



## 15 **Abstract**

16 Global quantification of protein abundances in single cells would provide more direct information on  
17 cellular function phenotypes and complement transcriptomics measurements. However, single-cell  
18 proteomics (scProteomics) is still immature and confronts technical challenges, including limited  
19 proteome coverage, poor reproducibility, as well as low throughput. Here we describe a nested nanoPOTS  
20 (N2) chip to dramatically improve protein recovery, operation robustness, and processing throughput for  
21 isobaric-labeling-based scProteomics workflow. The N2 chip allows reducing cell digestion volume to  
22 <30 nL and increasing processing capacity to > 240 single cells in one microchip. In the analysis of ~100  
23 individual cells from three different cell lines, we demonstrate the N2 chip-based scProteomics platform  
24 can robustly quantify ~1500 proteins and reveal functional differences. Our analysis also reveals low  
25 protein abundance variations (median CVs < 16.3%), highlighting the utility of such measurements, and  
26 also suggesting the single-cell proteome is highly stable for the cells cultured under identical conditions.

## 27 **Introduction**

28 With the success of single-cell genomics and transcriptomics, there is a growing demand for high-  
29 throughput single-cell proteomics (scProteomics) technologies. Global profiling of protein expressions in  
30 individual cells can potentially reveal specific protein markers accounting for heterogeneous populations,  
31 provide more concrete evidence of cellular function phenotypes, and help to identify critical post-  
32 translational modifications that regulate protein activities<sup>1-3</sup>. Despite this transformative potentials,  
33 scProteomics still lags behind single-cell transcriptomics in terms of coverage, measurement throughput,  
34 and quantitation accuracy.<sup>4</sup>

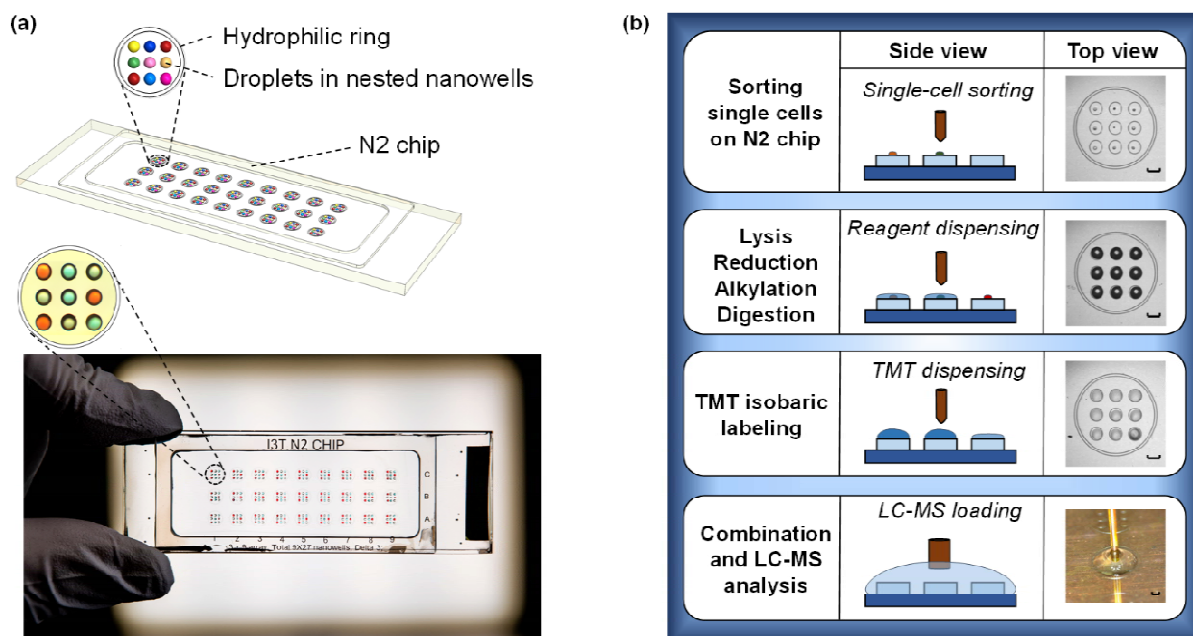
35 Most reported mass-spectrometry-based scProteomics technologies can be classified based upon whether  
36 they make use of isotopic labeling; i.e. they are either label free or use isobaric labeling. In the label-free  
37 methods<sup>5-10</sup>, single cells are individually processed and analyzed using liquid chromatography-mass  
38 spectrometry (LC-MS) signal intensity measurements (i.e. MS1 ion currents) to quantify protein

39 abundance. To improve proteome coverage, high-recovery sample preparation systems<sup>9-11</sup> and highly  
40 sensitive LC-MS systems<sup>7, 12, 13</sup> are usually employed. Although label-free approaches exhibit better  
41 quantification accuracy and higher dynamic range, their throughputs are limited, as each cell requires >  
42 0.5 hour-long LC-MS analysis. In the isobaric labeling approaches (e. g., tandem mass tags or TMT)<sup>14-18</sup>,  
43 single-cell digests are labeled with unique isobaric labels, that are then pooled together for a multiplex  
44 LC-MS analysis. Importantly, the peptides originating from different single cells appear as a single MS1  
45 peak. As a consequence, the pooled ions contributing to a given precursor peak is higher than from  
46 individual cells and their fragmentations result in a richer MS2 spectrum for peptide identification. The  
47 released reporter ions infer protein abundance in different single cells. A “carrier” sample containing a  
48 larger amount of peptides than individual cells (e.g. ~100×) is spiked into each isobaric labeling pool to  
49 maximize the peptide identification (SCoPE-MS)<sup>14, 15, 18</sup>. Currently, the isobaric-labeling approaches have  
50 enabled to analyze ~100 single cells per day. We anticipate the throughput will increase gradually with  
51 new releases of higher multiplex isobaric reagents, shorter LC gradients, and the inclusion of ion mobility  
52 in single-cell proteomics pipelines.

53 Analogous to single-cell transcriptomics, microfluidic technologies play increasing roles in sample  
54 preparation for scProteomics<sup>6, 9, 11</sup>. By minimizing the sample processing volumes in nanowells or  
55 droplets, the non-specific-binding-related protein/peptide loss is reduced, resulting in improved sample  
56 recovery. More importantly, both protein and enzyme concentrations increase in nanoliter volumes,  
57 enhancing tryptic digestion efficiency. For example, our lab developed a nanoPOTS (nanodroplet  
58 processing in one-pot for trace samples) platform for significantly improving proteomics sensitivity by  
59 minimizing the reaction volume to < 200 nL<sup>11</sup>. NanoPOTS allowed reliably identifying 600–1000  
60 proteins with label-free approaches<sup>7, 12, 13</sup>. When isobaric labeling approaches were used, ~1500 proteins  
61 could be quantified across 152 single cells and at a throughput of 77 per day<sup>6, 15</sup>. Despite this progress,  
62 challenges remain. In current microfluidic approaches, the sample processing volume is >10,000 larger  
63 than a single cell and gains would be expected from further miniaturizing the volumes, but it is presently

64 constrained by liquid handling operations, including reagent dispensing, sample aspirating, transferring,  
65 and combination. Among these, the nanoliter-scale aspirating and transferring steps, which are commonly  
66 performed in isobaric-labeling workflows, are challenging, time-consuming, and prone to sample losses.  
67 Additionally, most reported microfluidic approaches employed home-built nanoliter liquid handling  
68 systems, which limits their broad dissemination.

69 Herein, we describe a nested nanoPOTS (N2) chip to improve isobaric-labeling-based scProteomics  
70 workflow. Compared with our previous nanoPOTS chip,<sup>6,15</sup> where nanowells are sparsely distributed,  
71 we cluster arrays of nanowells in dense areas and use them for digesting and labeling single cells with  
72 single TMT sets. With the N2 chip, we eliminate the tedious and time-consuming TMT pooling steps.  
73 Instead, the single-cell samples in one TMT set are pooled by simply adding a microliter droplet on top of  
74 the nested nanowell area and retrieving it for LC-MS analysis. The N2 chip reduces the sample  
75 processing volumes by one order of magnitude and allows over 5× more numbers of nanowells in one  
76 microchip for high-throughput single-cell preparation. We demonstrate the N2 chip not only efficiently  
77 streamlines the scProteomics workflow, but also dramatically improves sensitivity and reproducibility.



78

79 **Figure 1.** (a) A 3D illustration (top) and a photo (bottom) of the nested nanoPOTS (N2) chip. Nine nanowells are  
80 nested together and surrounded by a hydrophilic ring for an TMT set. (b) Single-cell proteomics workflow using the  
81 N2 chip. The scale bar is 0.5 mm.

82

## 83 **Methods**

### 84 ***Fabrication and assembly of the N2 chips***

85 The chips were fabricated on glass slides using standard photolithography, wet etching, and silane  
86 treatment approach as described previously<sup>11,19</sup>. Briefly, as shown in Figure 1 and S1a, 27 (3 × 9)  
87 nanowell clusters with a distance of 4.5 mm between adjacent clusters are designed on single microscope  
88 slide (1 × 3 inch, Telic Company, Valencia, USA). In each cluster, 9 nanowells with 0.5-mm diameter  
89 and 0.75-mm well-to-well distance are nested together. To facilitate droplet combination and retrieval  
90 process, a micro-ring surrounds the nested nanowells. After photoresist exposure, development, and  
91 chromium etching, the glass slide was etched to a depth of ~5 μm with buffered hydrofluoric acid<sup>20</sup>. The  
92 freshly etched slide was dried by heating it at 120 °C for 2 h and then treated with oxygen plasma for 3  
93 min (AP-300, Nordson March, Concord, USA). To selectively pattern the chip, 2% (v/v) heptadecafluoro-  
94 1,1,2,2-tetrahydrodecyl-dimethylchlorosilane (PFDS, Gelest, Germany) in 2,2,4-trimethylpentane was

95 applied on the chip surface and incubate for 30 min. After removing the remaining chromium layer, all  
96 the chromium-covered regions are (nanowells and micro-rings) are hydrophilic and exposed areas are  
97 hydrophobic. Finally, a glass frame was attached to the nanowell chip with epoxy to create a headspace  
98 for reaction incubation.

