Ultra-streamlined single cell proteomics by all-in-one chip and data-independent-acquisition mass spectrometry

Sofani Tafesse Gebreyesus[1,2,3,†], Asad Ali Siyal[1,4,5,†], Reta Birhanu Kitata[1], Eric Sheng-Wen Chen[1], Bayarmaa Enkhbayar[6], Takashi Angata[6], Kuo-I Lin[7], Yu-Ju Chen[1,3,4,8,*] and Hsiung-Lin Tu[1,2,4,8,*]

1. Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan
2. Taiwan International Graduate program, Nano Science and Technology program, Academia Sinica, Taipei 115, Taiwan
3. Department of Chemistry, National Taiwan University, Taipei 106, Taiwan
4. Taiwan International Graduate Program, Chemical Biology and Molecular Biophysics Program, Academia Sinica, Taipei 115, Taiwan
5. Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan
6. Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan
7. Genomics Research Center, Academia Sinica, Taipei 115, Taiwan
8. Genome and Systems Biology Degree Program, Academia Sinica and National Taiwan University, Taiwan

†: Equal contribution

*: corresponding author

ABSTRACT

Single cell proteomics provides the ultimate resolution to reveal cellular phenotypic heterogeneity and functional network underlying biological processes. Here, we present an ultra-streamlined workflow combining an integrated proteomic chip (iProChip) and data-independent-acquisition (DIA) mass spectrometry for sensitive microproteomics analysis down to single cell level. The iProChip offers multiplexed and automated all-in-one station from cell isolation/counting/imaging to complete proteomic processing within a single device. By mapping to project-specific spectra libraries, the iProChip-DIA enables profiling of 1160 protein groups from triplicate analysis of a single mammalian cell. Furthermore, the applicability of iProChip-DIA was demonstrated using both adherent and non-adherent malignant cells, which reveals 5 orders of proteome coverage, highly consistent ~100-fold protein quantification (1-100 cells) and high reproducibility with low missing values (<16%). With the demonstrated all-in-one cell characterization, ultrahigh sensitivity, robustness, and versatility to add other functionalities, the iProChip-DIA is anticipated to offer general utility to realize advanced proteomics applications at single cell level.

INTRODUCTION

Rapidly developing single cell omics-based molecular measurements have revolutionized modern biological research[1, 2]. As proteins are functional workhorses of the cell, proteomic profiling provides a direct snapshot of the dynamic biological network and its alteration to complement the genomics and transcriptomics architecture[3]. However, the sensitivity of proteomic profiling is greatly limited due to the wide dynamic range of proteome constituents and the lack of a viable protein amplification strategy[4]. Targeted protein analyses have enabled sensitivity down to the single cell level, but their multiplexity is often limited and depend heavily on the antibody availability and quality[5-8]. Mass spectrometry (MS)-based proteomics approach, meanwhile, offers label-free analysis with high specificity and deep proteomic coverage, which theoretically extends to the single cell sensitivity[8-11]. However, multi-step processing in traditional MS workflows often results in significant sample loss, linking trade-offs between the high proteome coverage and the accessible sample size[8, 12].

Microproteomics workflows aiming at handling minute samples were widely developed to expand the horizon of MS-based proteomic analysis towards limited input samples (<1000 cells)[13].

Typical strategies developed reagents and methods that could integrate the entire or partial workflow compatible to small sample size. For example, filter-aided sample preparation (FASP), inStageTip (iST), integrated proteome analysis device (iPAD) and single-pot solid-phase-enhanced sample preparation (SP3) reported protocols that combines cell lysis, protein digestion, and/or detergent removal to improve proteome identification at the level of few hundreds cells[14-17]. Alternatively, sample preparation on nano-liter droplet using pre-deposited cells has been developed, including the nanoPOTs, oil-air droplet (OAD) and digital microfluidic (DMF-SP3) chip; these methods effectively reduced adsorptive loss and afforded sensitive proteome coverage of 1517, 1063 and 2500 proteins from 10, 100 and 500 cells, respectively[18-20]. A recent extension of nanoPOTs with ultra-low-flow nanoLC, high-field asymmetric ion mobility spectrometry (FAIMS) coupled to Orbitrap Eclipse instrument reported sensitive profiling of 1056 protein groups from a single cell[21]. With these advances, nonetheless, a fully automated all-in-one workflow, starting from multiplexed input cell capturing, counting and imaging, cell lysis, protein digestion to peptides desalting, all integrated within a single device to realize sensitive proteomics analysis for low-input samples is not yet established, although it is anticipated to most drastically minimize sample loss and achieve high reproducibility and sensitivity.

Microfluidic devices use custom chip integration and hydraulic actuations to achieve precise μL-to-nL fluid manipulation, and are ideal platforms to execute a complex protocol[22-24]. However, microfluidics has not been explored for streamlined proteomics workflow primarily due to challenges associated with the compatibility among reagents used in multi-step process for one-pot protocol, concerns of mixing in confined space and overall system integration. Towards highly sensitive and robust proteomic analysis, we developed herein an automated microfluidic chip (termed iProChip) coupled with data-independent acquisition (DIA) MS as a streamlined microproteomics pipeline. The iProChip was designed as a full automated station of the entire proteomic workflow, offering built-in features including readily quantifiable cell capture and imaging, complete cell lysis and protein digestion as well as effective peptide desalting. Following iProChip processing, we showed that DIA MS, which detects all precursors and fragments in the entire m/z range within isolation windows, enabled all retrospective peptides mapping against spectral libraries and offered superior coverage than conventional data-dependent-acquisition (DDA) mode by 2.3-fold[25]. Importantly, the iProChip-DIA workflow characterized 938, 980 and 1011 protein groups from a single cell with 1% FDR. The versatility of iProChip-DIA was

demonstrated using both the human adenocarcinoma cell (PC-9) and human chronic B cell leukemia cell (MEC-1), whose size differences were readily quantified using the built-in cell imaging feature. The results revealed superior performance of 5 orders of proteome coverage, >100-fold quantification range, high reproducibility (Pearson correlation of 0.88-0.98) and low between-run missing values (<16%). To the best of our knowledge, this is the first implementation of a miniaturized device with all-in-one functionality to achieve automated and streamlined cell isolation, imaging and proteome sample preparation, which offers ultra-high sensitivity and reproducibility for limited input samples. Further integration with DIA MS analysis achieved one of the highest proteome coverage for a single cell with good quantitation performance.

RESULTS

Design and characterization of the iProChip and streamlined microproteomics workflow

To provide a streamlined microproteomics analysis pipeline for extremely limited input samples, we designed a microfluidic device as an integrated proteomics chip (iProChip) to offer all-in-one functionality from cell input to complete proteomic sample processing. The iProChip has a two-layer, push-up geometry and allows accurate fluid manipulation via 34 valves controlled by a custom program, therefore offering an automated protocol for precise and systematic control (**Fig. 1a-f** and **Supplementary Fig. 1a**)[26]. The chip is composed of 9 units to enable multiplexed proteomics experiments running in parallel. Each unit contains a cell capture, imaging and lysis chamber, a protein reduction, alkylation and digestion vessel, and a peptide desalting column (**Fig.1c** and **Supplementary Fig. 1b**). All units share 9 common inlets and 2 outlets, allowing programmed delivery of reagents and simultaneous sample processing to increase assay throughput. The cell trap is made up of arrays of 10, 50 and 100 twin pillars spaced by 5 μm for rapid size-based cell capture (**Fig. 1e)**[27]. A circular chamber with a diameter of 2 mm (312 nL) was fabricated to accommodate the entire proteomic workflow, including cell lysis, protein reduction, alkylation and digestion in a single-step (**Fig. 1f** and **Supplementary Fig. 2**). A desalting column was fabricated by packing reversed phase C18 beads into the serpentine microchannel pre-patterned with 5 μm filters to perform on-chip clean-up of digested peptides prior to LC-MS/MS analysis (**Fig. 1e, f,** and **Supplementary video 1**). To increase the proteome coverage, we applied a deep single-shot profiling strategy that integrates direct- and library-based

DIA analysis using Orbitrap mass spectrometer. We developed a spectral library resource, complementarily established by hybrid DDA-DIA datasets using either cancer cell lines or immune cells consisting of different proteome composition, which can serve as a digital map to theoretically recover all peptides in the m/z and retention time domains of DIA data (**Fig. 1g**). Specifically, spectra libraries constructed from the cell lines with different cell numbers have been tested and optimized to maximize the number of protein identification and quantification.

