ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

One-STAGE Tip Method for TMT-Based Proteomic Analysis of a Minimal Amount of Cells

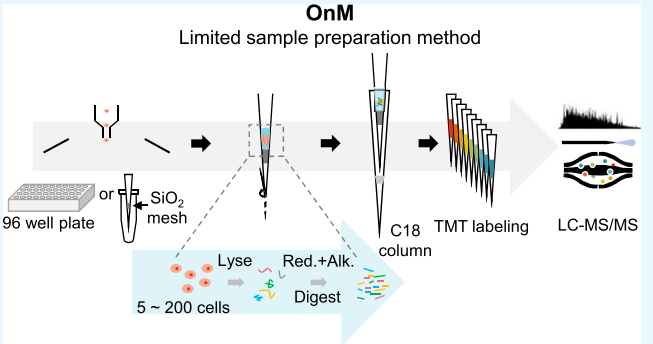
Narae Park,[⊥] Hankyul Lee,[⊥] Yumi Kwon, Shinyeong Ju, Seonjeong Lee, Seongjin Yoo, Kang-Sik Park, and Cheolju Lee*

Cite This: *ACS Omega* 2023, 8, 19741−19751

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Liquid chromatography-tandem mass spectrometry (LC−MS)-based profiling of proteomes with isobaric tag labeling from low-quantity biological and clinical samples, including needle-core biopsies and laser capture microdissection, has been challenging due to the limited amount and sample loss during preparation. To address this problem, we developed OnM (On-Column from Myers et al. and mPOP)-modified on-column method combining freeze−thaw lysis of mPOP with isobaric tag labeling of On-Column method to minimize sample loss. OnM is a method that processes the sample in one-STAGE tip from cell lysis to tandem mass tag (TMT) labeling without any transfer of the sample. In terms of protein coverage, cellular components, and TMT labeling efficiency, the modified On-Column (or OnM) displayed similar performance to the results from Myers et al. To evaluate the lower-limit processing capability of OnM, we utilized OnM for multiplexing and were able to quantify 301 proteins in a TMT 9-plex with 50 cells per channel. We optimized the method as low as 5 cells per channel in which we identified 51 quantifiable proteins. OnM method is a low-input proteomics method widely applicable and capable of identifying and quantifying proteomes from limited samples, with tools that are readily available in a majority of proteomic laboratories.



1. INTRODUCTION

Liquid chromatography-tandem mass spectrometry (LC−MS/MS)-based proteomics requires a large sample amount (e.g., ∼100−1000 μg of protein or ∼10^6−10^7 cells).[1−4] However, different types of samples, including needle-core biopsies, precious clinical samples, and laser capture microdissection (LCM) tissues, contain relatively low and limited amounts of proteome to be analyzed by previous LC−MS/MS methods. Furthermore, using a stable isotope tagging method, such as tandem mass tag (TMT), may require more samples due to sample loss. To overcome this shortcoming, there has been a growing need for an optimized method to obtain proteomes from a low-input amount of proteomic sample.

Current methods have been developed for preparing samples in nanoliter volumes.[5−7] An example of such a method is nanodroplet processing in one pot for trace samples (nanoPOTS). NanoPOTS processes nanoliters of sample volume to minimize loss due to adhesion to the surface of pipette tips or tubes and increase recovery rate. Using nanoPOTS, 1500−3000 proteins have been identified from 10 to 140 cells.[7] The method can produce robust results, employing a robotic pipetting system and nanowell array chips, but these instruments are not readily available in every lab compared to other broad benchtop applications. Cong et al. have identified over 1000 proteins in single-cell proteome (SCP) analysis without the aid of carrier boosting or TMT multiplexing.[8] However, the experiment was done with an ion mobility spectrometer, which majority of labs lack. Another low-input method using a TMT-multiplex, streamlined protocol by Myers et al. (On-Column), was able to identify and quantify 7000 proteins from 3 × 10^5 cells from each primary cell in 12 isolated immune cell types.[5] On-Column method conducts sample preparation in a single tip, decreasing sample loss and processing time.

One of the ways of minimizing sample loss, especially when managing a miniscale amount of sample (e.g., single cell), is by removing steps like desalting demonstrated by ProteOmic sample Preparation (mPOP).[6] By utilizing water for cell lysis, freeze-and-boil method can remove the desalting step in addition to avoiding probe sonication to prevent further sample loss.

Received: March 1, 2023
Accepted: May 15, 2023
Published: May 24, 2023

Table 1. MS Parameters

experiments		performance of OnM (1)	performance of OnM (2)	lower-limit processing with TMT-multiplex	optimization of MS parameters
relevant figure		Figure S2	Figures 2, 3, and S3	Figures 4, S4, and S6	Figure S5
MS operated		Q-exactive	Q-exactive	Q-exactive	Q-exactive
LC condition	column length (cm)	50	50	15	15
	gradient time (acetonitrile %)	4−32 over 160 min	5−40 over 160 min	4−35 over 70 min	4−35 over 70 min
	flow rate (nL/min)	300	300	300	300
	run time (min)	190	190	84	84
MS parameters	MS1 spectra (m/z)	300−1600	300−1600	400−1800	400−1800
	MS1 resolution	70,000	70,000	70,000	70,000
	top N	12	12	10	10
source ionization parameters	ionization source (kV)	2.1	1.9	2.1	2.1
	capillary temperature (°C)	275	275	275	275
MS2 parameters	MS2 resolution	35,000	35,000	35,000	35,000
	AGC	5.00×10^4	5.00×10^4	5.00×10^4	5.00×10^4
					5.00×10^5
					5.00×10^6
	IT (ms)	120	120	300	120, 300
	isolation width (m/z)	1.2	1.2	1.2	1.2
	NCE	32	32	32	32
	fixed first mass (m/z)	120.0	120.0	120.0	120.0
	preferred for peptide match	preferred	preferred	preferred	preferred
	dynamic exclusion (s)	60.0	60.0	60.0	60.0
	intensity threshold	2.50×10^4	2.50×10^4	1.00×10^4	2.50×10^4

We combined the one-STAGE (STop And Go Extraction) tip method for TMT-based proteomic analysis from On-Column and cell lysis method from mPOP (hence given the name OnM) to optimize the method to lower amounts of hundreds of cells. Adapting the methods above to minimize sample loss throughout the entire sample preparation, we were able to decrease the starting amount of sample to 5 cells without the assistance of an automated machine. Our OnM is a method that can be performed using instruments commonly available in proteomic laboratories from cell lysis to TMT labeling without a single transfer of sample minimizing the loss of sample.

2. EXPERIMENTAL SECTION

2.1. A549 Cell Culture. A549 human lung cancer cell line was grown in 15 cm diameter plates in RPMI (Gibco) medium supplemented with 10% fetal bovine serum (FBS) (Gibco) and 1% penicillin/streptomycin/fungizone (PSF) (Gibco) at 5% CO_2 and 95% air in a humidified incubator at 37 °C. Cells were grown to >90% confluence.

2.2. Harvesting Cells. After removal of cell culture media, cells were dissociated by trypsinization with 0.25% of trypsin−ethylenediaminetetraacetic acid (EDTA) (Gibco) by incubating for 5 min at 37 °C. Phosphate-buffered saline (PBS) was added to dissociated A549. Cell pellets were centrifuged at 500g for 3 min at 4 °C and washed three times with ice-cold PBS. Cell counts and viability were determined using trypan blue staining and an automated cell counter (Logos Biosystems).

