et al. Essential roles of S100A10 in Toll-like receptor signaling and immunity to infection. *Cell Mol Immunol* **17**, 1053–1062 (2020).
41. Zhou, X., Liao, W.J., Liao, J.M., Liao, P. & Lu, H. Ribosomal proteins: functions beyond the ribosome. *J Mol Cell Biol* **7**, 92-104 (2015).
42. Weismann, D. et al. Complement factor H binds malondialdehyde epitopes and protects from oxidative stress. *Nature* **478**, 76-81 (2011).
43. Chettimada, S., Yang, J., Moon, H.G. & Jin, Y. Caveolae, caveolin-1 and cavin-1: Emerging roles in pulmonary hypertension. *World J Respirol* **5**, 126-134 (2015).
44. Deady, L.E. et al. L-plastin is essential for alveolar macrophage production and control of pulmonary pneumococcal infection. *Infect Immun* **82**, 1982-1993 (2014).
45. Wang, Y. et al. Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. *Proc Natl Acad Sci U S A* **115**, 2407-2412 (2018).

46. Vasilopoulou, C.G. et al. Trapped ion mobility spectrometry and PASEF enable in-depth lipidomics from minimal sample amounts. *Nat Commun* **11**, 331 (2020).
47. Michelmann, K., Silveira, J.A., Ridgeway, M.E. & Park, M.A. Fundamentals of trapped ion mobility spectrometry. *J Am Soc Mass Spectrom* **26**, 14-24 (2015).
48. Wojcik, R. et al. SLIM Ultrahigh Resolution Ion Mobility Spectrometry Separations of Isotopologues and Isotopomers Reveal Mobility Shifts due to Mass Distribution Changes. *Anal Chem* **91**, 11952-11962 (2019).
49. Nagy, G. et al. Towards resolving the spatial metabolome with unambiguous molecular annotations in complex biological systems by coupling mass spectrometry imaging with structures for lossless ion manipulations. *Chem Commun (Camb)* **55**, 306-309 (2019).

Materials and Methods

Cell culture and single-cell sorting

All cell lines used in this study were maintained in a medium compatible with each cell line and incubated at 37 ℃ with 5% of CO_2. Of the three Acute Myeloid Leukemia (AML) cell lines, K562 and MOLM14 cells were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), and CMK cells were maintained in RPMI-1640 medium with 20% FBS added. For HeLa cells, DMEM supplemented with 10% FBS was added. RAW 264.7 cells were maintained in DMEM supplemented with 10% FBS followed to be stimulated with 100 ng/ul of LPS (Sigma Aldrich) in serum-free DMEM (Thermo Fisher scientific) for 24 hr or 48 hr. For the control of Raw264.7 cells (non-treated), ten million cells were collected before stimulation with LPS. In the same way, LPS stimulated cells were harvested after 24 hr or 48 hr of treatments. HeLa and RAW.264.7 cells were washed by chilled PBS and sorted on the nanoPOTS chips (4 × 12, 1.2 mm diameter per each well) using the Influx II cell sorter (BD Biosciences, San Jose, CA) as described previously [15]. To build the in-depth spectral library, 50 cells of each cell line (or equivalent peptides of ~10 ng) were loaded onto the microPOTS chip (3 × 9, 2.2-mm diameter per well).

Protein digestion

For the low-input mock samples (0.2 ng, equivalent amount peptides to a single-cell), AML cell lines were lysed in a tube with lysis buffer including 50 mM NH_4HCO_3 (pH8.0), 8 M UREA, and 1 % phosphatase inhibitor followed by sonicated in a cold bath for 3 min. After the measurements of the protein concentrations by BCA assay (Thermo Fisher Scientific), proteins equivalent to 200 μg were reduced in 5 mM DTT for 1 hr at 37 ℃ and alkylated with 10 mM iodoacetamide (IAA) in the dark for 1 hr at room temperature. Eight-fold diluted samples with 50 mM NH_4HCO_3 were digested with Lys-C peptidase at 37 ℃ with a ratio of 50:1 (w/w) for 3 hr followed by digesting with trypsin with a ratio of 50:1 (w/w) at 37 ℃ overnight. The tryptic digested peptides were acidified by 0.5% trifluoroacetic acid

(TFA) at final concentration, then desalted using C18 SPE tips. After concentrated, the BCA assay was performed to estimate the final concentration of the peptides. Using the nanoPOTS robot, 0.2 ng and 10 ng of the peptides from each AML cell line were loaded on the nanowell/microwell chips and completely dried by a vacuum system [10].