In the first step of streamlined proteomics workflow, the cell trapping efficiency was optimized and determined using non-small cell lung cancer (NSCLC) PC-9 cells (**method**). Using optimal cell density (5.0×10^5 cell/ml), desired numbers of cells (1-100) for each unit can be trapped in 10-60 s. The average percentage of cells captured from traps containing a single cell were 100 %, 92 ± 3 % and 89 ± 8 % for chambers with 10, 50 and 100 traps respectively. The targeted capture efficiency for all units achieves ~100% after counting traps containing 1 (90%) and 2 or 3 cells (~10%), establishing it as an absolute quantifiable module to perform simple and fast size-based cell isolation (**Fig. 2a, b, method** and **Supplementary Video 2**). Compared to external stand-alone cell sorting instruments, such a built-in module offers a simple, rapid and efficient cell isolation. Additionally, we also showed that by using lower input cell density of 2.5×10^4 cell/ml, such cell chambers allow precise capture of lower numbers of cells at the level of 1 and 5 cells (**Supplementary Fig. 3**). Furthermore, although unexplored in this study, such design is amenable to adapt different numbers of traps, as well as alternative cell sorting strategies, such as via the trap functionalization of cell surface markers.

Next, we sought to characterize if mixing of input reagents can efficiently occur in the closed vessel during the cell lysis and protein digestion. Three mixing approaches, including vortexing, shaking (by a plate shaker), and passive diffusion, were tested (**Supplementary Fig. 4**). Using imaging analysis, relative mixing index (RMI) was calculated to assess the mixing performance (**method**)[28]. The result showed that it took 11, 16, and 30 min for vortexing, shaking, and diffusion-mixing to reach 75% RMI, indicating that all three mixing strategies were sufficient to accommodate reactions within minutes to hours reaction kinetics, which fit the timescale of conducting proteomics workflow (**Fig. 2c, d** and **Supplementary Fig. 4**). Although vortex mixing was found to provide faster mixing, mixing by shaking was used in subsequent experiments due to its flexibility in handling and sufficient reaction timescale.

Another integration to the miniaturized device is the multiplexed on-chip peptide desalting module. Two sets of 5 μm spaced filters were embedded at both ends of the column to enable desalting materials such as C18 beads packing (**Fig 2e** and **method**). The loading capacity and peptide recovery efficiency of the desalting module were assessed through the BCA assay by using tryptic peptides of BSA protein (**method**). The quantitation result showed a linear correlation from 0.125 to 1 μg with ~89 % recovery (**Fig. 2f** and **method**). Assuming that a typical mammalian cell contains 200 pg proteins[14], the desalting column is thus anticipated to capture peptides from approximately 4000 cells. Furthermore, the concern of a compromised sample retrieval from circular reaction vessel to desalting column due to the preferential flow resulting from (1) circular shape of the chamber and (2) non-negligible flow resistance caused by compactly-packed C18 beads was studied by flowing a colored dye through the C18 beads-packed column (**method**). The result showed that 9 psi was the minimal flow pressure to overcome the preferential flow, and that 11 psi was used in our following workflow (**Supplementary Fig. 5** and **method)**.

Integration of iProChip with DIA MS

To operate the iProChip for proteomic workflow, precise numbers of cells including 1, 5, 10, 50 and 100 cells were prudently captured using built-in cell traps. Parallel sample processing of the cells trapped in multiple capture chambers were performed by dispensing and incubating with the cocktail buffer, containing the RapiGest, TCEP, and CAA, which was specifically adapted to achieve one-pot cell lysis, protein reduction and alkylation to minimize sample loss (**Supplementary Fig. S2, S6** and **video 3**). Subsequent protein digestion and acidification were carried out in the reaction vessel, and digested peptides were then subjected to fast and multiplexed desalting by passing through the C18 beads-loaded column for 15 min. For subsequent LC-MS/MS analysis following the iProChip, single-shot DIA-MS acquisition parameters, including isolation window, resolution, peptide amount and LC-MS/MS gradient time, were optimized and adapted to enhance proteomic profiling of low-abundant proteins in low numbers of cells (**method**).

To allow deep profiling and enhanced identification of low abundant and cancer-relevant proteins by DIA, high quality project-specific spectra libraries were constructed using lung cancer and human chronic lymphocytic leukemia cell line. The protein compositions and dynamic range may vary in bulk samples of thousands cells and a single cell, which likely affect the chromatographic

time domain and DIA acquisition. Thus, we constructed both large-scale (1 μg) and small-scale (~10 cells) libraries from respective cell types, i.e. PC-9 and MEC-1 cells, and implemented them to analyze different numbers of cells. Specifically, the large-scale project specific libraries of PC-9 and MEC-1 processed in bulk/dilution with data-dependent and data-independent acquisition mode consisted of 6,345 protein groups (83,305 peptides) and 6,261 protein groups (60,335 peptides) with 1% precursor and protein false discovery rates (FDRs), respectively. These large-scale libraries were used for analyzing higher cells (i.e. >10cells). For extremely low samples (i.e. ~5 and single cells), we reasoned that a small-scale specific library should be beneficial for identification and quantification. To maximize the proteome profiling sensitivity at the single cell level, we thus constructed small-scale spectra libraries from ~10 cells processed through iProChip as well as aliquots of ~1.5ng (~10 cells) processed through bulk/dilution, yielding a depth of 2,231 protein groups (14,054 peptides) and 2,440 protein groups (11,720 peptides), respectively.

To evaluate the performance of DIA-based quantitation, analytical merits in sensitivity, proteome coverage, and reproducibility were systematically investigated by using iProChip to process 13-14 PC-9 cells and compared to the conventional DDA method (**Fig. 3a**). By the conventional DDA method, only an average of 869 protein groups (3,280 peptides) were identified in triplicate analysis. In comparison, the direct DIA (*dir*DIA) approach using Spectronaut tool identified 1,409 protein groups (5,174 peptides), whereas the library-assisted DIA (*lib*DIA) approach using the large-scale PC-9 cell library showed significantly higher proteome coverage of 1,874 protein groups (6,929 peptides) of 1% FDR at peptide-to-spectrum match (PSM), peptide and protein level. Comparing the direct-DIA and spectral library-based results, the superior quantitation of *lib*DIA is likely due to more efficient detection of low intensity peptide ions in DIA mode to match corresponding peptide spectra in our library. By combining the complementary *dir*DIA and *lib*DIA results, the overall identification coverage further increased to 2,022 proteins (7,757 peptides). The identification of 2.3-fold and 2.4-fold protein groups and peptides, respectively, by DIA approach revealed its superior profiling coverage over DDA approach (**Fig. 3a**). These results demonstrated that a single shot DIA-based LC-MC/MC, complementarily processed by *dir*DIA and *lib*DIA, dramatically improved the proteome identification coverage for the small-scale sample from the fully automated sample preparation in iProChip.

To evaluate the reproducibility of our iProChip-DIA workflow, we calculated the percentage of overlapping proteins between the triplicate analysis of 14 PC-9 cells. The result showed 84.2% among 2,022 identified proteins and 71.7% among 869 identified proteins were reproducibly detected by DIA and DDA, respectively, indicating higher reproducibility and proteome coverage in iProChip-DIA approach (**Fig. 3b**). For evaluation of highly reproducible quantitation with CV ≤ 20%, DIA achieved significantly higher coverage of 1,160 quantifiable proteins compared to only 522 proteins quantified by DDA (**Fig. 3c**). Previous label-free quantification methods have commonly observed 10-50% between-run missing values, presenting a bottleneck for reproducible quantification across samples[29]. Additional comparison for run-to-run variabilities revealed fewer missing values, i.e. number of proteins only quantified in one of triplicate runs, in the DIA result (16%) compared to that of DDA (28%) (**Fig. 3d**). Meanwhile, the wide dynamic range of proteome compositions presents another major bottleneck for deep profiling, especially for low abundant proteins. Thus, we assessed the dynamic range based on protein abundance rank. By conventional DDA, the abundances of the 1,014 identified proteins were found to span ~4 orders of magnitude, whereas the 2,170 identified proteins in DIA span ~5 orders of magnitude with coverage of important and low-abundant oncoproteins related to cancer. Notably, FDA approved druggable targets for lung cancer, such as EGFR, MAP2K1, MAP2K2 and proteins involved in NSCLC pathway including EGFR, NRAS, MAP2K1, MAP2K2, MAPK1, MAPK3, CDK4, TP53 were readily identified in DIA, whereas only TP53 and CDK1 were detected in DDA using our approach (**Fig. 3e**)[30](*https://www.cancer.gov/about-cancer/treatment/drugs/lung*). In summary, these results showed that the DIA approach provided significantly higher proteome profiling coverage, lower missing values, more reproducible quantification and wider dynamic range compared to DDA at the level of 14 cells.

Single cell proteomic profiling by iProChip and DIA MS

Next, we systematically evaluated the iProChip-DIA performance for proteomic profiling of PC-9 cells in chambers with 10, 50 and 100 traps. As expected, the iProChip provided precise cell counting for each chamber of different cell traps to ensure unambiguous quantification (**Fig. 4a**). Combining *lib*DIA and *dir*DIA analysis, on average 4,722±10 protein groups (25,785 peptides) were identified at 1% FDR from 106±2 cells. At chambers with lower number of cell traps, an average number of 3,435±262, 2,022±114 and 1638±191 protein groups were identified from 55±1,

14±1 and as low as 5±1 cells in three replicate analysis, respectively. (**Fig. 4a** and **Supplementary Fig. 7**). The overlapping of identified protein groups in triplicate analyses is 77%-93% from all cells, illustrating the high reproducibility of the iProChip-DIA approach **(Supplementary Fig. 8**).

