Supporting Information

On-capillary Alkylation Micro-reactor: A Facile Strategy for Proteo-metabolome Profiling in the Same Single Cells

Yingyun He,[a,b] Huiming Yuan,*[a] Yu Liang,[a] Xinxin Liu,[a] Xiaozhe Zhang,[a] Yahui Ji,[a] Baofeng Zhao,[a] Kaiguang Yang,[a] Jue Zhang,[c] Shen Zhang,[c] Yukui Zhang,[a] and Lihua Zhang*[a]

a. CAS Key Laboratory of Separation Science for Analytical Chemistry, National Chromatographic Research and Analysis Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China.

b. University of Chinese Academy of Sciences, Beijing 100049, China

c. Clinical Research Center for Reproduction and Genetics in Hunan Province, Reproductive and Genetic Hospital of CITIC-XIANGYA, Changsha 410013, China

* Corresponding Authors: huimingyuan@dicp.ac.cn; lihuazhang@dicp.ac.cn

Table of Contents

Experimental section

Reagents and materials

Fused silica capillaries were purchased from Polymicro Technologies (Phoenix, AZ). 3-(trimethoxysilyl) propyl methacrylate (γ-MAPS, 98%), polyethylenimine (PEI, average Mw = 800, branched), sodium dodecyl sulfate (SDS), tris(2-carboxyethyl) phosphine (TCEP), myoglobin (from equine heart) and formic acid (FA, HPLC grade) were all obtained from Sigma-Aldrich (St. Louis, MO, USA). Glycidyl methacrylate (GMA, >95%) and N-succinimidyl iodoacetate (IAA-NHS, >95%) were ordered from TCI Shanghai Development Co., Ltd. (Shanghai, China). 2,2'-Azobisisobutyronitrile (AIBN, 99%, recrystallized) was purchased from Macklin (Shanghai, China). n-Propanol (analytical grade) were purchased from Tianjin Damao Chemical Reagent Factory (Tianjin, China). Acetonitrile (ACN, HPLC grade) and methanol (HPLC grade) were purchased from Merck (Darmstadt, Germany). Trypsin and Lys-C were purchased from Promega (Madison, WI, USA). Deionized water was used throughout the experiments (Millipore, Milford, MA). C18 column packing material (ReproSil-Pur 120 Å C18-AQ 1.9 μm) was purchased from Dr.Maisch company (Tübingen, Germany).

Preparation of OCAM

The first step was pretreatment and silanization of the fused silica capillary. Specifically, the capillary (200 μm i.d., 360 μm o.d., 5 m long) was flushed with 1 M NaOH and 1 M HCl to activate the inner wall, then the capillary was washed with water and methanol for 1 h, respectively, and dried with nitrogen. Subsequently, the solution of 50% (v/v) γ-MAPS in methanol was filled into the pretreated capillary. After both ends were sealed with rubber, the capillary was heated at 50 °C for 24 h, then the capillary was washed with methanol for 1 h and dried with nitrogen. After silanization, the solution containing 0.1% (m/m) AIBN and 10% (m/m) GMA in n-Propanol was injected into the capillary, and the capillary was sealed and incubated in water bath at 60 °C for 6 h, then washed with methanol and dried with nitrogen.

After polymerization, 10 mg/mL PEI was introduced into the open tubular capillary column (OTC). The capillary was sealed and reacted at 50 °C for 5 h, then washed to neutral with water, followed by washing with methanol and drying with nitrogen. Finally, the OTC was IAA-functionalized by IAA-NHS

to prepare the on-capillary alkylation micro-reactor (OCAM). 10 mg/mL IAA-NHS solution in methanol containing 16% (v/v) phosphate buffer (50 mM, pH = 8.0) was ultrasonically mixed and introduced into the OTC, sealed and reacted at 40 °C for 24 h. Then OCAM was washed with methanol and finally dried with nitrogen.

Before sample processing, OCAM was pretreated with myoglobin to saturate the possible adsorption sites to reduce sample loss. 2 mg/mL myoglobin solution was introduced into OCAM and kept at room temperature for 15 min. Then OCAM was washed with water and methanol for 1 h, respectively, and dried with nitrogen prior to use.

Characterization of OCAM

The morphology and elemental composition of the inner surface of OCAM were characterized by scanning electron microscope (SEM) and energy dispersive X-ray spectroscopy (EDS, JSM-7800F, JEOL, Japan). The protein binding capacity of OCAM was evaluated using BSA as a model protein. Briefly, 1 mg/mL BSA solution containing 0.05% (w/v) SDS and 5 mM TCEP was incubated in a water bath at 95 °C for 15 min, then introduced into OCAM (200 μm i.d. × 110 cm). After OCAM was sealed and reacted at room temperature for 6 h, the BSA solution was collected and the concentration was measured by a detergent-compatible Bradford protein assay kit. Finally, the protein binding capacity of OCAM was calculated using Equation 1.

$$protein\ binding\ capacity = \frac{(c_0 - c)v}{l} \quad (1)$$

where c_0 is the initial protein concentration; c is the protein concentration after treatment with OCAM; v is the volume of processed solution; l is the length of OCAM.

