HHS Public Access

Author manuscript

Anal Chem. Author manuscript; available in PMC 2021 August 04.

Published in final edited form as:

Anal Chem. 2020 August 04; 92(15): 10588–10596. doi:10.1021/acs.analchem.0c01551.

Automated Coupling of Nanodroplet Sample Preparation with Liquid Chromatography-Mass Spectrometry for High-Throughput Single-Cell Proteomics

Sarah M. Williams[1], **Andrey V. Liyu**[1], **Chia-Feng Tsai**[2], **Ronald J. Moore**[2], **Daniel J. Orton**[2], **William B. Chrisler**[2], **Matthew J. Gaffrey**[2], **Tao Liu**[2], **Richard D. Smith**[2], **Ryan T. Kelly**[1,3], **Ljiljana Pasa-Tolic**[1], **Ying Zhu**[*,1]

[1]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354 USA

[2]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354 USA

[3]Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, 84602, USA

Abstract

Single-cell proteomics can provide critical biological insight into the cellular heterogeneity that is masked by bulk-scale analysis. We have developed a nanoPOTS (nanodroplet processing in one pot for trace samples) platform and demonstrated its broad applicability for single-cell proteomics. However, because of nanoliter-scale sample volumes, the nanoPOTS platform is not compatible with automated LC-MS systems, which significantly limits sample throughput and robustness. To address this challenge, we have developed a nanoPOTS autosampler allowing fully automated sample injection from nanowells to LC-MS systems. We also developed a sample drying, extraction, and loading (DEL) workflow to enable reproducible and reliable sample injection. The sequential analysis of 20 samples containing 10-ng tryptic peptides demonstrated high reproducibility with correlation coefficients of > 0.995 between any two samples. The nanoPOTS autosampler can provide analysis throughput of 9.6, 16, and 24 single cells per day using 120 min, 60 min, and 30 min LC gradients, respectively. As a demonstration for single-cell proteomics, the autosampler was first applied to profiling protein expression in single MCF10A cells using a label-free approach. At a throughput of 24 single cells per day, an average of 256 proteins was identified from each cell and the number was increased to 731 when the Match Between Runs algorithm of MaxQuant was used. Using a multiplexed isobaric labeling approach (TMT-11plex), ~77 single cells could be analyzed per day. We analyzed 152 cells from three acute myeloid leukemia cell

[*]**Corresponding Author**: Ying Zhu, ying.zhu@pnnl.gov.

[†]Author Contributions

Y. Z., A. L., and R. J. M designed and assembled the autosampler system. S. M. W., C. T., R. J. M., D. J. O., W. B. C., M. J. G., and Y. Z. performed the sample preparation and LC-MS experiments. S. M. W., C. T., R. T. K., T. L., R. D. S., L. P. T. and Y. Z. wrote the manuscript. All authors have given approval to the final version of the manuscript.

DATA AVAILABILITY

The proteomics raw data generated in this work have been deposited to the ProteomeXchange Consortium via MassIVE (https://massive.ucsd.edu) with the dataset identifier MSV000085230.

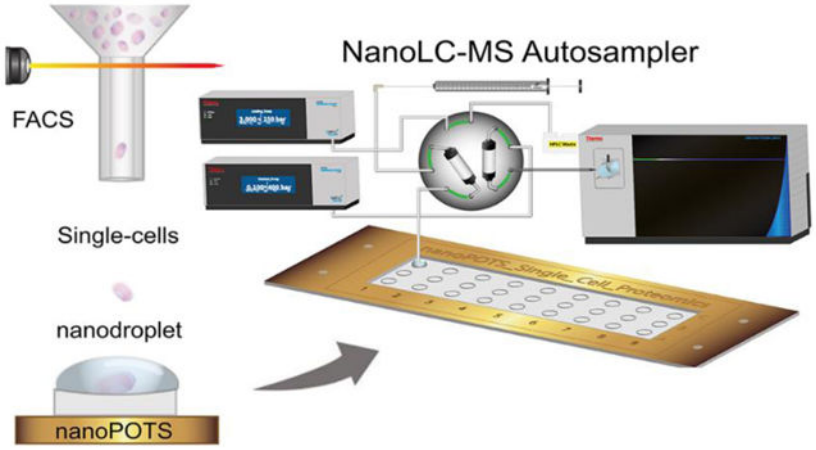
SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website.

Supporting Information PDF file including figures Figure S1-S3 and Table S1-S2.

lines, resulting in a total of 2,558 identified proteins with 1465 proteins quantifiable (70% valid values) across the 152 cells. These data showed quantitative single-cell proteomics can cluster cells to distinct groups and reveal functionally distinct differences.

Graphical Abstract



INTRODUCTION

Biological organs and tissues contain diverse cell types and tremendous cell-to-cell heterogeneity that dictates a multitude of functions for e.g. homeostasis and response to the physiological environment. Molecular measurements at the single-cell level can reveal mechanisms and unique features of cell populations, differentiation, impacts of microenvironments, and rare cell populations in complex cellular systems that are typically hidden from bulk-cell measurements[1,2]. While single-cell RNA sequencing (scRNA-seq), for example, allows studies of gene expression in thousands of single cells, RNA transcript levels often correlate poorly with protein abundance, and are frequently unreliable indicators of cell functions[3–5]. To this end, great efforts have been dedicated to developing mass spectrometry (MS)-based approaches to directly measure protein expression in single cells[6–13].

Because of the lack of a proteome amplification method and the fact that only picogram levels of protein are present in typical mammalian cells, single-cell proteomics constitutes a significant technical challenge. Two types of methods have recently been reported for the identification and quantification of protein expressed in single mammalian cells. In the first type of methods, single cells are processed using high-recovery sample preparation approaches, including an integrated proteome analysis device (iPAD)[6], a nanoliter-scale oil-air-droplet (OAD) chip[14], and our nanodroplet processing in one pot for trace samples (nanoPOTS) platform[15]. The processed samples are then directly analyzed in a label-free manner using LC-MS, with proteins identified and quantified from peptides. Compared with the labeling-based methods, the label-free methods have been shown to exhibit improved quantification accuracy and higher dynamic range,[16] albeit lower proteome coverage and sample throughput. In the second type of method, tryptic digests from single cells are

labeled with discrete isobaric tandem mass tags (TMT), together with a TMT-labeled peptide mixture (carrier peptides) prepared from hundreds of pooled cells[7,8,13]. During data-dependent acquisition, the high-abundance carrier peptides provide rich fragment information for peptide identification, while the signal intensities of individual reporter ions provide quantitative information for single cells. The key attraction of the TMT-based method is its throughput, as over 8 cells can be measured in a single LC-MS measurement.

Our laboratory has developed the nanoPOTS-based single-cell proteomics platform by integrating microfabricated nanowell chips with a robotic nanoliter liquid handling system[8,10,13,15,17–20]. We believe the nanoPOTS platform has three unique advantages over other approaches: (1) the use of nanowells for single-cell processing can significantly reduce adsorptive protein/peptide loss and improve overall recovery[15]; (2) because of reaction volume miniaturization, both protein and trypsin concentrations are increased by > 500-fold, resulting in enhanced tryptic digestion; and (3) the nanowell chips have open-space architecture, facilitating the integration with upstream sample isolation technologies such as fluorescence-activated cell sorting (FACS)[8,13,18,20] or laser capture microdissection (LCM)[21–23], as well as downstream LC-MS platforms. We have recently demonstrated that nanoPOTS can enable the reliable identification of 600–900 proteins from single HeLa cells[19,20] and quantify 330 proteins from dissociated primary cells of human lung using a label-free approach[20]. In the application to analyze inner hair cells from utricles of embryonic Day-15 chickens, single-cell proteomics enabled de novo reconstruction of a developmental trajectory using protein expression levels. It revealed the expression of several proteins was greatly altered during differentiation of hair cells from supporting cells[18]. In addition to label-free quantification, the nanoPOTS platform is also compatible with isobaric labeling; we demonstrated the nanoPOTS-TMT approach can quantify ~1200 to ~1500 proteins across single cells from three cell lines, and reveal protein markers of each[8,13].