Encouraged by the high sensitivity, we further pushed the profiling sensitivity at a single PC-9 cell using 10-cells capture chambers. An average of 976±37 protein groups (3,069 peptides) were identified from a single PC-9 cell (**Fig. 4b**). Triplicate measurements yielded identification of a total of 1160 protein groups (3,995 peptides) from a single PC-9 cell. To the best of our knowledge, the results present one of the highest reported coverage so far from independent replicates. Comparing the identification coverage showed 69% protein groups and 55% peptides were common between triplicate results. Intriguingly, with the built-in capability to directly image each captured cell, we observed that numbers of identified proteins and peptides are correlated with the approximate size of the individual captured cell (**Fig. 4b**).

To evaluate the analytical reproducibility of our approach for analysis of different numbers of cells, the protein abundances in triplicate datasets were quantitatively compared by pairwise correlation analysis. The result showed high reproducibility (Pearson's correlation of 0.88-0.98) in the measured protein abundance for protein quantification by our iProChip-DIA workflow (**Fig. 4c** and **Supplementary Fig. 9**). To assess the quantitative performance, next, the distribution of overall protein abundances quantified in each cell number were calculated, which showed a log-linear correlation across different cell numbers (**Fig. 4d**). The capability of quantitative proteomics analysis was further evaluated at the individual protein level. Representative examples were selected from lung cancer related oncoproteins and their abundances were computed by Spectronaut[31]. The average protein abundances among representative lung cancer proteins, EGFR, CDK1, and MAP2K1, revealed a good linearity between the measured protein abundance and increasing cell numbers (**Fig. 4e**). Most importantly, it is noted that many quantified proteins, such as selected examples of TP53, ITGB1, PGK1 and MAPK1 show good quantitative dependence (50-100 fold) between the absolution protein abundance and cell number (1-106 cells) (**Fig. 4e** and **Supplementary Fig. 10**). The quantification of 50-100 fold of magnitudes also demonstrated a dramatically wider dynamic range compared to conventional quantitative proteomic results in bulk scale. In line with the aforementioned quantitative performances, iProChip-DIA enabled high degree of robustness, excellent reproducibility and quantitative proteomic measurements down to

the level of single cells, a level of performance only achieved previously for ensemble measurements.

The identification of these proteins enabled us to map lung cancer-related signaling pathways searched against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [31]. Total of 329 pathways were enriched, such as NSCLC pathway, metabolic pathways, pathways in cancer, spliceosome, viral carcinogenesis, proteoglycans in cancer, MAPK signaling, and apoptosis (**Supplementary Fig. 11)**. The major lung cancer pathway, NSCLC pathway, was enriched with coverage of a total of 29 proteins across different numbers of cells (**Fig. 4f**). Even at low cell numbers (14±1 cells), 13 proteins including the drug targets EGFR, MAP2K1, MAP2K2, MAPK1, MAPK3, KIF5B, tumor suppressor TP53, and other key signaling components (KRAS, CDK4, CDKN2A, EML4, KIF5B, NRAS, BAX, RB1) were identified. In terms of sensitivity, EGFR, MAPK1, MAP2K1, MAP2K2, CDKN2A, TP53, KIF5B and GRB2 proteins were still detected down to as low as 5 cells, whereas MAP2K1, KRAS and TP53 were even identified at single-cell level. Based on the lung cancer model study, these results revealed the capability of the developed approach to provide protein identification coverage to study the cancer proteome and wide range of cellular pathways at extremely low cell numbers.

Application of iProChip-DIA for single leukemia cell proteomic profiling

The general applicability of our iProChip-DIA platform was next demonstrated on the proteomic profiling of the human B-chronic lymphocytic leukemia (B-CLL) MEC-1 cell line. From the methodology development aspect, the leukemia cells representing a highly heterogeneous cancer type are ideal models for developing highly sensitive proteomics tools, as they could readily complement various existing methods by delineating the system-wide profiles of the phenotypic functionality[32]. When processing MEC-1 cells on-chip, the imaging-based cell trapping feature of iProChip readily revealed that MEC-1 cells were noticeably smaller than PC-9 cells (**Supplementary Fig. 12**). Nevertheless, the cell trapping strategy towards the comparatively smaller MEC-1 cells still displayed good capture efficiencies within the targeted numbers. Combining dirDIA and libDIA using the MEC-1 spectral library, triplicate analyses of the MEC-1 cell by iProChip-DIA analysis yielded average of 3,811±362, 931±72 and 455±98 protein groups at 1% FDR from 117±1, 14±1 cells and a single cell, respectively (**Fig. 5a** and **Supplementary**

Fig. 13). The protein abundance was found to span ~5 orders of magnitude across different cell numbers, allowing the detection of important B-cell surface markers CD20, HLA-B, HLA-DRA, HLA-DRB1 and HLA-DRB5 from as little as a single cell (**Fig. 5b**). While other key proteins including CD19, CD22, CD47 and CD74 were identified from 14 and 117 cells in addition to the aforementioned list of proteins (**Fig. 5b**). Functional annotation using UniProt showed that many proteins related to adaptive immunity, innate immunity, kinases, phosphatases and Ig domain were identified from as little as a single MEC-1 cell, where the depth of protein coverage positively correlated with the cell number (**Fig. 5c**). By mapping 518 human kinases deposited in KinMap[33], 114 protein kinases were identified across all major branches of kinase phylogenetic tree, including tyrosine kinase (TK), tyrosine kinase-like (TKL) kinases, serine/threonine protein kinases (STE), casein kinase 1 (CDK1), Ca^{2+}/calmodulin-dependent protein kinase (CAMK), CDK-MAPK-GSK3-CLK families (CMGC), Atypical and AGC kinase groups family (consisting of cyclic-nucleotide and calcium-phospholipid-dependent kinases, ribosomal S6 kinases and G-protein-coupled kinases) in the dendrogram of the human kinome (**Fig. 5d**). It was also noted that although MEC-1 cells were smaller than PC-9 cells, protein identification achieved good coverage and overlap (61%-81%) using the iProChip-DIA approach, suggesting the versatility and robustness of our platform for different cell types **(Supplementary Fig. 7**).

B-cell receptor (BCR) signaling is crucial for mounting efficient adaptive immunity and involved in the survival and growth of malignant B cells in B cell leukaemias or lymphomas[34]. The B-cell activation is regulated via the interaction between the surface receptor complexes in the BCR and specific antigens or via the tonic-signaling that is independent of antigen ligation[35]. In B-CLL, antigen-dependent activation of BCR contributes to cell proliferation[34]. Using the iProChip-DIA approach, we were able to map 83% proteins within the BCR pathway and identify the key BCR co-receptors including CD19, CD21, CD22 and CD81 (**Supplementary Fig. 14**). In addition, proteins involved in the downstream pathway of BCR such as mitogen-activated protein kinase (MAP2K1, MAP2K2, MAPK3 and MAPK1,) were also confidently identified. Compared to large-scale human immune cell proteomics study at the depth of >10,000 proteins using 28 primary hematopoietic cell populations by Rieckmann et al., 93% of the 4,211 proteins identified from MEC-1 cells were in common with additional 309 proteins uniquely identified in this study, including a comparable number of key B cell surface receptors such as CD19, CD21, CD22, CD81, FcgRIIB, and Igβ[36]. Taken together, the results demonstrated that the iProChip-DIA approach is a

highly sensitive and reproducible method with precise cell and protein quantitation capability that is readily applicable to profile extremely limited numbers of leukemia cells even at single cell level.

DISCUSSION

We reported herein, to the best of our knowledge, the first all-in-one and fully automated device capable of cell isolation, counting, imaging and proteomic processing for in-depth microproteomics identification and quantification down to single cell level. A size-based sorting allows rapid and quantifiable cell isolation and imaging, obviating the necessity to use external cell sorting modules such as FACS that may not be easily accessible and incompatible to handle limited input cells. In addition, such an ultra-streamlined strategy circumvents limitations in sample loss during multi-step sample transfer and surface absorption on sample vials encountered in conventional workflows and substantially improves the sensitivity in a highly reproducible manner. The customized reagents for the one-pot iProChip protocol facilitates on-chip cell lysis, protein digestion, followed by multiplexed desalting to generate peptides ready for MS analysis. Such automated preparation effectively mitigated manual handling and thus resulted in enhanced proteome coverage and superior reproducibility compared to StageTip-based workflow. Importantly, in conjunction with the DIA-MS analysis, this integrated iProChip-DIA workflow demonstrated one of the most sensitive proteome coverage of 1160 and 553 protein groups from triplicate analysis of a single PC-9 and MEC-1 cell, respectively. An interesting observation on the correlation of cell size and number of proteins was noted by the unique functionality of cell and proteome quantification in iProChip-DIA workflow. Superior analytical merits including good quantitation linearity, wide dynamic range in protein abundances, between-replicate reproducibility and low between-run missing values were systematically benchmarked. Furthermore, important druggable targets, notable biomarkers and key signaling components related to either NSCLC or BCR signaling pathways were readily identified and quantified at low-scale samples, demonstrating the applicability of iProChip-DIA for both cell biology and clinical proteomics research.