2.3. Isolation of Limited Number of Cells Using Fluorescence-Activated Cell Sorter. For evaluation of lower limit of OnM using counted cells, harvested cells were isolated using an MA900 cell sorter (SONY) with a 100 μm nozzle for the exact cell count. Fluorescent beads were utilized for optimizing parameters for sorting and to confirm appropriate isolation in each well. 5−200 cells were collected with a semiyield mode in a 96-well plate (Eppendorf) into 4 μL of high-performance liquid chromatography (HPLC)-grade water and stored in a deep-freezer. The isolated cells were prepared for evaluating the lower limit of OnM.

2.4. OnM Sample Preparation. A 20 μL Eppendorf tip was packed with four punches of Whatman QM-A grade SiO_2 mesh with a 16-gauge needle at the bottom of the tip with about 3 mm of the tip left for cell capture. For cell capture, cells that were diluted with PBS from the harvested cells in ∼50 μL PBS were transferred onto the microreactor tip on top of the mesh by pipetting and centrifuging at 3500g in a 1.5 mL tube with an adapter for the tip to ensure the tip is not in contact with PBS after centrifugation. After cells were captured, all of the reagents were added onto the microreactor on top of the SiO_2 using 10 μL gel loading tips (QSP). 4 μL of H_2O was added, and the tip was kinked at the bottom manually. Then, the tip was placed in a 2 mL Eppendorf tube and frozen at -80 °C for storing until further experiment. The tip remained in the tube until digestion was completed. The frozen sample in the tube was heated at 95 °C for 5 min using a thermomixer for hypotonic lysis.

For both 0.1 M tris(2-carboxyethyl)phosphine (TCEP) and 0.4 M chloroacetamide (CAM), 0.44 μL was added to the microreactor to achieve a final concentration of 10 mM TCEP and 40 mM CAM for reduction and alkylation for 5 min at 45 °C. The tip was cooled down to room temperature, and 0.49 μL of 1 M triethylammonium bicarbonate (TEAB) was added to adjust to pH 8 for trypsin digestion. 0.49 μL of 0.1 μg/μL sequencing grade-modified trypsin (Promega) was added, and cell lysates were digested overnight at room temperature. Digestion was performed at room temperature instead of 37

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

°C to minimize evaporation since the handling volume was 5.5 μL, a miniscule amount that could have affected the protein concentration drastically if evaporated.

During digestion, a C18 column tip (DK-Tip C18, DIATECH KOREA) was activated with 50 μL of 80% acetonitrile (ACN) and 0.1% formic acid (FA) and equilibrated with 50 μL of 0.1% trifluoroacetic acid (TFA). Afterward, the kinked microreactor tip was removed from the 2 mL tube, straightened with a pincette, and placed onto the C18 column for the construction of a microreactor-C18 STAGE tip. The captured sample inside the microreactor was transferred by centrifuging at 3500*g* until the digested sample was loaded onto the STAGE tip. After removing the microreactor tip, the C18 STAGE tip was washed twice with 75 μL of 0.1% trifluoroacetic acid (TFA) each time. The digested sample was ready for TMT labeling.

TMT labeling process was based on the On-Column method from Myers et al.[5] For validation of digestion efficiency or LC−MS/MS analysis without TMT labeling, the elution was performed by adding 50 μL of 50% ACN/0.1% FA twice. For TMT labeling, 1 μL of TMT reagent (Thermo) in 100% ACN was added to 100 μL of freshly made HEPES, pH 8, and loaded onto the STAGE tip. The solution was centrifuged at 350*g* until all of the solution passed through C18 resin for 5−10 min. If TMT solution remained in the STAGE tip, the solution was removed by centrifuging at 3500*g*. The column was washed twice with 75 μL of 0.1% TFA for residual HEPES and TMT, and the peptides were eluted twice with 50 μL of 50% ACN/0.1% FA. The eluent was collected in a low-bind 1.5 mL tube with an adaptor for the STAGE tip by centrifuging at 2000*g* and dried by a SpeedVac vacuum concentrator and reconstituted with 0.1% FA (v/v) for LC−MS/MS.

2.5. LC−MS/MS Analysis. The samples in 0.1% FA were performed with an in-house reversed phase C18 column (75 μm i.d., 3 μm, 120 Å, Dr. Maisch GmbH) on an Eksigent nano LC-ultra 1D plus system at a flow rate of 300 nL/min. The HPLC system was connected to a Q-Exactive mass spectrometer (Thermo Scientific, Bremen, Germany), operating in data-dependent acquisition (DDA) mode. The MS parameters are detailed in Table 1.

2.6. Data Analysis. LC−MS/MS data were analyzed using Proteome Discoverer v2.2 (Thermo Fisher Scientific). LC−MS/MS raw files were searched against the reviewed UniProt human database (released in October 2020, 20,536 entries) and the common Repository of Adventitious Proteins (cRAP) (downloaded from thegpm.org/crap/) for the database of protein contaminants frequently found in laboratories with the SEQUEST-HT search engine (Thermo) with the following parameters: full tryptic specificity, two maximum missed cleavage sites, carbamidomethylation of cysteine (+57.021464 Da), TMT zero (+224.15248 Da) or TMT11 (+229.16293 Da) labeling of peptide N-terminus and lysine for fixed modifications, oxidation of methionine (+15.9949 Da), and N-terminal acetylation (+42.0106 Da) for variable modifications. Searches were performed using 15 ppm of precursor ion tolerance and 0.05 Da for fragment ion tolerance. The false discovery rate (FDR) was set to 0.01 at both the peptide and the peptide spectrum match (PSM) levels. Label-free quantification was performed using three nodes—Minora feature detector, feature mapper, and precursor ion quantifier. We used default settings for the Minora feature detector. For feature mapper, we adjusted the maximum RT shift to 10 min and the minimum signal-to-noise ratio to 5. With precursor ion quantifier, precursor abundance was based on intensity and normalized by the total peptide amount. For calculation of TMT labeling efficiency, TMT and carbamylation of peptide N-terminus and lysine side branch were set as dynamic modifications, while other settings remained the same. For overlabeling of TMT, TMT labeling for serine, threonine, tyrosine, and histidine was set to dynamic modification, while TMT labeling at peptide N-terminus and lysine was set to static modification, with other settings maintaining the same.

For the analysis of the data made accessible publicly by Myers et al.,[5] we utilized the raw file titled "M20150425_SAM_TipFxns_NoFxnation_1xgradient_01.raw" via the identifier MassIVE MSV000083180. The raw file was searched with SEQUEST-HT using the same database as Myers et al., Uniprot Mouse (released in October 2014, 41,309 entries). The search parameters were identical to the options used for OnM except carbamylation modification and TMT labeling of TMT11 (+229.16293 Da) were set as variable modifications.

2.7. Bioinformatics Analysis. Ingenuity pathway analysis (IPA) was used to predict the subcellular localization of proteins, and all bioinformatics analysis was performed using Perseus v1.6.6.0. Perseus was used to calculate Pearson correlation and to generate heat maps. Pearson correlation was calculated to evaluate the reproducibility of OnM on the peptide level. The heat map was made for the evaluation of a lower limit of OnM with 5−200 cells. Z-score normalization was performed for the TMT reporter ion intensity within the respective sample. Proteins identified in all samples were clustered based on the Euclidean distance and visualized into a heat map.

To compare the peptide LFQ intensities, Proteome Discoverer v2.2 was utilized. Precursor ion quantifier options for LFQ were as follows: scaling mode on all averages, protein abundance calculation with summed abundances, top 3 for top N, pairwise ratio-based protein ratio calculation, and background-based *t*-test for hypothesis test. The intensities were log2-transformed and normalized by subtracting the median of each sample.