For single-cell analysis, single and 50 FACS-sorted cells on the chip were processed on the nanoPOTS platform for single cells and spectral library, respectively. To extract proteins, we first added a lysis buffer containing 0.2% n-Dodecyl b-D-maltoside (DDM) and 5 mM DTT in 0.5× PBS and 25 mM NH_4HCO_3 buffer in each well, then incubated for 1 hr at 70 ℃. Denatured and reduced proteins were alkylated with 10 mM IAA in the dark for 30 min at RT. Double enzymatic digestions were performed by incubating with LysC (1 ng for single-cell, 5 ng for 50 cells) for 4 hr at 37 ℃ followed by treatment with trypsin (2 ng for single-cell, 10 ng for 50 cells) overnight. Peptides were acidified with 5% formic acid and completely dried using a vacuum system. All chips were stored in a -20 ℃ freezer until MS analysis.

LC-FAIMS-MS/MS analysis

In-house assembled nanoPOTS autosampler with an in-house packed SPE column (100 μm i.d., 4 cm, 5 μm, 300 Å C18 material, Phenomenex) and an LC column (50 μm i.d., 25 cm, 1.7 μm, 190 Å C18 material, Waters) heated to 50 ℃ using AgileSleeve column heater (Analytical Sales and services, Inc., Flanders, NJ) was used for sample analysis [10]. Briefly, samples were dissolved with Buffer A (0.1% formic acid in water) on the chip, then trapped on the SPE column for 5 min. After washing the peptides, samples were eluted at 100 nL/min and separated using a 60-min gradient from 8% to 35% Buffer B (0.1% formic acid in acetonitrile).

An Orbitrap Fusion Lumos Tribrid MS (Thermo Scientific) operated in data-dependent acquisition mode was used for all analyses. Peptides were ionized by applying a voltage of 2,000 V or 2,400 V for standard or FAIMS methods, respectively.

For the standard method, precursor ions with mass range 375-1600 m/z were scanned at 120,000 resolution with an ion injection time (IT) of 254 ms and an AGC target of 1E6. To analyze pooled samples for generating the spectral libraries, the selected precursor ions with +2 to +7 charges were fragmented by a 30% level of high energy dissociation (HCD) and scanned at 60,000 resolution with an IT of 118 ms and an AGC target of 1E5. When single-cell level (0.2 ng) peptides were injected, fragmented peptide ions were scanned at 120,000 resolution with an IT of 246 ms and an AGC target of 1E5.

For the TIFF method, the ionized peptides were fractionated by the FAIMSpro interface using a 2-CV (-45, -65 V) method or a 4-CV (-45, -55, -65, -75 V) method. Fractionated ions with a mass range 350-1500 m/z were scanned at 120,000 resolution with an IT of 254 ms and an AGC target of 1E6. For the pooled samples for generating a spectral library, a single CV was used for each LC-MS run. Precursor ions with intensities > 1E4 were selected for fragmentation by 30% HCD and scanned in an Ion trap with an AGC of 2E4 and an IT of 150 ms. For single-cell samples, cycle times of 1.5 s and 0.6 s were used for the 2-CV and 4-CV methods, respectively. Precursor ions with intensities > 1E4 were fragmented by 30% HCD and scanned with an AGC of 2E4 and an IT of 254 ms.

Data analysis

All raw files were processed by MaxQuant (Ver. 1.6.2.10) with the Uniport protein sequence database of *homo sapiens* (Downloaded in 03/12/2020 containing 20,364 reviewed sequences) and of *mus musculus* (Downloaded in 5/19/2020 containing 17,037 reviewed sequences) using the Andromeda search engine with a 6-ppm precursor ion tolerance after mass calibration [26]. Protein acetylation in N-terminal and oxidation at methionine were chosen as variable modifications. Carbamidomethylation of cysteine residues was set as a fixed modification. Both proteins and peptides were filtered with a false discovery rate (FDR) less than 0.01. Match between runs algorithm in Maxquant was activated with a matching window of 0.4 min and alignment windows of 10 min. For raw files with multiplex FAIMS CVs, we

converted them to multiple mzxml files corresponding to separate individual CVs using an in-house converting tool (https://github.com/PNNL-Comp-Mass-Spec/FAIMS-MzXML-Generator/releases). Those separated files were assigned to non-adjacent fractionation numbers (e.g. 1, 3, 5, 7) during the Maxquant search to ensure feature matching only occurs between the files with the same CV.

