

RESEARCH

Open Access



Optimized data-independent acquisition approach for proteomic analysis at single-cell level

Yuefan Wang¹, Tung-Shing Mamie Lih¹, Lijun Chen¹, Yuanwei Xu¹, Morgan D. Kuczler², Liwei Cao¹, Kenneth J. Pienta², Sarah R. Amend² and Hui Zhang^{1*}

Abstract

Background: Single-cell proteomic analysis provides valuable insights into cellular heterogeneity allowing the characterization of the cellular microenvironment which is difficult to accomplish in bulk proteomic analysis. Currently, single-cell proteomic studies utilize data-dependent acquisition (DDA) mass spectrometry (MS) coupled with a TMT labelled carrier channel. Due to the extremely imbalanced MS signals among the carrier channel and other TMT reporter ions, the quantification is compromised. Thus, data-independent acquisition (DIA)-MS should be considered as an alternative approach towards single-cell proteomic study since it generates reproducible quantitative data. However, there are limited reports on the optimal workflow for DIA-MS-based single-cell analysis.

Methods: We report an optimized DIA workflow for single-cell proteomics using Orbitrap Lumos Tribrid instrument. We utilized a breast cancer cell line (MDA-MB-231) and induced drug resistant polyaneploid cancer cells (PACCs) to evaluate our established workflow.

Results: We found that a short LC gradient was preferable for peptides extracted from single cell level with less than 2 ng sample amount. The total number of co-searching peptide precursors was also critical for protein and peptide identifications at nano- and sub-nano-gram levels. Post-translationally modified peptides could be identified from a nano-gram level of peptides. Using the optimized workflow, up to 1500 protein groups were identified from a single PACC corresponding to 0.2 ng of peptides. Furthermore, about 200 peptides with phosphorylation, acetylation, and ubiquitination were identified from global DIA analysis of 100 cisplatin resistant PACCs (20 ng). Finally, we used this optimized DIA approach to compare the whole proteome of MDA-MB-231 parental cells and induced PACCs at a single-cell level. We found the single-cell level comparison could reflect real protein expression changes and identify the protein copy number.

Conclusions: Our results demonstrate that the optimized DIA pipeline can serve as a reliable quantitative tool for single-cell as well as sub-nano-gram proteomic analysis.

Introduction

Cells from the same living organism have a similar genomic background, which are eventually differentiated into diverse cell types in different tissues or organs via the expression of different genes to proteins leading to cellular heterogeneity. Although the rapid development of genomic and transcriptomic methods made it possible to analyze genomic and transcriptomic alterations

*Correspondence: huizhang@jhu.edu

¹ Department of Pathology, Johns Hopkins University, Baltimore, MD 21287, USA

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of cellular heterogeneity at single-cell level [1, 2], the absence of protein amplification techniques hampered single-cell proteomic analysis. Originally, single-cell proteomic studies were limited to the detection of less than 15 targeted proteins from single mammalian cell by means of flow cytometry [3, 4], mass cytometry [5], and single-cell western blotting [6]. After decades of development, mass spectrometry (MS), a primary tool for analyzing proteome and protein post-translational modifications (PTMs) from bulk samples, was applied to the first hypothesis-free mammalian single-cell proteomic analysis known as Single Cell Proteomics by Mass Spectrometry (SCoPE-MS) [7] in 2018. SCoPE-MS utilizes isobaric tandem mass tags (TMT) to label peptides from single cells along with a carrier sample containing highly excessive number of peptides to increase the detection of peptide fragment ions, especially for low abundant peptides, during tandem mass spectrometry analysis (MS/MS). Several works have followed the aforementioned concept and greatly expended the single-cell proteomics field [8]. However, TMT carrier-based methods have the limitations that the data quality and quantitation are highly dependent on the extremely imbalanced carrier ratio and instrument dynamic [9].

Data independent acquisition (DIA)-MS is considered as a consistent proteomic analytical method that allows the fragmentation of all the precursor ions within selected isolation m/z range generating comprehensive MS/MS spectra [10]. DIA-MS can provide reproducible global quantitative data with minimal cost [10–13]. Various software tools have been developed [14, 15] to analyze DIA data that can be classified into spectral library-based approach [16] and library-free approach [17]. The spectral library-based DIA analysis is a peptide-centric method, which usually requires building spectral libraries either using corresponding DDA and/or DIA data from the same samples or using pre-built publicly available spectral libraries. However, the sample types and experimental conditions should be taken into consideration while building spectral libraries, especially when using external sources. Moreover, the spectral library size has direct impacts on DIA data search results [18], thus, an inappropriate library size would compromise the identification results [19, 20]. On the other hand, the library-free DIA analysis is a spectrum-centric method. There are several tools to conduct library-free analysis, including DIA-Umpire [17] and directDIA embedded in Spectronaut [21]. Library-free approach detects or deconvolutes chromatographic features of precursor-fragment ion groups to generate pseudo-MS/MS spectra, which allows multiple DIA raw files to be processed together. While generating pseudo-MS/MS spectra from one or more DIA raw files, an internal spectral library

is constructed like the pre-built library from DDA data/external sources. Therefore, the internal spectral library size could influence the DIA study. Nonetheless, library-free approach only relies on DIA data itself, thus, it is highly sample-specific compared to spectral library-based approach, which is more suitable for single cell global proteome identification.

In this study, we evaluated the performances of DIA-MS approach for the analysis at the nano and sub-nanogram peptide levels using MDA-MB-231 cancer cells and drug resistant PACCs induced by platinum or docetaxel treatment [22] to optimize the DIA-MS workflow for single-cell level proteomics. PACCs are a large cancer cell state with high genome content that are induced by stress and treatment. We evaluated the DIA performances in different liquid chromatography (LC), MS/MS, and data analysis settings. We found that a 15-min short LC gradient and library-free approach via directDIA for data analysis allowed the identification of 3260 and 1530 proteins from 2 ng (corresponding to 10 PACCs) and 0.2 ng (corresponding to a single PACC) peptides with good reproducibility, respectively. Therefore, the results demonstrate that our optimized DIA pipeline can serve as a reliable quantitative tool for single-cell proteomic analysis.

Experimental section

Cell culture and cell counting

All cells were maintained in RPMI-1640 (Gibco), supplemented with 10% fetal bovine serum (FBS) and 1% penicillin streptococcus, and cultured in standard tissue culture conditions (37 °C, 5% CO₂). MDA-MB-231 was originally obtained from ATCC. Cell lines are routinely authenticated and tested for mycoplasma.