### 99 ***Reagents and chemicals***

100 Urea, n-dodecyl- $\beta$ -D-maltoside (DDM), Tris 2-carboxyethyl phosphine (TCEP), Iodoacetamide (IAA),  
101 Ammonium Bicarbonate (ABC), Triethylammonium bicarbonate (TEAB), Trifluoroacetic acid (TFA),  
102 Anhydrous acetonitrile (a-ACN), and Formic acid (FA) were obtained from Sigma (St. Louis, MO, USA).  
103 Trypsin (Promega, Madison, WI, USA) and Lys-C (Wako, Japan) were dissolved in 100 mM TEAB  
104 before usage. TMTpro 16plex, 50% hydroxylamine (HA), Calcein AM, Acetonitrile (ACN) with 0.1% of  
105 FA, and Water with 0.1% of FA (MS grade) were purchased from Thermo Fisher Scientific (Waltham,  
106 MA, USA).

### 107 ***Cell culture***

108 Three murine cell lines (RAW 264.7, a macrophage cell line; C10, a respiratory epithelial cell line; SVEC,  
109 an endothelial cell line) were obtained from ATCC and cultured at 37°C and 5% CO<sub>2</sub> in Dulbecco's  
110 Modified Eagle's Medium supplemented with 10% fetal bovine serum and 1× penicillin-streptomycin  
111 (Sigma, St. Louis, MO, USA).

### 112 ***Bulk-scale proteomic sample preparation and mimic single-cell experiments***

113 The cultured cell lines were collected in a 15 ml tube and centrifuged at 1,000 × g for 3 min to remove the  
114 medium. Cell pellets were washed three times by 1× PBS buffer, then counted to obtain cell concentration.  
115 Ten million cells per cell population were lysed in lysis buffer containing 8M urea in 50 mM ABC in ice.  
116 Protein concentration was measured with BCA assay. After protein was reduced and alkylated by DTT  
117 and IAA, Lys-C (enzyme-to-protein ratio of 1:40) was added and incubated for 4 h at 37°C. Trypsin  
118 (enzyme-to-protein ratio of 1:20) was added and incubated overnight at 37°C. The digested tryptic

119 peptides were acidified with 0.1 % TFA, desalted by C18 SPE column, and completely dried to remove  
120 the acidic buffer.

121 After measuring the peptide concentration with BCA assay, samples from three different cell types were  
122 mixed at 1:1:1 ratio and used for boost and reference samples. All peptide samples were dissolved with 50  
123 mM HEPES (pH 8.5) followed by mixing with a TMT-16 reagent in 100% ACN. To maintain high  
124 labeling efficiency, a TMT-to-peptide ratio of 1:4 (w/w) was used. After 1-h incubation at room  
125 temperature, the labeling reaction was terminated by adding 5% HA and incubating for 15 min. The  
126 TMT-labeled peptides were then acidified with 0.1% FA and cleaned with C18 stage tips. Before use,  
127 different amounts of peptides (0.1 ng for mimic single cell, 0.5 ng for reference, 10 ng for boost) were  
128 diluted in 0.1% FA buffer containing 0.1% DDM (w/v) to prevent sample loss at low concentration  
129 conditions.

130 To mimic single-cell proteomics preparation in nanowell chips, 0.1-ng peptide samples in 200 nL buffer  
131 from the three cell lines were loaded into 1.2-mm nanowells using a nanoPOTS dispensing robot <sup>11</sup> and  
132 incubated for 2 h at room temperature. Next, samples from the same TMT set were collected and  
133 combined into a large-size microwell (2.2-mm diameter), which contained 10 ng and 0.5 ng TMT-labeled  
134 peptides for boost and reference samples, respectively.

135 To deposit these single-cell-level peptide samples to N2 chip, we employ a picoliter dispensing system  
136 (cellenONE F1.4, Cellenion, France) to dispense 0.1-ng peptide in 20 nL buffer in each nanowells (Figure  
137 S1b). After incubating the chip at room temperature for 2 h, mixed boost and reference samples (10 ng  
138 and 0.5 ng, respectively) were equally distributed into each nanowell.

139 Samples in both nanowell chip and N2 chip were completely dried out in a vacuum desiccator and stored  
140 in a -20°C freezer until analysis.

141 *scProteomics sample preparation using the N2 Chip*

142 The cellenONE system was used for both single-cell sorting and sample preparation on N2 chip. Before  
143 cell sorting, all the cells were labeled with Calcein AM (Thermo Fisher) to gate out dead cells and cell  
144 debris. After single-cell deposition, 10 nL lysis buffer containing 0.1% DDM and 5 mM TCEP in 100  
145 mM TEAB was dispensed into each nanowell. The N2 chip was incubated at 70°C for 45 min in a  
146 humidity box to achieve complete cell lysis and protein reduction. Next, 5 nL of 20 mM IAA was added,  
147 following by reaction incubation for 30 min in the dark. Proteins were digested to peptide by sequentially  
148 adding 0.25-ng Lys-C (5 nL) and 0.5-ng-trypsin (5 nL) into the nanowells and incubating for 3 hours and  
149 8 hours, respectively. For isobaric labeling, we added 50 ng TMT tag in 10 nL ACN into each of the  
150 corresponding nanowells according to experimental design. After 1-hr incubation at room temperature,  
151 the remaining TMT reagents was quenched by adding 5 nL of 5% HA. Finally, TMT labeled boost (10 ng)  
152 and reference (0.5 ng) peptide was distributed into nanowells. The samples were acidified with 5 nL of 5%  
153 FA and dried for long-term storage.

#### 154 *LC-MS/MS analysis*

155 All the samples are analyzed with a nanoPOTS autosampler<sup>6</sup> equipped with a C18 SPE column (100 µm  
156 i.d., 4 cm, 300 Å C18 material, Phenomenex) and an LC column (50 µm i.d., 25 cm long, 1.7 µm, 130 Å,  
157 Waters) heated at 50°C using AgileSleeve column heater (Analytical Sales and Services Inc., Flanders,  
158 NJ). Dried samples from nanowell chips or N2 chips were dissolved with Buffer A (0.1% FA in water),  
159 then trapped on the SPE column for 5 min. Samples were eluted out from the column using a 120-min  
160 gradient from 8% to 45% Buffer B (0.1% FA in ACN) and a 100 nL/min flow rate.

161 An Orbitrap Eclipse Tribrid MS (Thermo Scientific) operated in data-dependent acquisition mode was  
162 employed for all analyses for peptides. Peptides were ionized by applying a voltage of 2,200 V and  
163 collected into an ion transfer tube at 200°C. Precursor ions from 400-1800 m/z were scanned at 120,000  
164 resolution with an ion injection time (IT) of 118 ms and an AGC target of 1E6. During a cycle time of 3 s,  
165 precursor ions with >+2 charges and > 2E4 intensities were isolated with a window of 0.7 m/z, an AGC



166 target of 1E6, and an IT of 246 ms. The isolated ions were fragmented by high energy dissociation (HCD)  
167 level of 34%, and fragments were scanned in an Orbitrap at 120,000 resolution.

### 168 *Database searching*

169 All the raw files from the Thermo MS were processed by MaxQuant<sup>21</sup> (Ver. 1.6.14.0) with the  
170 *UniProtKB* protein sequence database of *Mus musculus* species (downloaded on 05/19/2020 containing  
171 17,037 reviewed protein sequences). Reporter ion MS2 was set as the search type and TMT channel  
172 correction factors from the vendor were applied. The mass tolerance for precursor ions and fragment ions  
173 was using the default value in MaxQuant. Protein acetylation in N-terminal and oxidation at methionine  
174 were chosen as variable modifications, and protein carbamidomethylation in cysteine residues was set as  
175 fixed modification. Both peptides and proteins were filtered with a false discovery rate (FDR) of 1% to  
176 ensure identification confidence.

### 177 *Single-cell proteomics data analysis*

178 The corrected reporter ion intensities from MaxQuant were imported into Perseus (Ver. 1.6.14.0)<sup>22</sup> and  
179 were log<sub>2</sub>-transformed after filtering out the reverse and contaminant proteins. Proteins containing >70%  
180 valid values in each cell type were considered as quantifiable proteins, and the report ion intensities of the  
181 quantified proteins were normalized via the quantile-normalization followed by replacing the missing  
182 values based on a standard distribution of the valid values (width: 0.3, downshift: 1.8). To minimize the  
183 batch effect from multiple TMT sets, we adjusted the batch effects using the Combat algorithm<sup>23</sup> in SVA  
184 package, which is embedded in Perseus. Next, the data matrix was separated by cell types and grouped by  
185 TMT channel. Combat algorithm is applied to minimize the TMT channel effect. The combined matrix  
186 was further applied for statistical analysis, including principal component analysis (PCA) and heatmap  
187 hierarchical clustering analysis. ANOVA test was performed to determine the proteins showing  
188 statistically significant differences across the three cell types (Permutation-based  $FDR < 0.05$ ,  $S_0 = 1$ ), and

189 a 2-way student t-test was applied to explain the significant differences between two groups (*p*-value <  
190 0.05). The processed data were visualized with Graphpad and Perseus.  
191 Proteins intensities without missing values in each cell type in intra-batch or inter-batches were used to  
192 calculate the coefficient of variations (CVs). Briefly, for intra-batches, the CVs were calculated using raw  
193 protein intensities inside each TMT set and then pooled together to generate the box plots. For inter-  
194 batches without batch correction, the CVs were calculated using raw protein intensities across all the  
195 TMT sets. To calculate the CVs of intra-batches with batch corrections, raw protein intensities were log<sub>2</sub>  
196 transformed and followed by imputing missing values. After normalization and batch correction using  
197 Combat algorithm<sup>23</sup>, proteins with imputed values were replaced to 'NaN' and filtered out. The protein  
198 intensities were exponentially transformed to calculate the CVs.