Apart from current study, size-based cell isolation of iProChip could be combined with affinity-based cell isolation such as through surface marker functionalization to enable microproteomics analysis from a sub-population of input cells. Meanwhile, the current workflow collected peptides from the iProChip and manually transferred them to conventional autosampler to load ~80% of

desalted peptides for LC-MS/Ms analysis. We believe such peptide transfer efficiency could be improved through the online interface of the iProChip directly with the autosampler on LC-MS/MS, which would reduce the sample loss and increase overall sensitivity substantially. Additional improvement in the proteomic coverage, especially for single cell, is anticipated by reducing total processing volume of current iProChip, as well as by using ultra-low-flow nanoLC to enhance chromatographic resolution and FAIMS to effectively filtering singly charged ions[21].

In addition to the demonstrated high performance in single cell analysis, iProChip is designed as a highly versatile, scalable and robust device in which existing proteomic methods can be readily integrated to fit for different applications. Conceivably, in addition to static proteome analyses, iProChip may serve as a particularly powerful platform for studying dynamic proteomic alteration upon cell stimulation, e.g. by ligand triggering, using cell imaging directly on-chip. Additional steps such as multiplex isobaric labeling and sub-proteome characterizations, can be readily integrated to achieve sensitive and multiplexed quantitative proteomics profiling under desired context. Further incorporation of peptide enrichment components in the device can be designed to extend the single cell proteomics beyond the post-translational modification of proteome. The iProChip-DIA approach is anticipated to find a variety of applications where only limited input samples are available, e.g. rare cell population from clinical specimens. We believe our new approach can open a new avenue to bringing distinct functionalities into a single miniaturized platform, to enable new ways of proteomic research that have been hampered previously, at single cell level.

METHODS

Materials and reagents

Triethylammonium bicarbonate (TEABC), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), chloroacetamide (CAA), trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA) and formic acid was bought from Honeywell Fluka and were freshly prepared in ddH_2O each day before use. LC-MS grade Acetonitrile (ACN) was purchased from Thermo Fisher Scientific. RapiGest SF surfactant (Waters, MA, USA) was dissolved in a fresh 50 mM triethylammonium bicarbonate buffer with a concentration of 0.3% (w/v), aliquoted and stored at -30°C until further use. Lys-C (MS grade) and Trypsin ((MS grade) were bought from Promega

(Madison, WI, USA). 5 μm C18 beads (300-Å pore size) were purchased from VDSpher® (VDS optilab, Chromatographie Technik GmbH, Berlin, Germany). Pierce™ BCA Protein Assay Kit (Cat. no. 23227, Thermo Scientific, USA), AZ-40XT photoresist and its developer (MicroChem, USA), SU-8 3025 photoresist and its developer (MicroChem, USA), polydimethylsiloxane (RTV-615, Momentive Specialty Chemicals, USA), chlorotrimethylsilane (Cat. No. 92360, Sigma-Aldrich, USA), hexamethyldisilazane (Cat. No. 999-9-3, Sigma-Aldrich, USA). Deionized water was purified using a Milli-Q Ultrapure Water Purification System (Millipore, Billerica, MA, USA).

Cell culture

The human lung adenocarcinoma cell line PC-9 were cultured in RPMI-1640 medium supplemented with fetal bovine serum (FBS) (10% v/v), sodium bicarbonate (2% w/v), 1mM sodium pyruvate, 100 units/ml penicillin, and 100 μg/ml streptomycin at 37 ˚C in a humidified 5% CO2 incubator. The human chronic lymphocytic leukemia cell line MEC-1 was obtained from DSMZ GmbH (Germany) and cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% heat-inactivated FBS, 100 units/ml penicillin, and 100 μg/ml streptomycin in a humidified incubator at 37 ˚C and 5% CO2.

Design and fabrication of the microfluidic chip

The integrated proteomics chip (iProChip) is a two-layer PDMS device with a top flow layer and a bottom control layer. The device layouts were designed using the AutoCAD software (Autodesk, USA). The flow layer contains a channel network which includes cell and buffer inlets, cell capture chambers, reaction vessels, desalting columns, and sample collection outlets (see chip schematics in **Supplementary Fig. 1**). Triplicate operational units are designated for 10, 50 and 100, cells. The control layer contains 34 hydraulic microvalves to control the flow layer. To account for PDMS shrinkage, the flow layer layout was expanded by 1.5 % relative to the control layer. The photo-mask for the flow layer and control layer was fabricated by the Advanced Nano/Micro Fabrication and Characterization Laboratory in Academia Sinica (Taiwan) and Taiwan Kong King Co., Ltd. respectively.

To fabricate master molds for the iProChip, regular photolithographic protocols were followed and performed on silicon wafers using a EVG-620 mask aligner[23]. Briefly, a 4-inch silicon wafer was

cleaned thoroughly using acetone, isopropanol, and DI water, followed by dehydration (105 ºC; 5 min) and hexamethyldisilazane (HMDS) coating to promote photoresist adhesion. The mold for the control layer was generated by spinning the negative photoresist SU 8-3025 (MicroChem, USA) at 4200 rpm to obtain the height of 25 μm, followed by standard photo-patterning, developing and baking protocols. Meanwhile, for the flow layer mold, valve structures were firstly fabricated using the positive photoresist AZ 40XT (MicroChem, USA) spinning at 3500 rpm to achieve the height of 25 μm, followed by standard protocols including photo-patterning, developing, baking and reflow to generate rounded features for effective valve closures[23]. Then, onto the same wafer the rest of flow layer features were generated using SU 8-3025 spinning at 4200 rpm to obtain the height of 25 μm. To obtain a 100 μm-height reaction chamber, the aforementioned SU 8-3025 protocol was repeated with a spin speed of 1100 rpm to cast additional 75 μm photoresist at the reaction chamber region using a dedicated photo-mask. Mercury match light at exposure dose of 250 mJ/cm^2 was used for patterning all features.

The iProChip was prepared by casting an optically transparent soft elastomer PDMS onto patterned master molds. The wafer with either flow or control layer layout was pre-treated with trimethylchlorosilane in a fume hood for 15 min to ensure chip features were non-sticky to PDMS, and thus allowing multiple rounds of usage. The device has a push up design, with a thick flow layer binding over a thin control layer. To make the flow layer, 60 g PDMS base and 6 g curing agent were thoroughly mixed by a mixer (Thinky ARE-310 Planetary Centrifugal Mixer) for 3 min at 2000 rpm and 1.5 min at 2200 rpm, followed by degassing for 1 h using a desiccator before pouring onto the mold. To prepare the thin control layer, 10 g PDMS base and 1 g curing agent were thoroughly mixed, followed by spinning on the control layer mold in three steps: 300 rpm for 20 s, 1800 rpm for 50 s and 0 rpm for 10 s (Laurell WS-650HZ-23NPP/UD2 Spin coater). The control layer was then allowed to level-off for 15 min on a horizontal surface, and both layers were baked in a 80 ºC oven for at least 45 min. The thick flow layer was then peeled off from the wafer, followed by cutting and hole punching (710 μm inner diameter biopsy puncher; Syneoco, USA). It was then activated using the oxygen plasma at highest RF power for 1 min (Harrick plasma cleaner PDC-001-HP), before being aligned and bound to the thin control layer using a custom stereo-microscope with independent x-, y- and z- alignment controller (Nikon-SMZ18). After baking in the 80 ºC oven overnight, the bounded chip was trimmed, peeled off and holes-punched,

before binding to a freshly plasma-treated 75x50x1 mm glass slide. The bounded chip was then placed in the 80 ºC oven for at least 48 h before following experimental use.

Preparation and characterization of the SPE column using BCA assay

The desalting columns were prepared by slurry packing with 5 μm C18 beads in acetone (5 g/ml) with an input pressure of 13 psi, which typically takes 10 to 12 min (**Supplementary Video 1**), followed by washing with methanol for 4-6 min and activated by buffer A (100% ACN + 0.1% TFA), buffer B (50% ACN + 0.1% TFA) and buffer C (0.1% TFA, *dd*H2O) for 15 min. To examine the sample recovery efficiency of the on-chip desalting column, protein quantification was performed by BCA protein assay (Thermo Scientific, USA). Briefly, two batches of digested BSA peptides were prepared through serial dilutions to make final concentrations of 1.0, 0.5, 0.25 and 0.125 μg in 10μL Buffer C. One batch was then subjected into the pre-activated SPE columns, and resulting samples were then collected, speedvac dried and resuspended in 10 μL buffer C. Both samples were individually pipetted out (10 μL each) into a 96 well plate, and 190 μL of the BCA working reagent was added to each well, followed by thorough mixing on a plate shaker for 30 s and further incubation at 37 ^{O}C for 30 min. Finally, the plate was cooled to room temperature and the absorbance for each sample was measured at 562 nm by a plate reader (EnSpire™ Plate Reader). The sample recovery was determined by plotting the recovered peptide concentration versus input peptide concentration (**Fig. 2e**).