Parental MDA-MB-231 samples were prepared by seeding 1×10^6 cells and incubating for 24 h. Adherent cells were washed with PBS prior to being lifted with Cell Dissociation Buffer (Thermo Fisher Scientific). Lifted cells were re-suspended in 20 mL of culture medium and applied to a primed 15 μ M pluriStrainer® (The Cell Separation Company) and the flow-through retained. 1×10^6 cells were collected and washed twice in PBS. Following the final wash, supernatant was removed and the pellet was snap-frozen and stored at -80 °C.

Drug-induced PACC samples were prepared by seeding 1×10^6 MDA-MB-231 cells. After 24 h incubation, cultures were treated with IC₅₀ cisplatin or docetaxel for 72 h. After 72 h, surviving adherent cells were washed with PBS and lifted with Cell Dissociation Buffer (Thermo Fisher Scientific). Lifted cells were re-suspended in 20 mL of culture medium and filtered through primed 15 μ M pluriStrainer® (The Cell Separation Company). The flow-through from each filter was discarded and the

cells caught by the filter were harvested by flipping the filter upside-down and washing with 15 mL media. The PACC sample was pelleted at $1000\times g$ for 5 min, counted, and washed twice in PBS. Following the final wash, supernatant was removed and the pellet was snap-frozen and stored at -80°C .

Sample preparation

One million PACCs and parental MDA-MB-231 cancer cells (three samples from each cell type) were lysed in 60 μL lysis buffer containing 8 M urea as described in CPTAC protocol [23]. Briefly, cell lysates were centrifuged at $16,000\times g$ for 12 min at 4°C and protein concentrations were determined by PierceTM BCA protein assay (Thermo). The samples were reduced by 6 mM dithiothreitol for 1 h at 37°C and then alkylated by 12 mM iodoacetamide for 45 min at room temperature in dark place. The samples were diluted to 2 M urea concentration with 50 mM Tris buffer (pH 8.0). In the 2 M urea buffer, the samples were digested with Lys-C (Wako) at 1 mAU: 10 mg enzyme to substrate ratio for 2 h at room temperature, followed by the addition of trypsin (Promega) at the same ratio for overnight digestion at room temperature. After the digestion, the mixtures were acidified by 50% formic acid to get 1% formic acid as final concentration with $\text{pH}<3$. The digested peptides were desalted on C18 stage tips (3 M) and dried with Speed-Vac (Thermo). The dried peptides were redissolved in 3% acetonitrile with 0.1% formic acid and used NanoDropTM (Thermo) to determine the peptide concentration. Based on cell count and peptide yields (Additional file 1: Table S1), each PACC sample was 0.2 ng peptides per cell and each parental MDA-MB-231 sample was about 0.05 ng peptides per cell. Starting from 1 μg aliquoted peptides, a serial of dilution was performed to obtain 100 cells, 10 cells and 1 cell populations for two different sizes of cells corresponding to 20 ng, 2 ng, and 0.2 ng digested peptides for PACC, and 5 ng, 0.5 ng, and 0.05 ng digested peptides for parental cells. All the injections were spiked in $0.5\times \text{iRT}$ peptides (Biognosys) to calibrate the internal retention time.

NanoLC-MS/MS analysis

The aliquoted peptides equivalent to 100 cells, 10 cells, and a single cell of the PACCs and parental MDA-MB-231 cells were analyzed using two different LC gradient time, 15 min and 120 min. All the samples, from 1 μg to 0.05 ng of peptides, were separated by EASY-nLCTM 1200 instrument (Thermo) with hand-packed analytical column (75 μm i.d. \times 26.5 cm length packed with Repro-Sil-Pur 120 C18-AQ 1.9 μm beads) and Picofrit 10 μm opening (New Objective). The column was heated to 50°C with Nanospray FlexTM Ion Sources (Thermo). The

elution flow rate was 200 nL/min with 0.1% formic acid in 97% H_2O and 3% CH_3CN as buffer A, and 0.1% formic acid in 90% CH_3CN and 10% H_2O as buffer B. Peptides were separated using 4–30% buffer B in 15 min gradient and 7–30% buffer B in 120 min gradient. All the samples were analyzed via Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) and the parameters for the DIA method are as follows: resolution at 120,000, mass range of 350–1650 m/z , and maximum injection time of 60 ms for MS1 scan; resolution at 30,000, HCD collision energy of 34%, mass range of 300–1600 m/z , and maximum injection time of 80 ms for MS2 scan. For both MS1 and MS2, RF Lens 30% and normalized AGC Target 250% were applied. Total of 30 DIA raw files were acquired using two LC gradient settings from the aliquoted peptide samples of 100 cells, 10 cells and a single cell population of the PACC and the parental cancer cells (Additional file 1: Table S2).

DIA data analysis

The single-cell level DIA runs of global proteome were analyzed via library-free directDIA approach embedded in Spectronaut (version 14.10, Biognosys) with precursor and protein Qvalue cutoff at 1%. For the bulk sample analyses, five raw files (one from each of PACC samples and one from each of parental cell line samples) acquired from 1 μg injections were analyzed together in one directDIA search. For the single-cell analyses, the directDIA searches were conducted on five co-searching groups that each with different combination of raw files. Thus, each co-searching group generated an internal library containing different number of precursors (i.e., different library size). The five co-searching groups of single-cell analyses are as follows (Additional file 1: Table S3): one single-cell raw file (271 precursor from 0.05 ng peptides of a single MDA-MB-231 cell, referred to **GS-1r_M**); 2258 precursors from 0.2 ng peptides of a single PACC cell, referred to as **GS-1r_P**), 10 raw files of all the single-cell injections from two LC gradients (2687 precursors, referred to as **GS-10r**), 16 raw files from the combination of **GS-10r** and six injections (two LC gradients) from 0.5 ng peptides (10 cells) of MDA-MB-231 cells (5787 precursors, referred to as **GS-16r**), 20 raw files from the combination of **GS-10r** and ten-cell injections (two LC gradients) of PACC samples (16,496 precursors, referred to as **GS-20r**), and all 30 raw files (47,374 precursor, referred to as **G-30r**).