The mechanism of protein binding on OCAM was further validated using FITC-labeled BSA. 60 ng of FITC-BSA in 50 mM TCEP solution was injected into OCAM and the bare OTC, respectively. After incubation at 95 °C for 5 min, OCAM and OTC were washed sequentially with 50% (v/v) methanol and 50 mM ammonium bicarbonate (ABC) buffer (pH = 8.0) to remove unreacted proteins. Then the fluorescence characterizations of OCAM and OTC were performed using a fluorescence microscope (Nikon ECLIPSE Ti-S, Japan) in dark, and FITC-labeled BSA was visualized under a green laser.

HeLa cell preparation

The HeLa cells were cultured at 37 °C under 5% CO_2 in the minimum essential medium (MEM), which was supplemented with 10% fetal bovine serum (FBS) and 1000 units/mL penicillin/streptomycin. After cell culture, the adherent cells were harvested by treatment with 0.25% (w/v) trypsin solution, and then washed with PBS for three times. Subsequently, the cell concentration was measured using a hemocytometer, then the cells were diluted into concentrations of 5×10^5 and 2.5×10^5 cells/mL using PBS buffer as soon as possible to minimize cell lysis, and the diluted HeLa cells were immediately processed by OCAM.

Sample processing of 10-100 HeLa cells with OCAM

One end of OCAM (5 cm long) was first connected to a 1 μL syringe (Hamilton) for sample loading by manual manipulation or with a syringe pump (Harvard) operated in the withdraw mode. For sample processing of 100 HeLa cells, 100 nL lysis buffer containing 0.05% (w/v) SDS and 50 mM TCEP in 50 mM ABC buffer, 200 nL HeLa cell solution with concentration of 500 cells/μL, and 100 nL of the same lysis buffer were sequentially injected into OCAM. The sampling process was performed carefully to avoid introducing air bubbles between the above solutions. There were approximately 100 cells in OCAM with a total processing volume of 400 nL. After both ends were sealed with rubbers, OCAM was sonicated at 4 °C for 5 min, followed by incubation in a water bath at 95 °C for 5 min. Protein denaturation, reduction and covalent binding were performed in a single step. Then OCAM was flushed with 300 μL 50% (v/v) methanol to remove SDS, and washed with 300 μL ABC buffer (50 mM, pH = 8.0), which was pushed out completely before digestion. Subsequently, a mixed solution of 40 ng Trypsin and Lys-C (1:1, m/m) in 50 mM ABC buffer was introduced into OCAM, the capillary was sealed and incubated in a water bath at 37 °C overnight for digestion. Finally, OCAM was rinsed with 5 μL loading buffer containing 0.1% (v/v) FA in water to elute the peptides for LC-MS/MS analysis.

For sample processing of 50 HeLa cells, 100 nL HeLa cell solution (500 cells/μL) and 100 nL lysis buffer were introduced into OCAM to generate approximately 50 cells with a total processing volume of 300 nL. For fewer HeLa cells, 40 nL HeLa cell solution with concentrations of 500 and 250 cells/μL, and 40-60 nL lysis buffer were introduced into OCAM for 20 and 10 cells, respectively, with a total processing volume of 120-160 nL. All processing procedures were the same as described above.

Moreover, the solution of Trypsin and Lys-C with different concentrations (20 ng/μL, 10 ng/μL and 4 ng/μL) were introduced for digestion at an enzyme-to-protein ratio of 2:1 (m/m). We further performed a blank experiment to evaluate the effect of leaked proteins and contaminates in the cell suspension. The supernatant of the cell suspension was collected by centrifugation and processed with OCAM, and the processing steps of blank samples were the same as those for 10-100 cell samples.