3. RESULTS AND DISCUSSION

3.1. Rationale for the Development of Sample Preparation Methods with Limited Sample Amount (OnM). Various sample preparation methods have been developed for limited input proteomics samples for quantification of a single cell to thousands of cells.[9−16] An example of such methods is the "On-Column" method (Myers et al.), which quantified over 7000 proteins from 3 × 10^5 FACS-sorted primary cells.[5] In this method, one-STAGE tip was used for the entire digestion protocol and a C18 tip was used for TMT labeling. It minimized sample loss by performing the TMT labeling step and the desalting step in a single C18 column and established a simple and rapid sample preparation method, reducing the time of the original protocol of ∼1 h to 7 min. However, in On-Column method, ∼10,000 cells (2 μg of proteins) were physically pipetted for cell disruption. While sufficient in breaking the cells, pipetting could increase sample loss due to the sample remaining on tip walls.

In addition, using On-Column method to deeply profile proteomes of hundreds to thousands of cells can cause complications, one of them being low protein concentration for trypsin digestion. When applying On-Column to prepare

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

A

			Optimal method	
	Number of cells	100	100	300,000
	Predicted protein (μg)	0.015	0.015	45
Myers et al.	Starting vol. (μL)	0.625	10	10
	Final vol. (μL)	5	80	80
	Protein conc. (μg/μL)	0.003	0.00019	0.56

: inappropriate

Modified lysis step

	Number of cells	100	200	400	800	4,000
	Predicted protein (μg)	0.015	0.03	0.06	0.12	0.6
Our platform	Starting vol. (μL)	4	4	4	4	4
	Final vol. (μL)	5.42	5.42	5.42	5.42	5.42
	Protein conc. (μg/μL)	0.003	0.006	0.011	0.022	0.11
	Trypsin (ng/μL)	10	10	10	10	10

B

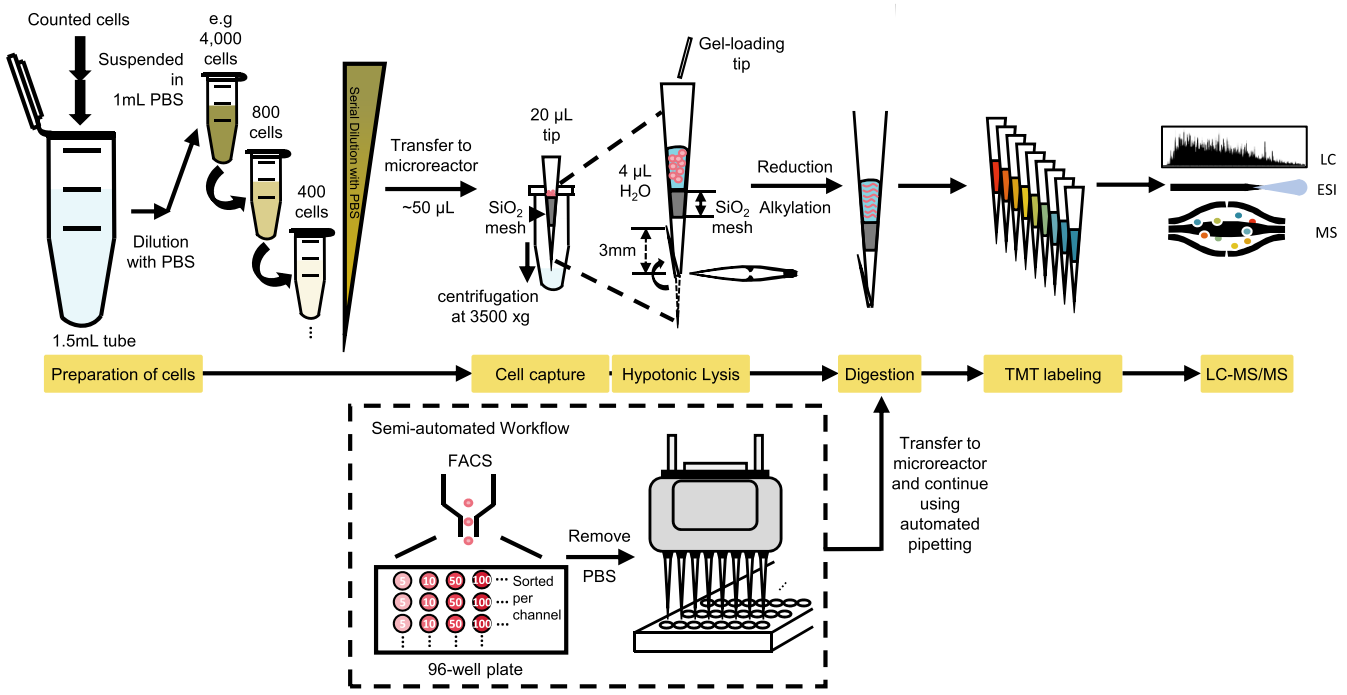


Figure 1. Development of the OnM sample preparation for TMT-based proteomics. (A) When applying the method from Myers et al. to hundreds of cells, either the starting volume would be too small for cell disruption by manual pipetting or the protein concentration would be too low for an efficient trypsin digestion. OnM is the modified On-Column method with freeze-and-thaw lysis and higher trypsin concentration to address these issues. (B) OnM with conventional and suggested semiautomated workflow.

100 cells for LC−MS/MS, the protein concentration for trypsin digestion is 1.9×10^{-4} μg/μL (Figure 1A). According to the Michaelis−Menten equation for enzymatic kinetics, the rate of reaction is proportional to the concentration of substrate. For trypsin, the digestion rate is proportional to protein concentration and low protein concentration results in low digestion rate.[17] Therefore, additional dilution required when utilizing urea as lysis buffer for in-solution digestion with trypsin decreases protein concentration and can cause low digestion efficiency. The digestion efficiency could be increased by preparing a sample with higher protein concentration, but the starting volume would be meager. When adjusting the volume to achieve 3×10^{-3} μg/μL, the same concentration for preparing 100 cells with OnM results in the starting volume of 0.625 μL. This starting volume would be incredibly difficult to manually pipet up and down for cell disruption. Our claim on the relationship between protein concentration and trypsin digestion is exhibited clearly through