Although nanoPOTS efficiently addresses sample preparation challenges of single-cell proteomics, it is not compatible with current automated LC-MS systems due to the nanoliter-scale sample volumes. Most commercial LC autosamplers are only capable of sampling from conventional vials or multiwell plates containing at least 5 μL of sample, and thus are ineffective for nanoPOTS-generated samples. Consequently, LC-MS measurements were performed manually by aspirating nanodroplet samples to a section of capillary, attaching the capillary and eluting the sample onto a solid-phase extraction (SPE) column, and finally inserting the SPE column in-line with an analytical column for gradient separation and MS detection[8,15,20]. These procedures involved the depressurizing and re-pressurizing of the LC system and repeated disconnecting and reconnecting of high-pressure fittings. Such manual operation requires extensive expertises to, e.g. avoid leaks and achieve reproducible sample loading, is highly labor-intensive, and limits throughput to ~6 LC-MS analyses per day.

In this work, we designed and implemented a nanoliter-scale autosampler to seamlessly integrate nanoPOTS-based sample preparation with automated LC-MS platforms. The autosampler consisted of a high-precision robotic system, a Peltier cooler, a 10-port valve, a syringe pump for sample loading, and two LC pumps for SPE loading and separation. We also developed a drying, extraction, and loading (DEL) approach to achieve reliable and

reproducible sample transfer from nanowell to LC column. We systematically evaluated the reproducibility and throughput of the nanoPOTS autosampler for nanoscale proteomic samples. Finally, we demonstrated its capability in both label-free and TMT-labeling based single-cell proteomics.

EXPERIMENTAL SECTION

Reagents and chemicals

Deionized water (18.2 MΩ) was prepared using a Barnstead Nanopure Infinity system (Los Angeles, CA) and used throughout the study. Proteomic reagents including dithiothreitol (DTT), iodoacetamide (IAA) triethylammonium bicarbonate (TEAB), TMT11plex, and hydroxylamine were purchased from ThermoFisher (Rockford, IL). N-Dodecyl-β-D-Maltoside (DDM) was a product of Sigma-Aldrich (St. Louis, MO). Trypsin and Lys-C (both MS grade) were obtained from Promega (Madison, WI).

Cell Culture

The bacterium *Shewanella oneidensis* MR-1 was cultured under fed-batch mode using a Bioflow 3000 fermentor (New Brunswick Scientific, Enfield, CT). The horse blood agar (HBa) culture media was supplemented with 0.5 mL/L of 100 mM ferric NTA, 1 mL/L of 1 mM Na_2SeO_4, and 1 mL/L of 3 M $MgCl_2{\cdot}6H_2O$ as well as vitamins and amino acids.

All mammalian cells were cultured at 37 °C and 5% CO_2 and split every 3 days following standard protocol. MCF-10A cell line was obtained from American Type Culture Collection (ATCC) and cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1× penicillin streptomycin. Three Acute Myeloid Leukemia (AML) cell lines (MOLM-14, K562, and CMK) were kindly provided by Dr. Anupriya Agarwal at Oregon Health & Science University. MOLM-14 and K562 cells were grown in RPMI-1640 medium supplemented with 10% FBS and 1× penicillin streptomycin, and CMK cells were maintained in RPMI-1640 medium supplemented with 20% FBS and 1× penicillin streptomycin.

Bulk-scale proteomic sample preparation

Shewanella oneidensis MR-1 cells were harvested at steady state and pelleted at 11,900 ×g for 8 min at 4 °C. Cells were lysed by homogenization with 0.1 mm zirconia/silica beads in the Bullet Blender (Next Advance, Averill Park, NY), speed 8, for 3 min. Proteins were extracted and denatured with 8 M urea and reduced with 10 mM DTT. The denatured proteins were digested with Trypsin at a protein/Trypsin ratio of 50:1 and incubated at 37 °C for 3 hours. The digested peptides were purified on a C18 SPE column and aliquoted at a concentration of 100 ng/μL for long-term storage.

AML cells were collected and washed with ice-cold phosphate-buffered saline (PBS) three times. The cell pellets were lysed in a lysis buffer containing 50 mM NH_4HCO_3 (pH 8.0), 8 M urea, and 1% phosphatase inhibitor, and sonicated in an ice bath for 3 min. After determining the protein concentrations with BCA assay (Pierce, Thermo Fisher Scientific), diluted protein solutions were digested with Lys-C at 37 °C (protein:enzyme, 50:1, w/w) for

3 h and trypsin (protein: enzyme, 50:1, w/w) at 37 °C overnight. The digested tryptic peptides were acidified by trifluoroacetic acid (TFA) with a final concentration of 0.5%, and then purified by C18 SPE tips. Tryptic peptides were dissolved in 200 mM HEPES (pH 8.5) and then mixed with TMT-126 reagent in 100% ACN for peptide labeling. The TMT-to-protein ratio was maintained at 1:1 (w/w) to achieve optimal labeling efficiency. After incubation for 1 hour at room temperature, extra TMT reagents were quenched with 5% hydroxylamine for 15 min. Finally, the TMT-labeled peptides were acidified with 0.5% formic acid (FA) and purified by C18 tips.

Single-cell proteomic preparation using nanoPOTS

Single cells were prepared using both label-free and TMT-labeling based nanoPOTS workflows.

The label-free workflow was similar to our previous work with minor modifications[15,22]. Briefly, FACS was used to isolate and deposit single cells in nanowells with a diameter of 1.2 mm. Cells were lysed by adding 100 nL lysis buffer containing 0.2% (w/v) DDM, 5 mM DTT, 0.5× PBS and 25 mM NH_4HCO_3 (pH 8.5). The chip was incubated at 65 °C for 1 h to extract, denature, and reduce proteins. Next, 50 nL of 30 mM IAA was added into each nanowell to alkylate proteins. The proteins were first digested by 0.5 ng Lys-C at 37 °C for 3 hours followed by 1 ng trypsin at 37 °C for 8 hours. Digested peptides were acidified by adding 50 nL of 5% (v/v) FA and incubating for 15 min. Finally, the sample droplets were dried out in a vacuum desiccator for 15 min. The nanoPOTS chips were stored at −20°C until analysis.

We simplified the TMT-labeling-based sample preparation workflow compared with our previous report[8]. The on-chip sample/cell arrangements and TMT channel designs are the same as our previous study[13]. Briefly, cells were lysed by adding 100 nL of lysis buffer containing 0.1% (w/v) DDM, 0.5× PBS and 25 mM TEAB (pH 8.5) and incubating the chip at 65 °C for 1 h. Next, 1 ng trypsin in 50 nL TEAB (50 mM) was directly added into each nanowell for overnight digestion (8 hours) at 37 °C. For TMT labeling, 1 μg TMT reagent (in 100 nL anhydrous ACN) was added and allowed to react for 1 hour at room temperature. Unreacted TMT reagents were quenched by adding 50 nL of 0.5% hydroxylamine (HA) in each reaction and incubated for 15 min. The droplet samples were acidified by adding 50 nL of 5% FA solution. Next, the TMT-labeled single-cell samples in nanowells from one TMT set were combined to a microliter-volume well (microwell with 2.2-mm diameter), and mixed with a 10-ng TMT126-labeled boosting peptide mixture and 0.2-ng TMT127N-labeled reference peptide mixture prepared in bulk as described above. Finally, the combined samples in microwells were dried out in a vacuum desiccator for 15 min and stored at −70°C until use.

Assembling of the nanoPOTS autosampler

A schematic diagram of the nanoPOTS autosampler is shown in Figure 1A. Briefly, one nanoflow LC pump (Dionex UltiMate NCP-3200RS with a picoflow flow meter, Thermo Scientific, Waltham, MA) was used for gradient elution. One microflow LC (NCP-3200RS with a microflow flow meter) was used for SPE loading. A home-made miniature syringe

pump was used to withdraw samples from nanowells as well as for cleaning the loading capillary and sample loop. A high-precision robotic system was built by assembling two linear actuators (X-LSQ200B and X-LSQ100B, Zaber, Vancouver, Canada) and one vertical stage (X-VSR40A, Zaber). A Peltier cooler (Part Number: 1487–1013-ND, Digi-Key, Thief River Falls, MN) was integrated under the nanoPOTS chip to maintain the temperature at ~ 8 °C. The cooler can efficiently minimize sample denaturation during large-scale and long-term sample analysis. A 10-port valve (C72MFSH-4570ED, Valco Instruments Co. Inc, Houston, TX) was employed to integrate sample loading, SPE cleanup, and LC separation functions by switching the valve. A central-control Labview program was developed to synchronize all the autosampler operations and trigger MS acquisition. A photo of the autosampler running in front of a Fusion Lumos Orbitrap MS, as well as a close-up view of the 10-port valve, a nanoPOTS chip, and a Peltier cooler are presented in Figure S1.