Collection of mouse oocytes

All animal procedures were carried out under the rules and guidelines of the Science and Ethics Committee of Dalian Institute of Chemical Physics, Chinese Academy of Sciences (Approval number: DICPEC2323). The specific pathogens free (SPF)-grade C57BL mice (4-week-old) were purchased from Hunan STA Laboratory Animal Co., Ltd. and housed in Department of Laboratory Animals in Central South University (Changsha, P. R. China). Oocytes were obtained from the ovaries of 4-week-old female mice 48 h after intraperitoneal injection of 5 IU of pregnant mare serum gonadotropin (PMSG, Ningbo Sansheng Pharmaceutical Co., Ltd., China). Fully grown oocytes were released by puncturing large antral follicles with a fine needle in 37 °C pre-warmed M2 medium (MR-015, Sigma-Aldrich). Fully grown oocytes cultured in minidrops of M16 medium (M7292, Sigma-Aldrich) covered with mineral oil (M5310, Sigma-Aldrich) at 37 °C in a 5% CO_2 atmosphere. The oocytes at GVBD stage were gathered at the time points of 4 h during the process of culture. For Metaphase II (MII) oocytes collection, 4-week-old female mice were superovulated by injection of 5 IU pregnant mare serum gonadotropin (PMSG) followed by 5 IU human chorionic gonadotropin (hCG, Livzon Pharmaceutical Group Co., Ltd. China) after 48 h. The oocyte-cumulus cell complexes were collected from the ampulla of the oviduct, then the granulosa cells were mechanically removed, and the denuded oocytes were collected.

Proteo-metabolome sample processing of single mouse oocytes

GV, GVBD, and MII oocytes were washed three times with PBS buffer containing 1% (v/v) phosphatase inhibitor cocktail, 1 mM PMSF and 10 mM EDTA. 100 nL of 50 mM TCEP solution was firstly injected into OCAM by a 1 μL syringe. The washed single oocytes in PBS buffer were immediately aspirated into a glass pipette under the microscope, and pushed into OCAM. Then 100 nL of 50 mM TCEP solution was injected into OCAM after oocyte sampling. The sampling process was performed

carefully to avoid introducing air bubbles. After both ends were sealed, the micro-reactor was sonicated at 4 °C for 5 min for cell lysis, then incubated in a water bath at 95 °C for 5 min. Protein denaturation, reduction and covalent binding were performed in a single step. To extract metabolites from single oocytes, 5 μL of 80% (v/v) acetonitrile was added into OCAM, and the eluted metabolites were collected for LC-MS/MS analysis. After washing with 300 μL 50 mM ABC buffer, the mixed solution of 40 ng Trypsin and Lys-C (1:1, m/m) was introduced into OCAM, which was sealed and incubated in a water bath at 37 °C for on-column digestion. Finally, OCAM was rinsed with 5 μL loading buffer containing 0.1% (v/v) FA to elute the peptides for LC-MS/MS analysis.

NanoLC-MS/MS and data analysis

For proteome analysis of 10-100 HeLa cells, LC-MS/MS analysis was performed using a Q-Exactive mass spectrometer coupled with an Easy-nano LC 1000 system (Thermo Fisher Scientific). A homemade C18 analytical column (50 μm i.d. × 15 cm) was used for peptide separation at a flow rate of 0.13 μL/min, the nanoESI tip was prepared using a micropipette puller (P-2000, Sutter Instrument, CA), then the column was packed with 1.9 μm C18 particles (120 Å; Dr.Maisch GmbH). The 85 min LC gradient was performed as followings: 7~23% B for 45 min, 23~40% B for 20 min, 40~80% B for 5 min, 80% B for 15 min, the mobile phase A and B were composed of 0.1% FA in 98% water and 98% ACN, respectively. MS measurement was performed in data-dependent acquisition (DDA) mode, and the electrospray voltage was set to 2.1 kV. The full MS scans were acquired from m/z 300-1800 at the resolution of 70,000 with the AGC target of 3e6 and the maximum injection time of 60 ms. Precursor ions of +2 to +6 charge were selected for MS/MS analysis, which were fragmented by higher-energy collision dissociation (HCD) with the normalized collision energy of 28%. For 100 HeLa cells, MS/MS scans were performed at the resolution of 17,500 with the AGC target of 5e4 and the maximum injection time of 60 ms. The AGC target and the maximum injection time were set to 5e5 and 200 ms for fewer HeLa cells.

An Orbitrap Exploris 480 (Thermo Fisher Scientific) mass spectrometer with a FAIMS Pro Interface (San Jose, CA) was employed for LC-MS/MS analysis of single HeLa cells. The electrospray voltage was set to 2.2 kV and the temperature of ion transfer tube was set to 320 °C. The Orbitrap resolution was set to 60,000 for full MS scans with the normalized AGC target of 300%, and the maximum injection

time was set to 100 ms. The exclusion duration was set to 45 s, and the intensity threshold was set to 5000. The HCD collision energy was 30% and the isolation window was set to 1.6 m/z. MS/MS scans were performed at the resolution of 15,000 with the AGC target of 5e5 and the maximum injection time of 150 ms. Default settings were used for FAIMS with a total carrier gas flow of 4 L/min and compensation voltages (CV) of -45V and -65V. In addition, data-independent acquisition (DIA) mode was applied for single-cell analysis. The precursor ions were collected for m/z 400-1200 with an isolation window of m/z 20, followed by fragmentation with the normalized collision energy of 30%, and MS/MS scans were performed at the resolution of 30,000. Other parameters were the same as the DDA method.