For the ten-cell analyses, the directDIA searches were conducted on five co-searching groups distinct from the co-searching groups of the single-cells as follows (Additional file 1: Table S4): one ten-cell raw file (3803 precursor from 0.5 ng peptides of a MDA-MB-231 cell sample, referred to as **G10c-1r_M**); 10,756 precursors from 2 ng

peptides of a PACC sample, referred to as **G10c-1r_P**), 10 raw files of all the ten-cell injections from two LC gradients (17,118 precursors, referred to as **G10c-10r**), 16 raw files by combining **G10c-10r** and all 100-cell injections of MDA-MB-231 cells (26,191 precursors, referred to as **G10c-16r**), all 30 raw files (47,374 precursor, referred to as **G-30r**), and 25 raw files as a combination of **G10c-10r**, all 100-cell injections of PACC samples and five 1 μg peptide injections (88,012 precursors, referred to as **G10c-25r**).

The PTM analyses were conducted by searching global DIA data (nano-gram and single-cell levels) against the pre-built PTM spectral libraries.

PTM spectral library generation

To analyze PTMs, the spectral libraries for Fig. 4a were generated from patient-derived xenograft (PDX) samples, phosphopeptide spectral library was built using IMAC-enriched phosphorylation data of PDX samples (followed CPTAC standard protocol [24]), acetylation and ubiquitination spectral libraries were constructed using the data of antibody-enriched PDX samples [25, 26]. All PTM spectral libraries from PDX samples were built by single DIA run, individually. Additionally, three phosphopeptide spectral libraries with different sizes (~42,000 precursors, ~87,000 precursors and ~142,000 precursors) were built using IMAC-enriched phosphorylation data of tumor tissues from clear cell renal cell carcinoma (ccRCC) tissues [24]. All the PTM libraries were built by using Pulser embedded in Spectronaut.

Results and discussion

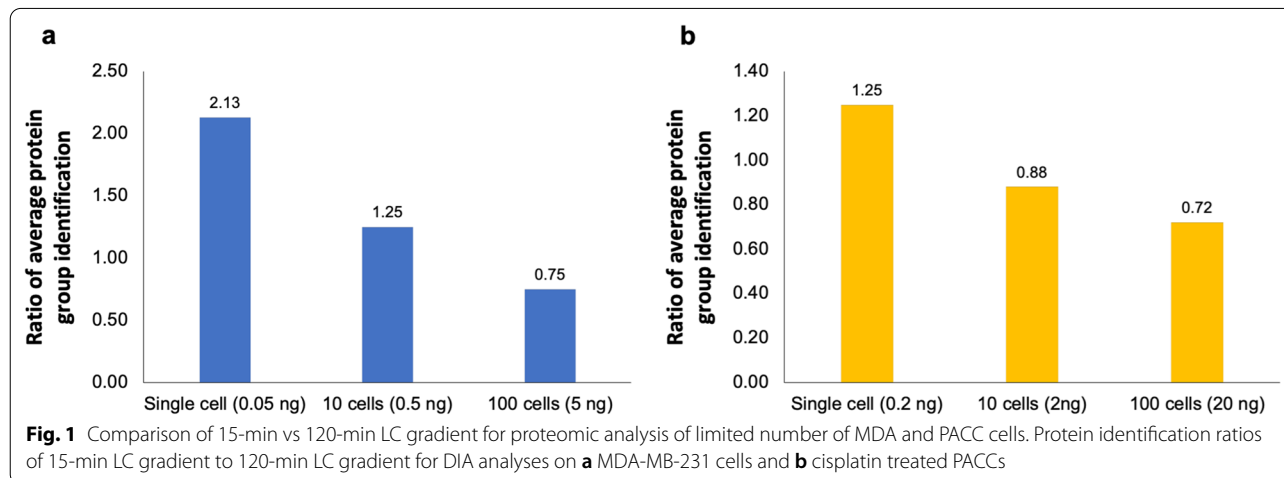
LC gradient time for protein identification at single-cell levels

In general, an optimal protein identification can be achieved using 1 μg peptides for DIA-MS analysis in

combination with 120-min LC gradient for large cell population ($\geq 20,000$ cells) and bulk cell samples [27]. However, such approach may not be ideal for small cell population. Therefore, we conducted a comparative analysis on different LC gradient settings for peptides equivalent to hundred-cell, ten-cell, and single-cell levels. Of note, we used identical MS setting for both gradients. We first compared two LC gradient settings for a single cell (0.05 ng peptides), 10 cells (0.5 ng) and 100 cells (5 ng) from MDA-MB-231 cell samples by computing the identification ratio based on the average number of identified proteins from 15-min LC gradient to the average number of identified proteins from 120-min LC gradient from single run by direct DIA searches. We found the number of protein identification rate was more in 15-min LC gradient for single-cell and 10-cell levels. As shown in Fig. 1a, the protein identification ratios of 15 min LC gradient to 120-min LC gradient were more than 1 for single and ten MDA-MB-231 cells. We observed similar result for the single cisplatin treated PACC cell (0.2 ng of peptides) (Fig. 1b), where the ratio of 15-min to 120-min LC gradient was also more than 1, indicating that less proteins were identified from 120-min LC gradient and an improvement in overall protein identification using a short LC gradient time at single-cell level of peptide injection amount < 2 ng. Therefore, we chose 15 min as our optimal LC gradient setting for single-cell level DIA analysis.