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

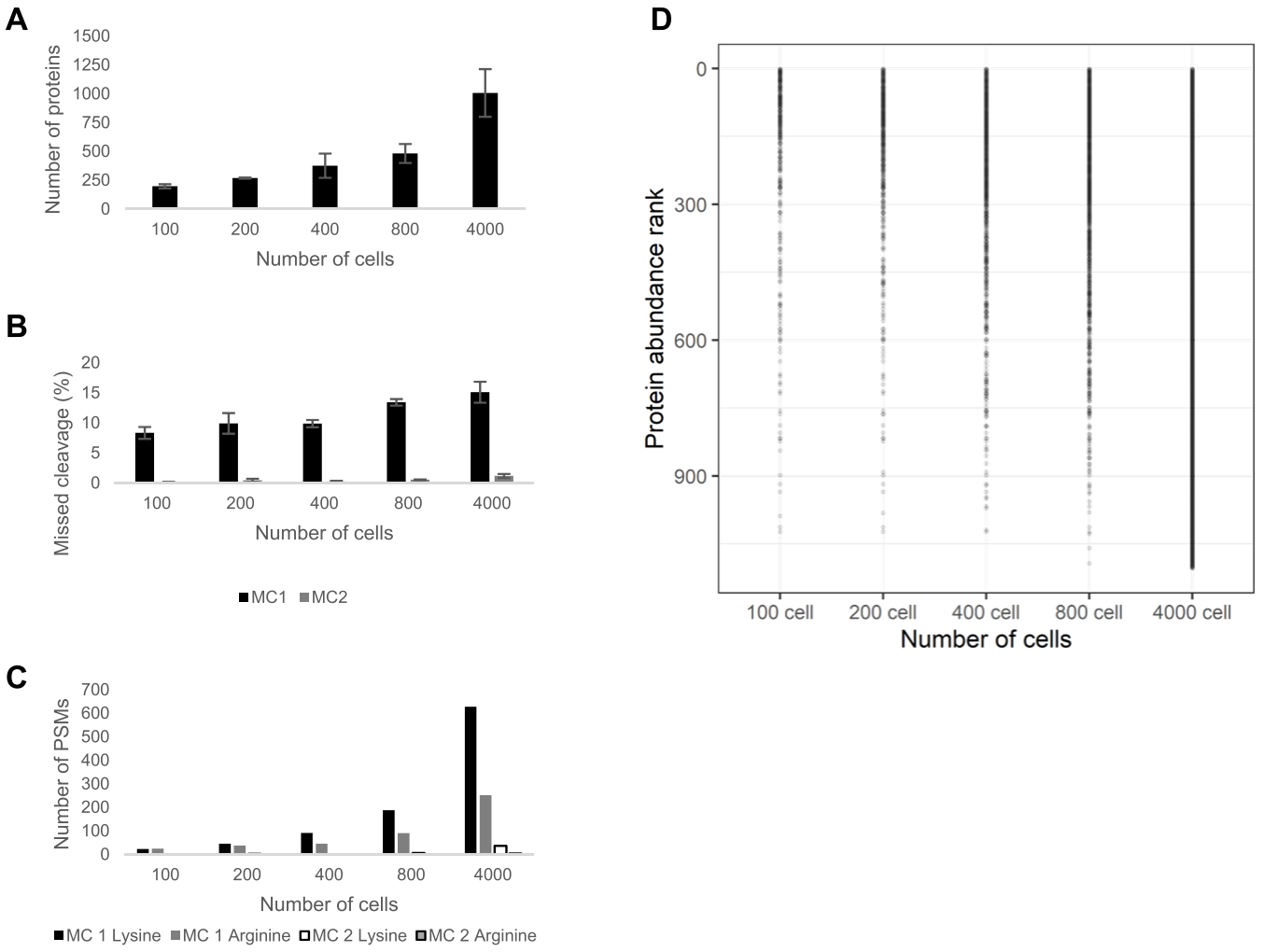


Figure 2. Performance of OnM with various cell amounts. (A) The number of identified proteins in each cell count. Error bar represents standard deviation (SD) of triplicate. (B) The rate of missed cleavages for each sample. MC1 and MC2 represent one missed cleavage and two missed cleavages, respectively. Error bar represents SD of triplicate. (C) Missed cleavages caused by lysine and arginine were distinguished. As the sample amount increases, missed cleavages at lysine are more frequent than at arginine according to the characteristics of trypsin digestion. (D) Abundance rank of proteins identified in all samples based on the LFQ intensity of 4000 cells. As the sample amount decreased, proteins with higher abundance rank were identified.

the experiments in which the concentration of trypsin was adjusted (Figure S1 and Supporting Information).

Therefore, we developed a method that resembles On-Column but can be applied to low-input protein amount as little as 5 cells. We modified the lysis to freeze-and-thaw method and the concentration of trypsin to 10 ng/μL across all samples to address the issue of low protein concentration for digestion, which can result in decreased digestion efficiency and increased missed cleavages and the difficulty in manually handling miniscule volumes (Figure 1A). The protein extraction rate of freeze-and-thaw cell disruption might be lower than the extraction rate of urea-based lysis method, but freeze-and-thaw reduces sample loss since it does not require desalting as it is performed in water.[6] We expect that the combination of freeze-and-thaw method (mPOP) and On-Column method can be used to identify more proteins from a TMT-based multiplexed sample with low number of cells and call this combined method OnM (Figure 1A). The sample preparation by OnM method can be performed either manually or in a semiautomated workflow using FACS (Figure 1B).

First, the performance of OnM method was evaluated by comparing the list of proteins from OnM and the list from On-Column identified by the same search method, as described in Section 2.6. We used the results of triplicate samples with 20,000 A549 cells prepared by OnM and the sample with primary murine immune cells prepared by On-Column without fractionation. Notably, our OnM sample originated from human cell line, while the sample prepared by Myers et al. was from primary cells. We compared results from each of the triplicate to the result of Myers et al. We identified about 3000 proteins in the triplicate prepared by OnM and about 1800 proteins from Myers et al. (Figure S2A). There is a difference in number of identifications between the result of Myers et al. and the result of OnM search method. This may be because Myers et al. utilized Spectrum Mill (Agilent) for search engine, while OnM method utilizes SEQUEST-HT (Thermo).

We evaluated the protein coverage from each method to assess digestion efficiency and observed similar coverage (Figure S2B). Next, we performed gene ontology analysis for cellular components (GO-CC) to examine if proteins from