For label-free quantification of single-cell-level peptides (0.2 ng) for three AML cell lines and dissociated human lung single-cell, Perseus (Ver. 1.6.12.0) was utilized for the data clean and statistical analysis. The iBAQ algorithm was used for the single-cell analysis because the iBAQ values are proportional to the molar quantities of the proteins. We log2 transformed the iBAQ values after filtering out contaminants and reverse identifications. Missing values were imputed based on a standard distribution (width: 0.3, downshift: 1.8) to simulate signals for low-abundance proteins. Data were normalized using width adjustment, which subtracts medians and scales for all values in a sample to show equal interquartile ranges. Two-way t-tests were performed for the pairwise comparison of the AML cell lines proteomes utilizing the threshold of Benjamini-Hochberg FDR < 0.05 and S_0=0.1, while ANOVA tests were employed for multiple sample tests of dissociated human lung single cells with Permutation based FDR < 0.05. To clarify cell populations from dissociated lung cells, multiple steps including principal components analysis (PCA) and hierarchical clustering were employed using Perseus. Gene ontology analysis for the biological process of the molecules was performed in DAVID web-based bioinformatic tools (database version 6.8, https://david.ncifcrf.gov/summary.jsp).

The processing of the macrophage single-cell data was performed using an R package; RomicsProcessor v1.1.0 (https://github.com/PNNL-Comp-Mass-Spec/RomicsPro). Briefly, the "proteingroups.txt" output file of the MaxQuant search was imported as a multilayered R object with its associated metadata to extract iBAQ values of the identified proteins. The iBAQ values were then log2 transformed and filtered to allow maximal missingness of 50% within at least one given condition. After median normalization, batch correction was applied to remove the batch effects between chips using ComBat algorithm from the SVA package (v3.36.0). The missing values were imputed using the function of imputeMissing() and

UMAP (the uniform manifold approximation and projection)-based dimensional reduction analysis was performed using the romicsUmapPlot() function in the RomicsProcessor package. For the statistics, ANOVA test was applied with a Benjamini-Hochberg FDR < 0.001 and a S_0=5; we applied a highly significant level to a large number of macrophage cells data in which the group was clearly distinguished by the duration of LPS treatment to give a statistical role to the difference between the median value.

For trajectory analysis of stimulated macrophage cells using CellTrails (V1.7.1) [35], the missing values after batch correction were replaced by zero to avoid imputation artifacts. The most variable features were filtered by three steps: detection levels (threshold=2), coefficient of variation (threshold=0.5) and Fano factors (threshold=1.7); 26 proteins with higher variance were kept for the response trajectory analysis. Then the selected proteins were used to fit four states (*p-value*<0.001 and S_0=1) and the trajectory was constructed based on the four states.

Supplementary Materials

Supplementary Figure 1. Comparison of different MS acquisition methods

Supplementary Figure 2. Benchmarking of detection sensitivity using different MS acquisition methods

Supplementary Figure 3. Quantification of single-cell-level peptides between CMK and K562 cell lines

Supplementary Figure 4. Quantification of single-cell level peptides between K562 and MOLM14 cell lines

Supplementary Figure 5. Quantification of single-cell level peptides between CMK and MOLM14 cell lines

Supplementary Figure 6. ScProteomics analysis of LPS-stimulated single macrophage cells

Supplementary Figure 7. Comparison of quantitative protein markers for human lung cells.

Supplementary Table 1. Comparison of the number of identified proteins in single mammalian cells

Supplementary Table 2. List of identified proteins from 19 single lung cells

Supplementary Table 3. List of 402 quantifiable proteins of 19 single lung cells

Supplementary Table 4. A list of statistically significantly abundant proteins classifying three cell populations