Evaluation of global proteomic analysis at single-cell level

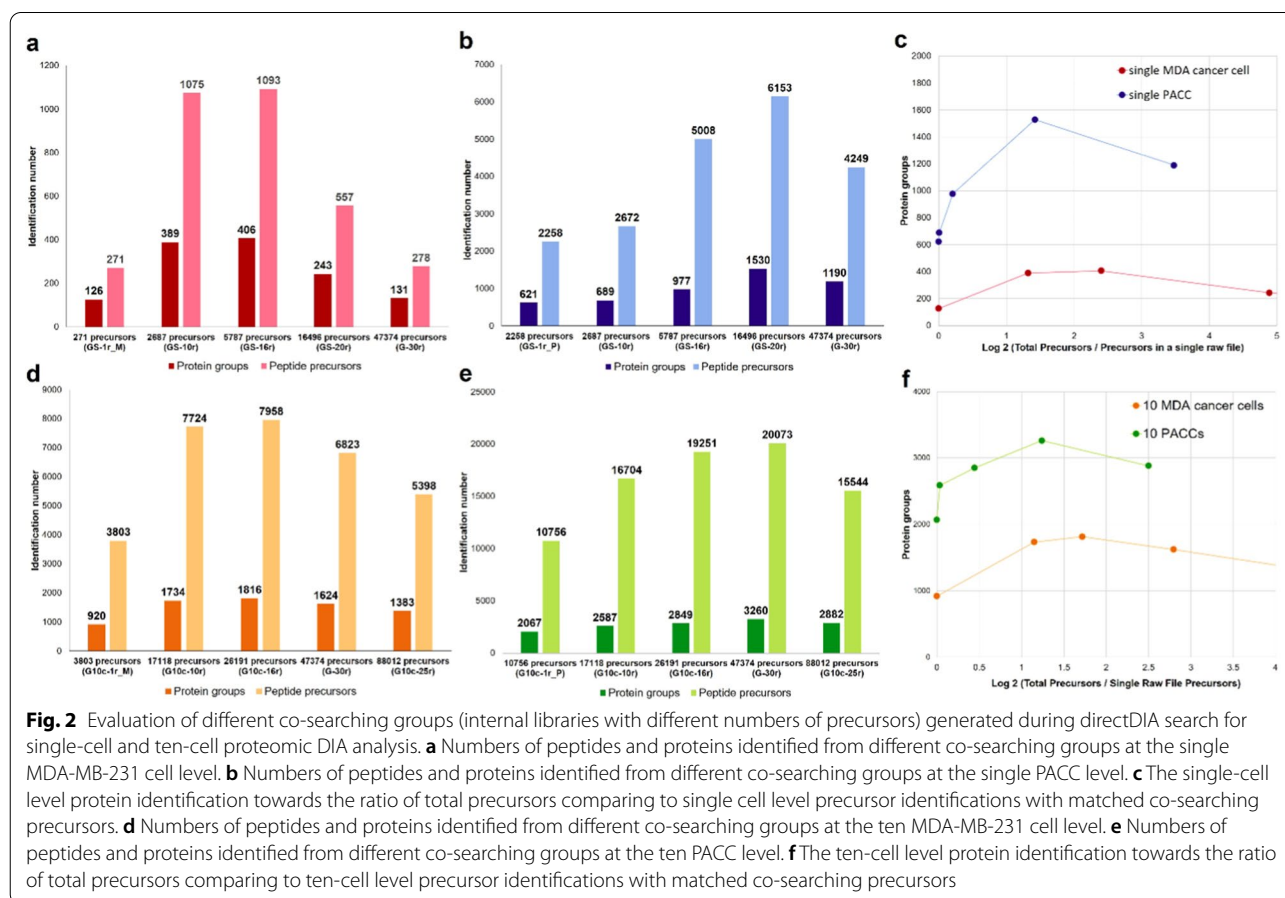
Besides investigating the suitable LC gradient for acquiring DIA-MS data at single-cell level, the search space of single-cell DIA data (i.e., the size of internal library generated during directDIA search) was also evaluated among the established co-searching groups. The DIA data of MDA-MB-231 cell samples (0.05 ng



peptides equivalent to the single cell level (0.05 ng), searching GS-1r_M (total of 271 peptide precursors from the raw file of a single MDA-MB-231 cell), and 126 protein groups were identified (Fig. 2a). As the size of the internal library increased to 5787 precursors (i.e., GS-16r), we observed the highest peptide and protein coverage for the single MDA-MB-231 cancer cell. Total of 1093 peptide precursors and 406 protein groups were identified using GS-16r (Fig. 2a), corresponding to gains of 303% and 222% at peptide and protein levels compared to the results obtained by using GS-1r_M only. For cisplatin-treated PACC at single-cell level (0.2 ng of peptides), 621 protein groups were identified using the directDIA approach to search against the internal library of GS-1r_P (2258 peptide precursors) (Fig. 2b). Moreover, we found that using co-searching group of GS-20r (16,496 precursors) yielded the best identification, where 6153 peptide precursors and 1530 protein groups were identified (Fig. 2b). By using GS-20r, 172% and 146% gains at peptide precursor and protein levels relative to using GS-1r_P, respectively. Of note, the number of identified proteins/peptides was not necessarily increased as the search space expanded.

As shown in Fig. 2c, the best precursor identification is within the range of two- to four-fold difference between the total number of precursors in an internal library and total number of precursors detected in the sample of interest. Our results suggested that the internal library size was critical to protein identification at single-cell and sub-nano-gram levels via DIA-MS approach.

Furthermore, we evaluated the co-searching methods at ten-cell level. We identified 1816 protein groups (Fig. 2d) and 3260 protein groups (Fig. 2e) from 10 MDA-MB-231 cells and 10 PACCs, respectively. Similarly, the best peptide precursor and protein identifications were also fallen into the two- to four-fold changes between the internal library size and detected precursors (Fig. 2f). In addition, we observed the optimal protein identification via directDIA search for the single-cell and ten-cell injections when co-searched with injected peptide amount that were about 10-fold difference within the similar samples (Additional file 1: Table S5). Overall, it was essential to use a co-searching internal library generated from similar samples during the directDIA search to enhance



protein identification at the single-cell and nano-gram levels.

Reproducibility on single-cell level proteomic analysis using DIA

After evaluation of LC gradient and size of internal library for DIA analysis at single-cell level, we investigated the inter-person reproducibility using our established workflow. Two sets of samples, each contained cisplatin- and docetaxel-treated PACC and parental MDA-MB-231 cancer cells, were prepared and analyzed a month apart by two researchers following the procedures stated in Experimental section. Here, we use the cisplatin treated PACC sample and one MDA-MB-231 sample to demonstrate the reproducibility of our DIA method. We observed pairwise Spearman correlation >0.80 for the MDA-MB-231 cell sample between the two sets at single-cell, ten-cell, and hundred-cell levels (Fig. 3a, b, c). We found similar results for cisplatin treated PACC samples,

where Spearman correlation ≥ 0.83 were observed in three different cell populations at single-cell, ten-cell and hundred-cell levels (Fig. 3d, e, f). Taken together, these results demonstrated that our optimized DIA workflow provided robust quantitative global proteome profiling for single-cell DIA analysis, and larger cell population further improved the reproducibility.