LC-MS/MS procedures

SPE columns (100-μm-*i.d.,* 4 cm long) were packed with 5-μm C18 packing material (300-Å pore size, Phenomenex, Torrance, CA). Self-Pack PicoFrit capillary tubings (50-μm-*i.d.,* New Objective, Woburn, MA) with integrated 10-μm-*i.d.* ESI emitters were used to pack LC columns. The 60-cm-long LC columns were packed with 3-μm, C18 packing material (300-Å pore size, Phenomenex). To reduce column back pressures, the LC columns were heated to 50°C using AgileSleeve column heater (Analytical Sales and Services, Inc., Flanders, NJ).

The sample loading procedures are illustrated in Figure 1B. Unlike typical sample injection procedures, a DEL approach was developed for the nanoPOTS autosampler. Briefly, an aqueous droplet (0.1% FA in water) was applied to a nanowell and then allowed it to sit for 4 min and extract the sample into the droplet (Figure 1B1). Next, the droplet was withdrawn into the sample loop. To increase sample recovery, two additional washing droplets were applied to the nanowell and withdrawn into the sample loop. Air bubbles were used to segment the sequentially generated sample plugs. Then, the 10-port valve was switched to the SPE loading position to concentrate the sample at the head of the SPE column, as well as to desalt the sample (Figure 1B3). Finally, the valve was switched to separation position and the peptide separation and MS detection were started (Figure 1B4). Detailed autosampler operation procedures are shown in Table S1 of Supporting Information.

For the LC separations, water with 0.1% FA was used as mobile phase A (Buffer A) and ACN with 0.1% FA was used as mobile phase B (Buffer B). Four different gradient profiles were used in this study. For example, for a typical 60-min gradient, peptides were separated by a linear gradient from 8% to 22% Buffer B in 50 min, followed by linearly raising to 35% Buffer B in 10 min. The column was washed by maintaining at 80% Buffer B for 10 min. Finally, the column was equilibrated for 5 min using 2% Buffer B. It should be noted that the actual column equilibration time was ~17 min, because the gradient pump was maintained at 2% Buffer B during the sample injection process.

An Orbitrap Fusion Lumos Tribrid MS (ThermoScientific) operated at data-dependent acquisition mode was used for analyses with 2 kV used for electrospray ionization (ESI) of the LC separated peptides. The ion transfer capillary was heated to 200 °C to accelerate desolvation, and the RF lens was set at 30%. MS spectra were acquired in the Orbitrap

analyzer with a 120,000 resolution (m/z 200), an AGC target of 1E6, and a maximum ion injection time of 246 ms. For *Shewanella oneidensis* MR-1 peptides, precursors were isolated at an m/z window of 2 and fragmented by higher-energy collisional dissociation (HCD) with collision energy (CE) at 35%. The Orbitrap was used to collect MS/MS spectra at a resolution of 60,000, an AGC target of 2E5, and maximum ion injection times of 118 ms, 250 ms, and 350 ms for peptide loadings of 10 ng, 5 ng, and 1 ng, respectively. For MCF10A cells, an ion trap analyzer was used to collect MS/MS spectra. The precursors were isolated at an m/z window of 2 and fragmented by collision-induced dissociation (CID) with 35% CE. An AGC target of 2E4 was used and two maximum ion injection times including 120 ms for 50 cells and 250 ms for 1–10 cells. For single AML cells, precursors were isolated at an m/z window of 0.7 and fragmented by HCD with 38% CE. MS/MS spectra were collected in Orbitrap at a resolution of 120,000, an AGC target of 1E6, and a maximum ion injection time of 246 ms[13].

Data Analysis

MaxQuant (version 1.6.2.10) was used to process all the raw files[24]. Briefly, the Andromeda engine was used to search MS/MS spectra against an UniProtKB database for *Shewanella oneidensis* MR-1 (Downloaded in 2/23/2017 containing 645 reviewed and 3426 unreviewed sequences) and an UniProtKB human database (Downloaded in 12/29/2016 containing 20,129 reviewed sequences). N-terminal protein acetylation and methionine oxidation were selected as variable modifications. Both peptides and proteins were filtered with a maximum false discovery rate (FDR) of 0.01. For label-free quantification for *Shewanella oneidensis* MR-1 and MCF10A samples, Match-Between-Runs (MBR) was activated with a matching window of 0.4 min and alignment window of 10 min. The LFQ intensities were extracted, filtered to contain ≥ 70% valid values, and visualized using Perseus (Version 1.6.2.1)[25]. For TMT-based quantification, corrected reporter ion intensities were extracted. Reporter ion intensities from single cells were normalized to the reference channel containing 0.2 ng of a peptide mixture from the three AML cell lines at PSM level using MaxQuant (version 1.6.12.0)[26]. To minimize the batch effect from multiple TMT experiments, the relative abundances from 19 TMT plexes were log2-transformed and the data matrices from all the TMT experiments were combined after sequentially centering the column and row values to their median values. A criterion of >70% valid values and at least two identified peptides per protein were required to generate the “quantifiable” protein list. The data matrix was further analyzed by Perseus for statistical analysis including principle component analysis (PCA) and heatmap analysis[26].

RESULT AND DISCUSSION

Design of the nanoPOTS autosampler and the injection procedures.

Although the nanoPOTS platform significantly improved the overall sensitivity of single-cell proteomics analysis, nanoliter-scale samples pose a challenge for current automated LC-MS analysis systems. First, the small-size and customized nanowells are not compatible with commercially available LC autosamplers. Second, nanoliter-scale aqueous droplet samples are prone to evaporation during long-term storage before analysis (weeks to months), as well as during automated LC analysis (often days to weeks). Such evaporation could lead to

incomplete sample injection and poor reproducibility in injection volumes. To this end, we designed and assembled a prototype autosampler using a programmable robotic system (Figure S1). The home-built high-precision robot allowed us to precisely align the sampling capillary with the nanoPOTS chips for reproducible injection. A Labview software was developed to drive the automated analysis (extraction, loading to sample loop, loading to SPE column, multi-step washing, LC and MS triggering).

To address the challenge associated with droplet evaporation, we developed a DEL (drying, extraction, and loading) approach (Figure 1B); instead of keeping the sample hydrated, we completely dried the samples in a vacuum desiccator after sample preparation. During the drying process, all the peptides precipitated together with the salts and surfactants. To bring back the peptides to solution phase, a 4-min droplet extraction step was applied. To maximize peptide recovery, two additional washing steps were also included. All the sample and washing solutions were loaded into the sample loop and enriched into a SPE column.

Reproducibility of sample injection.

The key aspect of a LC autosampler is the ability to perform reproducible sample injection. To assess this, we continuously injected and analyzed 20 samples of 10-ng tryptic peptides from *Shewanella oneidensis* MR-1 cells. As shown in Figure 2A and 2B, the 20 analyses give similar levels of peptide and protein identifications. An average of 10268 unique peptides were identified from each analysis, leading to an average of 1683 protein groups. 1390 out of total 1911 identified proteins were quantifiable with >70% valid values and 2 peptides required across the 20 analyses. To evaluate the reproducibility, we performed pairwise correlation for all 20 injections using log2-transformed LFQ intensities provided by MaxQuant. High run-to-run reproducibility with Pearson's correlation coefficients of > 0.995 (median value of 0.998) was achieved (Figure 2C). These results demonstrated the nanoPOTS autosampler and the DEL injection procedures can achieve reproducible sample injections for low-nanogram peptide samples.

Sensitivity of the nanoPOTS-LC-MS system.

To evaluate the sensitivity of the nanoPOTS autosampler coupled with a LC-MS system, we preloaded serially diluted tryptic peptides from 1 ng to 20 ng in nanowells and measured them using a 50-μm-*i.d.* LC and an Orbitrap Fusion Lumos MS. With the increased sample loadings, the average peptide identifications increased from 4544 to 10268, and protein identifications from 1082 to 1628 (Figure 3A and 3B). Using MaxQuant's MBR capability, the average numbers of protein identifications were increased to 1674, 1832, and 1872 for 1 ng, 5 ng, and 20 ng peptide samples, respectively. As expected, protein intensities increased linearly with the peptide loadings (Figure 3C). More importantly, for all the peptide loading groups, excellent reproducibility with median coefficients of variation (CVs) below 4.5% (protein level) was achieved (Figure 3D), indicating the autosampler can reliably analyze low nanogram tryptic peptides prepared in nanowells.