Figures and Table

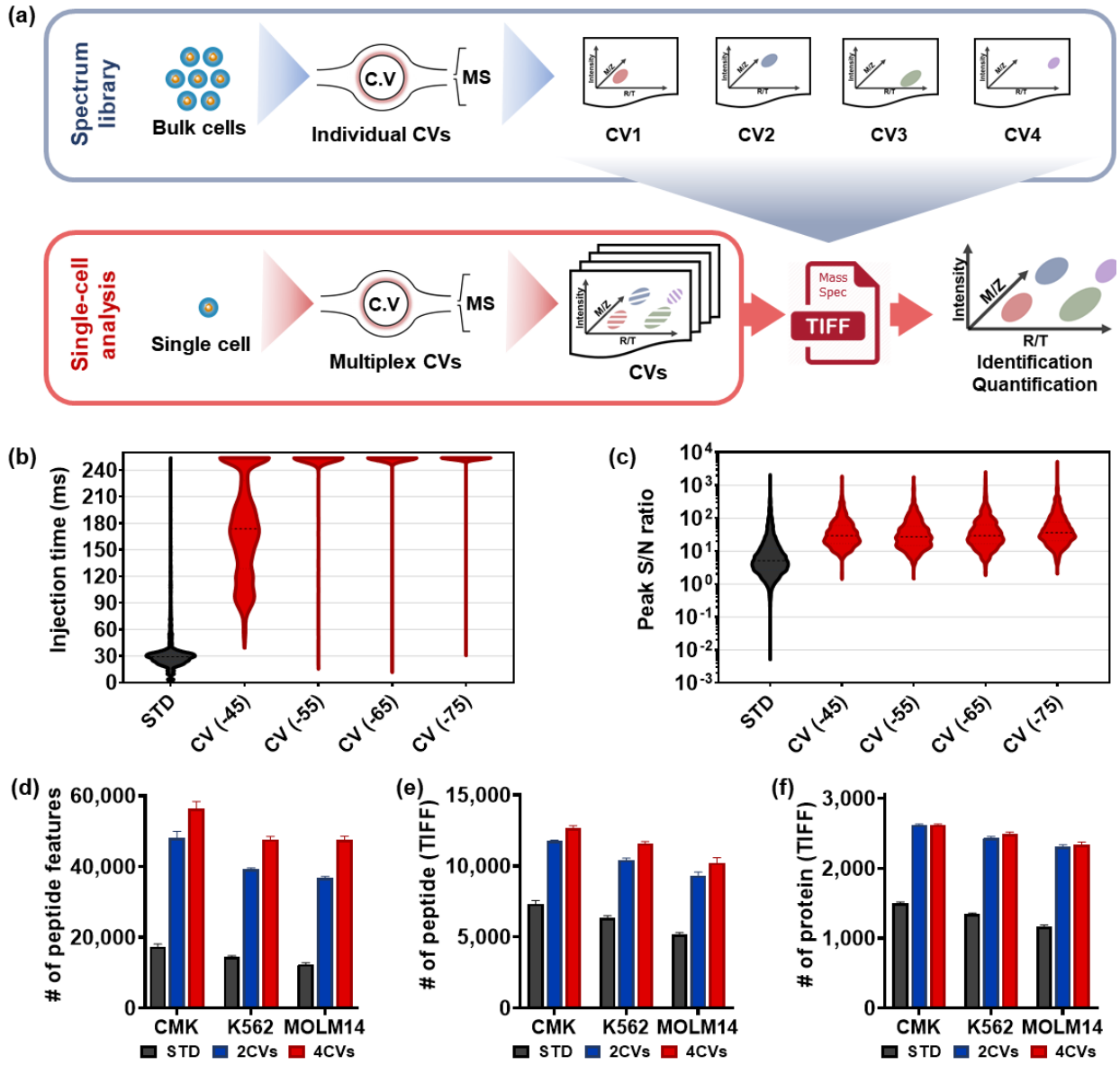

Figure 1. The concept of TIFF method.

(a) Workflow of the TIFF method (Transferring Identification based on FAIMS Filtering). High-input samples (usually from 50 to100 cells) are analyzed by LC-FAIMS-MS with each LC-MS analysis utilizing a discrete FAIMS CV to generate a spectral library; Single-cell samples are analyzed by cycling through multiple FAIMS CVs for each LC-MS analysis. Peptide features in single cells are identified by

matching to the spectral library based on three-dimensional tags including retention time, m/z, and FAIMS CVs. **(b)** Injection time distributions of MS1 for the single-cell level peptides (0.2 ng, CMK cell) in the standard MS (STD) method and FAIMS method with four different CVs. **(c)** The distributions of signal to noise ratios (S/N) of LC-MS features for the 0.2-ng peptides in STD run or FAIMS run with 4 CVs. **(d-f)** The mean number of peptide features (charge>+1), unique peptides, and unique proteins using single-cell level (0.2 ng) peptide digests from three cell lines (CMK, K562, and MOLM14). Standards deviation error bars include triplicate of the sample. Benchmarking was performed between samples analyzed by the standard method, 2-CV TIFF (-45 and -65 V), and a 4-CV TIFF (-45, -55, -65 and -75 V) methods.