MaxQuant (version 2.0.3.0) was employed for database searching and label-free quantification of the DDA data, the MS raw data was searched against UniProt human database (UP000005640, downloaded on June 14th, 2023, containing 20422 reviewed sequences). Trypsin and LysC were selected as digestion enzymes with maximum two missed cleavage sites. Oxidation (M) and Acetyl (Protein N-term) were set as variable modifications. The match between runs (MBR) algorithm was used to improve proteome coverage. Both proteins and peptides were filtered with a false discovery rate (FDR) of 0.01. The common contaminant proteins were removed from the protein identification list. The DIA raw files were loaded to DIA-NN software (version 1.8) and searched with library-free mode. The precursor and fragment mass tolerances were set to 15 and 20 ppm, and proteins were filtered with 1% FDR. Other searching parameters were the same as described previously.

DIA-MS mode was used for proteome analysis of single mouse oocytes. The oocyte samples were spiked with iRT standards (Biognosys) for retention time calibration. The LC-MS analysis was performed using an Orbitrap Exploris 480 mass spectrometer coupled with a FAIMS Pro Interface. The precursor ions were collected for m/z 400-1200 with an isolation window of m/z 20, followed by fragmentation with the normalized collision energy of 30%, and MS/MS scans were performed at the resolution of 30,000 with the AGC target of 5e5 and the maximum injection time of 50 ms. The DIA raw files were loaded to Spectronaut (versions 18.1, Biognosys, Schlieren, Switzerland) and searched against UniProt mouse proteome database (UP000000589, downloaded on July 13th, 2023, containing 17167 reviewed sequences). Library-free mode (directDIA) was performed using standard settings. Both PSM and proteins were filtered with 1% FDR. Normalized quantitative intensity values at MS2 level were loaded into Perseus (versions 1.5.8.5), log 10-transformed and filtered to retain proteins detected in more than 50% samples of each group. ANOVA tests were employed for multiple sample tests of oocytes at

different developmental stages with P value < 0.01. Student's t-tests were performed for the pairwise comparison of different stages with P value < 0.05. PCA and unsupervised hierarchical cluster analysis were employed using ClustVis web tool (https://biit.cs.ut.ee/clustvis/).[1] Gene ontology analysis for the biological process of the differentially expressed proteins was performed using DAVID bioinformatic tool (https://david.ncifcrf.gov/).[2]

For metabolite samples extracted from single oocytes, LC separation was conducted using a homemade C18 column (50 μm i.d. × 15 cm) with a flow rate of 0.15 μL/min. Metabolite elution was performed using a linear 10 min gradient of 20~70% buffer B (80% ACN, 0.1% FA), then the percentage was increased to 98% over 17 min and maintained for 3 min. The full MS scans were acquired from m/z 100-1200 at the resolution of 60,000 in positive mode, and the spray voltage was set at 2.5 kV. Precursor ions with charges of +1 to +6 were selected for MS/MS analysis. Other MS/MS parameters were at default settings. The raw data obtained from LC-MS/MS analysis were loaded into MoloSight System (Mosian Technology Ltd CHN) for peak extraction, peak alignment, and peak identification. Features contained in the blank samples were removed to subtract metabolite background. All features were initially screened by the following steps: first, features with more than 80% missing values across all samples were eliminated; second, the remaining missing values were filled with the minimum values; third, features with relative standard deviation (RSD) less than 20% were reserved; finally, Quantile standardization, Log10 transformation and Pareto normalization were performed.

The normalized data were analyzed using machine learning models including ANOVA tests, Student's t-test, Fold change and Partial least squares discriminant analysis (PLS-DA) among different sample groups. The variable importance in projection (VIP) value > 1.00 and P value < 0.05 were set as the cutoff threshold of the statistical significance. Metabolite database was built based using Mus musculus metabolites collected from Kyoto Encyclopedia of Genes and Genomes (KEGG) to assist identification. Differential features were searched through the database for candidates by matching mass within Δm = 20 ppm. The differential features without MS2 were excluded. The candidate compounds for the remaining features were scored and predicted with in silico fragmentation by Metfrag algorithm (http://ipb-halle.github.io/MetFrag/). The integration analysis of metabolomics and proteomics data was conducted using MetaboAnalyst platform (https://www.metaboanalyst.ca/).[3]