PTM analyses for nano-gram levels of peptides without enrichment

Protein modifications are important for the regulation of various protein activities and cellular signaling events, and alternation in PTMs are associated with many diseases, including cancer [28]. When conducting MS-based PTM analysis, PTM enrichment is an essential procedure; however, there is very limited report of nano-gram/single-cell level of enrichment strategies for MS analysis. Thus, we established an alternative approach for PTM analysis at such level

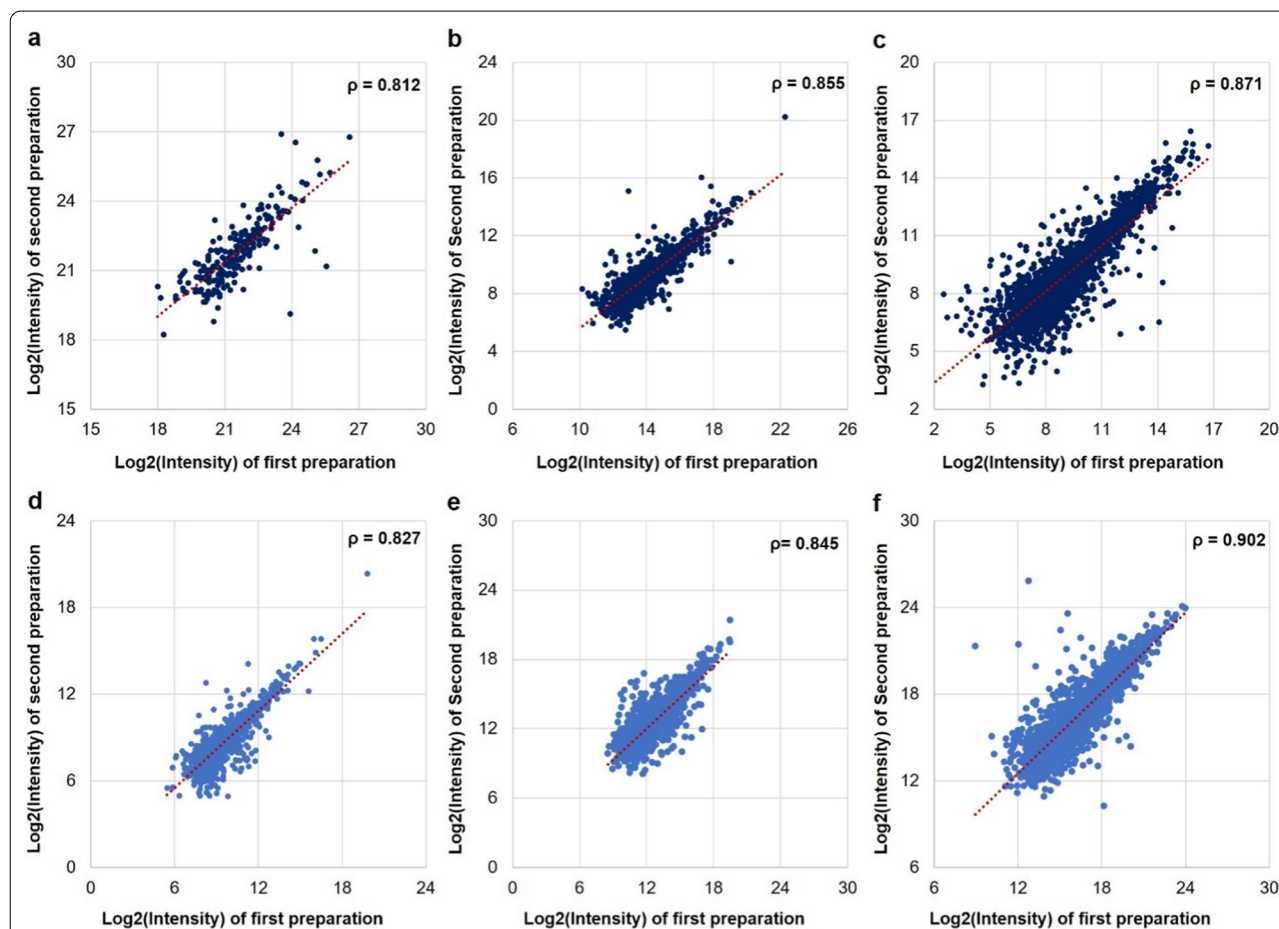


Fig. 3 Reproducibility of the DIA-MS runs where the sample preparations were done month apart by two different researchers. Pairwise Spearman correlation (ρ) of **a, b, c** protein intensities between two preparations and DIA-MS analyses of MDA-MB-231 cells at single-cell, ten-cell, hundred-cell levels, and **d, e, f** protein intensities between two preparations and DIA-MS analyses of cisplatin treated PACCs single-cell, ten-cell, and hundred-cell levels

by utilizing global proteomic DIA data and spectral libraries built from bulk samples. Unlike DDA-MS, DIA-MS allows that all the peptide precursors are co-fragmented within a selected m/z range to produce comprehensive MS2 spectra. The information of modified peptides should be retained in the global data even without PTM enrichment. Therefore, we were able to directly identify PTMs from the nano-gram level (i.e., 100 cells) of global proteomic DIA data using customized PTM spectral libraries for phosphorylation, acetylation, and ubiquitination.

We firstly explored the possibility of identifying phosphorylation, acetylation, and ubiquitination from global DIA data of PDX samples at nano-gram level (Fig. 4a). We identified 72 phosphorylated peptides, 35 acetylated peptides, and 99 ubiquitinated peptides from 100 PACCs, indicating the possibility of finding PTMs without using enrichment.