Throughput and robustness.

In addition to reproducibility and sensitivity, another important aspect of the automated LC-MS system for single cell proteomics is sample throughput, as a large number of single cells

usually need to be profiled to reveal rare cell populations. The throughput of proteomics analysis is mainly limited by the LC separation step, where hours-long gradients were required to reduce the proteome complexity and achieve good coverage. To assess the proteome coverage with different gradient times, we analyzed 10-ng tryptic peptides from *Shewanella oneidensis* MR-1 cells using 30 min, 60 min, and 120 min LC gradients.

The overall designs of LC operation programs are illustrated in Figure 4A. We found the use of SPE column can significantly improve the throughput in three aspects: (1) Sample can be loaded to the SPE column at a high flow rate (3 μL/min); (2) Because the column equilibration step is parallelized with sample injection and SPE loading steps, only 5-min is required after column washing step; (3) The SPE efficiently purified the sample and improved the robustness of LC-MS.

The automated LC-MS system can analyze 9.6, 16, and 24 samples per day when 120 min, 60 min, and 30 min LC gradients were used. Figure 4B shows the total numbers of MS/MS spectra, peptide spectrum matches (PSMs), and peptide identifications at the three LC gradients. Reducing gradient times from 120 min to 30 min significantly reduces the MS/MS spectra from 24,354 to 12,314, leading to a decrease of PSMs from 14,757 to 7,578. However, only a 15% decrease in unique peptides and 6% in protein groups was observed when reducing gradient times from 120 min to 60 min. The unique peptides and protein groups decreased by 33% and 22%, respectively, when reducing gradient time from 60 min to 30 min. These results suggest the 60-min gradient provides a good balance between proteome coverage and sample throughput for the present application, and a starting point for low-nanogram sample analyses. It should be noted that the optimal gradient time is expected to vary with other parameters such as peptide loading amounts and MS scan rates, as well as specific applications.

System robustness is another key concern for large-scale proteomic studies. The autosampler was constructed using robust components, including robotic stages (Zaber), ThermoFisher's Dionex pumps, and Valco's UPLC valves, which have been widely adopted and demonstrated in various analytical systems. To improve the operation robustness of the automated LC-MS system, we also employed 50-μm *i.d.* LC columns instead of more sensitive (i.e. lower flow rate) 30-μm *i.d.* LC columns used in our previous nanoPOTS-related studies[8,15,18,20,23]. The autosampler has been extensively applied in various projects involving single mammalian cells[13], as well as studies of nano-scale mouse and human tissue sections, and plant tissue sections (unpublished). Over 500 LC-MS analyses have been performed without system failure, although some runs failed due to clogged electrospray emitters and LC columns. For nanogram purified peptide samples and single cell samples, the autosampler can work continuously for >2 weeks and >100 injections.

Label-free single cell proteomics.

We next evaluated the performance of the nanoPOTS autosampler for label-free single-cell proteomics using FACS-sorted MCF10A cells. As demonstrated previously, MaxQuant's MBR algorithm of can significantly improve the proteome coverage[15,18,20], by allowing the single cell samples to be matched using peptide libraries generated from samples comprising 50 MCF10A cells.

We first evaluated the impact of LC gradients on the proteome coverage of single cells. As shown in Figure 5A, the average numbers of protein identifications by MS/MS increased from 234 to 269 (15%), when reducing gradient lengths from 60 min to 30 min. If MBR was used, the average number of protein identifications increased to 764 for the 30-min LC gradient. The improved proteome coverage with the 30-min gradient can be ascribed to increased peak sharpness and intensity with the faster separation. The 75th percentile of full width at half maximum (FWHM) were reduced from 11.0 ± 0.33 s (n=5) to 9.1 ± 0.20 s (n=5). More importantly, the use of short gradient also significantly increases the analysis throughput to 24 single cells per day (Figure 4A1).

Next, we evaluated the proteome coverage at different cell loadings, including controls (blank droplets from FACS), single cells, 3 cell, and 10 cells. Only 16 proteins were identified from controls, and the numbers of protein identifications increased to 252, 582, and 894 for single cells, 3 cells, and 10 cells, respectively. Using MBR, the identification numbers were increased to 88, 727, 1216, and 1580, respectively (Figure 5B). Most proteins identified from control samples are included in single cells (Figure S2A), indicating that the background contaminations were possibly from the lysed cells in cell suspension or the carry-over of LC columns. We also compared the summed protein intensities between control and single cells (Figure S2B). The total protein intensity of control sample only accounts for 2.07% of single cells.

The proteome coverage for single cells was on a par with our previous reports using manual sample injection and a lower flow-rate LC system (30-μm-*i.d.* LC column), where an average of 669 proteins was identified[20]. The proteome coverage across the 10 single cells are similar, both with and without MBR (Figure 5C). When 70% valid iBAQ (intensity-based absolute quantification) values were required, 602 out of total 1093 proteins are quantifiable across the 10 cells. Figure 5D shows the pair-wise correlation matrix between the 10 single cells. High Pearson's correlation coefficients with a median value of 0.92 were obtained. The low heterogeneity in protein expression from these single cells can be attributed to the fact that all the cells were cultured in identical conditions.

TMT labeling-based single cell proteomics.

We next evaluated the performance of the nanoPOTS autosampler for large-scale quantitative single-cell proteomic analysis using FACS-sorted AML cells (57 MOLM-14 cells, 57 K562 cells, and 38 CMK cells) and TMT-based isobaric labeling approach[7,8,13]. In addition to 104 cells that were analyzed in our previous study[15], we further incorporated 48 additional single cells for this study, resulting in total of 152 cells. To improve quantification precision, an improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) strategy was used[13]. TMT126 and TMT127N were used as boosting (10 ng peptide mixture) and reference (0.2 ng peptide mixture) channels, respectively. Single cells were labeled with the remaining TMT channels (MOLM-14:TMT128N, TMT129C and TMT131N; K562:TMT128C, TMT130N and TMT131C; CMK:TMT129N and TMT130C) (Table S2). The coverage of single-cell proteome mainly relied on the high peptide loadings in boosting channels. We observed similar number of peptides and proteins were identified across the 19 LC-MS runs. As shown in Figure S3A, an average of 9,845±115 unique peptides were

identified from each run (8 cells per run), leading to an average of 1,821±32 proteins (total 2,558 protein groups and 2,071 proteins having two or more peptides). To assess the reproducibility in sample preparation and injection, we extracted the protein intensities based on MS1-level quantification. As shown in Figure S3B, high correlations with median values of 0.98 were obtained across the 19 LC-MS runs (the median value is 0.98), demonstrating high reproducibility in both nanoPOTS-based sample preparation and automated sample injection.

We used PSM-level ratio normalization approach[26] (available in MaxQuant 1.6.12.0) to mitigate the batch effect from TMT-based quantification. Among the 2,558 total identified protein groups, 1,424 proteins were quantifiable with more than 70% valid values across all the 152 AML cells. The correlation matrices show the expected higher correlations between the same type of cells and lower correlations between different cell types (Figure 6A). Moreover, PCA projection shows the single cells from same cell lines clearly cluster together and well separate from one another (Figure 6B).

CONCLUSION

We have demonstrated the nanoPOTS autosampler can reproducibly and robustly perform nanoliter-scale sample injection from nanowells to LC-MS system for automated analysis. Compared with the previous manual sample injection approach, the autosampler greatly improved the operation robustness and sample throughput by avoiding the labor-intensive manual procedures and enabling 24-hour non-stop LC-MS operation. We also showed the DEL approach is an efficient and reproducible way for storing and injecting low-volume samples for LC-MS. The performance of the autosampler was demonstrated in both label-free and TMT labeling-based single cell proteomics, resulting in 602 (label-free) and 1,424 (TMT) quantifiable proteins across all the studied single cells with a rigorous criterion (>70% valid values). High-throughput single-cell analysis at throughput of 24 cells per day and ~77 cells per day was achieved when label free and TMT labeling approaches were used, respectively.