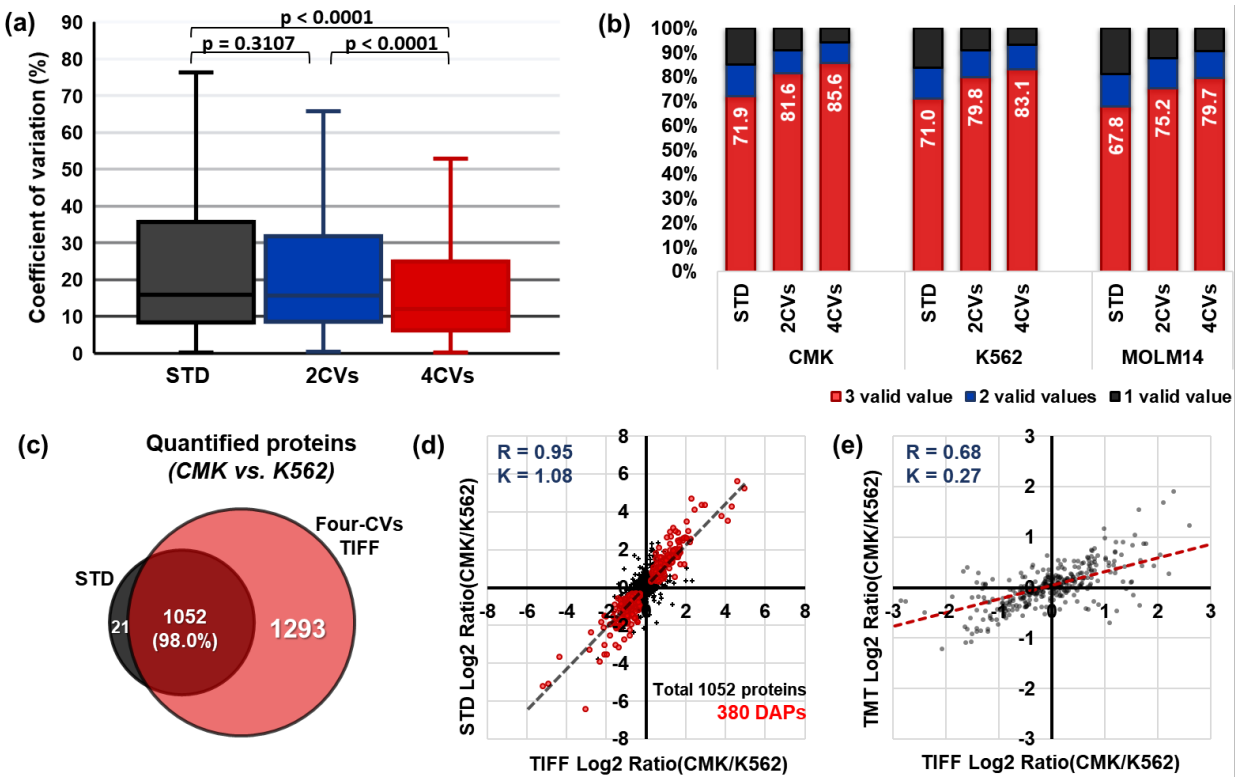

Figure 2. The evaluation of quantification performance of the TIFF method

(a) Distributions of the coefficient of variations for quantified proteins using the three data acquisition methods (STD, 2-CV-TIFF, and 4-CV-TIFF). **(b)** Percentage of valid values using the three methods. Red, blue, and black color indicate the percentages of proteins with valid values of 3, 2, and 1 across the triplicate, respectively. **(c)** Overlap of quantifiable proteins between CMK and K562 samples measured by standard and 4-CV-TIFF methods. **(d)** The linear correlation of log2-transformed fold changes of CMK and K562 cells between the TIFF method (4 CVs) and the STD method. Red dots indicate differentially abundant proteins (DAPs) in both methods calculated by t-test (FDR<0.05, S_0=0.1). **(e)** The linear correlation of the protein fold changes of CMK and K562 cells between the TIFF method and a TMT method [11]. The trendline on the plots (d and e) was generated by Excel with linear regression.