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

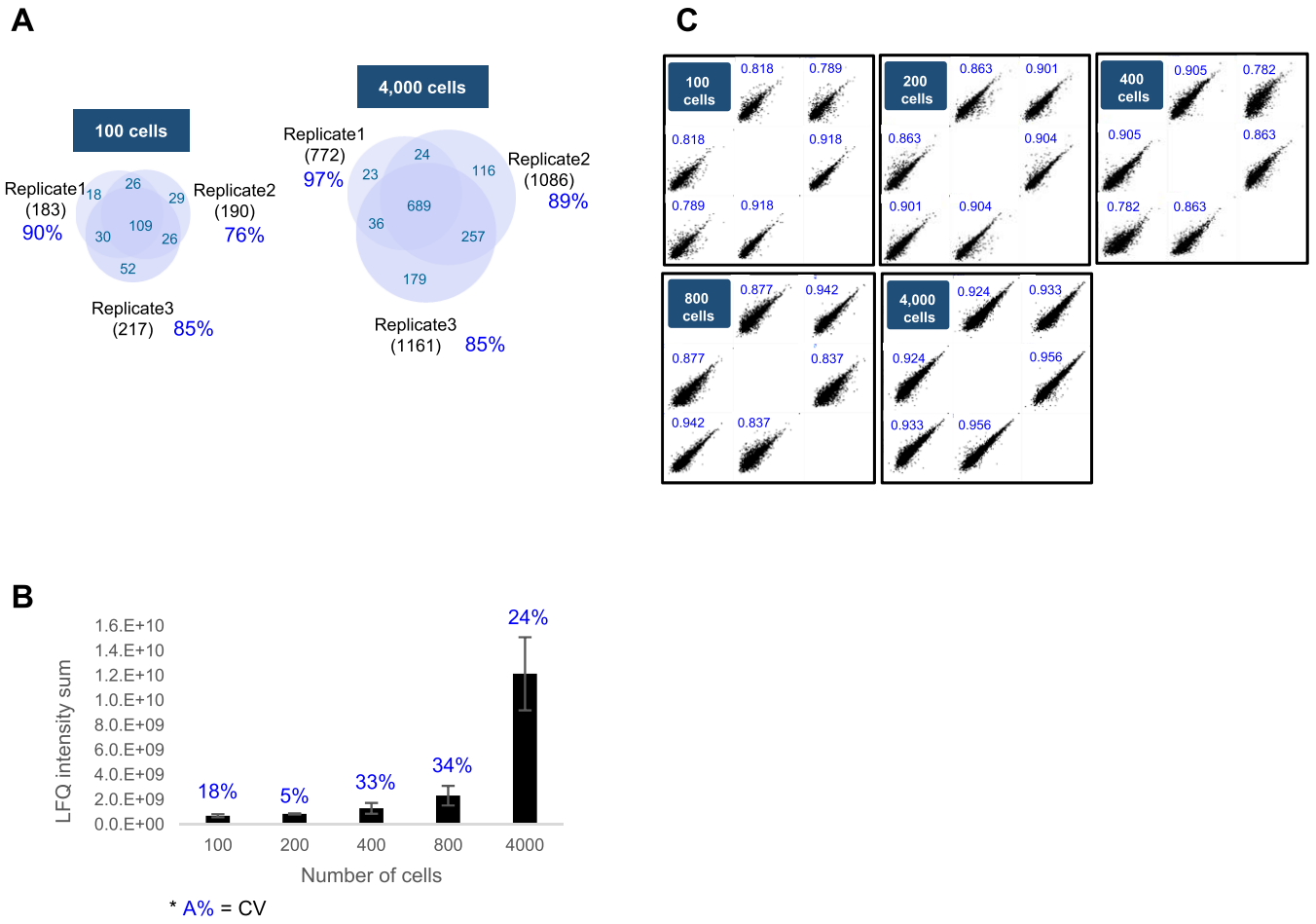


Figure 3. To evaluate the reproducibility of OnM, the overlap of proteins, %CV of intensity sum, and Pearson correlation were analyzed. (A) The overlap of identified proteins with 100 cells and 4000 cells. (B) %CV of LFQ intensity sum for each amount of cells. The error bar represents the standard deviation of triplicate. (C) Pearson correlation of triplicate for each sample.

different cellular components were still identified after the modification of lysis step. The GO-CC results show an almost identical distribution of cellular components (Figure S2C). Since OnM utilizes freeze-and-thaw instead of mechanical disruption such as probe sonication and pipetting up and down for cell disruption, the proportion of hydrophobic proteins, especially membrane proteins, might decrease. However, GO-CC analysis revealed that the number of proteins related to plasma membrane was similar.

Urea, the chaotropic agent used for denaturing proteins, is known to react with primary amines for carbamylation.[18,19] We compared the rate of carbamylation as OnM utilizes H_2O instead of concentrated urea solution. The carbamylation rate was 0.05−0.09% for OnM and 0.47% for On-Column (Figure S2D). The rate of 0.47% may not be significant in profiling proteomes but it highlights the advantage of modifying the lysis step. Comparison of OnM and On-Column methods shows that the modified lysis step and adjusted concentration of trypsin were applicable to the workflow for preparing samples of 20,000 A549 cells.

3.2. Performance of the OnM Platform with Various Cell Amounts. After confirming that the modified method gave comparable results to On-Column, we performed the OnM sample preparation with 100−4000 cells using TMT zero followed by LC−MS/MS with Q-Exactive. The numbers of identified proteins, peptides, and PSMs were 200−1000, 500−5000, and 570−6200, respectively (Figures 2A and S3A).

We examined the number of missed cleavages because the low concentration of protein substrate can decrease the digestion efficiency of trypsin. Peptides without missed cleavage were 84−92% (Figure 2B). Peptides with one missed cleavage or two missed cleavages were 8−15 and 1% or less, respectively. The rate of missed cleavage was below the rate reported by the manufacturer for all samples.[20] Other studies including Zhu et al. and Huang et al. have reported similar missed cleavage rates of trypsin around in both low-input sample and bulk amount.[7,21] We further investigated the ratio of missed cleavages from lysine and arginine. With 100 cells, missed cleavages from lysine and arginine were in similar proportions. With 200−4000 cells, missed cleavages with lysine were more prominent than missed cleavages with arginine (Figure 2C). Missed cleavages from trypsin digestion are known to occur more frequently at lysine than at arginine.[20] The same result was observed when hundreds to thousands of cells were prepared by OnM.

With 100−4000 cells, we inspected traits of the identified proteins as the starting amount of cells were decreased. Since the samples with various cell amounts were labeled with TMT zero, the proteins from each sample were quantitated and compared using LFQ. All identified and quantitated proteins from the sample with 4000 cells were ranked from high to low

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

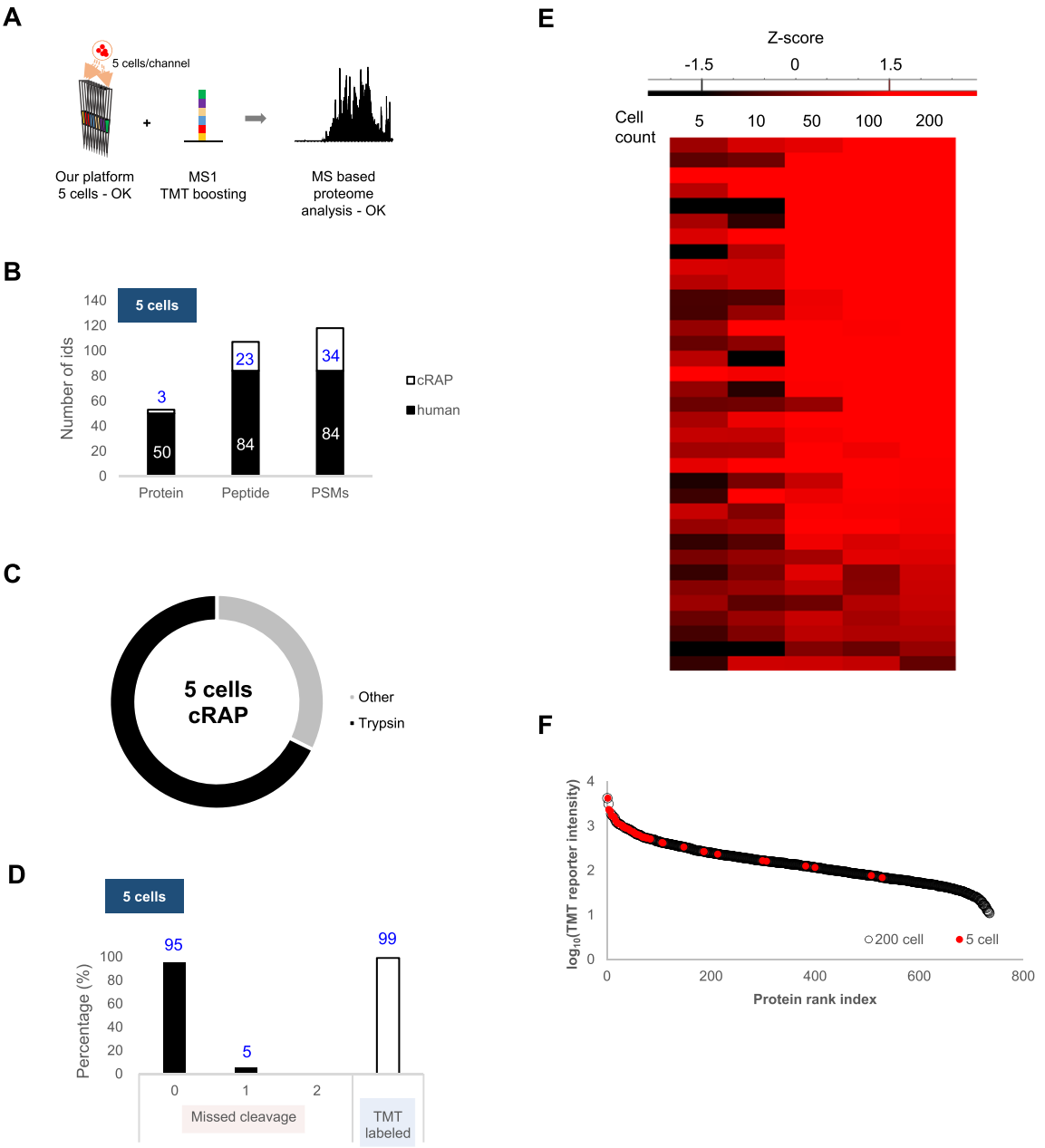


Figure 4. Evaluation of the lowest sample amount that can be processed by OnM. (A) OnM with TMT multiplexing enables LC−MS/MS of sample as low as 5 cells by MS1 boosting. (B) Number of identified proteins, peptides, and PSMs with 5 cells per channel. Since excessive amount of trypsin was used, the identified list was divided into matches from human database versus cRAP. (C) Proportion of trypsin-related PSMs from all PSMs matched with cRAP database. (D) Rate of missed cleavage and efficiency of TMT labeling. (E) Z-score normalization of proteins identified in all samples from 5 to 200 cells. Z-score normalization was first performed for TMT reporter ion intensities within each sample. Then, Z-score values for proteins identified in all samples were extracted and plotted on a single heat map. Proteins of high abundance in bright red are found across 5−200 cells. (F) The abundance rank of proteins identified in 200 cells. Red dots representing proteins identified in both 200 cells and 5 cells are shown to examine the abundance of overlapping proteins.

in terms of abundance. The proteins from samples with other cell amounts were dotted if they were identified in both the respective sample and the sample with 4000 cells. As a result, more abundant proteins were observed, especially when starting amount decreased (Figure 2D).