## 199 **Results**

### 200 *Design and operation of the N2 chip*

201 The N2 chip is distinct from previous nanoPOTS chips<sup>6, 11, 15, 16</sup>. We cluster an array of nanowells in high  
202 density and use each cluster for one multiplexed TMT experiment. In this proof-of-concept study, we  
203 designed 9 (3×3) nanowells in each cluster and 27 (3×9) clusters, resulting in total 243 nanowells on one  
204 chip (Figure 1a and S1a). Additionally, we designed a hydrophilic ring surrounding the nested nanowells  
205 to confine the droplet position and facilitate the TMT pooling and retrieval steps. Compared with previous  
206 nanoPOTS chips<sup>6, 8, 15</sup>, we reduced the nanowell diameters from 1.2 mm to 0.5 mm, corresponding to an  
207 82% decrease in contact areas and an 85% decrease in total processing volumes (Table 1). The  
208 miniaturized volume could result in a ~45× increase in trypsin digestion kinetics because both trypsin and  
209 protein concentrations are increased by 6.67×. Both the reduced contact area and increased digestion  
210 kinetics are expected to enhance scProteomics sensitivity and reproducibility.

211 **Table 1.** Comparison of technical characteristics between N2 and nanowell chips

	N2 chip	Nanowell chip
<b>Diameter (mm)</b>	0.5	1.2
<b>Contact area (mm<sup>2</sup>)</b>	0.20	1.13
<b>Total volume (nL)</b>	30	200
<b>Digestion kinetics</b>	45×	1×
<b>Capacity (cells/chip)</b>	243	44
<b>Measured running time (min/chip, min/cell)</b>	18, 0.07	36.5, 0.83

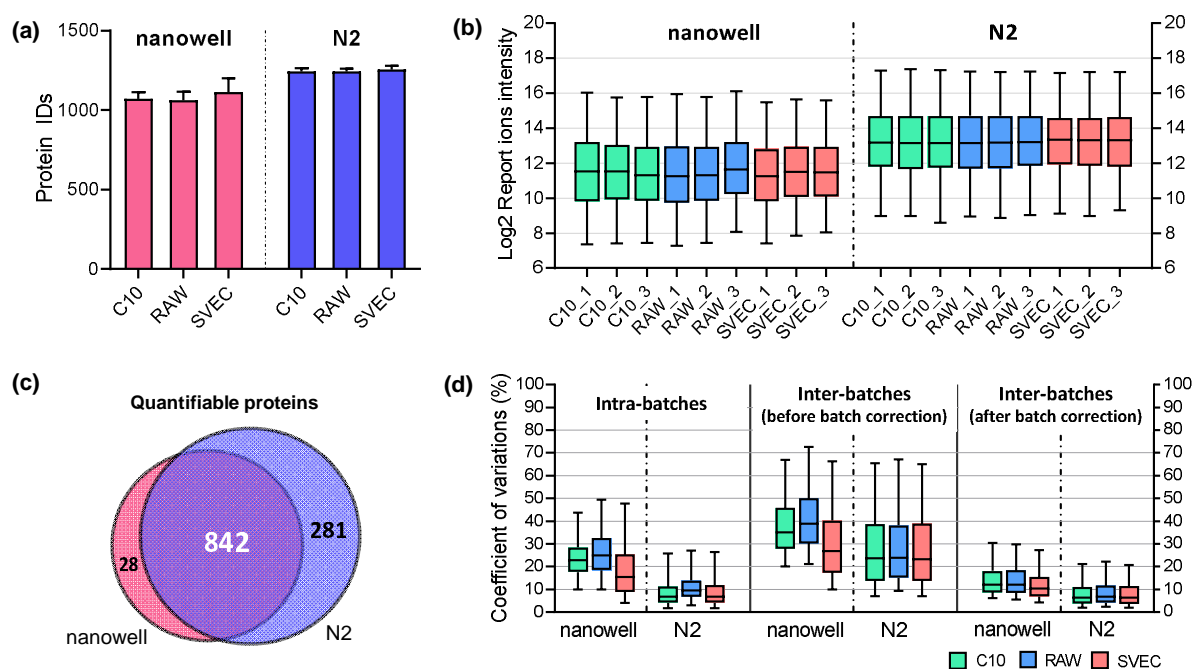
212  
213 The scProteomics sample preparation workflow using the N2 chip is illustrated in Figure 1b. To sort  
214 single cells in the miniaturized nanowells, we employed an image-based single-cell isolation system  
215 (IBSCI, cellenONE F1.4). The cellenONE system also allowed us to dispense low nanoliter reagents for  
216 cell lysis, protein reduction, alkylation, and digestion. After protein digestion, TMT reagent is dispensed  
217 to label peptides in each nanowell uniquely. Finally, we distributed 10 ng boosting/carrier peptide and 0.5  
218 ng reference peptide in each nanowell cluster to improve the protein identification rate (Figure S1b)<sup>14</sup>. To  
219 integrate the N2 chip in our LC-MS workflow, we loaded the chip in a nanoPOTS autosampler<sup>6</sup>. We  
220 applied a 3- $\mu$ L droplet on the rested nanowells, combined the TMT set, and extracted the peptide mixture  
221 for LC-MS analysis (Figure 1b). Compared with our previous nanoPOTS-TMT workflow<sup>6, 15, 16</sup>, the total  
222 processing time of each chip was reduced from 36.5 min to 18 min (Figure S1c), which is equivalent to  
223 the reduced time from 0.83 min to 0.07 min for each single cell. As such, the N2 chip increases the single-  
224 cell processing throughput by >10 $\times$ .

225 It should be noted that the N2 chip can be directly coupled with conventional LC system without the  
226 use of the customized nanoPOTS autosampler. As shown in Figure S1d, the user can simply add a 8- $\mu$ L  
227 droplet on the chip and aspirate it back into an autosampler vial for LC injection.

### 228 ***Sensitivity and reproducibility of the N2 chip***

229 We first benchmarked the performance of N2 chip with our previous nanowell chip using diluted peptide  
230 samples from three murine cell lines (C10, Raw, SVEC). To mimic the scProteomics sample preparation

231 process, we loaded 0.1-ng peptide in each nanowell of both N2 and nanowell chips (Figure S1b) and then  
 232 incubated the chips at room temperature for 2 h. The long-time incubation would allow peptides to absorb  
 233 on nanowell surfaces and lead to differential sample recoveries. The combined TMT samples were  
 234 analyzed by the same LC-MS system. When containing at least 1 valid reporter ion value was considered  
 235 as identified peptides, an average of 5706 peptides was identified with N2 chip, while only 4614 were  
 236 achieved with nanowell chip (Figure 2a). The increased peptide identifications result in a 15%  
 237 improvement in proteome coverage; the average proteome identification number was increased from  $1082$   
 238  $\pm 22$  using nanowell chips to  $1246 \pm 6$  using N2 chips (Figure 2b). Indeed, we observed significant  
 239 increases in protein intensities with N2 chips. The median  $\log_2$ -transformed protein intensities are 13.21  
 240 and 11.49 for N2 and nanowell chips, respectively, corresponding to  $\sim 230\%$  overall improvement in  
 241 protein recovery (Figure 2b). Together, these results demonstrated the N2 chips can dramatically improve  
 242 the sample recovery and proteomics sensitivity.



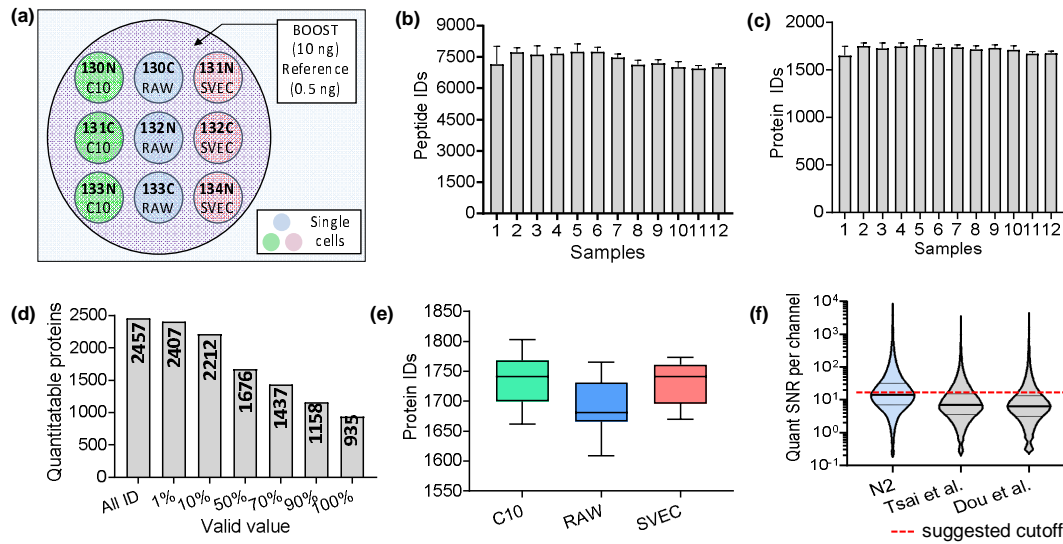
243  
 244 **Figure 2.** (a) The numbers of protein identifications using nanowell and N2 chips. The error bars indicate standard  
 245 deviations of 12 samples each containing 0.1 ng tryptic peptides from three cell lines. (b) The distributions of  $\log_2$   
 246 transformed protein intensities in each TMT channel. (c) Venn diagram of quantifiable proteins between nanowell  
 247 and N2 chips. (d) The distributions of the coefficient of variations (CVs) for proteins identified in each cell type.