Supplementary figures

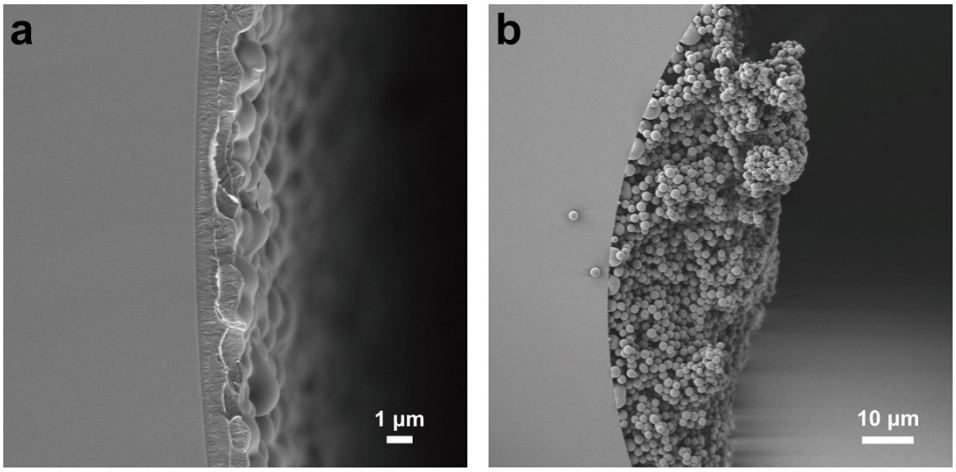

Figure S1. SEM images of the OTC polymer structures prepared at polymerization temperatures of (a) 60 °C and (b) 70 °C. Scale bars represent 1 µm and 10 µm, respectively.

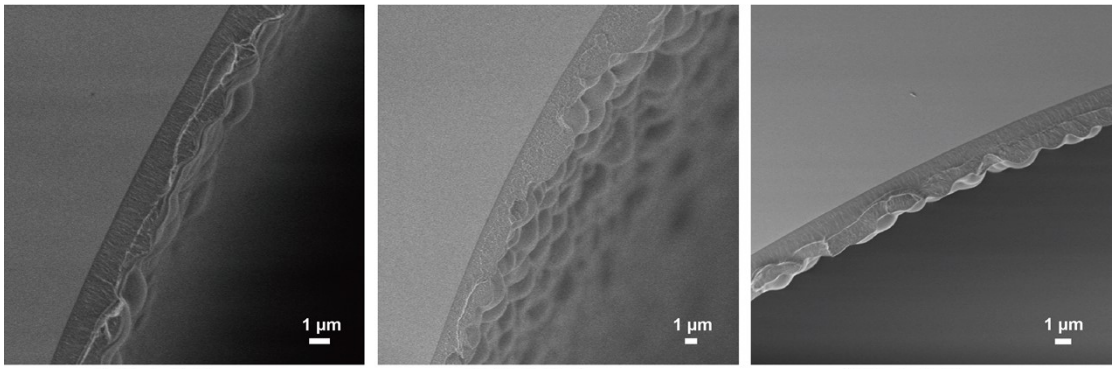

Figure S2. SEM images and thicknesses of the OTC polymer layers obtained from triplicate preparations with 10% (m/m) GMA at 60 °C. Scale bars represent 1 µm.

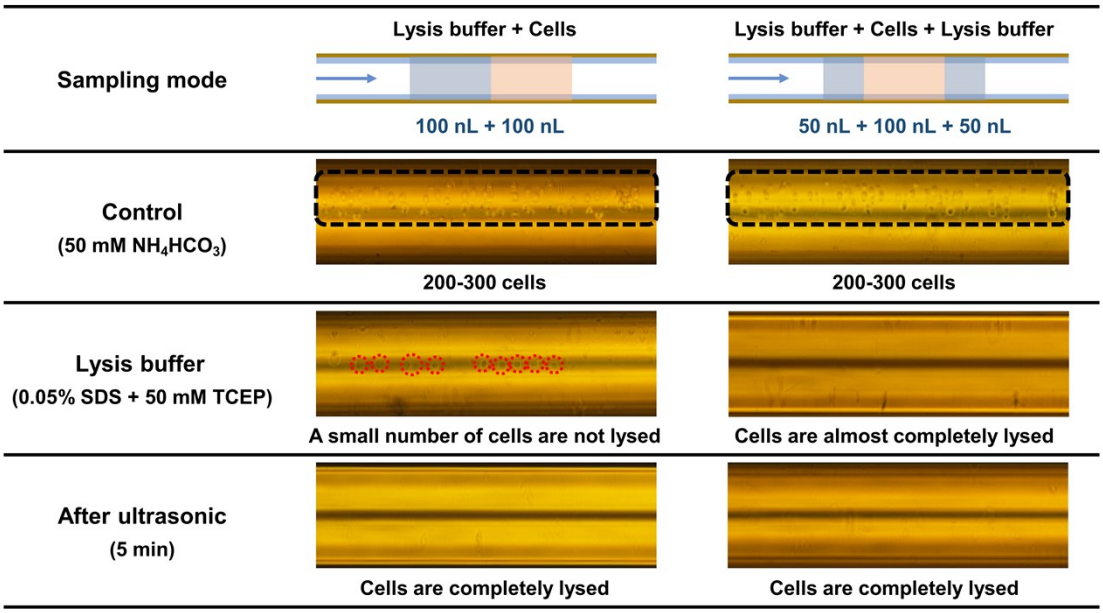

Figure S3. Evaluation of cell lysis efficiency under two sampling modes (lysis buffer + cells and lysis buffer + cells + lysis buffer) assisted by ultra-sonication.