3.3. Reproducibility of the OnM. After confirming that OnM method is compatible with preparing a low amount of proteins as low as 100 cells, we examined if the method is reproducible. First, we checked the overlap between identified proteins. The proportion of proteins identified in at least two of the replicates with the most amount of cells, 4000 cells, was 76 and 60% for 100 cells, the least amount of cells (Figure 3A). From 200 to 800 cells, the proportion of overlap was similar from 60 to 70% (Figure S3B). To examine if proteins are recovered in relatively similar amounts for each sample, LFQ intensities of all identified proteins within the respective sample were summed and coefficient variant (CV) values of LFQ intensity sum with triplicate were compared for each cell amount. The lowest %CV value was 5% with 200 cells, and 18−34% CV values were observed for other cell amounts. Furthermore, a linear increase in the LFQ intensity sum was observed as the starting amount increased (Figure 3B). Lastly,

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

Pearson correlation values of LFQ intensities for proteins between each pair of the triplicate were 0.78−0.96. With 4000 cells, high correlation values over 0.9 were observed, while the correlation values declined below 0.8 with a decreased starting amount of cells. A correlation value of 0.789 was from 100 cells, the lowest starting amount, indicating that the reproducibility diminished along with the initial sample amount (Figure 3C). Despite having 24% CV for the whole proteome LFQ intensity sum, the Pearson correlation values for 4000 cells were 0.92−0.96. Looking at the overlapping proteins within triplicate, LFQ intensity sum, and Pearson correlation, we can conclude that OnM is a method that can generate reproducible results.

3.4. Evaluation of the Lower-Limit Processing Capabilities. We further tested the lowest amount of cells OnM can prepare with 5−200 cells (Figure 4). Using isobaric tags such as TMT can boost MS detection sensitivity since the intensity of identical peptide from different samples will be summed in MS1. Then, the fragmentation of the reporter ions enables the relative quantification of such peptide from different samples in MS2. Therefore, merging multiple samples labeled with TMT can be used to increase the starting amount. Fluorescence-activated cell sorting (FACS) is a flow cytometry technique that can sort various cells into any desired amount according to specific phenotypes of the cells. So, we utilized FACS and TMT to measure the lowest number of cells for a quantifiable starting amount by preparing samples from 5 cells to 200 cells.

We used TMT 10-plex for labeling the samples emptying TMT 126 channel, which was equivalent to using a 9-plexed sample. Preparing 5 cells per TMT channel, we were able to profile proteomes using the combination of FACS and TMT 9-plex (Figure 4A). Since the amount of trypsin was relatively greater than the starting amount of protein, we utilized cRAP database, a list of common contaminant proteins, in our search to see whether trypsin caused any interference during profiling. Given that the modified trypsin methylated at N^{ε}-amine of lysine was used, we included lysine methylation for a variable modification for our search to identify as many contaminants as possible. With 5 cells, we identified 34 PSMs that were matched to the cRAP database (Figure 4B). Out of 34 PSMs, 23 were related to trypsin consisting of 19% of all identified PSMs (Figure 4C). The number of identified proteins, peptides, and PSMs for 5−200 cells can be found in Figure S4A and Tables S1 and S2. For 5 and 10 cells, trypsin-related PSMs were 19 and 22% of all of the identified PSMs, respectively, while for 50−200 cells trypsin-related PSMs were only 6−1% (Figure S4B). Despite about one-fifth of all PSMs being matched to trypsin due to the excess amount of trypsin for digestion with samples below 10 cells, out of 53 identified proteins, 50 proteins were identified as human proteins unrelated to trypsin (Figure 4B). The list of identified proteins and peptide groups for 5−200 cells can be found in Tables S3 and S4A.

Digestion efficiency for samples with 5−200 cells was evaluated by calculating the rate of missed cleavages to ensure that the modified lysis method and fixed concentration of trypsin were applicable to decreased amounts of cells. The number of PSMs without a missed cleavage for 5 cells was 95% and over 84% for 10−200 cells, showing high digestion efficiency (Figures 4D and S4C).

The labeling efficiency was 99−100% for all samples with a decreased amount of cells (Figures 4D and S4D). When analyzing the overlabeling rate of TMT at serine, threonine, and tyrosine residues, with 5 and 10 cells, no overlabeling was observed. With 50−200 cells, serine was the most overlabeled residue. However, the proportion of overlabeling for 50−200 cells was 3%, which was only a small fraction of labeled proteins (Figure S4E).

To examine reproducibility of the sample preparation method, all TMT reporter ion intensities of all quantified proteins were added for each channel and the %CV values of this TMT reporter intensity sum for channels of the same starting amount were assessed. The %CV for 5−200 cells were between 9 and 21% (Figure S4F). Also, there was a steady increase in the sum of intensity of precursor ion for each channel as the starting sample amount was increased (Figure S4G). These results show that OnM is a reproducible sample preparation method that is applicable for a low starting amount of sample since the difference in %CV is relatively small and the number of identified proteins decreased as the starting amount of cell decreased. For quantitative analysis, we performed z-score normalization for proteins in each sample based on TMT reporter ion intensities within each sample. We then evaluated the abundance distribution of 35 proteins identified in all different starting amounts. For 35 proteins, Z-score values within each sample were extracted and plotted onto a heat map (Figure 4E). Out of 35 proteins, the ones quantified as high abundance in channels with 200 cells are also quantified in channel with 5 cells. Note that most of the proteins found in 200 cells are marked red indicating high abundance in Figure 4E. When the abundance of the same 35 proteins was examined in the sample with 200 cells, the abundances of these proteins were mostly high (Figure 4F).