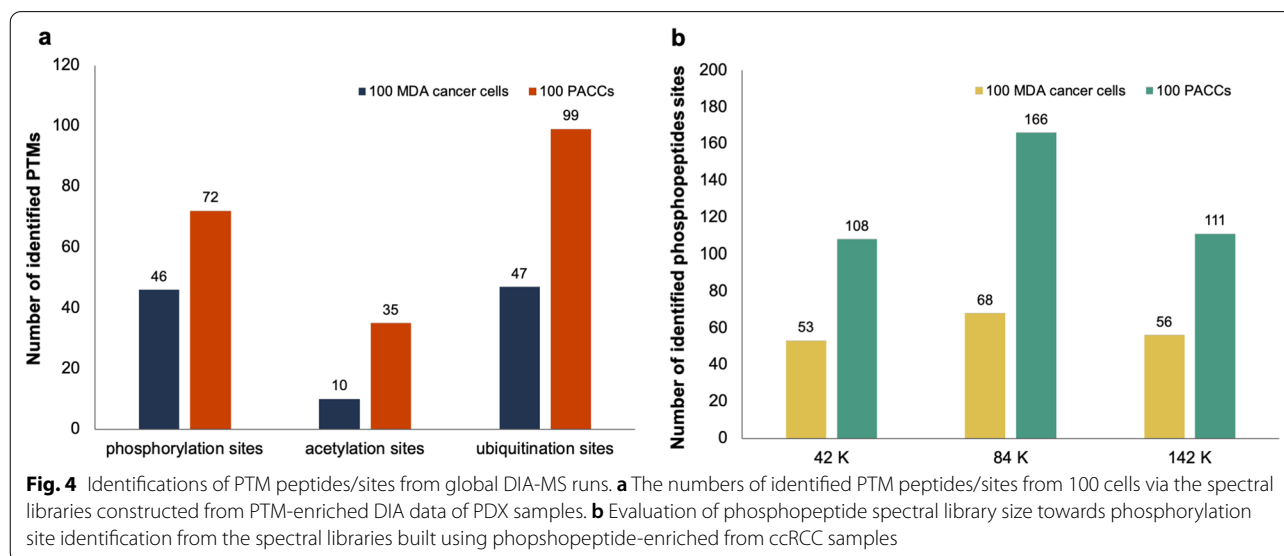
We further evaluated the association between PTM spectral library size and PTM identification by examining the alteration in phosphopeptide identification rate from the nano-gram level of global DIA data, since a large collection of phosphopeptide-enriched DDA and DIA raw files from CPTAC study [24] allowed the construction of spectral libraries with various sizes ranging from ~42 to ~141 K precursors. As shown in Fig. 4b, among the three phosphopeptide spectral libraries, the library containing ~84 K precursors contributes to the highest identification number for 100 MDA-MB-231 cells (5 ng of peptides) and 100 PACCs (20 ng of peptides) of which 68 and 166 phosphopeptides with localized sites are identified, respectively. These results suggested that PTM analysis of

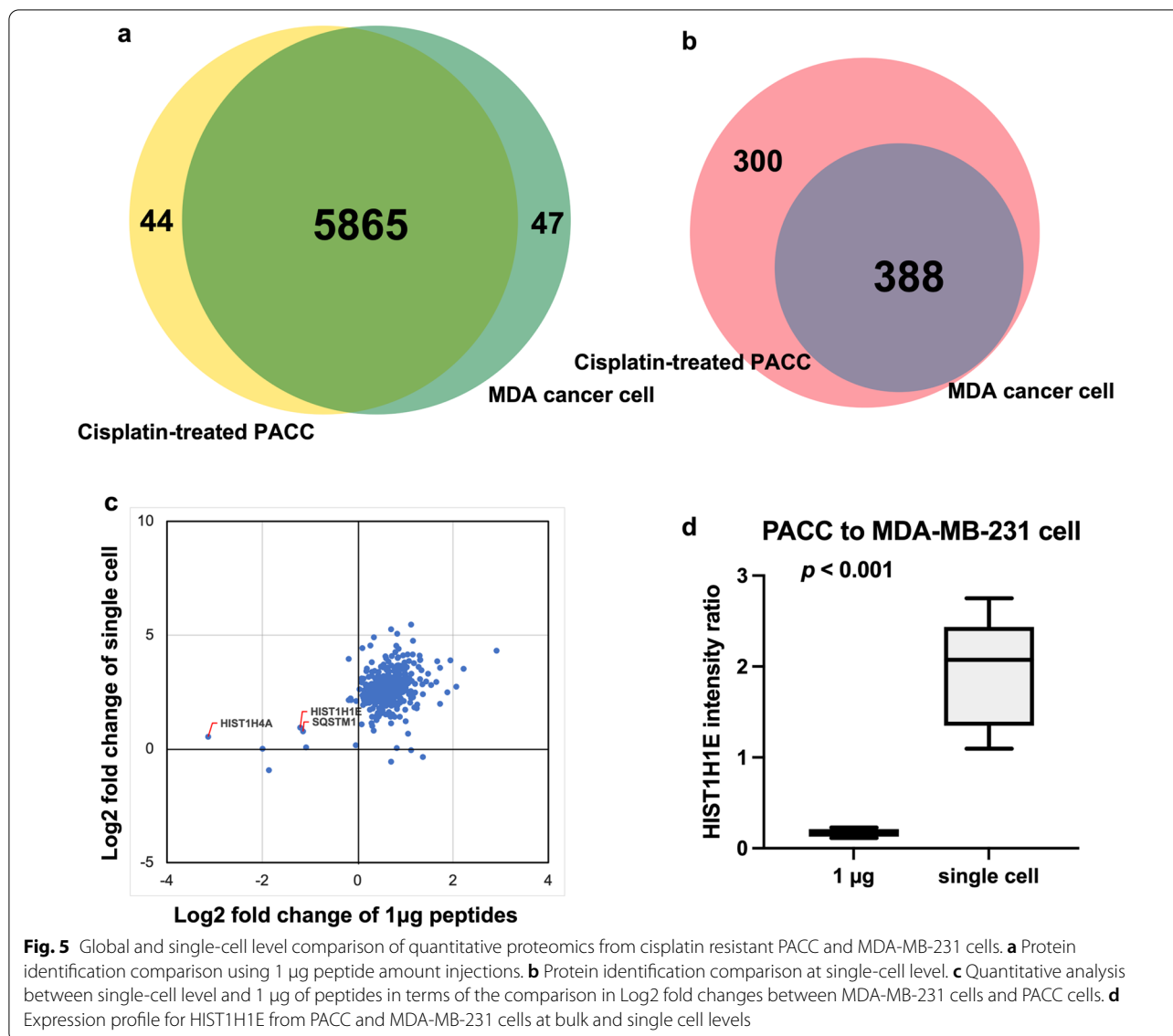
nano-gram scale could be achieved by utilizing global DIA data along with a suitable PTM library built from bulk samples.