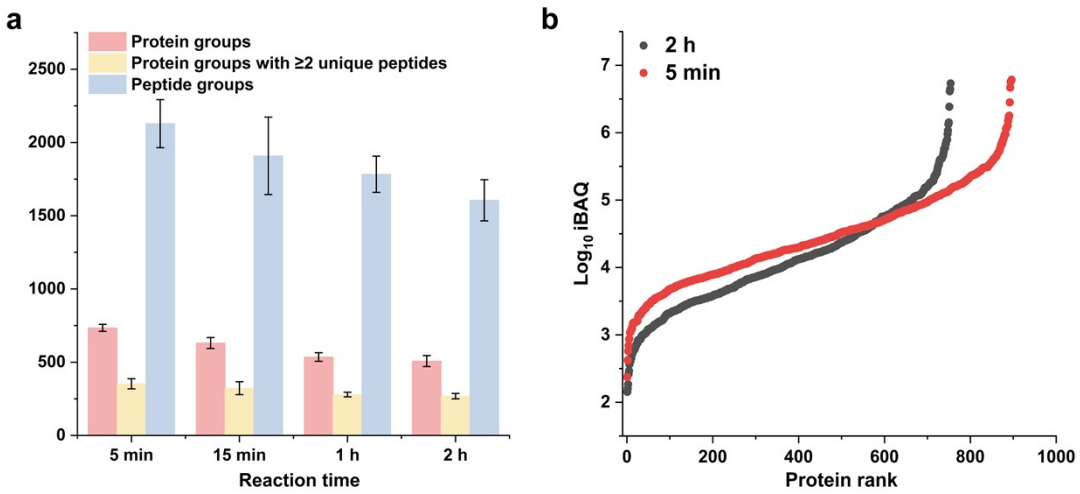

Figure S4. Optimization of reaction time for sample processing using OCAM. (a) The number of proteins and peptides identified from 100 SH-SY5Y cells using OCAM at different reaction times (5 min to 2 h). The error bars represent the standard deviations of the number of proteins and peptides from triplicate analyses. (b) Comparison of the dynamic range of protein expression from 100 cells processed within 5 min and 2 h.

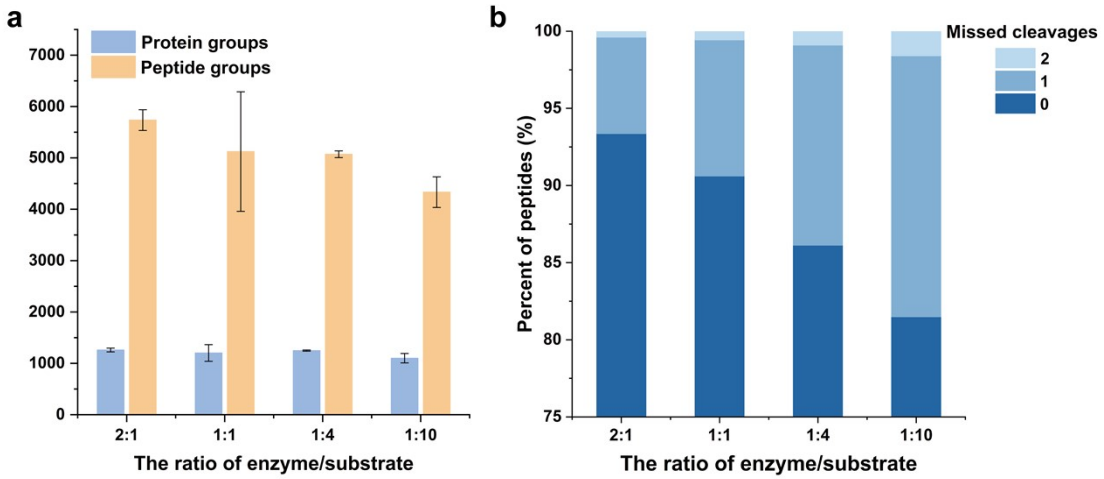

Figure S5. Optimization of digestion condition for sample processing using OCAM. (a) The number of proteins and peptides identified from 20 ng HeLa cell lysate using OCAM under different ratios of enzyme/protein (2:1 to 1:10, m/m). The error bars represent the standard deviations of the number of proteins and peptides from triplicate analyses. (b) Comparison of the missed cleavages of identified peptides under different ratios of enzyme/protein (2:1 to 1:10, m/m).