We believe the autosampler will be an indispensable part of the nanoPOTS platform for single-cell proteomic analysis. Although single-cell proteomics was demonstrated in this work, the autosampler is also broadly applicable to other nanoscale samples including small numbers of cell populations enriched by FACS and micromanipulator, on-demand subcellular capillary microsampling from live cells or vertebrate embryos,[27] and cell-type-specific and spatially resolved tissue mapping in combination with LCM[14,21,22].

The performance of the automated nanoPOTS-LC-MS platform could be improved with further technical developments. For example, the sample throughput can be doubled using parallel LC columns to improve the duty cycle[28], because ~50% of analysis time was spent on column washing and sample injection procedures when 30-min gradient was used (Figure 4). Our previous study showed the peptide recoveries in nanowells during a single-step DEL procedure were in the range of 79–83%[29]. The sample recovery could be further improved using smaller nanowells or low-binding chemical coatings on nanowell surfaces[30]. The LC-MS sensitivity can be enhanced using lower-flow-rate and higher-resolution LC columns

such as micropillar array-based nanoLC cartridges[31]. Advanced Orbitrap mass spectrometer[19], trapped ion mobility spectrometry (TIMS)[32] and SLIM IMS with time of flight (TOF)[10,33] should further improve the sensitivity and protein identifications. With these ongoing developments, we expect the fully automated nanoPOTS-LC-MS platform will become an indispensable tool in biomedical research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Matthew Monroe for helping with data deposition. This work was supported by a Laboratory Directed Research and Development award from PNNL (to YZ). Portions of the research were supported by a P41 GM103493 grant from the National Institute of General Medical Sciences to RDS and a U24CA210955 grant from the NCI Clinical Proteomic Tumor Analysis Consortium (CPTAC) to TL and RDS. The research was performed using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at PNNL.

REFERENCES

(1). Kolodziejczyk AA; Kim JK; Svensson V; Marioni JC; Teichmann SA The Technology and Biology of Single-Cell RNA Sequencing. Mol. Cell 2015, 58 (4), 610–620. [PubMed: 26000846]

(2). Wu AR; Wang J; Streets AM; Huang Y Single-Cell Transcriptional Analysis. Annu. Rev. Anal. Chem 2017, 10 (1), 439–462.

(3). Tian Q; Stepaniants SB; Mao M; Weng L; Feetham MC; Doyle MJ; Yi EC; Dai H; Thorsson V; Eng J; et al. Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. Mol. Cell. Proteomics 2004, 3 (10), 960–969. [PubMed: 15238602]

(4). Zhang B; Whiteaker JR; Hoofnagle AN; Baird GS; Rodland KD; Paulovich AG Clinical Potential of Mass Spectrometry-Based Proteogenomics. Nat. Rev. Clin. Oncol 2019, 16 (4), 256–268. [PubMed: 30487530]

(5). Liu Y; Beyer A; Aebersold R Review On the Dependency of Cellular Protein Levels on MRNA Abundance. Cell 2016, 165 (3), 535–550. [PubMed: 27104977]

(6). Shao X; Wang X; Guan S; Lin H; Yan G; Gao M; Deng C; Zhang X Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. Anal. Chem 2018, 90 (23), 14003–14010. [PubMed: 30375851]

(7). Budnik B; Levy E; Slavov N SCoPE-MS: Mass Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity during Cell Differentiation. Genome Biol. 2018, 19, 161. [PubMed: 30343672]

(8). Dou M; Clair G; Tsai C-F; Xu K; Chrisler WB; Sontag RL; Zhao R; Moore RJ; Liu T; Pasa-Tolic L; et al. High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. Anal. Chem 2019, 91 (20), 13119–13127. [PubMed: 31509397]

(9). Song Y; Shi Y; Zhu Z; Yang C Bioinspired Engineering of Multivalent Aptamer- Functionalized Nanointerface to Enhance Capture and Release of Circulating Tumor Cells. 2018, 361005, 1287–1290.

(10). Couvillion SP; Zhu Y; Nagy G; Adkins JN; Ansong C; Renslow RS; Piehowski PD; Ibrahim YM; Kelly RT; Metz TO New Mass Spectrometry Technologies Contributing towards Comprehensive and High Throughput Omics Analyses of Single Cells. Analyst 2019, 144 (3), 794–807. [PubMed: 30507980]

(11). Lombard-Banek C; Moody SA; Nemes P Single-Cell Mass Spectrometry for Discovery Proteomics: Quantifying Translational Cell Heterogeneity in the 16-Cell Frog (Xenopus) Embryo. Angew. Chemie - Int. Ed 2016, 55 (7), 2454–2458.

(12). Sun L; Dubiak KM; Peuchen EH; Zhang Z; Zhu G; Huber PW; Dovichi NJ Single Cell Proteomics Using Frog (Xenopus Laevis) Blastomeres Isolated from Early Stage Embryos, Which Form a Geometric Progression in Protein Content. Anal. Chem 2016, 88 (13), 6653–6657. [PubMed: 27314579]

(13). Tsai C-F; Zhao R; Williams SM; Moore RJ; Schultz K; Chrisler W; Pasa-Tolic L; Rodland K; Smith RD; Shi T; et al. An Improved Boosting to Amplify Signal with Isobaric Labeling (IBASIL) Strategy for Precise Quantitative Single-Cell Proteomics. Mol. Cell. Proteomics 2020, 19 (5), 828–838. [PubMed: 32127492]

(14). Li ZY; Huang M; Wang XK; Zhu Y; Li JS; Wong CCL; Fang Q Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. Anal. Chem 2018, 90 (8), 5430–5438. [PubMed: 29551058]

(15). Zhu Y; Piehowski PD; Zhao R; Chen J; Shen Y; Moore RJ; Shukla AK; Petyuk VA; Campbell-Thompson M; Mathews CE; et al. Nanodroplet Processing Platform for Deep and Quantitative Proteome Profiling of 10–100 Mammalian Cells. Nat. Commun 2018, 9 (1), 882. [PubMed: 29491378]

(16). Hogrebe A; von Stechow L; Bekker-Jensen DB; Weinert BT; Kelstrup CD; Olsen JV Benchmarking Common Quantification Strategies for Large-Scale Phosphoproteomics. Nat. Commun 2018, 9 (1), 1045. [PubMed: 29535314]

(17). Zhu Y; Piehowski PD; Kelly RT; Qian WJ Nanoproteomics Comes of Age. Expert Rev. Proteomics 2018, 15 (11), 865–871. [PubMed: 30375896]

(18). Zhu Y; Scheibinger M; Ellwanger DC; Krey JF; Choi D; Kelly RT; Heller S; Barr-Gillespie PG Single-Cell Proteomics Reveals Changes in Expression during Hair-Cell Development. Elife 2019, 8, e50777. [PubMed: 31682227]

(19). Cong Y; Liang Y; Motamedchaboki K; Huguet R; Truong T; Zhao R; Shen Y; Lopez-Ferrer D; Zhu Y; Kelly RT Improved Single Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry. Anal. Chem 2020, 92 (3), 2665–2671. [PubMed: 31913019]

(20). Zhu Y; Clair G; Chrisler WB; Shen Y; Zhao R; Shukla AK; Moore RJ; Misra RS; Pryhuber GS; Smith RD; et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. Angew. Chemie - Int. Ed 2018, 57 (38), 12370–12374.

(21). Piehowski PD; Zhu Y; Bramer LM; Stratton KG; Zhao R; Orton DJ; Moore RJ; Yuan J; Mitchell HD; Gao Y; et al. Automated Mass Spectrometry Imaging of over 2000 Proteins from Tissue Sections at 100-Mm Spatial Resolution. Nat. Commun 2020, 11 (1), 8. [PubMed: 31911630]

(22). Zhu Y; Dou M; Piehowski PD; Liang Y; Wang F; Chu RK; Chrisler WB; Smith JN; Schwarz KC; Shen Y; et al. Spatially Resolved Proteome Mapping of Laser Capture Microdissected Tissue with Automated Sample Transfer to Nanodroplets. Mol. Cell. Proteomics 2018, 17 (9), 1864–1874. [PubMed: 29941660]

(23). Zhu Y; Podolak J; Zhao R; Shukla AK; Moore RJ; Thomas GV; Kelly RT Proteome Profiling of 1 to 5 Spiked Circulating Tumor Cells Isolated from Whole Blood Using Immunodensity Enrichment, Laser Capture Microdissection, Nanodroplet Sample Processing, and Ultrasensitive NanoLC–MS. Anal. Chem 2018, 90 (20), 11756–11759. [PubMed: 30269481]

(24). Tyanova S; Temu T; Cox J The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. Nat. Protoc 2016, 11 (12), 2301–2319. [PubMed: 27809316]

(25). Tyanova S; Cox J Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. Methods Mol. Biol 2018, 1711, 133–148. [PubMed: 29344888]

(26). Yu S-H; Kiriakidou P; Cox J Isobaric Matching between Runs and Novel PSM-Level Normalization in MaxQuant Strongly Improve Reporter Ion-Based Quantification. bioRxiv 2020, DOI: 2020.03.30.015487.