248 *Protein CVs were calculated inside single TMT batches (left), among different TMT batches without batch*  
249 *corrections (middle), and with batch correction (right).*

250 We assessed if the N2 chip could provide comparable or better quantitative performance compared with  
251 nanowell chips. As expected, more proteins are quantifiable with N2 chip if 70% valid values in each cell  
252 line were required; the quantifiable protein numbers were 870 and 1123 for nanowell and N2 chips,  
253 respectively (Figure 2c). For nanowell chips, pairwise analysis of any two samples showed Pearson's  
254 correlation coefficients from 0.97 to 0.99 between the same cell types and from 0.87 to 0.95 between  
255 different cell types (Supplementary figure 2a and 2b). When N2 chips were used, Pearson's correlation  
256 coefficients were increased to a range of 0.98-0.99 for the same cell types, and a range of 0.91-0.96 for  
257 different cell types. We next evaluated the quantification reproducibility by measuring the coefficient of  
258 variations (CV) of samples from the same cell types. In intra-batch calculations, we obtained median  
259 protein CVs of <9.6% from N2 chips, which were dramatically lower than that from nanowell chips  
260 (median CVs of < 24.9%) (Figure 2d). Higher CVs were obtained between different TMT batches, which  
261 was known as TMT batch effect<sup>24</sup>. When Combat algorithm<sup>23</sup> was applied to remove the batch effect, the  
262 median protein CVs from N2 chip dropped to < 6.7%. Such low CVs are similar or even better than other  
263 bulk-scale TMT data, demonstrating the N2 chip could provide high reproducibility for robust protein  
264 quantification in single cells.

### 265 ***Proteome coverage of single cells with the N2 chip***

266 We analyzed total 108 single cells (12 TMT sets) from three murine cell lines, including epithelial cells  
267 (C10), immune cells (Raw264.7), and endothelial cells (SVEC) (Figure 3a). Noteworthy, these three cell  
268 types have different sizes, which allows us to evaluate if the workflow presents a bias in protein  
269 identification or quantification based on cell sizes. Specifically, Raw cells have a diameter of 8  $\mu\text{m}$ ,  
270 SVEC of 15  $\mu\text{m}$  and C10 of 20  $\mu\text{m}$ . (Figure S3a). A total of 108 individual cells were analyzed,  
271 corresponding to 12 separate multiplexed TMT sets.



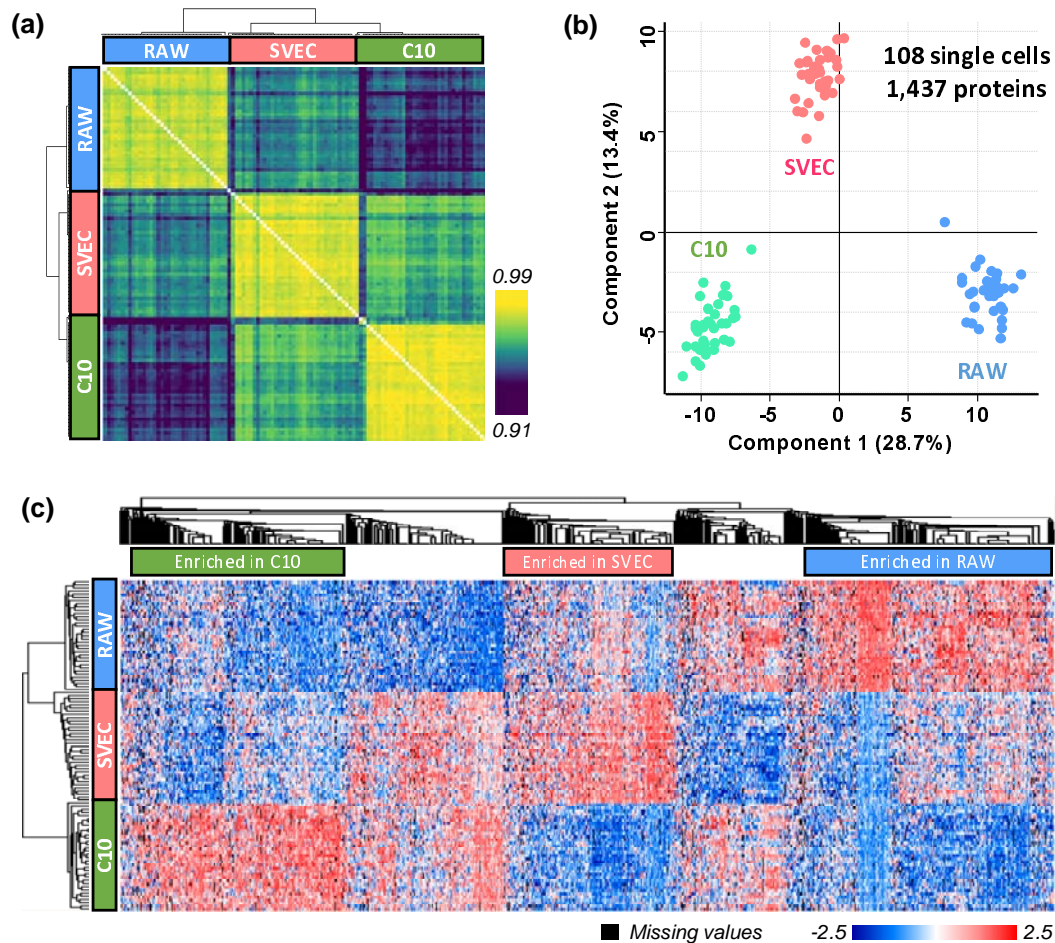
272

273 **Figure 3.** (a) Experiment design showing single-cell isolation and TMT labeling on the N2 chip. (b, c) The average  
 274 numbers of identified peptides and proteins for the 12 TMT sets. At least 1 valid value in the 9 single-cell channels is  
 275 required to count as an identification. Error bars show the standard deviations from 9 single cells. (d) The numbers  
 276 of quantifiable proteins based on different percentages of required valid values. (e) Box plot showing the  
 277 distributions of protein identification numbers. (f) Violin plot showing the distributions of signal to noise ratio (SNR)  
 278 per channel for raw single-cell signals calculated by SCPCompanion.<sup>17</sup> Two published TMT scProteomics datasets  
 279 from our group using nanowell chip<sup>15,16</sup> were used to benchmark the data generated in this study.

280 Among the 12 TMT sets, our platform identified an average of ~7369 unique peptides and ~1716 proteins  
 281 from each set with at least 1 valid value in the 9 single-cell channels (Figure 3b and 3c). We identified  
 282 total 2457 proteins, and 2407 proteins had reporter ion intensities in at least 1 single cells across the 108  
 283 cells (Figure 3d). When a stringent criteria of >70% valid values was applied, the number of proteins  
 284 dropped to 1437. As expected, we observed the numbers of proteins identified for three cell types ranked  
 285 according to the cell sizes (Figure S3a). Average 1735, 1690, and 1725 proteins were identified in C10,  
 286 RAW, and SVEC cells, respectively (Figure 3e). In addition, similar trends were also observed in the  
 287 distribution of protein intensities (Figure S3b).

288 Cheung and coworkers<sup>17</sup> recently introduced the software SCPCompanion to characterize the quality of  
 289 the data generated from single-cell proteomics experiments employing isobaric stable isotope labels and a  
 290 carrier proteome. SCPCompanion extracts signal-to-noise ratio (SNR) of single-cell channels and  
 291 provides cutoff values to filter out low-quality spectra to obtain high-quality protein quantitation. In line  
 292 with our experimental design, SCPCompanion estimated ~0.1 ng proteins were contained in single cells

293 and the boost-to-single ratio is ~100 (Supplementary Table 1), indicating minimal peptide losses in the  
294 N2 chip. More importantly, the median SNR per single-cell sample was 14.4, which is very close to the  
295 suggested cutoff value of 15.5, corresponding to ~50% of raw MS/MS spectra can provide robust  
296 quantification. We also compared the data quality generated with previous nanowell chips and similar  
297 LC-MS setup<sup>15,16</sup>. The median SNR values per sample were 7.0<sup>15</sup> and 6.4<sup>16</sup>, which were significantly  
298 lower than the data generated by the N2 chip (Figure 3f).



299

300 **Figure 4.** (a) Clustering matrix showing Pearson correlations across 108 single cells using log<sub>2</sub>-transformed  
301 protein intensities, and (b) PCA plot showing the clustering of single cells by cell types. Total 1437 proteins were  
302 used in the PCA projection. (c) Heatmap with hierarchical clustering showing 1,127 significant proteins based on  
303 ANOVA test. Three protein clusters used for pathway analysis were labeled and highlighted.