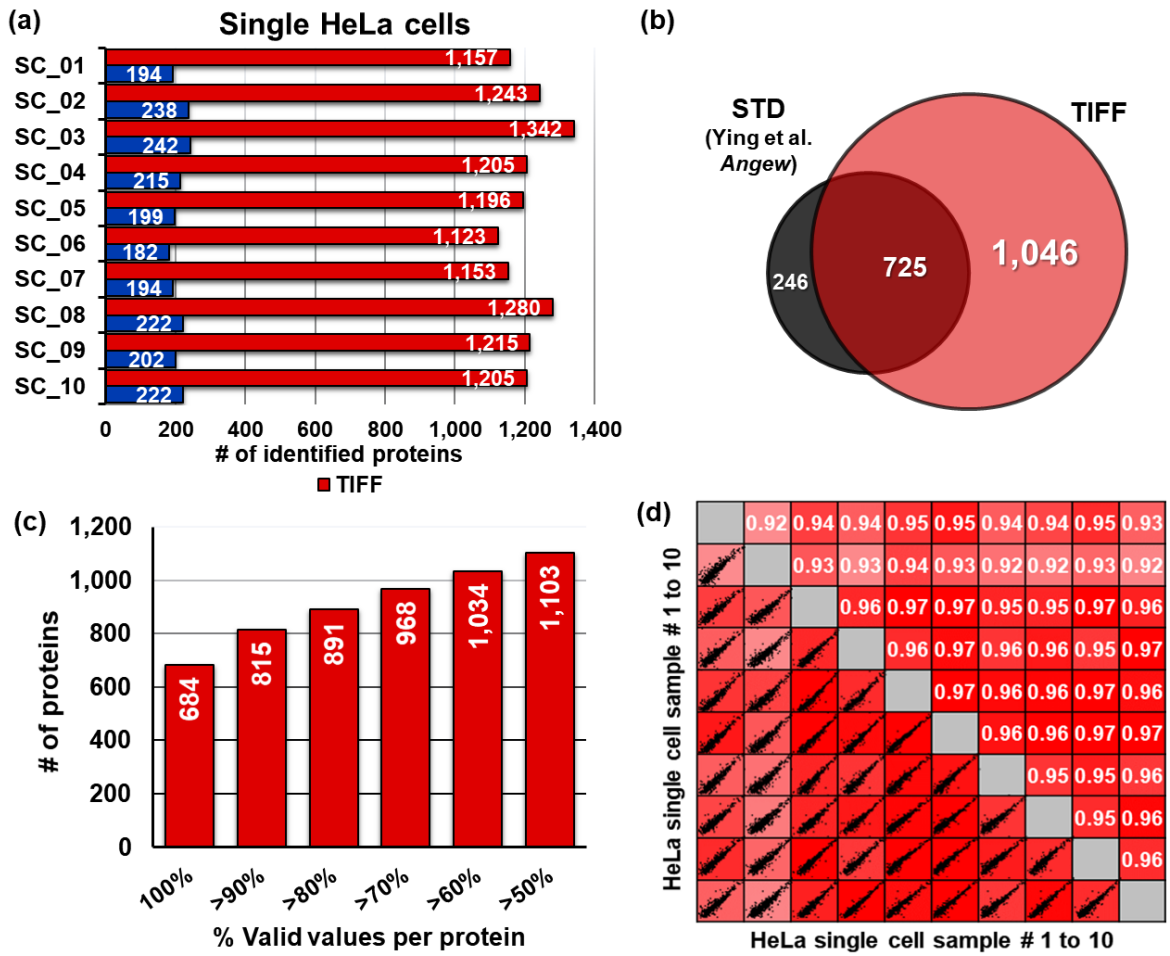

Figure 3. ScProteomics of HeLa cells using TIFF method

(a) Number of protein groups in single HeLa cells identified by MS/MS only (blue) and by the 4-CV TIFF method (red). **(b)** The overlap of identified proteins in single HeLa cells obtained in this study and a previous study with a similar LC-MS setting but without FAIMS [15]. **(c)** The numbers of proteins having valid values from 50% to 100% across the 10 single cells. **(d)** Pair-wise correlations of protein iBAQ intensities between the 10 cells. Proteins containing >70% valid values were required.