(27). Lombard-Banek C; Moody SA; Manzini MC; Nemes P Microsampling Capillary Electrophoresis Mass Spectrometry Enables Single-Cell Proteomics in Complex Tissues: Developing Cell Clones in Live Xenopus Laevis and Zebrafish Embryos. Anal. Chem 2019, 91 (7), 4797–4805. [PubMed: 30827088]

(28). Livesay EA; Tang K; Taylor BK; Buschbach MA; Hopkins DF; LaMarche BL; Zhao R; Shen Y; Orton DJ; Moore RJ; et al. Fully Automated Four-Column Capillary LC-MS System for Maximizing Throughput in Proteomic Analyses. Anal. Chem 2008, 80 (1), 294–302. [PubMed: 18044960]

(29). Dou M; Chouinard CD; Zhu Y; Nagy G; Liyu AV; Ibrahim YM; Smith RD; Kelly RT Nanowell-Mediated Multidimensional Separations Combining NanoLC with SLIM IM-MS for Rapid, High-Peak-Capacity Proteomic Analyses. Anal. Bioanal. Chem 2018, 411 (21), 5363–5372. [PubMed: 30397757]

(30). Chandradoss SD; Haagsma AC; Lee YK; Hwang JH; Nam JM; Joo C Surface Passivation for Single-Molecule Protein Studies. J. Vis. Exp 2014, No. 86, 1–8.

(31). Stadlmann J; Hudecz O; Krššáková G; Ctortecka C; Van Raemdonck G; Op De Beeck J; Desmet G; Penninger JM; Jacobs P; Mechtler K Improved Sensitivity in Low-Input Proteomics Using Micropillar Array-Based Chromatography. Anal. Chem 2019, 91 (22), 14203–14207. [PubMed: 31612716]

(32). Meier F; Brunner A-D; Koch S; Koch H; Lubeck M; Krause M; Goedecke N; Decker J; Kosinski T; Park MA; et al. Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. Mol. Cell. Proteomics 2018, 17 (12), 2534–2545. [PubMed: 30385480]

(33). Deng L; Ibrahim YM; Hamid AM; Garimella SVB; Webb IK; Zheng X; Prost SA; Sandoval JA; Norheim RV; Anderson GA; et al. Ultra-High Resolution Ion Mobility Separations Utilizing Traveling Waves in a 13 m Serpentine Path Length Structures for Lossless Ion Manipulations Module. Anal. Chem 2016, 88 (18), 8957–8964. [PubMed: 27531027]

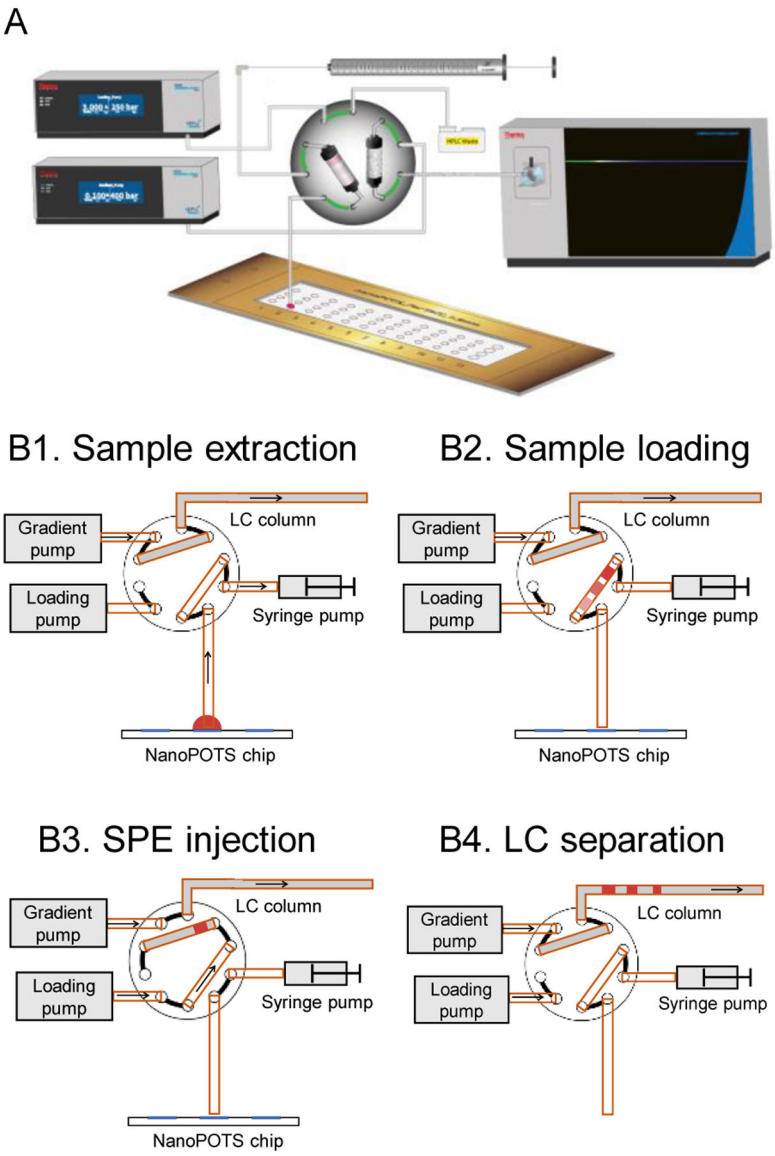


Figure 1.

(A) Schematic diagram of the nanoPOTS autosampler. A photo of the autosampler system is attached in Figure S1. (B) Step-by-step sample injection and separation procedures using the DEL approach. After sample preparation, droplet samples in nanoPOTS chip are dried under vacuum. (B1) A nanoliter droplet is applied to a nanowell and allowed to sit for several minutes to extract the sample; (B2) The droplet is withdrawn into the sample loop of 10-port valve, followed by two additional washing droplets; (B3) The sample plugs in the loop are concentrated into a SPE column and desalted; (B4) The 10-port valve is switched to the separation position. LC gradient is applied to elute peptides for MS detection.