Examination of mixing efficiency and preferential flow in the digestion chamber

The mixing efficiency of the chip was apprehended through introducing blue and yellow food dyes sequentially into the digestion chambers, followed by applying either active mixing techniques including vortex-mixing (Unico-L-VM2000) and shaker-mixing (Eppendorf-Thermomixer F), or passive mixing through diffusion processes alone. To visualize and quantify the mixing performance, time-lapsed images of the digestion chamber filled with blue and yellow dyes were recorded every 30 s for all conditions (**Supplementary Fig. S4**). Using imaging analysis (ImageJ), the standard deviation of the pixel intensity was determined for image of initial unmixed state, σ_o, and for individual time-lapsed images taken throughout the process, σ. The mixing efficiencies were then quantified as a non-dimensional parameter, Relative Mixing Index (RMI), according to the following equation:[28]

$$RMI = 1 - \frac{\sigma}{\sigma_o} = 1 - \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i - \langle I \rangle)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_{oi} - \langle I \rangle)^2}}$$

where, N is the total number of pixels, I_{oi} is the local pixel intensity in the unmixed state, I_i is the local pixel intensity in mixed state, and $\langle I \rangle$ is the average intensity.

The fluid dynamics, especially the possibility of preferential flow as a result of (1) the circular shape of the digestion chamber, (2) dominated laminar flow inside the microfluidic channels and (3) non-negligible flow resistance from the downstream integrated desalting column, were carefully examined to determine the optimal condition for subsequent sample retrieval. The time and corresponding pressures needed to push the colored fluid through C18 beads-packed desalting columns were found to be 30, 20, 13, 9 and 8 min for 9, 10, 11, 12 and 13 psi respectively (**Supplementary Fig. 5a**). On the other hand, for an empty desalting column it took less than 35 s for dye molecules to be transferred for pressure larger than 4 psi, confirming there was no preferential flow as a result of circular geometry of the digestion chamber (**Supplementary Fig. 5b**). Based on such characterization, input pressure of 11 psi was used in the proteomics workflow for transferring digested peptides to the packed SPE column during the final sample clean-up.

Microproteomics workflow in iProChip

The PMDS chip was hooked up to the control system via 34 stainless steel connectors attached to tygon tubings and then mounted onto an inverted microscope (Nikon-ECLIPSE-Ts2). Working pressures of 28 to 30 psi was used to operate the control layer. To minimize non-specific absorption during proteomic processing, all modules of the iProChip except the desalting column were freshly coated with 0.1 % BSA for 1 h, followed by PBS rinsing for 10 min and dried under the nitrogen stream. Afterwards, desalting columns were packed with C18 beads (0.01mg/ml in acetone), and cell capture chambers were degassed using PBS solution in order to achieve well dispersed cell flow into the chambers.

The proteomic processing steps start by introducing a cell suspension (5 x 10^5 cell/ml) into cell capture chambers under flow pressure of 4 psi, with which quantifiable number of cells in the range of 1, 5, 10, 50 and 100 were prudently trapped in a cell trapping chamber by controlling the

injection time through real-time imaging. In the second step, cocktail lysing buffer (42 nL) consisting of 0.3% RapiGest, 10mM TCEP, 40 mM CAA was infused to individual cell capture chamber, followed by incubation for 30 min at 75 ^{O}C on a plate shaker (Eppendorf-Thermomixer F; 400 rpm) (**Supplementary Fig. 6**). In the third step, Lys C (42 nL, protein/Lysine-C 20:1 w/w) and trypsin (42 nL, protein/trypsin 10:1 w/w) were sequentially infused into individual digestion vessels, followed by further incubation at 40 ^{O}C for 16 h on a plate shaker (400 rpm). Finally, 25% FA (42 nL, final 5% v/v) was infused to the individual chamber and incubated for 55 min at 40 ^{O}C to quench the enzymatic digestion. The peptide clean-up was carried out by the SPE columns preconditioned and equilibrated with buffer A, buffer B and buffer C running for 15 min each. Then, the processed peptides were pushed from the digestion vessels to the activated SPE columns by buffer C at 11 psi for 15 min of desalting. Finally, buffer B was passed through SPE columns to elute the peptides into lo-binding vials, speedvac dried prior to subsequent LC-MS/MS analysis.

LC-MS/MS analysis

The Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) coupled with an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific) was used for LC-MS/MS analysis in this study. The desalted peptides were resuspended to 5ul in the loading buffer (0.1% formic acid) spiked with iRT peptides (Biognosys, Schlieren, Switzerland) and 4µl was loaded to autosampler for LC-MS/MS analysis. The nanoflow Ultimate 3000 UHPLC (Thermo Fisher Scientific) with a capillary C18 column (Waters, nanoEase, 130Å, 1.7 µm, 75 µm X 250 mm) was employed for peptide separation at 300 nL/min using buffer A (0.1% FA in water) and buffer B (0.1% FA in ACN). The peptides were separated through gradient from 3% to 25% ACN in 137.5 min, followed by 4 min increase to 40% and 2 min increase to 95% ACN. After washout for 5 min at 95% ACN, the C18 column was re-equilibrated at 1% ACN for 10 min. The MS instrument was operated in the positive ion mode with spray voltage set to 1.75 kV, RF lens level set at 30%, and ion transfer tube heated at 305 °C. For DDA mode, top N multiply charged precursors were automatically isolated and fragmented according to their intensities within the cycle time of 3 s. Intensity threshold was set to 8E3. Full MS was scanned at a resolution of 120000 with an automatic gain control (AGC) target of 1E6 and a max injection time of 50 ms. Mass range was set to 375-1500 m/z and isolation width for MS/MS analysis was set to 1.4 m/z with advanced peak determination. Normalized collision energy (CE) of high-energy collision dissociation (HCD) was set to 30%. MS/MS was

scanned in orbitrap at a resolution of 120000 with an AGC target of 1.25E5 and a max injection time of 254 ms. For DIA mode, the full MS resolution, AGC target, and max injection time are the same with DDA mode. The mass range for DIA MS/MS analysis was set to 400-800 m/z and overall 40 scan events of 10 m/z isolation window were employed with an overlap of 1 m/z. The MS/MS scan was performed in HCD mode using the following parameters: normalized CE = 30%; resolution = 30000; AGC target = 4E5; max injection time =54 ms. All data were acquired in profile mode using positive polarity.

Spectral library construction

A set of project specific spectra libraries (both at large-scale and small-scale) were constructed for lung cancer cell line (PC-9), and human chronic lymphocytic leukemia cell line (MEC-1) processed in iProChip or in vial-based processing. The raw files for library generation were acquired in both data-dependent as well as data-independent acquisition mode to obtain project specific hybrid libraries. Large-scale libraries were constructed by using bulk samples (~1ug peptide, n=10 raw files for PC-9) while small-scale libraries were generated using ~10 cells processed through iProChip (n=9 raw files for PC-9) and aliquots of 1.5 ng (~10 cells, n=13 raw files for PC-9) obtained from dilution. To further maximize the identification results, another small-scale library using 5 and 50 cells DIA runs (n=6 raw files) processed through iProChip was also constructed. Similarly, for MEC-1 cells, large-scale libraries from bulk (n=8 raw files) and small-scale one of ~10 cells (n=12 raw files) were constructed. Protein identification was performed using Spectronaut Pulsar software (v13; Biognosys, Switzerland)[31] [38] filtering at 1% FDR PSM, peptide and protein level. Trypsin was selected as a digestion enzyme with maximum two miscleavage sites. The minimum and maximum allowed peptide length in search space were 7 and 52, respectively. Variable modifications of acetylation on protein N-terminus and oxidation on methionine were included, while carbamidomethylation of cysteine was set as a fixed modification. For database search, the SwissProt/UniProt human proteome database (2015_12 release, Homo Sapiens= 20,193 entries) with inclusion of 11 iRT peptide sequence was used. Project-specific spectral libraries were generated using standard parameters in Spectronaut. Briefly, Normalized retention time was obtained using segmented regression to determine iRT in each run by the precision iRT function. The six most intense fragment ions were included with iRT retention time normalization. Fragment ions of minimum m/z 300, maximum m/z 1800, minimal relative

intensity of 5% were included. Fragment ions with less than three amino acid residues were not considered.