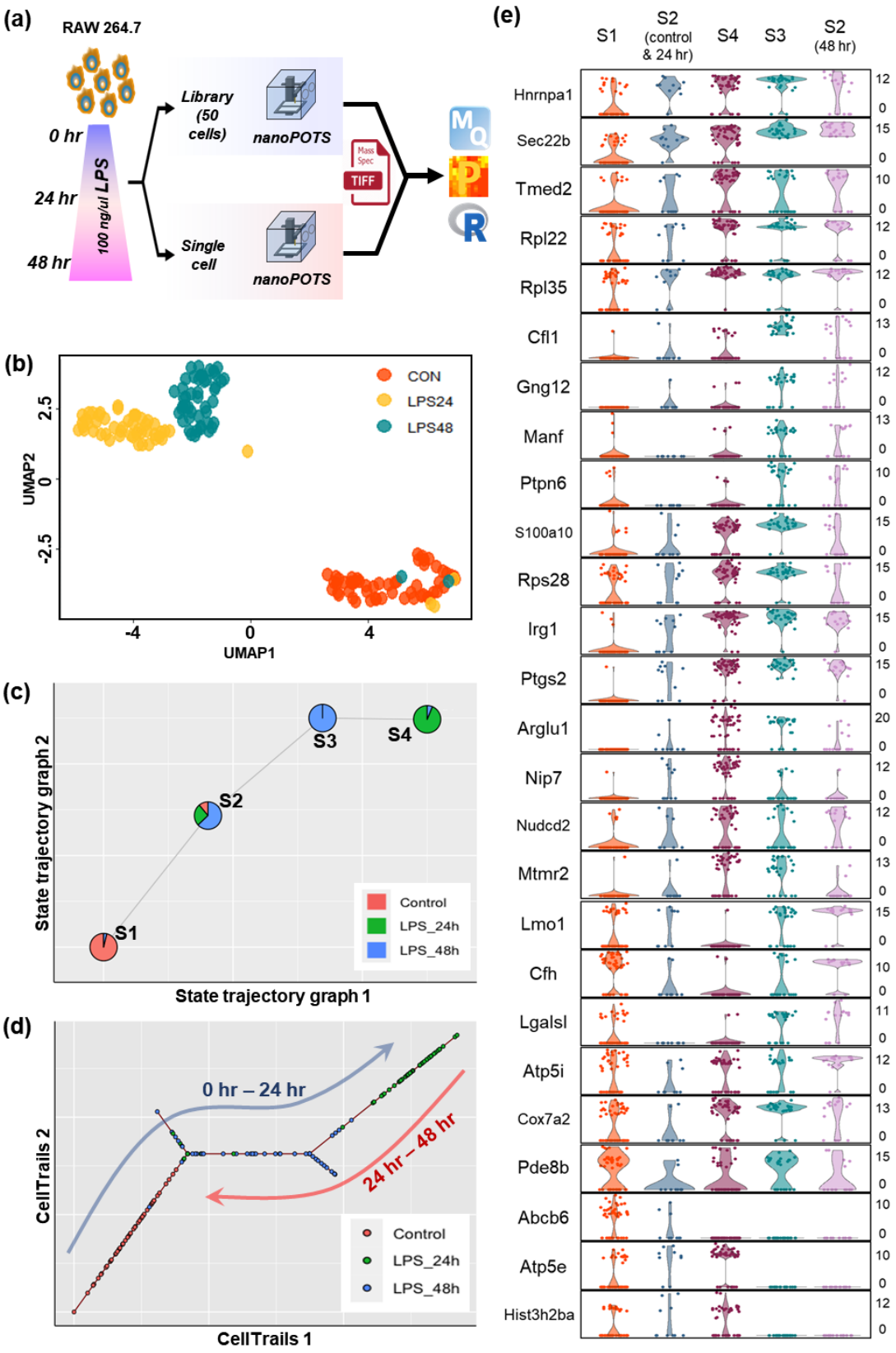

Figure 4. Large-scale proteomics analysis of single macrophage cells

(a) Illustration of workflow for scProteomics analysis of 155 macrophages containing the control cells and the cells stimulated by LPS for 24 and 48 hrs. **(b)** Visualizing dimensionality reduction of 155 single cells using unsupervised UMAP analysis. **(c)** Four representative states with the integrative cell conditions are placed in the area based on CellTrails analysis. **(d)** Trajectory mapping of the 155 single macrophage cells as a function of pseudotime indicating bifurcated changes of cell states. **(e)** Violin plot overview of abundance levels of 26 protein features used in CellTrails analysis in each state.