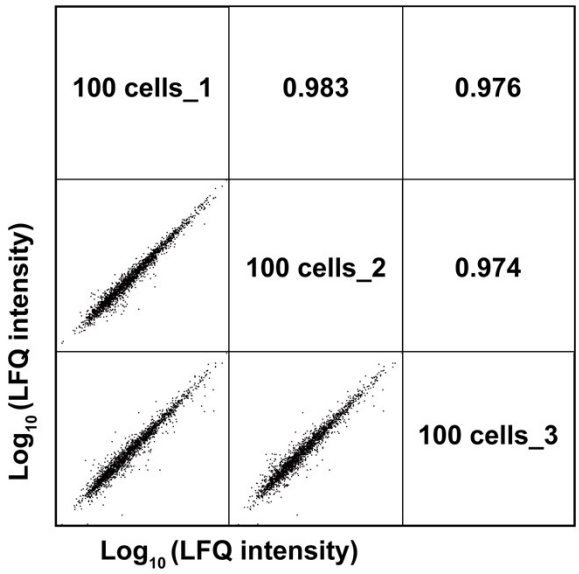

Figure S6. Reproducibility of protein LFQ intensity of 100 HeLa cells. Pearson correlation coefficients between triplicate analyses were labeled.

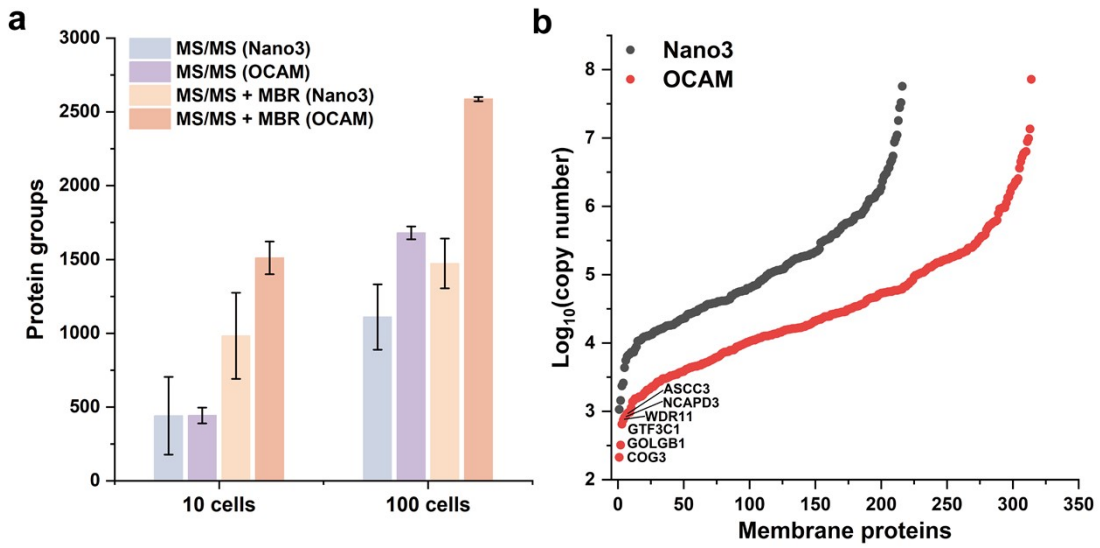

Figure S7. Comparison of the OCAM method with the reported Nano3 method.[4] (a) The number of protein groups identified from 10 and 100 HeLa cells using OCAM and the Nano3 method. The error bars represent the standard deviations. (b) Comparison of the dynamic range of membrane proteins identified by OCAM and the Nano3 method. Low-abundance membrane proteins identified by OCAM were highlighted.