To maximize the reporter ion intensity of nanoscale samples, we optimized automatic gain control (AGC) and maximum injection time (maximum IT) of mass spectrometer (Figure S5).[22] For the optimization, we utilized various amounts of peptide from the same sample from 2.5 to 10 ng aliquoted appropriately. Optimization of AGC and maximum IT can increase the reporter ion intensity of nanoscale sample for MS analysis.[23] So, we increased maximum IT from 120 to 300 ms and AGC from 5×10^4 to 5×10^6 (Figure S5A) to observe any changes in the number of protein IDs (Figure S5B). The number of protein IDs almost doubled when increasing the maximum IT while maintaining AGC using 2.5 ng. For 5 ng and 10 ng, increasing AGC and maximum IT increased TMT reporter ion intensities (Figure S5C), but most of the protein IDs did not increase (Figure S5B). Therefore, with a starting amount of 2.5 ng or less, increasing maximum IT can be beneficial in increasing the number of identifications and TMT reporter ion intensity. Since the amount of protein in 5 or 10 cell is expected to be below 2.5 ng,[24] the maximum IT was adjusted to 300 ms for assessing OnM method with 5−200 cells. The average IT across all sample amounts was around 100 ms in contrast to our expectation of using full IT, 300 ms, especially with 5 and 10 cells (Figure S5D,E). We further analyzed this result to find the reason behind the unexpected outcome. We first identified the overlapping proteins between 2.5 ng of the purified peptide and 5 cells prepared by OnM. Within the overlapping proteins, we found the average IT of PSMs to be 209 ms and 126 ms for 2.5 ng and 5 cells, respectively (Figure S6). The most notable difference was the MS/MS identification rate for method 2 (AGC 5×10^4 and maximum IT 300 ms). About 37.8% of MS/MS spectra were matched as PSMs for 2.5 ng, whereas only about 2.6% of MS/

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

MS spectra were matched as PSMs for 5 cells. Furthermore, a 2.5 ng sample displays the distribution of isolation interference with smaller values of interference in comparison to samples prepared with cells, indicating that the purity of 2.5 ng sample is relatively greater (Figure S7). These two data, identification rate and the distribution of isolation interference, are two of the few that can explain the discrepancy in the results when adjusting the maximum IT. We suspect the difference in matching rate may be due to the origins of the sample—one from purified peptides and the other from sorted cells (see the Supporting Information).

There has been a number of studies for proteomic analysis of limited amount of sample as small as a single cell.[7,8,25] TMT-based single-cell proteomics (SCP) employs one of two sample preparation methods: the usage of automated machines including robotics or using excessive carrier. Our OnM method may have quantified less proteins as the method above. However, what sets OnM apart is that the OnM method handles a minute number of cells without the use of a carrier. Usage of excessive carrier limits the identification of proteins to the proteins present in the carrier and therefore restricting the possibilities of the study sample. We expect that OnM can be utilized when a carrier cannot represent the various study samples. Also, the amount of carrier is crucial when using excessive carrier. If the carrier cannot be prepared in excessive amount relative to the sample, the carrier method cannot be used. In this case, we believe our OnM method can be applied for sample preparation.

4. CONCLUSIONS

In this work, we optimized the OnM sample preparation method for TMT-based deep profiling of proteome with limited sample amount. Compared to commonly used bulk sample preparations, the OnM method minimizes sample loss by performing all of the processing steps in a STAGE tip with limited input amount of cells and combining TMT labeling with desalting procedure. All of the equipment and instruments used in OnM—STAGE material and tip, pipette, centrifuge—are frequently utilized in proteomic laboratories, making the OnM method viable in numerous labs with a need for processing a limited amount of sample without the assistance of an automated system.

In single-cell proteomics (SCP) and low-input methods for quantification, various methods of MS signal boosting with excessive amount of carrier are being used. Notably, a TMT SCP method has quantified over 1000 proteins within a single cell. Even though our OnM method quantified less proteins than other SCP methods, we expect that OnM could be used for samples with limited amount when carriers cannot be utilized.

Furthermore, another advantage of OnM is the removal of the need to transfer the sample from the STAGE tip for TMT labeling, which decreases sample loss. For assessment of OnM, we performed various analyses. First, we compared the results from our workflow with the modified lysis step to results from Myers et al. to assess the performance of the modified workflow. Next, we evaluate the capacity of proteome profiling when the starting amount was decreased to 100−4000 cells. Finally, we tested the lowest amount of sample (5−200 cells) that can be prepared by OnM by utilizing TMT 9-plex. We were able to identify and quantify ∼51 proteins from 5 cells without using excess amount of carriers. Despite that 51 proteins identified in 5 cells are not as many as the number of proteins identified in single cell by other low-input methods, we expect the number of identified proteins prepared by OnM can increase with a more recent mass spectrometer instead of Q-Exactive. Furthermore, we anticipate that OnM can be performed in a semiautomated workflow with FACS, 96-well plate, and automated pipetting system, as described in Figure 1B.

We believe that the OnM method increases throughput and sensitivity for limited input samples and can potentially enable the analysis of various biomedical and biological samples with limited quantity.

■ ASSOCIATED CONTENT

Data Availability Statement

The raw files have been deposited in the ProteomeXchange Consortium via the MassIVE (http://massive.ucsd.edu) under the identifier MSV000091734 and the K-BDS (https://kbds.re.kr/) under the identifier PRJKA2085526.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c01392.

Figure S1: Digestion efficiency based on substrate (protein) concentration and enzyme (trypsin) concentration; Figure S2: Comparison of method performance for OnM and On-Column (Myers et al.); Figure S3: Performance of OnM; Figure S4: Evaluating the lowest amount of sample processable with OnM; Figure S5: Optimization of IT and AGC to maximize the intensity of reporter ion with nanoscale samples; Figure S6: The distribution of IT for proteins identified in both 2.5 ng and 5 cells; Figure S7: The distribution of interference for precursors of identified PSMs; Table S1: Number of protein, peptide group, and PSM identifications using 5−200 cells; Table S2: Proportion of trypsin identified from the cRAP database; Table S3: List of identified proteins from 9-plex of 5−200 cells in Figure S4; and Table S4: List of identified peptides from 9-plex of 5−200 cells in Figure S4 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Cheolju Lee − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea; Division of Bio-Medical Science & Technology, KIST School, University of Science and Technology, Seoul 02792, Republic of Korea;* orcid.org/0000-0001-8482-4696; Phone: +82-2-958-6788; Email: clee270@kist.re.kr; Fax: +82-2-958-6919

Authors

Narae Park − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea; KHU-KIST Department of Converging Science and Technology, Kyung Hee University, Seoul 02447, Korea*

Hankyul Lee − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea; KHU-KIST Department of Converging Science and Technology, Kyung Hee University, Seoul 02447, Korea*

Yumi Kwon − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul*

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751

02792, Republic of Korea; orcid.org/0000-0003-0523-6197

Shinyeong Ju − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea;* orcid.org/0000-0001-5483-4690

Seonjeong Lee − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea; Division of Bio-Medical Science & Technology, KIST School, University of Science and Technology, Seoul 02792, Republic of Korea*

Seongjin Yoo − *Chemical & Biological Integrative Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea; Division of Bio-Medical Science & Technology, KIST School, University of Science and Technology, Seoul 02792, Republic of Korea*

Kang-Sik Park − *KHU-KIST Department of Converging Science and Technology, Kyung Hee University, Seoul 02447, Korea; Department of Physiology, School of Medicine, Kyung Hee University, Seoul 02447, Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c01392

Author Contributions

⊥Equal contribution.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The study was supported by a grant (2022M3H9A2096187) from the National Research Foundation of Korea and grants (intramural, Great Challenge) from the Korea Institute of Science and Technology.

■ REFERENCES

(1) Mertins, P.; Mani, D. R.; Ruggles, K. V.; Gillette, M. A.; Clauser, K. R.; Wang, P.; Wang, X.; Qiao, J. W.; Cao, S.; Petralia, F.; Kawaler, E.; Mundt, F.; Krug, K.; Tu, Z.; Lei, J. T.; Gatza, M. L.; Wilkerson, M.; Perou, C. M.; Yellapantula, V.; Huang, K. L.; Lin, C.; McLellan, M. D.; Yan, P.; Davies, S. R.; Townsend, R. R.; Skates, S. J.; Wang, J.; Zhang, B.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Ding, L.; Paulovich, A. G.; Fenyo, D.; Ellis, M. J.; Carr, S. A.; Nci, C. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **2016**, *534*, 55−62.

(2) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359−362.