Data analysis

The DIA raw files were analyzed using Spectronaut software (version 13.11200127.43655) against home-built project spectra libraries as well as in library-free mode (directDIA) using standard settings. For library-free strategy (direct DIA), protein identification from DIA dataset was performed by database search against the SwissProt/UniProt human proteome database (Homo Sapiens: 20,193 entries) using Pulsar search engine in Spectronaut. For library-based strategy using our project spectra libraries, protein identification was performed using Spectronaut Pulsar software[31] filtering at 1% FDR PSM, peptide and protein level described in the previous session (See ***Spectral library construction***). For protein quantitation, peak area at MS2 level was calculated using top 3 peptides per protein and minor peptide grouping was adapted based on striped sequence. For comparison, the DDA raw files were analyzed with Maxquant software (version 1.5.6.5)[37] using standard settings for Orbitrap MS and LFQ protein quantification. The first search and main search peptide tolerance were set as 20 and 4.5 ppm, respectively. The protein and peptide were both filtered at 1% FDR. Variable modifications of acetylation on protein N-terminus and oxidation on methionine were included, while carbamidomethylation of cysteine was set as a static modification. The SwissProt/UniProt human proteome database (downloaded on 2015/12/15, Homo Sapiens: 20,193 entries) with iRT peptides sequence was used. The pathway analyses were performed using the KEGG Mapper from the KEGG online database (https://www.kegg.jp/kegg/mapper.html). The kinome tree was drawn using KinMap online tool from Kinhub database platform (http://www.kinhub.org/kinmap).

Data and code availability

The mass spectrometry raw data sets, spectral libraries, and Spectronanut quantification outputs were deposited in Japan ProteOme Standard Repository (jPOST; http://repository.jpostdb.org/) and can be accessed through ProteomeXchange (http://www.proteomexchange.org/). The accession number is JPST000971 for JPOST and PXD023325 for ProteomeXchange and the data can be accessed through https://repository.jpostdb.org/preview/13812480635feaaf61067ae with access key 5844. Code and scripts for chip control are available upon request.

ACKNOWLEDGEMENTS

We would like to thank Jing Lin for developing an initial chip control program, Yi-Wen Fang and Pin-Lian Jiang for assisting the PC-9 cell culture, Dr. Ying Li for SEM measurements and Pei-Yi Lin for assisting LC-MS/MS measurement. This work was funded by the Ministry of Science and Technology (Grant MOST 107-2113-M-001-032-MY3; MOST 107-2113-M-001-023-MY3) and Academia Sinica (AS-TP-108-ML06; AS-iMATE-108-21). The master molds for the iProChip were fabricated at the Advanced Nano/Micro Fabrication and Characterization Laboratory at Institute of Physics, Academia Sinica, Taiwan. The LC-MS/MS data acquisition and analysis by Orbitrap ECLIPSE mass spectrometer was performed in the Mass Spectrometry Facility located at the National Biotechnology Research Park (NBRP), Academia Sinica, Taiwan.

AUTHOR CONTRIBUTION

S.T.G. and A.A.S. performed experiments, acquired and analyzed the data; A.A. S., R.B.K. and E.S.-W. C. generated spectral libraries, performed LC-MS/MS measurement and analyzed the data; B.E. T.A. and K.-I.L. provided reagents; H.-L.T. and Y.-J.C. conceived and supervised the work. S.T. G., A.A.S., Y.-J. C. and H.-L. T. wrote the initial draft. All authors commented and contributed to the final editing of the manuscript.

COMPETING INTEREST

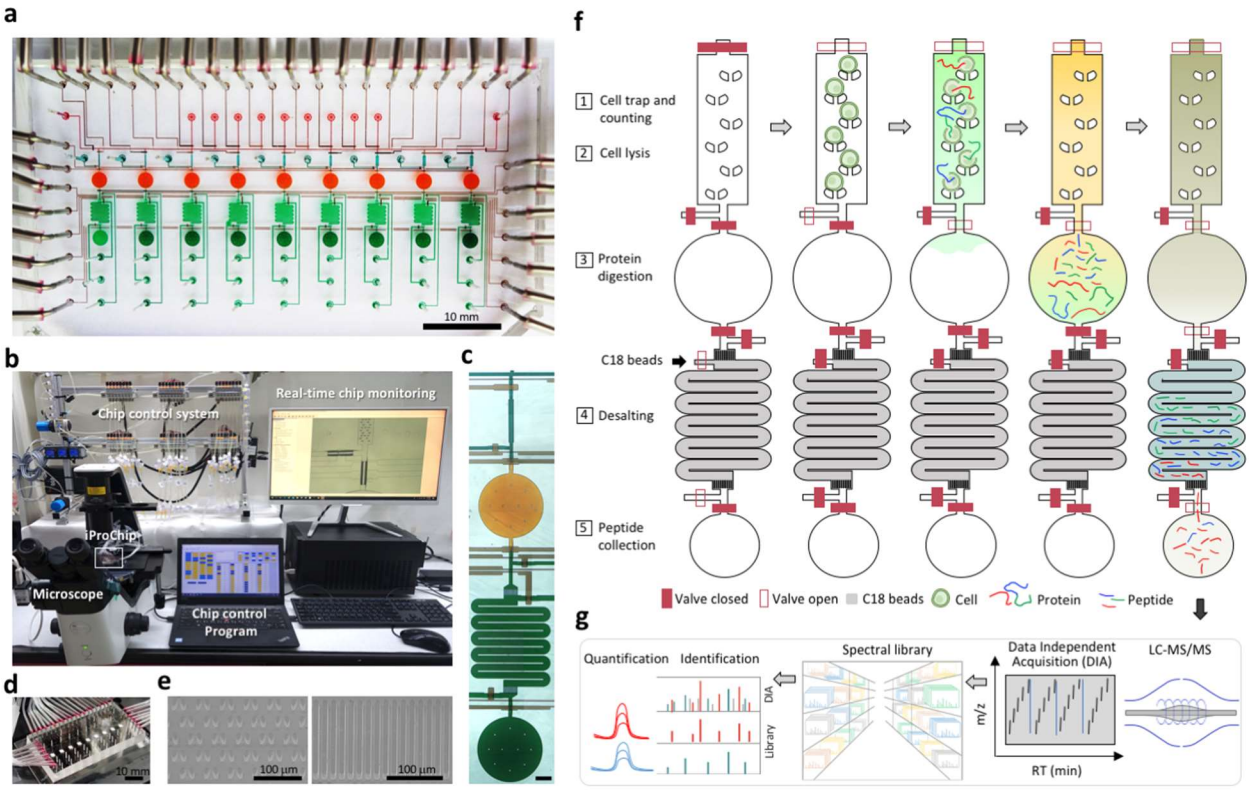
The authors declare no competing financial interests.

REFERENCES

1. Stuart, T. & Satija, R. Integrative single-cell analysis. *Nat Rev Genet* **20**, 257-272 (2019).
2. Chappell, L., Russell, A.J.C. & Voet, T. Single-Cell (Multi)omics Technologies. *Annu Rev Genomics Hum Genet* **19**, 15-41 (2018).
3. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355 (2016).
4. Schubert, O.T., Rost, H.L., Collins, B.C., Rosenberger, G. & Aebersold, R. Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* **12**, 1289-1294 (2017).

5. Bandura, D.R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem* **81**, 6813-6822 (2009).
6. Shi, Q. et al. Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells. *Proc Natl Acad Sci U S A* **109**, 419-424 (2012).
7. Sinkala, E. et al. Profiling protein expression in circulating tumour cells using microfluidic western blotting. *Nat Commun* **8**, 14622 (2017).
8. Labib, M. & Kelley, S.O. Single-cell analysis targeting the proteome. *Nature Reviews Chemistry* **4**, 143-158 (2020).
9. Angel, T.E. et al. Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem Soc Rev* **41**, 3912-3928 (2012).
10. Slavov, N. Single-cell protein analysis by mass spectrometry. *Curr Opin Chem Biol* **60**, 1-9 (2020).
11. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**, 161 (2018).
12. Gillet, L.C., Leitner, A. & Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem (Palo Alto Calif)* **9**, 449-472 (2016).
13. Couvillion, S.P. et al. New mass spectrometry technologies contributing towards comprehensive and high throughput omics analyses of single cells. *Analyst* **144**, 794-807 (2019).
14. Kulak, N.A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods* **11**, 319-324 (2014).
15. Sielaff, M. et al. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J Proteome Res* **16**, 4060-4072 (2017).
16. Hughes, C.S. et al. Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **10**, 757 (2014).
17. Chen, Q., Yan, G., Gao, M. & Zhang, X. Ultrasensitive Proteome Profiling for 100 Living Cells by Direct Cell Injection, Online Digestion and Nano-LC-MS/MS Analysis. *Anal Chem* **87**, 6674-6680 (2015).
18. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nat Commun* **9**, 882 (2018).
19. Li, Z.Y. et al. Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. *Anal Chem* **90**, 5430-5438 (2018).
20. Leipert, J. & Tholey, A. Miniaturized sample preparation on a digital microfluidics device for sensitive bottom-up microproteomics of mammalian cells using magnetic beads and mass spectrometry-compatible surfactants. *Lab Chip* **19**, 3490-3498 (2019).
21. Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies >1000 protein groups per mammalian cell. *Chemical Science* (2021).
22. Whitesides, G.M. The origins and the future of microfluidics. *Nature* **442**, 368-373 (2006).
23. Unger, M.A., Chou, H.P., Thorsen, T., Scherer, A. & Quake, S.R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113-116 (2000).