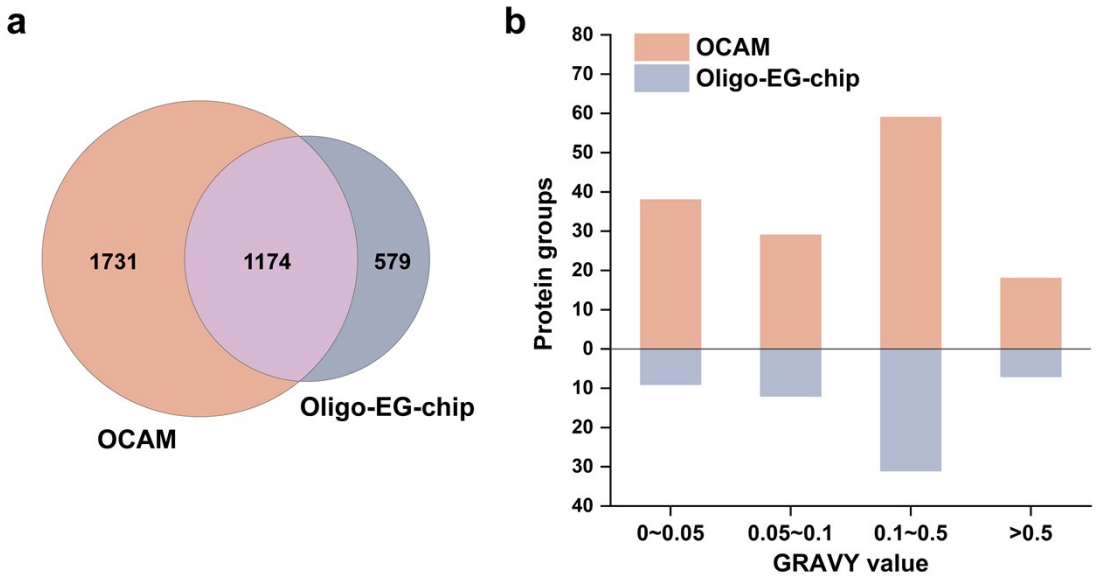

Figure S8. Comparison of the OCAM method with the reported Oligo-EG modified microwell chip-based method.[5] (a) Venn diagram showing the overlap of identified protein groups from single HeLa cells using OCAM and the Oligo-EG-chip. (b) Comparison of the number of hydrophobic proteins identified by OCAM and the Oligo-EG-chip.

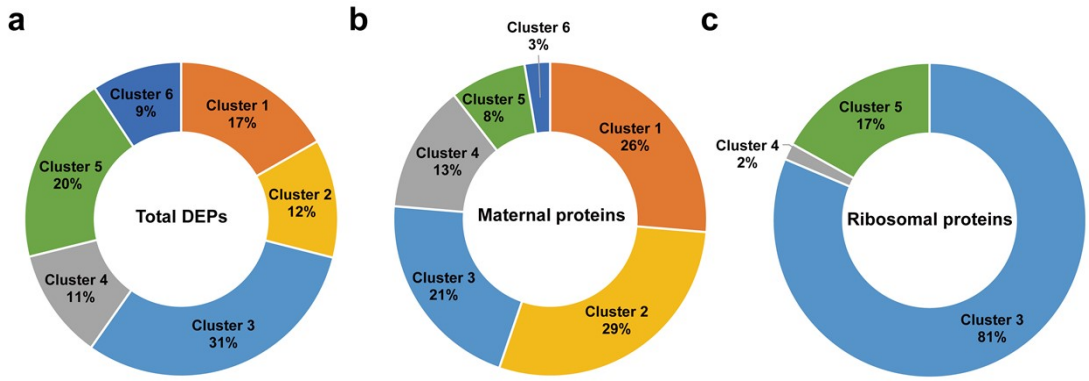

Figure S9. The percentage of (a) total differentially expressed proteins (DEPs) (b) maternal proteins and (c) ribosomal proteins in each cluster.

Supplementary tables

Table S1. The identified protein groups from 100 HeLa cells and single HeLa cells. See the attached excel file.

Table S2. The quantified protein groups from single mouse oocytes. See the attached excel file.

Table S3. The differentially expressed proteins between GV, GVBD and MII stages (ANOVA test, $p < 0.01$). See the attached excel file.

Table S4. The identified metabolites from single mouse oocytes. See the attached excel file.

Table S5. The identified differential metabolites between GV, GVBD and MII stages ($p < 0.05$ and VIP > 1.00). See the attached excel file.

Table S6. Multi-omics pathway analysis of differential metabolites and proteins between GV, GVBD and MII stages. See the attached excel file.

References

1. T. Metsalu and J. Vilo, *Nucleic Acids Res.*, 2015, **43**, W566-570.
2. B. T. Sherman, M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi and W. Chang, *Nucleic Acids Res.*, 2022, **50**, W216-221.
3. Z. Pang, G. Zhou, J. Ewald, L. Chang, O. Hacariz, N. Basu and J. Xia, *Nat. Protoc.*, 2022, **17**, 1735-1761.
4. Z. Yang, Z. Zhang, D. Chen, T. Xu, Y. Wang and L. Sun, *Anal. Chem.*, 2021, **93**, 10568-10576.
5. Y. Li, H. Li, Y. Xie, S. Chen, R. Qin, H. Dong, Y. Yu, J. Wang, X. Qian and W. Qin, *Anal. Chem.*, 2021, **93**, 14059-14067.