(3) Zhang, B.; Wang, J.; Wang, X.; Zhu, J.; Liu, Q.; Shi, Z.; Chambers, M. C.; Zimmerman, L. J.; Shaddox, K. F.; Kim, S.; Davies, S. R.; Wang, S.; Wang, P.; Kinsinger, C. R.; Rivers, R. C.; Rodriguez, H.; Townsend, R. R.; Ellis, M. J.; Carr, S. A.; Tabb, D. L.; Coffey, R. J.; Slebos, R. J.; Liebler, D. C.; Nci, C. Proteogenomic characterization of human colon and rectal cancer. *Nature* **2014**, *513*, 382−387.

(4) Zhang, H.; Liu, T.; Zhang, Z.; Payne, S. H.; Zhang, B.; McDermott, J. E.; Zhou, J. Y.; Petyuk, V. A.; Chen, L.; Ray, D.; Sun, S.; Yang, F.; Chen, L.; Wang, J.; Shah, P.; Cha, S. W.; Aiyetan, P.; Woo, S.; Tian, Y.; Gritsenko, M. A.; Clauss, T. R.; Choi, C.; Monroe, M. E.; Thomas, S.; Nie, S.; Wu, C.; Moore, R. J.; Yu, K. H.; Tabb, D. L.; Fenyo, D.; Bafna, V.; Wang, Y.; Rodriguez, H.; Boja, E. S.; Hiltke, T.; Rivers, R. C.; Sokoll, L.; Zhu, H.; Shih, I. M.; Cope, L.; Pandey, A.; Zhang, B.; Snyder, M. P.; Levine, D. A.; Smith, R. D.; Chan, D. W.; Rodland, K. D.; Investigators, C. Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* **2016**, *166*, 755−765.

(5) Myers, S. A.; Rhoads, A.; Cocco, A. R.; Peckner, R.; Haber, A. L.; Schweitzer, L. D.; Krug, K.; Mani, D. R.; Clauser, K. R.; Rozenblatt-Rosen, O.; Hacohen, N.; Regev, A.; Carr, S. A. Streamlined Protocol for Deep Proteomic Profiling of FAC-sorted Cells and Its Application to Freshly Isolated Murine Immune Cells. *Mol. Cell. Proteomics* **2019**, *18*, 995−1009.

(6) Specht, H.; Harmange, G.; Perlman, D. H.; Emmott, E.; Niziolek, Z.; Budnik, B.; Slavov, N. Automated sample preparation for high-throughput single-cell proteomics. *bioRxiv* **2018**, No. 399774.

(7) Zhu, Y.; Piehowski, P. D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R. J.; Shukla, A. K.; Petyuk, V. A.; Campbell-Thompson, M.; Mathews, C. E.; Smith, R. D.; Qian, W. J.; Kelly, R. T. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat. Commun.* **2018**, *9*, No. 882.

(8) Cong, Y.; Motamedchaboki, K.; Misal, S. A.; Liang, Y.; Guise, A. J.; Truong, T.; Huguet, R.; Plowey, E. D.; Zhu, Y.; Lopez-Ferrer, D.; Kelly, R. T. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chem. Sci.* **2020**, *12*, 1001−1006.

(9) de Graaf, E. L.; Pellegrini, D.; McDonnell, L. A. Set of Novel Automated Quantitative Microproteomics Protocols for Small Sample Amounts and Its Application to Kidney Tissue Substructures. *J. Proteome Res.* **2016**, *15*, 4722−4730.

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

(11) Martin, K.; Zhang, T.; Lin, T. T.; Habowski, A. N.; Zhao, R.; Tsai, C. F.; Chrisler, W. B.; Sontag, R. L.; Orton, D. J.; Lu, Y. J.; Rodland, K. D.; Yang, B.; Liu, T.; Smith, R. D.; Qian, W. J.; Waterman, M. L.; Wiley, H. S.; Shi, T. Facile One-Pot Nano-proteomics for Label-Free Proteome Profiling of 50-1000 Mammalian Cells. *J. Proteome Res.* **2021**, *20*, 4452−4461.

(12) Muller, T.; Kalxdorf, M.; Longuespee, R.; Kazdal, D. N.; Stenzinger, A.; Krijgsveld, J. Automated sample preparation with SP3 for low-input clinical proteomics. *Mol. Syst. Biol.* **2020**, *16*, No. e9111.

(13) Petelski, A. A.; Emmott, E.; Leduc, A.; Huffman, R. G.; Specht, H.; Perlman, D. H.; Slavov, N. Multiplexed single-cell proteomics using SCoPE2. *Nat. Protoc.* **2021**, *16*, 5398−5425.

(14) Specht, H.; Emmott, E.; Petelski, A. A.; Huffman, R. G.; Perlman, D. H.; Serra, M.; Kharchenko, P.; Koller, A.; Slavov, N. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **2021**, *22*, 50.

(15) Weke, K.; Singh, A.; Uwugiaren, N.; Alfaro, J. A.; Wang, T.; Hupp, T. R.; O'Neill, J. R.; Vojtesek, B.; Goodlett, D. R.; Williams, S. M.; Zhou, M.; Kelly, R. T.; Zhu, Y.; Dapic, I. MicroPOTS Analysis of Barrett's Esophageal Cell Line Models Identifies Proteomic Changes after Physiologic and Radiation Stress. *J. Proteome Res.* **2021**, *20*, 2195−2205.

(16) Wu, R.; Xing, S.; Badv, M.; Didar, T. F.; Lu, Y. Step-Wise Assessment and Optimization of Sample Handling Recovery Yield for Nanoproteomic Analysis of 1000 Mammalian Cells. *Anal. Chem.* **2019**, *91*, 10395−10400.

(17) Nelson, D. L.; Cox, M. *Principles of Biochemistry*; W. H. Freeman and Company: New York, 2017.

(18) Stark, G. R. S.; H, W.; Moore, S. Reactions of the Cyanate Present in Aqueous Urea with Amino Acids and Proteins. *J. Biol. Chem.* **1960**, *235*, 3177−3181.

(19) Sun, S.; Zhou, J. Y.; Yang, W.; Zhang, H. Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. *Anal. Biochem.* **2014**, *446*, 76−81.

(20) Saveliev, S.; B, M.; Zubarev, R.; Szapacs, M.; Budamgunta, H.; Urh, M. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nat. Methods* **2013**, *10*, i−ii.

(21) Huang, E. L.; Piehowski, P. D.; Orton, D. J.; Moore, R. J.; Qian, W. J.; Casey, C. P.; Sun, X.; Dey, S. K.; Burnum-Johnson, K. E.; Smith, R. D. SNaPP: Simplified Nanoproteomics Platform for

Reproducible Global Proteomic Analysis of Nanogram Protein Quantities. *Endocrinology* **2016**, *157*, 1307−1314.
(22) Kwon, Y.; Lee, S.; Park, N.; Ju, S.; Shin, S.; Yoo, S.; Lee, H.; Lee, C. Phosphoproteome Profiling Using an Isobaric Carrier without the Need for Phosphoenrichment. *Anal. Chem.* **2022**, *94*, 4192−4200.
(23) Tsai, C. F.; Zhao, R.; Williams, S. M.; Moore, R. J.; Schultz, K.; Chrisler, W. B.; Pasa-Tolic, L.; Rodland, K. D.; Smith, R. D.; Shi, T.; Zhu, Y.; Liu, T. An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy for Precise Quantitative Single-cell Proteomics. *Mol. Cell. Proteomics* **2020**, *19*, 828−838.
(24) Volpe, P.; Eremenko-Volpe, T. Quantitative studies on cell proteins in suspension cultures. *Eur. J. Biochem.* **1970**, *12*, 195−200.
(25) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **2018**, *19*, 161.

https://doi.org/10.1021/acsomega.3c01392
ACS Omega 2023, 8, 19741−19751