Application of single-cell level DIA approach to the drug resistant cancer cell study

To investigate whether the difference in cell size affected identification and protein expression patterns, we conducted a comparative analysis between PACCs (large cells) and MDA-MB-231 cells (smaller cells) at 1 μg peptide injection and single-cell level of peptide injection. We observed 98.5% of overlap in protein identification between cisplatin treated PACC and MDA-MB-231 samples (Fig. 5a), suggesting that they shared similar proteome profile, regardless of cell size. At single-cell level, we examined the identified protein groups from the co-searching via all single-cell raw files (i.e., GS-10r, Fig. 2a). We observed 388 protein groups identified in the single MDA-MB-231 cell and 688 proteins were identified from the cisplatin-treated PACC at single cell level (Fig. 5b).

Although more proteins were identified using 1 μg peptide injections, we speculated that single-cell proteomic analysis could capture the real protein expression changes comparing to bulk proteomic analysis, which would assist in the study of cellular heterogeneity between PACC and parental MDA-MB-231 cells. We compared the protein fold changes as shown in Fig. 5c. At single-cell level, we found majority of the proteins showing higher expression in cisplatin resistant PACC relative to the parental MDA-MB-231 cell with Log2 fold changes ranged between 2 and 5, indicating increased protein levels for these proteins after treating MDA-MB-231 cells with cisplatin and transitioning to a PACC state, except a





small set of proteins such as selective autophagy receptor protein p62/SQSTM1 and histone proteins (e.g., HIST1H4A and HIST1H1E) displayed similar expression profiles between PACCs and untreated MDA-MB-231 cells. In contrast, in bulk analysis of the MDA-MB-231 and cisplatin-treated PACCs, most of the proteins showed similar abundances at the same injection amount (1.0 µg) level; however, we found more than two-fold decrease in PACC compared to MDA-MB-231 cancer cells for p62/SQSTM1, HIST1H4A and HIST1H1E. If we only run the samples at 1 µg peptide level, we could only observe the decrease in p62/SQSTM1, HIST1H4A, and HIST1H1E proteins in PACCs. However, at the single-cell level analyses, we noticed these proteins maintaining

in the same expression levels in both drug-treated PACCs and untreated MDA-MB-231 cells.

Single-cell level analysis may help to understand the mechanism of forming drug resistant PACCs after drug treating MDA-MB-231 cells. As lack of amplification of histone proteins during cell division is reported to lead to cell cycle elongation [29, 30], drug-treatment of MDA-MB-231 cells inhibits amplification of histone proteins such as HIST1H4A and HIST1H1E are unable to undergo division, which results in PACCs that have much bigger size than untreated MDA-MB-231 cells. The protein p62/SQSTM1 is a selective autophagy receptor in a ubiquitylation-dependent approach [31, 32]. Lack of amplification of p62/SQSTM1 protein might lead the cells insensitive to drug-associated stress and

escape from autophagy [33, 34]. We also observed the similar p62/SQSTM1, HIST1H4A, and HIST1H1E protein expression patterns from docetaxel resistant PACC. These findings in single cell level analysis reveal PACCs formation and survival mechanism.

The difference in protein abundance was further evaluated by comparing drug-induced PACCs performed in replicates to MDA-MB-231 cells without treatment performed in triplicates. The protein intensity ratios of the HIST1H1E protein from PACC samples to MDA-MB-231 samples were summarized in box plot, showing significantly statistical significance ($p < 0.001$) (Fig. 5d). Taken together, quantitative analysis of single-cell proteome via DIA approach can benefit our understanding of cellular heterogeneity and provide more accurate protein expression profiles which may be misinterpreted at bulk population.

Conclusions

Single-cell proteomic analysis provides insights into cellular heterogeneity allowing the characterization of cellular microenvironment, whereas proteomic analysis using bulk samples only captures a population average hindering our understanding of the diversity in cellular functions. In this study, we established and optimized a single-cell proteomic analysis workflow which utilized DIA-MS and directDIA method to analyze the global proteome of the MDA-MB-231 cancer cells and the matched drug resistant PACC cells. We first systematically evaluated aliquoted peptide samples from a single cell to 100 cell levels (0.05–20 ng) of the two cell types under different LC gradient settings. We found that 120-min LC gradient was more suitable for peptide injection amount > 2 ng (~ 10 PACCs), whereas 15-min LC gradient produced was more appropriate for the injection amount < 2 ng (~ 10 PACCs). By applying and investigating directDIA search method using different co-searching groups (i.e., internal libraries), we observed approximately a four-fold difference between the internal library size and total number of detected precursors of a DIA raw file produced the highest protein identification rate with good reproducibility [35]. Of note, 1500–3000 proteins were identified from 10 to 140 mammalian cells (equaled to 0.5–7 ng of peptides) by using narrow-bore columns (i.d. 30 μm or 20 μm) coupled with low flow rate separation [36], where < 2000 proteins were identified from 2 ng aliquoted peptide samples by using a special LC system [37, 38]. Nevertheless, using our optimized workflow, 2 ng (~ 10 PACCs) of peptides allowed an identification of ≥ 3200 proteins and 1500 proteins were identified from a single PACC cell (~ 0.2 ng of peptides) even with normal bore column (i.d. 75 μm) and normal LC system. Furthermore, identification of PTMs at nano-gram/single-cell level without any PTM enrichment was achieved by

directly searching the nano-gram level of global DIA data against pre-generated PTM libraries. More PTM sites may be identified if the PTM libraries constructed by PTM enrichments from the same samples are used to search the global proteomic data [39]. Additionally, we were able to detect the cellular heterogeneity between PACCs and their parental MDA-MB-231 cells at single-cell level using our established workflow. In summary, we developed a novel approach to study small cell population, including single cell, by using DIA-MS coupled with a short LC gradient and the directDIA search with an internal library with an appropriate size, which can support quantitative single-cell proteomic and PTM analyses at high-throughput.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12014-022-09359-9>.

Additional file 1. Table S1. Large amount peptide amount and peptide amount per cell. **Table S2.** Detail information of 30 MS runs. **Table S3.** Information of single-cell level co-searching libraries. **Table S4.** Information of ten-cell level co-searching libraries. **Table S5.** Identified protein groups towards library size.

Author contributions

HZ, YW, TML, LWC, MDK, SRA, and KJP conceived and designed the experiments; HZ, YW and TML interpreted the results; MDK performed the cell culture experiments; YW and L-JC performed the experiments of mass spectrometry part; YX prepared the PTM libraries; YW, TML and HZ analyzed data and prepared figures; YW drafted the manuscript; TML, HZ, MDK, KJP, and SRA edited the manuscript; HZ oversaw the execution of the project. All authors read and approved the final manuscript.