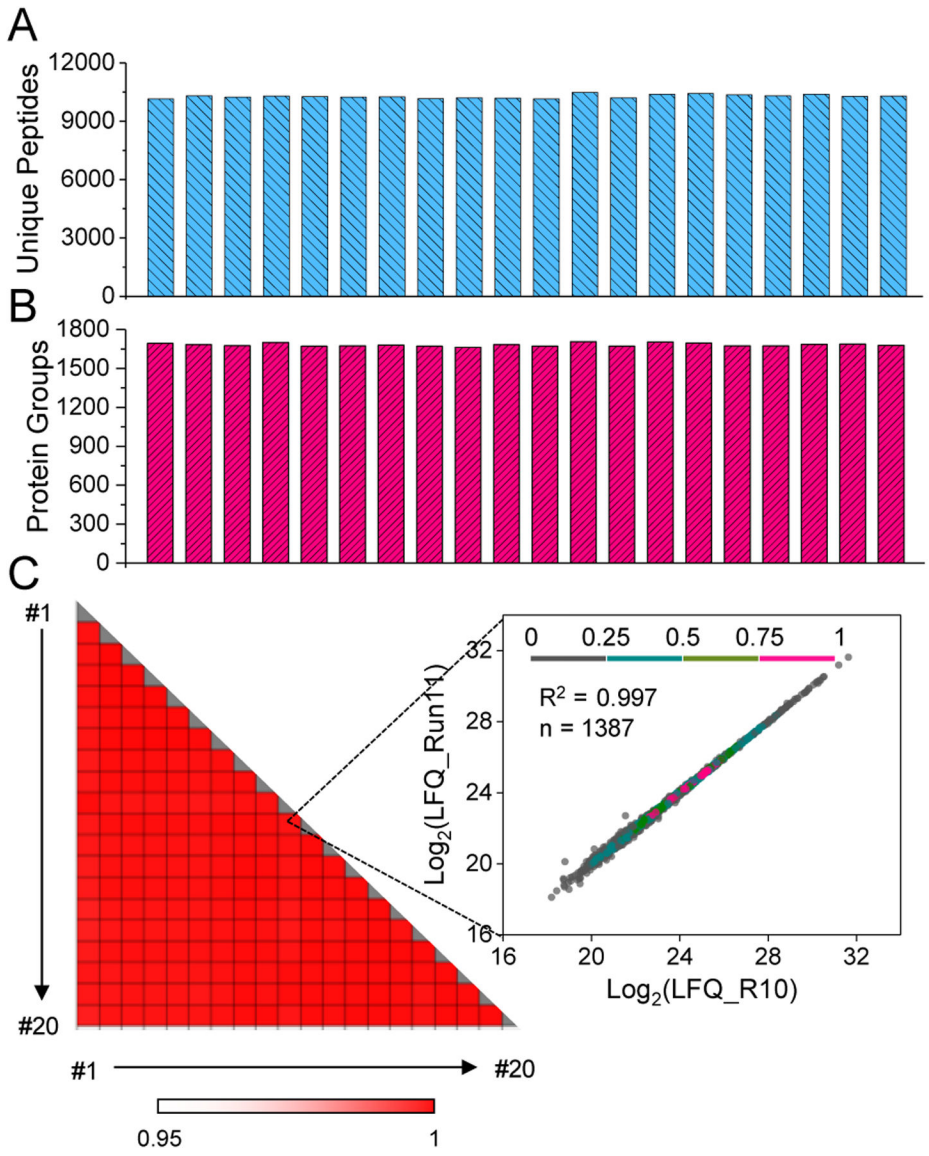


Figure 2.

(A) Unique peptide and (B) protein groups identified from consecutive injection of 10-ng tryptic peptides from *Shewanella oneidensis* MR-1 cells based on tandem mass spectra. (C) Pairwise correlations of log2-transformed LFQ intensities for the 20 injections. The inset graph shows a typical correlation between the 10th and 11th injections.

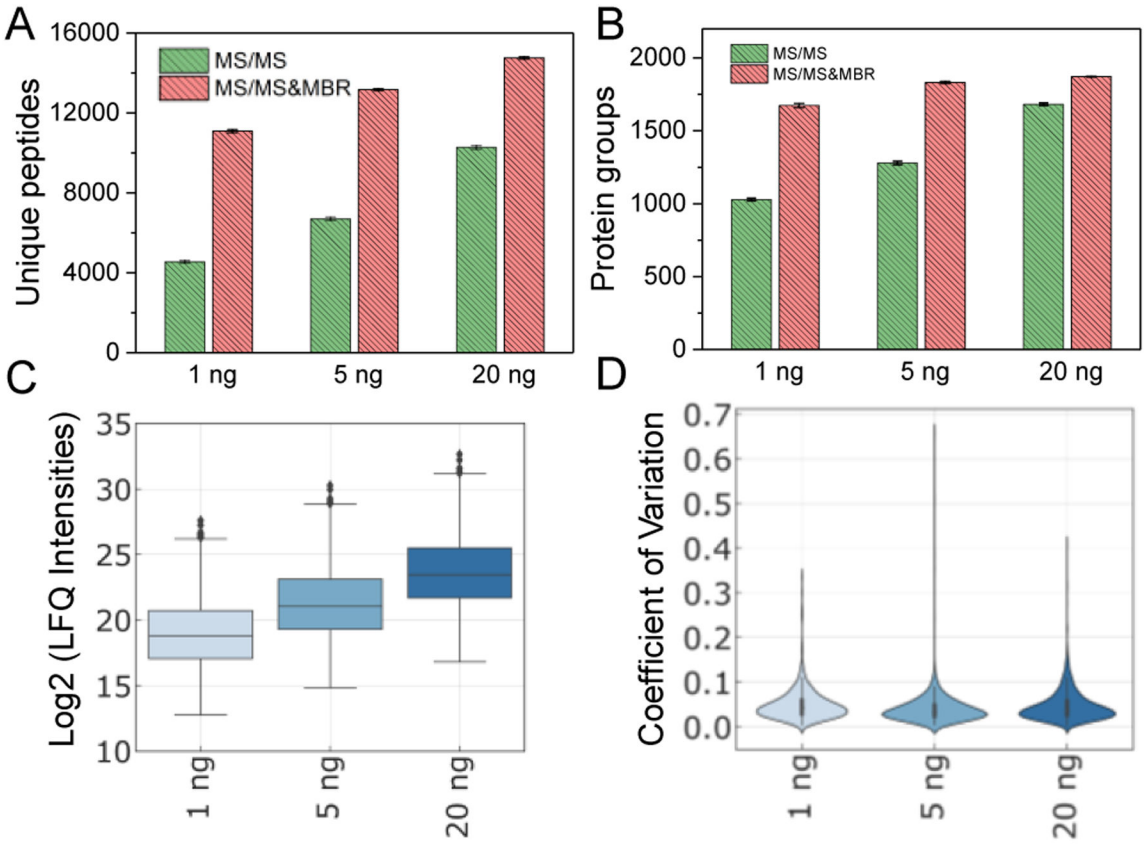


Figure 3.

(A) Unique peptide and (B) protein groups identified from serially diluted tryptic peptides (1 ng, 5 ng, and 20 ng) from *Shewanella oneidensis* MR-1 cells with and without MBR. Error bars indicate standard deviations from 5 technical replicates. (C) The distributions of log2-transformed LFQ intensities for the three peptide loading samples. Center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles. (D) Violin plots show the distributions of coefficients of variation (CVs) of protein LFQ intensities in each loading group. LFQ intensities are generated by MaxQuant.