24. Sackmann, E.K., Fulton, A.L. & Beebe, D.J. The present and future role of microfluidics in biomedical research. *Nature* **507**, 181-189 (2014).
25. Gillet, L.C. et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **11**, O111 016717 (2012).
26. Gomez-Sjoberg, R., Leyrat, A.A., Pirone, D.M., Chen, C.S. & Quake, S.R. Versatile, fully automated, microfluidic cell culture system. *Anal Chem* **79**, 8557-8563 (2007).
27. Sarioglu, A.F. et al. A microfluidic device for label-free, physical capture of circulating tumor cell clusters. *Nat Methods* **12**, 685-691 (2015).
28. An Le, N.H. et al. Ultrafast star-shaped acoustic micromixer for high throughput nanoparticle synthesis. *Lab Chip* **20**, 582-591 (2020).
29. Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J Proteome Res* **15**, 1116-1125 (2016).
30. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
31. Bruderer, R. et al. Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol Cell Proteomics* **14**, 1400-1410 (2015).
32. How Can Systems Biology Test Principles and Tools Using Immune Cells as a Model? *Cell Systems* **6**, 146-148 (2018).
33. Eid, S., Turk, S., Volkamer, A., Rippmann, F. & Fulle, S. KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics* **18**, 16 (2017).
34. Burger, J.A. & Wiestner, A. Targeting B cell receptor signalling in cancer: preclinical and clinical advances. *Nat Rev Cancer* **18**, 148-167 (2018).
35. Myers, D.R., Zikherman, J. & Roose, J.P. Tonic Signals: Why Do Lymphocytes Bother? *Trends Immunol* **38**, 844-857 (2017).
36. Rieckmann, J.C. et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol* **18**, 583-593 (2017).
37. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).



682

683 **Figure 1 | Schematics of integrated proteomic chip and ultra-streamlined workflow for**
684 **microproteomics.** (a) A bright-field image of the integrated proteomic chip (iProChip), where cell
685 capture chambers (cyan), reaction vessels (orange), on-chip SPE columns (green), sample
686 collection ports (dark green) and control layers (brown) are shown. (b) The entire system setup for
687 iProChip operation. (c) A close-up view of a single operation unit. Scale bar: 300 μm. (d) A ready-
688 to-use iProChip mounted on the microscope. (e) SEM images of cell capturing pillars (left) and
689 C18 filters in the SPE column (right). (f) Operation procedures of iProChip for streamlined sample
690 preparation, including (1) cell trapping, imaging and counting, (2) cell lyse, (3) protein digestion,
691 (4) desalting and (5) peptide collection. (g) Proteomic analysis using data-independent-acquisition
692 mass spectrometry and spectral library search.

693

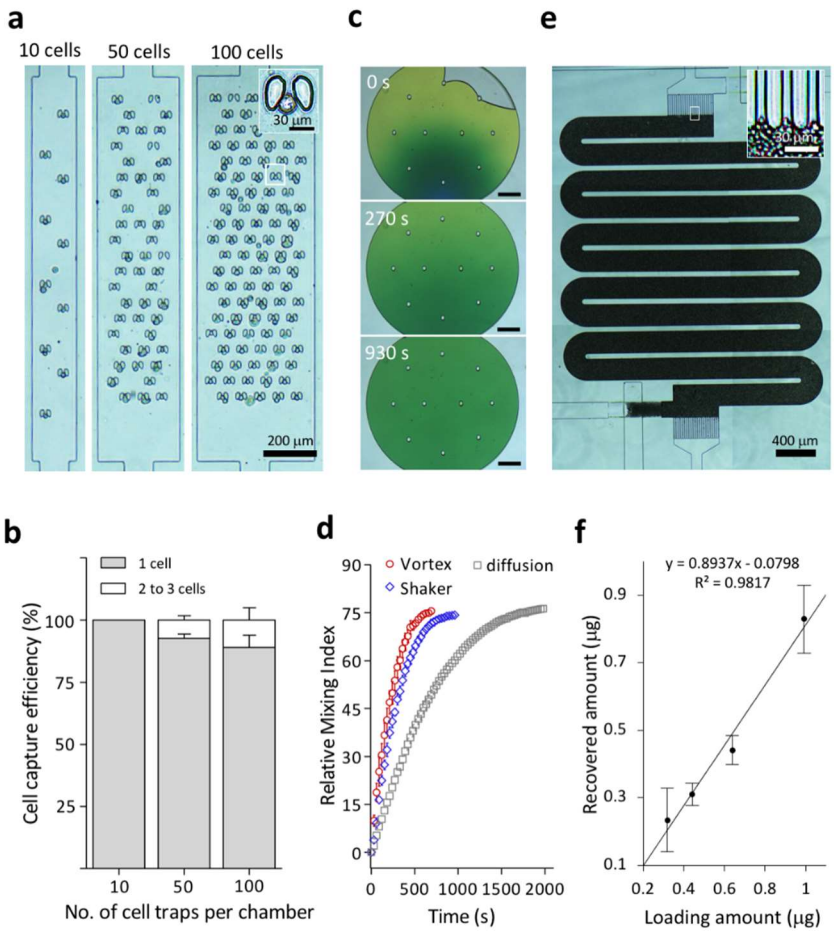


Figure 2 | Performance characterization of the iProChip. (a) Bright-field images of non-small lung cancer PC-9 cells captured in 10, 50 and 100 cell chambers. Top right: a zoom-in image of a trapped cell. (b) Characterization of the cell capture efficiency for different capture chambers by 3 independent experiments. (c) Representative time-lapsed images of a reaction vessel filled with green dye during mixing-by-shaking characterization. Scale bars: 300 μm. (d) Comparison of mixing efficiency in the reaction vessel. Error bars: s.e.m. (n = 3 independent experiments). (e) A bright-field image of the SPE column packed with C18 beads. Top right: a close-up view near the C18 filter. (f) Desalting recovery efficiency of the on-chip SPE column. Error bars: s.d. (n = 3 independent experiments).

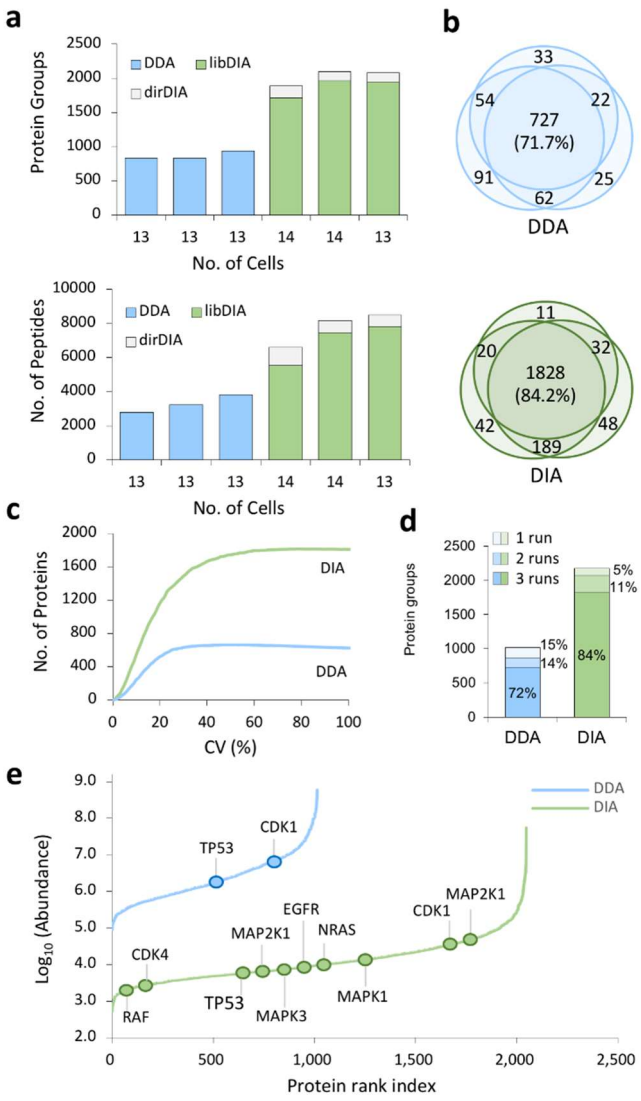


Figure 3 | Comparison of identification coverage and quantitation performance of proteomics profiling of PC-9 cells by DIA and DDA methods. (a) Comparison of protein groups and peptides identified by triplicate analysis using DDA and DIA. (b) Overlap of protein groups identified by DDA and DIA. (c) Distribution of coefficient of variation (CV%) for quantified protein groups by DDA and DIA. (d) Evaluation of missing values (%) of proteins identified and quantified in triplicate LC-MS/MS runs by DDA and DIA. (e) Assessment of dynamic range based on protein abundance rank and annotation of selected proteins related to lung cancer.