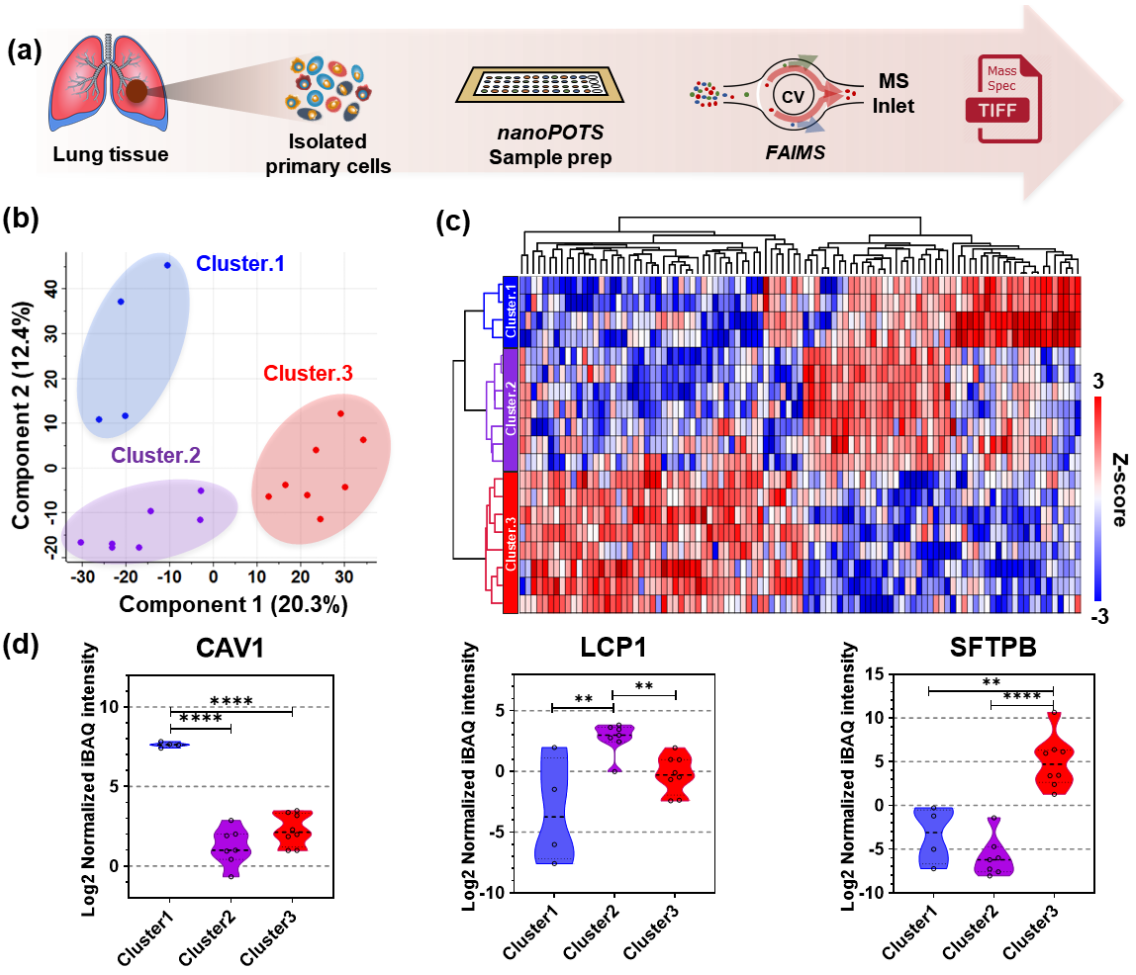

Figure 5. ScProteomics for classifying cell populations of a human lung.

(a) Schematic workflow for scProteomic analysis of dissociated human lung tissue from a 2-year old donor. The tissue was dissociated into the single cells followed by applying the scProteomics pipeline including FACS isolation, nanoPOTS processing, autosampler-LC, and TIFF MS method. **(b)** PCA plot of un-defined cell types. **(c)** Heatmap of differentially abundant proteins (DAPs) by ANOVA test. **(d)** Representative proteins of putative cell types. Cluster 1, 2, and 3 were predicted as lung endothelial, immune, and epithelial cells, respectively (**<0.01, ****<0.0001).