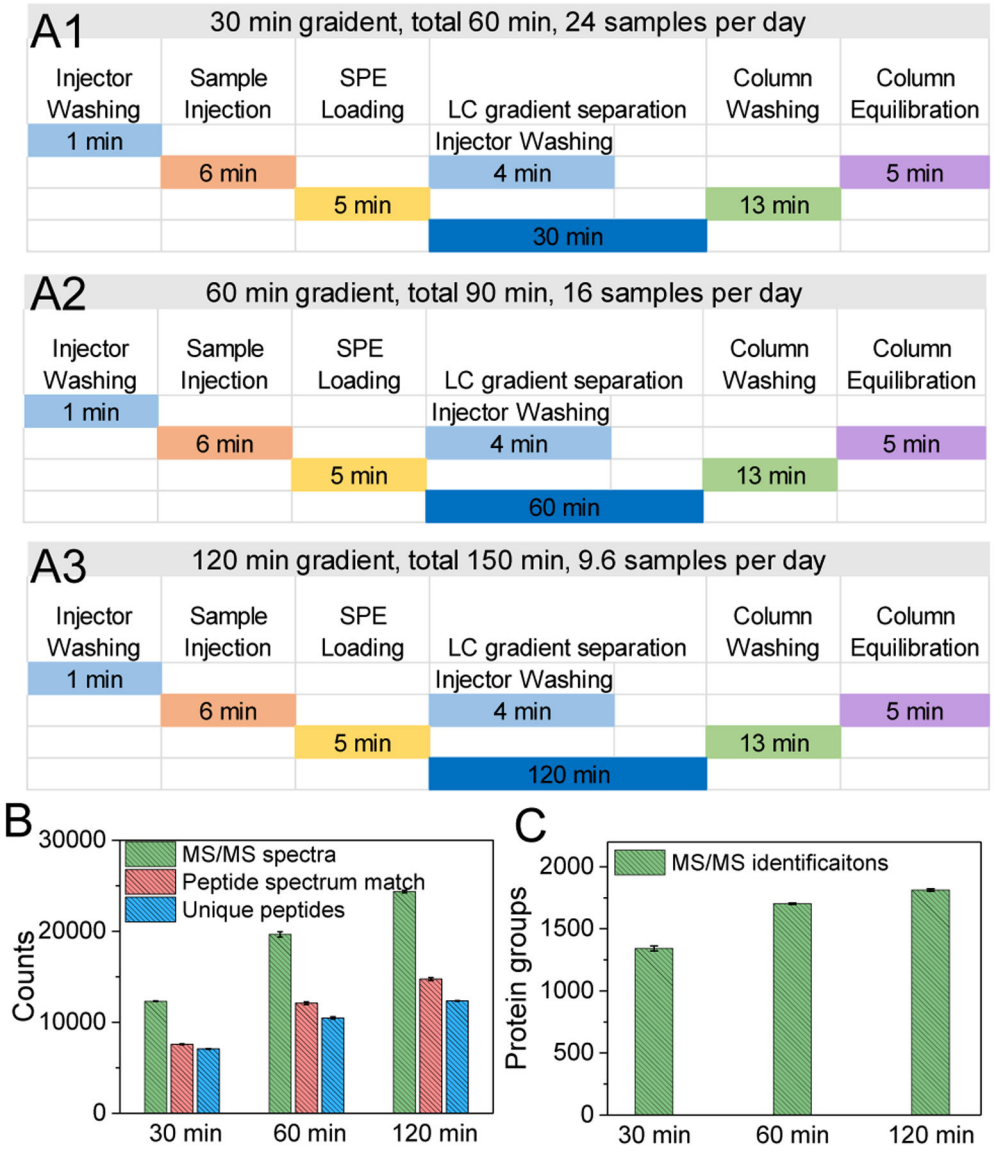


Figure 4.

(A) Operation programs of nanoPOTS autosampler at different LC gradient times. (B) Average of MS/MS spectra, PSM, and unique peptides by analyzing 10-ng tryptic peptides from *Shewanella oneidensis* MR-1 cells using the 30-min, 60-min, and 120-min gradients. (C) The corresponding protein identifications. Error bars indicate standard deviations from 5 technical replicates.

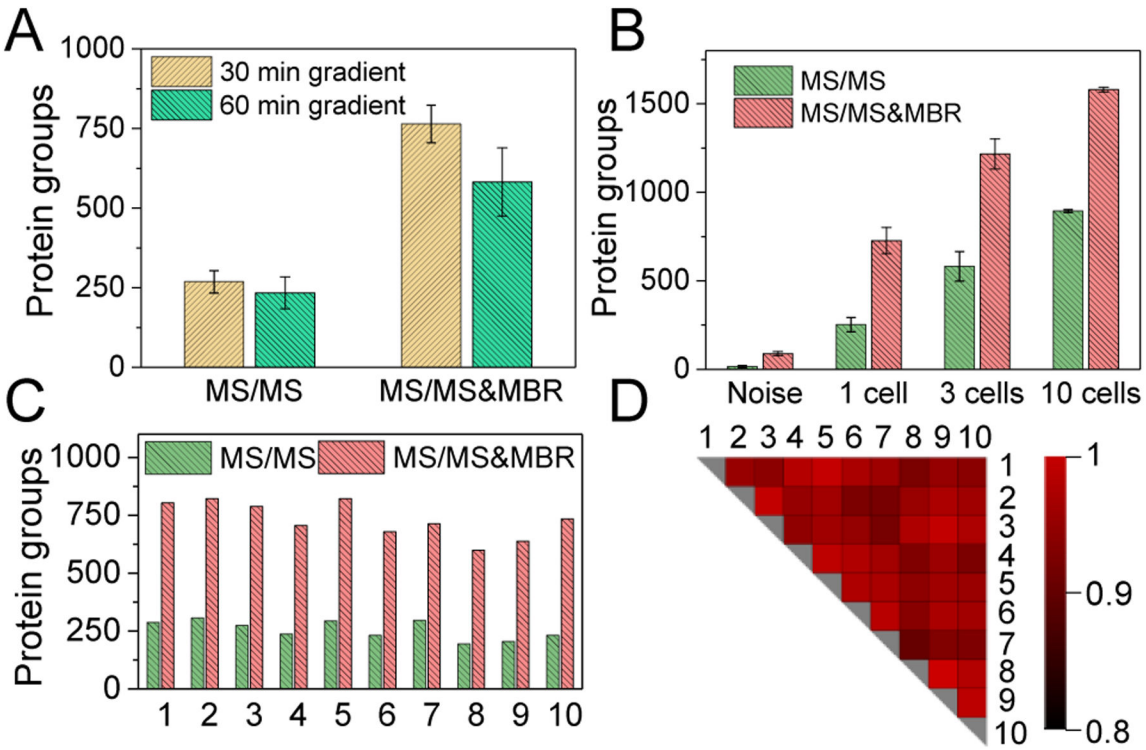


Figure 5.

(A) Average proteome coverage of single MCF10A cells using 30-min and 60-min LC gradients. Error bars indicate standard deviations from 5 biological replicates (5 different single MCF10A cells). (B) Average proteome coverages of MCF10A samples containing noise (0 cell), 1 cell, 3 cells, and 10 cells using a 30-min LC gradient. Error bars indicate standard deviations from 3 replicates for noise, 10 replicates for 1 cell, and 4 replicates for 3 and 10 cells. (C) The numbers of protein identifications from 10 different single MCF10A cells. For (A-C), the MBR identifications were obtained by matching to a peptide library prepared from 50 cells. (D) Pair-wise correlation matrix of the 10 MCF10A cells using log2-transformed iBAQ intensities. The median Pearson's correlation coefficient is 0.92.

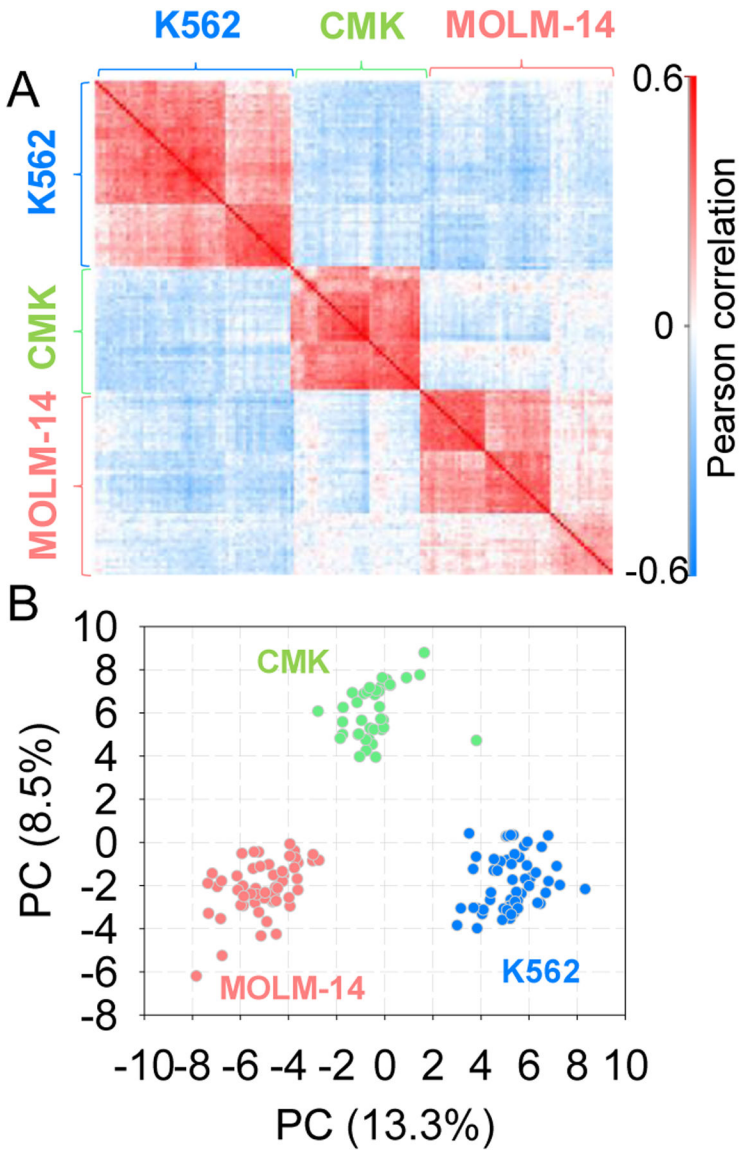


Figure 6.

(A) The clustering matrix showing pair-wise Pearson correlations between the 152 single AML cells using normalized protein intensities. (B) PCA shows the clustering of single cells and the separation of cells from different cell lines based on quantitative single-cell proteomics.