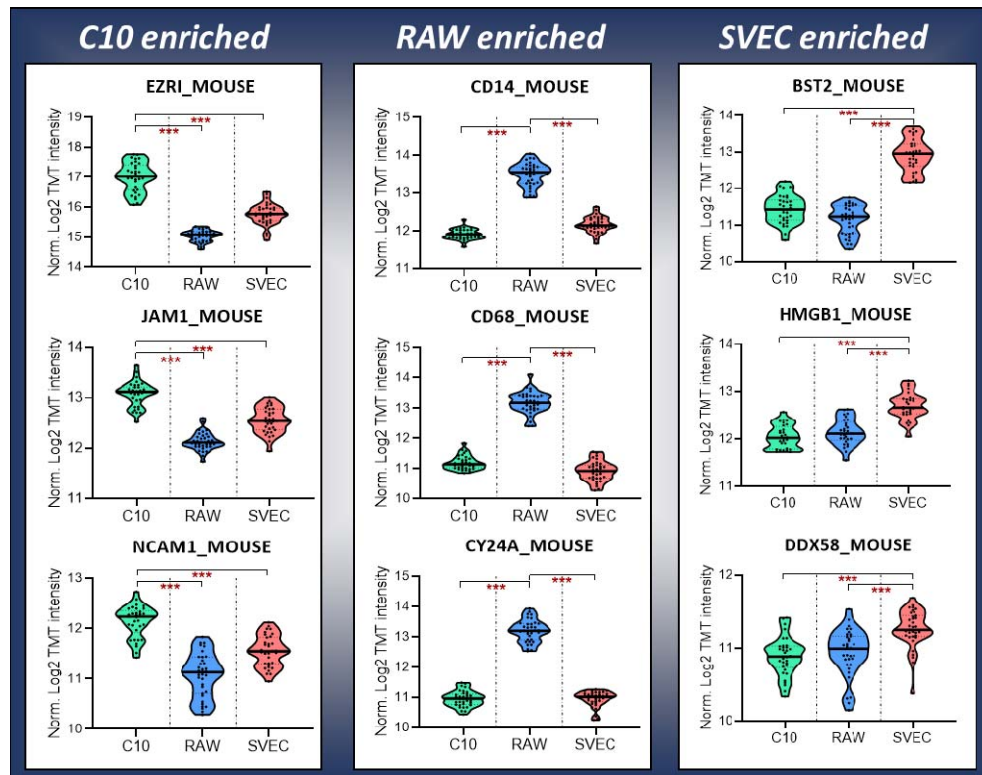
304 **Cell typing with scProteomics**

305 To assess the quantitative performance of the N2 chip-based scProteomics platform, we first performed a  
306 pair-wise correlation analysis using the 1437 proteins across the 108 single cells. As expected, higher  
307 correlations were observed among the same types of cells and lower correlations among different types of  
308 cells (Figure 4a). The median Pearson correlation coefficients are 0.98, 0.97, and 0.97 for C10, RAW, and  
309 SVEC cells, respectively. We next calculated the coefficient of variations (CVs) using protein abundances  
310 for the three cell populations. Interestingly, we see very low variations with median CVs < 16.3% (Figure  
311 S4), indicating protein expression are very stable for cultured cells under identical condition. Principal  
312 component analysis (PCA) showed strong clustering of single cells based on cell types and the three  
313 clusters were well separated from one another (Figure 4b). We compared the our previous PCA result for  
314 the same three cell types using nanowell-based platform (Figure S5)<sup>16</sup>. The N2 chip not only increased  
315 the proteome coverages, but also dramatically improved the classification power.

316 To identify proteins leading the clustering of the three cell populations, an ANOVA test was performed  
317 (Permutation-based FDR < 0.05, S0 = 1). Of the total 1437 proteins, 1127 proteins were significantly  
318 differentially changed across three cell types (Figure 4c). Among them, 237 proteins were enriched in  
319 C10 cells, 203 proteins were enriched in SVEC cells, and 275 proteins were enriched in RAW cells.  
320 Proteins enriched in each cell type revealed differences in molecular pathways based on the REACTOME  
321 pathway analysis (Figure S6). For example, the proteins higher in abundance in C10 cells were  
322 significantly enriched in REACTOME terms such as “vesicle-mediated transport”, “membrane  
323 trafficking”, “innate immune system”, or “antigen processing-cross presentation”. These functions are in  
324 line with the known functions of lung epithelial cells, of which the C10 are derived from<sup>25</sup>. The protein  
325 more abundant in RAW cells, which derive from murine bone marrow macrophages, were enriched in  
326 REACTOME terms associated with “neutrophil degranulation”, “innate immune system” in line with  
327 their immune function. Other REACTOME terms related to the “ribosome” and the “pentose phosphate  
328 pathway” were also enriched. These pathways not only suggest that there is intricate cooperation between  
329 macrophages and neutrophils to orchestrate resolution of inflammation and immune system<sup>26</sup>, but also



330 show that system metabolism strongly interconnects with macrophage phenotype and function<sup>27</sup>.  
331 Proteins more abundant in SVEC cells (murine endothelial cells) were enriched in pathways include  
332 “processing pre-mRNA”, “cell cycle”, or “G2/m checkpoints”. This suggests its proliferation, migration,  
333 or coalescing of the endothelial cells to form primitive vascular labyrinths during angiogenesis<sup>28</sup>.



334  
335 **Figure 5.** Violin plots showing nine putative plasma membrane proteins enriched in three cell types.  
336 Protein in each column are statistically significant ( $p$ -value  $< 0.001$ ) expressed in the specific cell type.

### 337 **Identifying cell surface markers with scProteomics**

338 One of most unique advantages of scProteomics over single-cell transcriptomics is the capability to  
339 identify cell surface protein markers, which can be used to enrich selected cell populations for further  
340 studies. We next assessed if we can use our scProteomics data to identify cell-type-specific membrane  
341 proteins for the three cell populations. We matched the enriched protein lists to a subcellular-component  
342 database on UniProtKB, which consists 2,871 of reviewed plasma membrane proteins for *Mus musculus*  
343 (updated on 01/04/2021). We generated a list containing 63 plasma membrane proteins (Supplementary  
344 Table 2). Among them, 16 plasma membrane proteins were highly expressed in C10 compared to RAW

345 and SVEC cells, while 34 and 13 plasma membrane proteins were significantly enriched in RAW and  
346 SVEC cells, respectively. For example, EZRI<sup>29</sup>, JAM1<sup>30</sup>, and NCAM1<sup>31</sup>, which were previously known  
347 to protect the barrier function of respiratory epithelial cells by enhancing the cell-cell adhesion, were  
348 highly expressed in C10 cells (Figure 5, left panel). For RAW enriched membrane proteins, CD14<sup>32</sup> and  
349 CD68<sup>32,33</sup> are mainly produced in macrophage cells and widely used as a histochemical or cytochemical  
350 marker for inflammation-related macrophages (Figure 5, middle panel). CY24A is a sub-component of  
351 the superoxide generating NOX2 enzyme on macrophage membrane<sup>34</sup>. In term of SEVC enriched protein  
352 markers, BST2 is known to highly express in blood vessels throughout the body as an intrinsic immunity  
353 factor (Figure 5, right panel)<sup>35</sup>. HMGB1 and DDX58, which were found to be highly expressed in  
354 endothelial cells in lymph node tissue based on tissue microarray (TMA) results in human protein atlas,  
355 could also be used as protein markers to differentiate SVEC cells with other two cell types. We also  
356 attempted to compare with our previous results using nanowell chips (Figure S7)<sup>16</sup>. Only 5 out of the 9  
357 membrane proteins could be classified as cell surface markers and 3 proteins were not detected, likely due  
358 to low sensitivity and reproducibility. Together, these data indicate the feasibility of using scProteomics  
359 to identify cell-type-specific membrane proteins for antibody-based cell enrichment.

## 360 **Discussion**

361 We have developed a high-throughput and streamlined scProteomics sample preparation workflow based  
362 on nested nanoPOTS array (N2) chips. The N2 chips reduce nanowell volumes to ~30 nL and improve the  
363 protein/peptide sample recovery by 230% compared with our previous nanoPOTS chips<sup>15,16</sup>. The N2  
364 design also significantly simplifies the TMT-based isobaric labeling workflow by eliminating the tedious  
365 sample pooling step (e.g., aspirating, transferring, and combining). Using the N2 chip, 243 single cells  
366 can be analyzed in a single microchip, representing 5× more numbers than our previous chips. In the near  
367 future, we envision the development of higher capacity N2 chips and/or stable isotope isobaric labeling  
368 reagents to enabling higher multiplexing scProteomics experiments (e.g. over 1000 cells per chip  
369 containing 5×5 array and 40 total clusters).

370 Using a recently-developed software, SCPCCompanion<sup>17</sup>, we observed single-cell SNRs were dramatically  
371 improved with N2 chip-based scProteomics workflow. The improvement results in high Pearson  
372 correlations (median R of ~ 0.97) of single cells from the same cell lines. Importantly, we observed low  
373 protein expression variations (median CVs of ~16.3%), suggesting the proteome is highly stable for single  
374 cells under identical culture conditions. These observations suggest cultured cells are good models to  
375 evaluate and benchmark the quantitative performance of scProteomics technologies.

376 In the analysis of three different cell lines, we verified the scProteomics can robustly classify single cells  
377 based on protein abundances and reveal functional differences among them. We also showed it was  
378 possible to directly identify cell surface markers by leveraging established subcellular-component  
379 databases.

380 It should be noted that all the single-cell isolation and sample preparation were performed using a  
381 commercially available system (cellenONE). The microchip fabrication can be readily implemented in a  
382 typical cleanroom facility. Thus, we believe our N2 chip-based scProteomics workflow can be rapidly  
383 disseminated.

384 In summary, we believe the N2 chip provides a universal scProteomics platform with broad applications  
385 in studying cell differentiation, tumor heterogeneity, rare cells from clinical specimens.