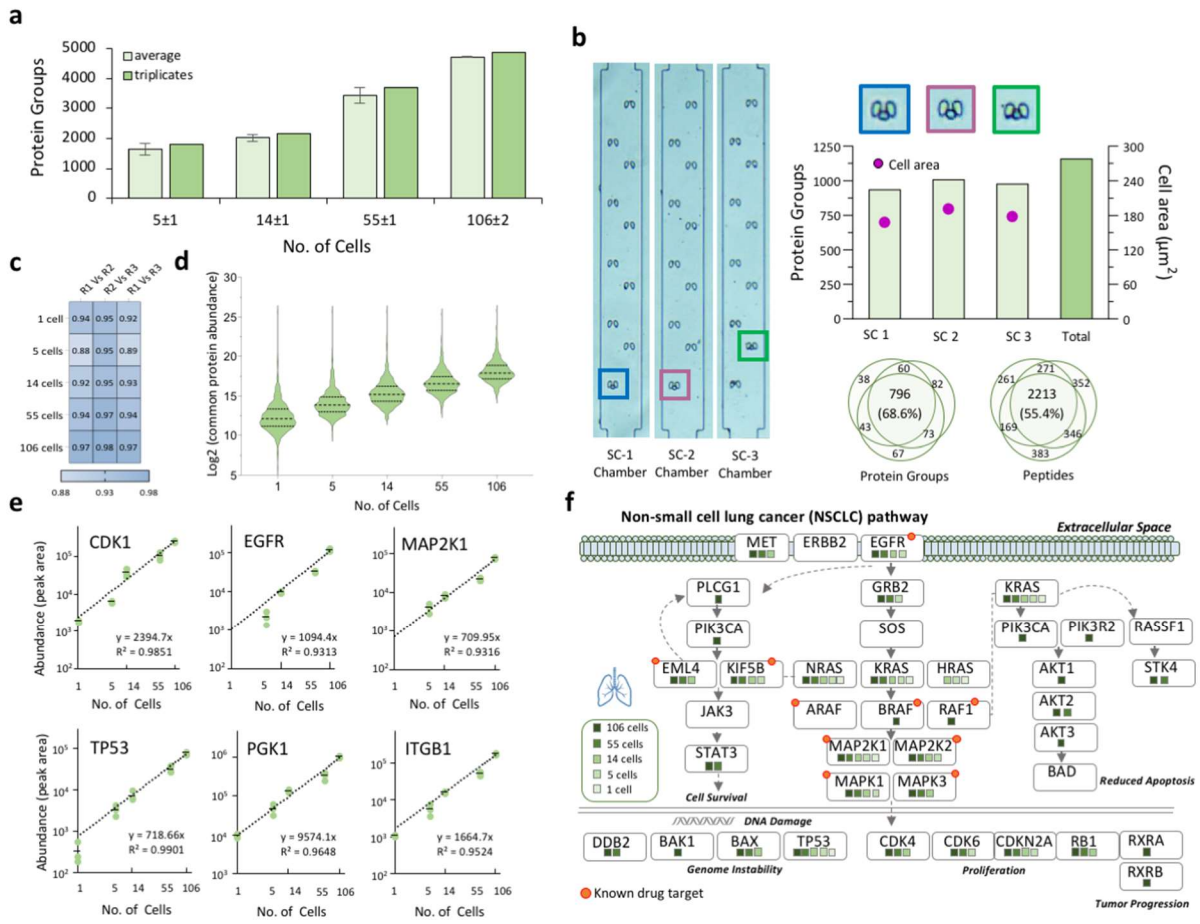
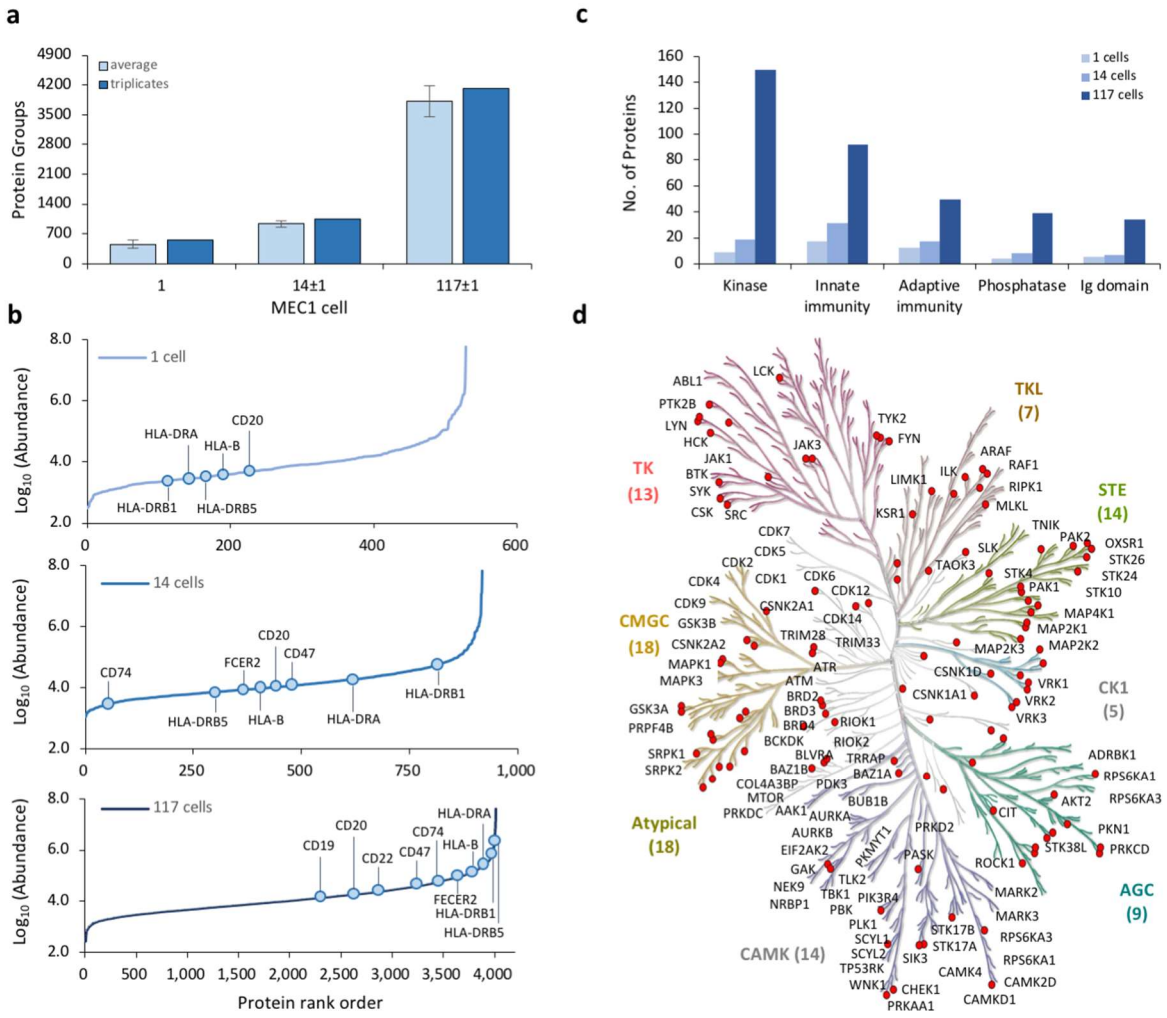


Figure 4 | Analytical performance in sensitivity, reproducibility, and quantitation in proteomic profiling of lower number of cells by iProChip-DIA. (a) Identification summary of protein groups across different PC-9 cell numbers by iProChip-DIA. Error bars: s.d. (n = 3 independent experiments). (b) Single PC-9 cell trapping using 10 cell capture chambers and corresponding cell image, cell size, and triplicate analysis results of identified protein groups and coverage. (c) A heatmap showing reproducibility of protein abundances obtained among different cell numbers. (d) Distributions of total protein abundances in commonly identified proteins across different cell numbers. (e) Representative examples of lung cancer related proteins showing quantitation of protein abundance calculated from peak area. (f) Identification coverage of proteins within the NSCLC pathway under different cell numbers.



762 **Figure 5 | Application of iProChip-DIA for proteomic profiling of MEC-1 cells.** (a)
763 Identification summary of protein groups across different MEC-1 cell numbers by iProChip
764 coupled to DIA-MS. Error bars: s.d. (n = 3 independent experiments). (b) Assessment of dynamic
765 range based on protein abundance rank and annotation of selected proteins related to immune
766 cancer markers. (c) Enrichment of immune-related and other functional classes against UniprotKB
767 database. (d) Kinase tree for mapping 114 kinases from the total cell numbers.
768