## 386 **Acknowledgments**

387 We thank Matthew Monroe for helping with data deposition. This work was supported by a Laboratory  
388 Directed Research and Development award (Y.Z.) from Pacific Northwest National Laboratory and a  
389 Intramural program (Y.Z.) at EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by  
390 the Office of Biological and Environmental Research and operated under Contract No. DE-AC05-  
391 76RL01830. Part of this work is also supported by NIH grants U01 HL122703 (C.A.) and P41  
392 GM103493 (R.D.S.).

## 393 **Contributions**

394 Y.Z conceptualized and designed the research. S.M.W, A.V.V. and H.S.M. fabricated N2 microchips.  
395 J.W., R.L.S, L.M.M., and J.C.B. performed the cell culture and sample preparation. S.M.W. and R.J.M  
396 performed LC-MS analysis. J.W., G.C., and Y.Z. analyzed data. J.W., G.C., R.D.S., L.P.T., and Y.Z.  
397 wrote the manuscript.

398 **Competing interests** J.C.B. is an employee of Scienion.

399 **Data Availability** The mass spectrometry proteomics data have been deposited to the  
400 ProteomeXchange Consortium via the MassIVE partner repository with the dataset identifier  
401 MSV000086809.

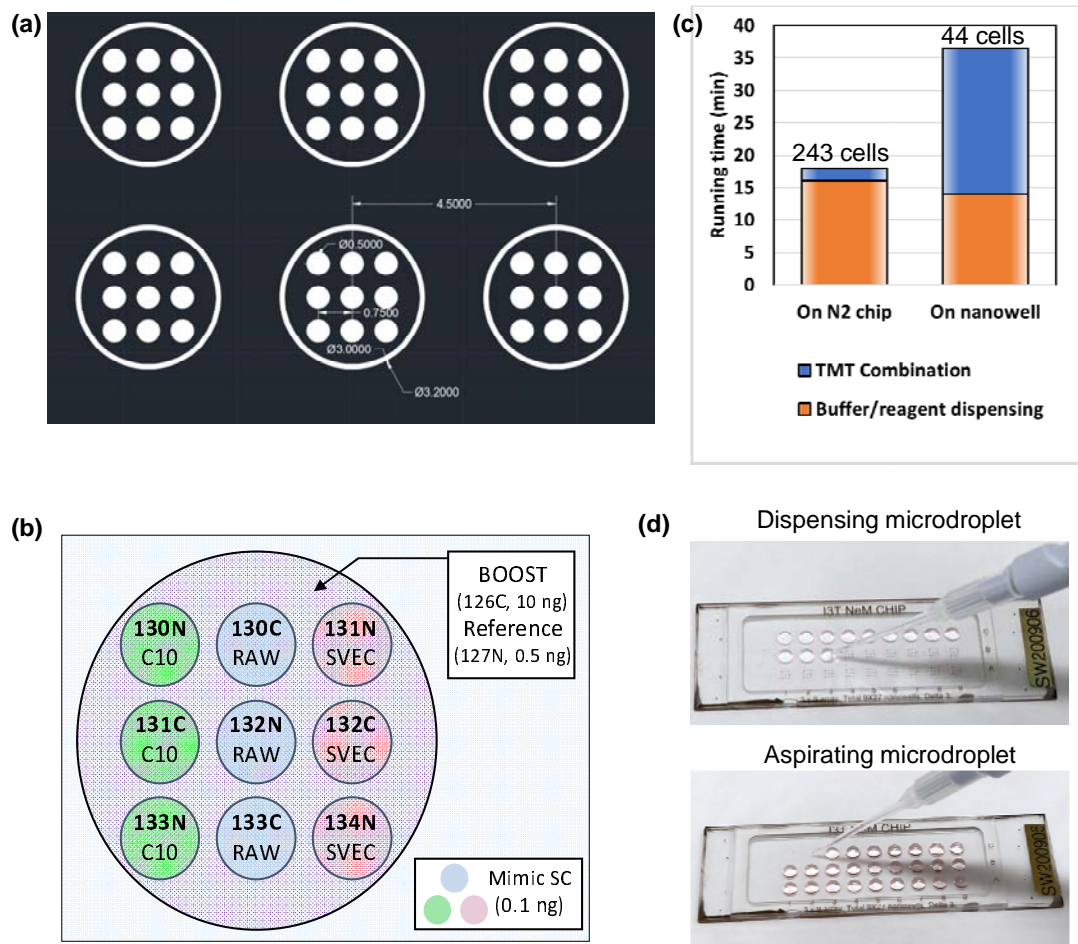
402

## 403 **References**

- 404 1. Labib, M. & Kelley, S.O. Single-cell analysis targeting the proteome. *Nat Rev Chem* **4**, 143-158  
405 (2020).
- 406 2. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function.  
407 *Nature* **537**, 347-355 (2016).
- 408 3. Lundberg, E. & Borner, G.H.H. Spatial proteomics: a powerful discovery tool for cell biology.  
409 *Nature Reviews Molecular Cell Biology* **20**, 285-302 (2019).
- 410 4. Kelly, R.T. Single-cell Proteomics: Progress and Prospects. *Mol Cell Proteomics* **19**, 1739-1748  
411 (2020).
- 412 5. Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per  
413 mammalian cell. *Chemical Science* **12**, 1001-1006 (2021).
- 414 6. Williams, S.M. et al. Automated Coupling of Nanodroplet Sample Preparation with Liquid  
415 Chromatography-Mass Spectrometry for High-Throughput Single-Cell Proteomics. *Anal Chem* **92**,  
416 10588-10596 (2020).
- 417 7. Zhu, Y. et al. Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet  
418 Sample Preparation and Ultrasensitive NanoLC-MS. *Angew Chem Int Ed Engl* **57**, 12370-12374  
419 (2018).
- 420 8. Zhu, Y. et al. Single-cell proteomics reveals changes in expression during hair-cell development.  
421 *Elife* **8** (2019).
- 422 9. Li, Z.Y. et al. Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. *Anal*  
423 *Chem* **90**, 5430-5438 (2018).

- 424 10. Shao, X. et al. Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. *Anal*  
425 *Chem* **90**, 14003-14010 (2018).
- 426 11. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of  
427 10-100 mammalian cells. *Nat Commun* **9**, 882 (2018).
- 428 12. Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per  
429 mammalian cell. *Chemical Science* (2021).
- 430 13. Cong, Y.Z. et al. Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC  
431 Columns and Ultrasensitive Mass Spectrometry. *Anal Chem* **92**, 2665-2671 (2020).
- 432 14. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single  
433 mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**,  
434 161 (2018).
- 435 15. Tsai, C.F. et al. An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy  
436 for Precise Quantitative Single-cell Proteomics. *Mol Cell Proteomics* **19**, 828-838 (2020).
- 437 16. Dou, M. et al. High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in  
438 a Nanodroplet Sample Preparation Platform. *Anal Chem* **91**, 13119-13127 (2019).
- 439 17. Cheung, T.K. et al. Defining the carrier proteome limit for single-cell proteomics. *Nat Methods*  
440 **18**, 76-83 (2021).
- 441 18. Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity  
442 using SCoPE2. *Genome Biology* **22**, 50 (2021).
- 443 19. Liu, W.W., Zhu, Y., Feng, Y.M., Fang, J. & Fang, Q. Droplet-Based Multivolume Digital Polymerase  
444 Chain Reaction by a Surface-Assisted Multifactor Fluid Segmentation Approach. *Anal Chem* **89**,  
445 822-829 (2017).
- 446 20. Liang, Y., Truong, T., Zhu, Y. & Kelly, R.T. In-Depth Mass Spectrometry-Based Single-Cell and  
447 Nanoscale Proteomics. *Methods Mol Biol* **2185**, 159-179 (2021).
- 448 21. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-  
449 based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).
- 450 22. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of  
451 (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
- 452 23. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing  
453 batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**,  
454 882-883 (2012).
- 455 24. Brenes, A., Hukelmann, J., Bensaddek, D. & Lamond, A.I. Multibatch TMT Reveals False Positives,  
456 Batch Effects and Missing Values. *Mol Cell Proteomics* **18**, 1967-1980 (2019).
- 457 25. Waters, C.M., Roan, E. & Navajas, D. Mechanobiology in lung epithelial cells: measurements,  
458 perturbations, and responses. *Compr Physiol* **2**, 1-29 (2012).
- 459 26. Yang, W. et al. Neutrophils promote the development of reparative macrophages mediated by  
460 ROS to orchestrate liver repair. *Nat Commun* **10**, 1076 (2019).
- 461 27. Baardman, J. et al. A Defective Pentose Phosphate Pathway Reduces Inflammatory Macrophage  
462 Responses during Hypercholesterolemia. *Cell Rep* **25**, 2044-2052 e2045 (2018).
- 463 28. Wang, S. et al. Control of endothelial cell proliferation and migration by VEGF signaling to  
464 histone deacetylase 7. *Proc Natl Acad Sci U S A* **105**, 7738-7743 (2008).
- 465 29. Jia, M. et al. Ezrin, a Membrane Cytoskeleton Cross-Linker Protein, as a Marker of Epithelial  
466 Damage in Asthma. *Am J Respir Crit Care Med* **199**, 496-507 (2019).
- 467 30. Mandell, K.J., Babbin, B.A., Nusrat, A. & Parkos, C.A. Junctional adhesion molecule 1 regulates  
468 epithelial cell morphology through effects on beta1 integrins and Rap1 activity. *J Biol Chem* **280**,  
469 11665-11674 (2005).
- 470 31. Ulm, C. et al. Soluble polysialylated NCAM: a novel player of the innate immune system in the  
471 lung. *Cell Mol Life Sci* **70**, 3695-3708 (2013).

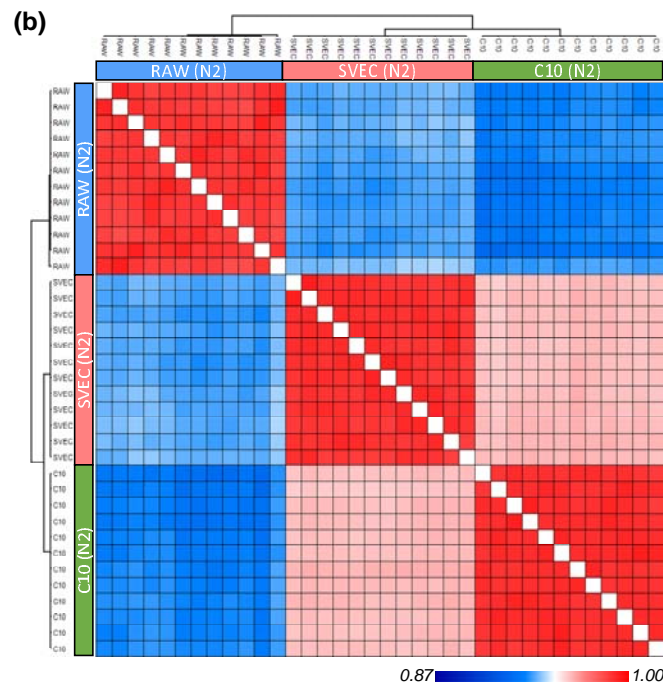
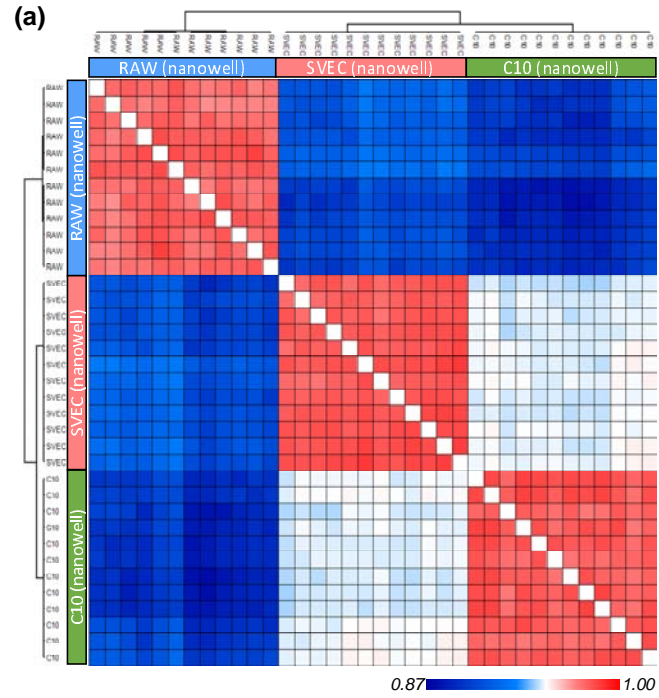
- 472 32. Khazen, W. et al. Expression of macrophage-selective markers in human and rodent adipocytes.  
473 *FEBS Letters* **579**, 5631-5634 (2005).
- 474 33. Chistiakov, D.A., Killingsworth, M.C., Myasoedova, V.A., Orekhov, A.N. & Bobryshev, Y.V.  
475 CD68/macrosialin: not just a histochemical marker. *Laboratory Investigation* **97**, 4-13 (2017).
- 476 34. Kim, Y.R. et al. Identification of highly potent and selective inhibitor, TIPTP, of the p22phox-  
477 Rubicon axis as a therapeutic agent for rheumatoid arthritis. *Sci Rep* **10**, 4570 (2020).
- 478 35. Erikson, E. et al. In vivo expression profile of the antiviral restriction factor and tumor-targeting  
479 antigen CD317/BST-2/HM1.24/tetherin in humans. *Proc Natl Acad Sci U S A* **108**, 13688-13693  
480 (2011).
- 481



482

483 **Supplementary figure 1.** (a) Schematic illustration showing the N2 chip design. The white areas are designed for nanowells and  
 484 hydrophilic rings, while the rest is hydrophobic surface. (b) Experiment design for mimic single-cell sample (0.1 ng peptide) on  
 485 the N2 chip. (c) Estimated robot operation time for single cell proteomics using N2 chip and nanowell chip. (d) Photographs  
 486 showing TMT-based samples can be pooled together by spotting a 6- $\mu$ L droplet using a micropipette. Similarly, the pooled  
 487 sample can be retrieved and loaded into an autosampler vial for LC-MS analysis.

488

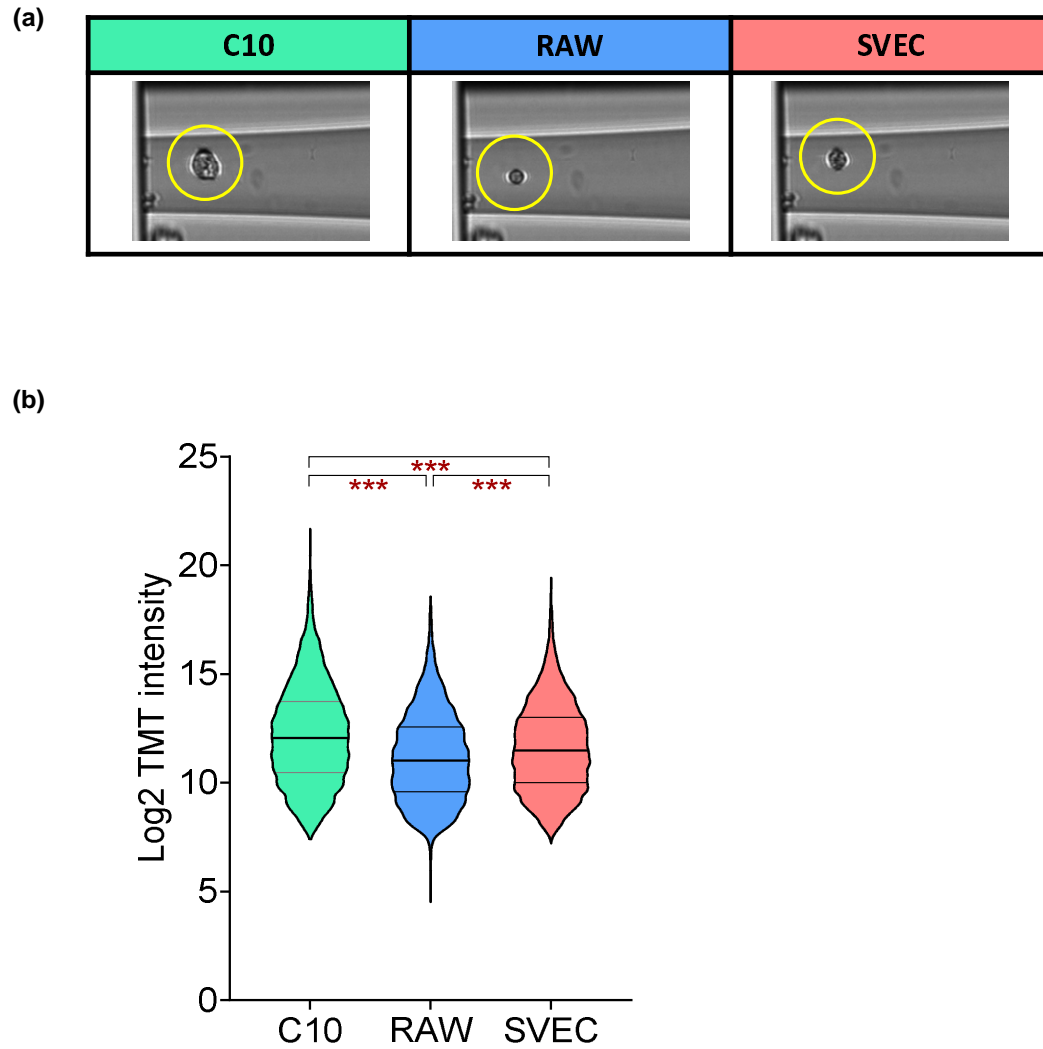


489

490 **Supplementary figure 2.** Heatmap of pairwise Pearson correlations among individual samples in nanowell chip (a) and N2 chip  
491 (b).

492



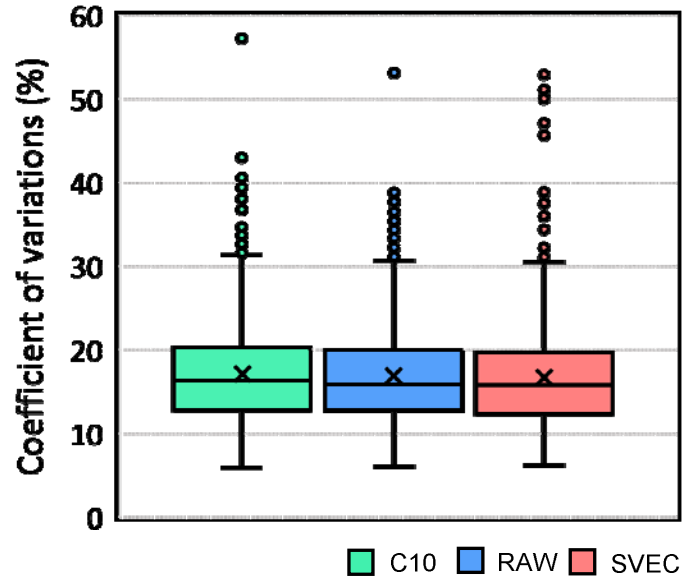


493

494 **Supplementary figure 3.** (a) Representative images of single cells. The measured cell sizes in diameter are 18-20  $\mu\text{m}$  for C10  
495 cells, 7-10  $\mu\text{m}$  for RAW cells, and 13-15  $\mu\text{m}$  for SVEC cells. (b) Violin plots showing the distribution of log<sub>2</sub> transformed protein  
496 intensities for the three cell types.

497

498

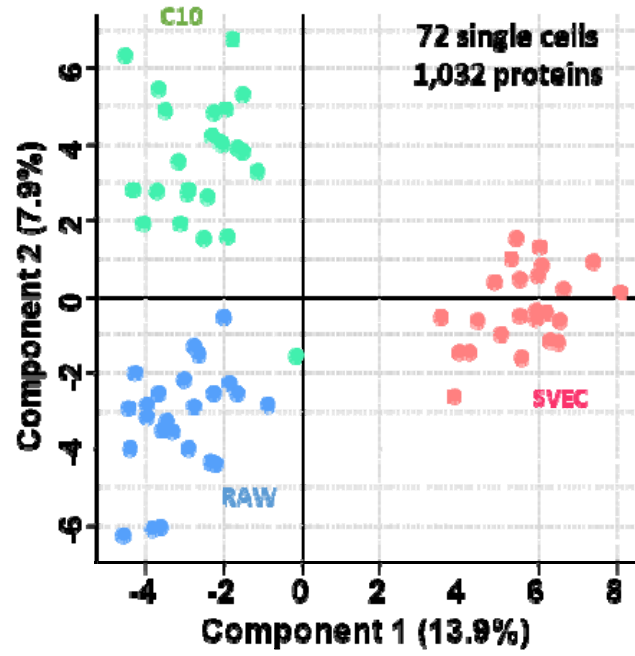


499

500 **Supplementary figure 4.** The distribution of coefficient of variations (CVs) for protein abundances in single cells among inter  
501 TMT batches with batch correction. For each cell type, 36 single cells from 12 TMT sets were used for the calculation.

502

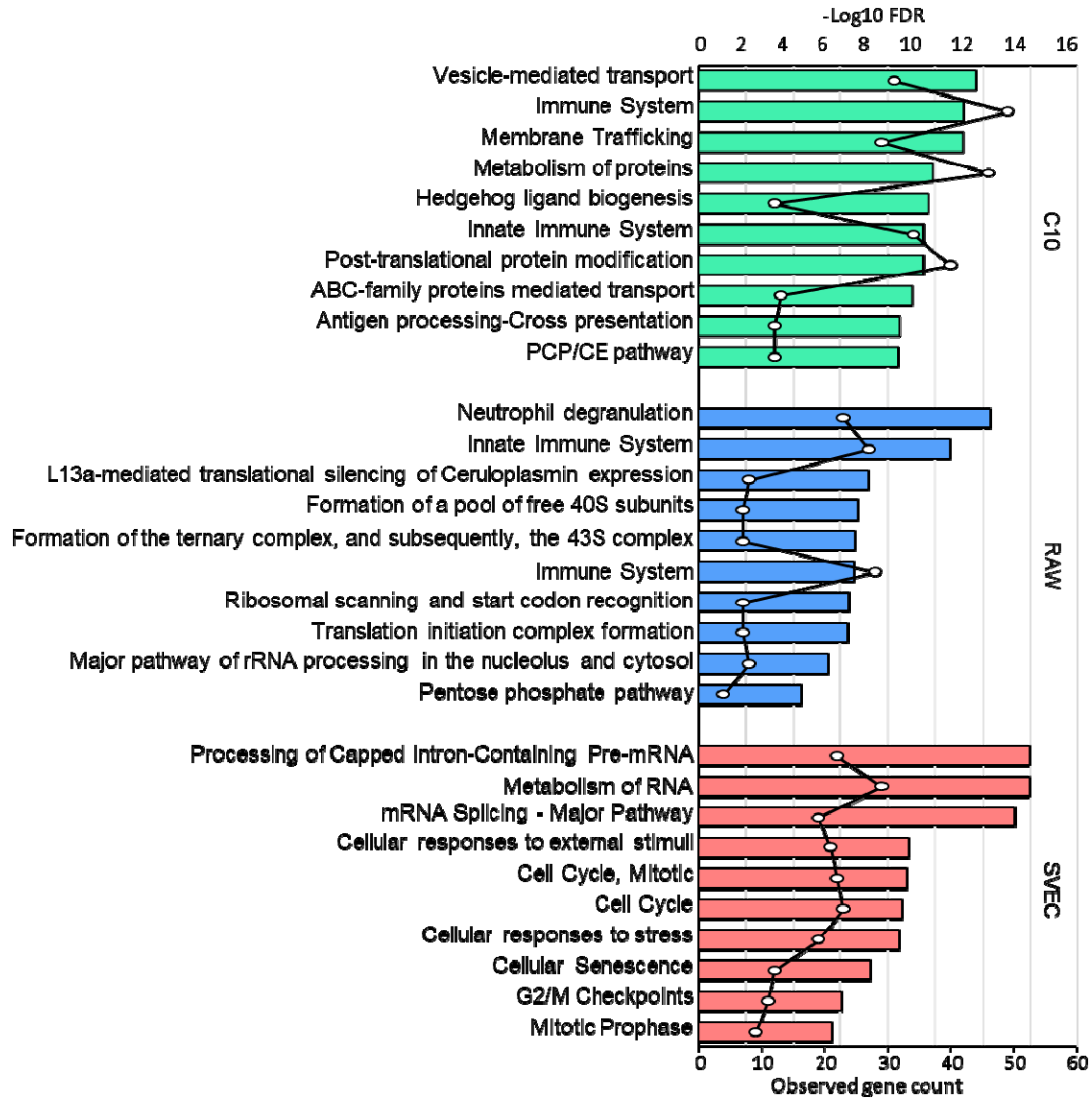
503



504

505 **Supplementary figure 5.** PCA plot showing the clustering of 72 single cells using nanowell chips<sup>16</sup>. Total 1032 proteins were  
506 used.

507



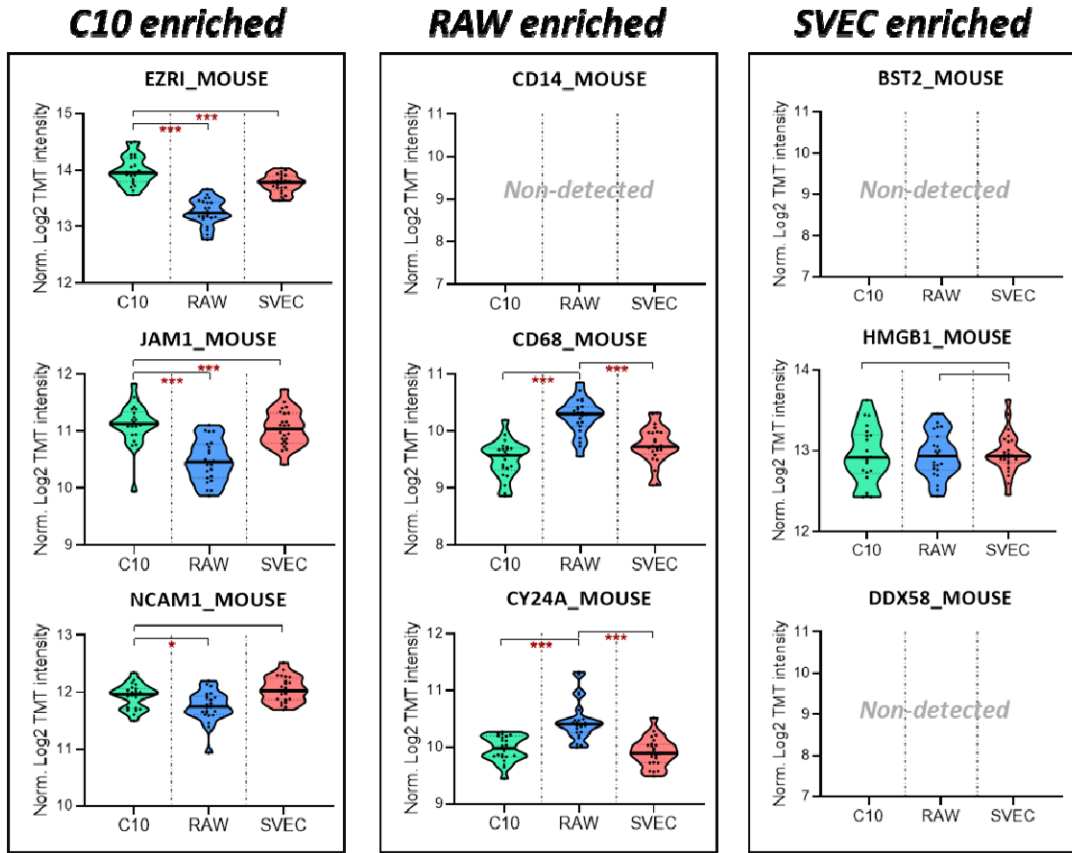
508

509 **Supplementary figure 6.** Reactome pathway analysis of enriched proteins in each cell type by hierarchical clustering analysis.

510 The top 10 of the pathways per each homogeneous cell type based on adjusted p-value were listed with the number of

511 observed protein count.

512



513

514 **Supplementary figure 7.** Violin plots showing the intensity distributions of putative plasma membrane protein markers for  
515 specific cell types using previous nanowell-chip-based single-cell proteomics data<sup>16</sup>. The protein lists were selected based on  
516 the data from the N2 chips and shown here to compare the improved proteome coverage and quantitation performance of the  
517 N2 chip.