Funding

This work was supported by fundings from: National Cancer Institute, the Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA210985) and National Cancer Institute, Early Detection Research Network (EDRN, U01CA152813) to H Zhang; US Department of Defense CDMRP/PCRP (W81XWH-20-10353), the Patrick C. Walsh Prostate Cancer Research Fund and the Prostate Cancer Foundation to SR Amend; and National Cancer Institute grants U54CA143803, CA163124, CA093900, and CA143055, and the Prostate Cancer Foundation to KJ Pienta.

Availability of data and materials

Available for MassIVE database. <ftp://massive.ucsd.edu/MSV000088274/>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

KJ Pienta is a consultant for CUE Biopharma, Inc., is a founder and holds equity interest in Keystone Biopharma, Inc., and receives research support from Progenics, Inc. SR Amend also holds equity interest in Keystone Biopharma, Inc. The other authors declare no conflicts of interest.

Author details

¹Department of Pathology, Johns Hopkins University, Baltimore, MD 21287, USA. ²Cancer Ecology Center, The Brady Urological Institute, Johns Hopkins School of Medicine, 600 N. Wolfe Street, Baltimore, MD 21287, USA.

Received: 8 December 2021 Accepted: 26 May 2022
Published online: 09 July 2022

References

- Macaulay IC, Ponting CP, Voet T. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 2017;32:155–68.
- Lee J, Hyeon DY, Hwang D. Single-cell multiomics: technologies and data analysis methods. *Exp Mol Med.* 2020;52:1428–42.
- De Rosa SC, Herzenberg LA, Herzenberg LA, Roederer M. 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. *Nat Med.* 2001;7:245–8.
- Perez OD, Nolan GP. Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat Biotechnol.* 2002;20:155–62.
- Bandura DR, et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem.* 2009;81:6813–22.
- Hughes AJ, Spelke DP, Xu Z, Kang C, Schaffer DV, Herr AE. Single-cell western blotting. *Nat Methods.* 2014;11:749–55.
- Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* 2018;19:161.
- Ctortea C, Mechtler K. The rise of single-cell proteomics. *Anal Sci Adv.* 2021;2:84–94.
- Cheung TK, Lee C, Bayer FP, McCoy A, Kuster B, Rose CM. Defining the carrier proteome limit for single-cell proteomics. *Nat Methods.* 2021;18:76–83.
- Gillet LC, Navarro P, Tate S, Röst H, Sellevsek N, Reiter L, Bonner R, Aebersold R. 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.* 2012;11:O111.016717.
- Collins BC, et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat Commun.* 2017;8:291.
- Muntel J, Kirkpatrick J, Bruderer R, Huang T, Vitek O, Ori A, Reiter L. Comparison of protein quantification in a complex background by DIA and TMT workflows with fixed instrument time. *J Proteome Res.* 2019;18:1340–51.
- Thomas SN, Friedrich B, Schanubelt M, Chan DW, Zhang H, Aebersold R. Orthogonal proteomic platforms and their implications for the stable classification of high-grade serous ovarian cancer subtypes. *iScience.* 2020;23:101079.
- Navarro P, et al. A multicenter study benchmarks software tools for label-free proteome quantification. *Nat Biotech.* 2016;34:1130–6.
- Zhang F, Ge W, Ruan G, Cai X, Guo T. Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics.* 2020;20:1900276.
- 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.* 2015;14:1400–10.
- Tsou C, Avtonomov D, Larsen B, Tucholska M, Choi H, Gingras A, Nesvizhskii AI. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods.* 2015;12:258–64.
- Parker SJ, Venkatraman V, Van Eyk JE. Effect of peptide assay library size and composition in targeted data-independent acquisition-MS analyses. *Proteomics.* 2016;16:2221–37.
- Wu JX, Song X, Pascovici D, Zaw T, Care N, Krisp C, Molloy MP. SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. *Mol Cell Proteomics.* 2016;15:2501–14.
- Barkovits K, et al. Reproducibility, specificity and accuracy of relative quantification using spectral library-based data-independent acquisition. *Mol Cell Proteomics.* 2020;19:181–97.
- Muntel J, Gandhi T, Verbeke L, Bernhardt OM, Treiber T, Bruderer R, Reiter L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol Omics.* 2019;15:348–60.
- Pienta KJ, Hammarlund EU, Axelrod R, Brown JS, Amend SR. Polyaneuploid cancer cells promote evolvability, generating lethal cancer. *Evolut Appl.* 2020;13:1626–34.
- Mertins P, et al. Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. *Nat Protocols.* 2018;13:1632–61.
- Clark DJ, et al. Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell.* 2019;179:964–83.
- Wang L, et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell.* 2020;39:509–528.e20.
- Udeshi ND, et al. Rapid and deep-scale ubiquitylation profiling for biology and translational research. *Nat Commun.* 2020;11:539.
- Cho K, et al. Deep proteomics using two dimensional data independent acquisition mass spectrometry. *Anal Chem.* 2020;96:4217–25.
- Chen L, Liu S, Tao Y. Regulating tumor suppressor genes: post-translational modifications. *Signal Transduct Target Ther.* 2020;5:90.
- Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet.* 2008;9:843–54.
- Günesdogan U, Jäckle H, Herzig A. Histone supply regulates S phase timing and cell cycle progression. *eLife.* 2014;3:e02443.
- Tsukada M, Ohsumi Y. Isolation and characterization of autophagy-defective mutants of *Saccharomyces cerevisiae*. *FEBS Lett.* 1993;333:169–74.
- Dikic I, Elazar Z. Mechanism and medical implications of mammalian autophagy. *Nat Rev Mol Cell Biol.* 2018;19:349–64.
- Sánchez-Martin P, Komatsu M. p62/SQSTM1—steering the cell through health and disease. *J Cell Sci.* 2018;131:222836.
- Kageyama S, et al. p62/SQSTM1 droplet serves as a platform for autophagosome formation and anti-oxidative stress response. *Nat Commun.* 2021;12:16.
- Siyal AA, et al. Sample size-comparable spectral library enhances data-independent acquisition-based proteome coverage of low-input cells. *Anal Chem.* 2021;93:17003–11.
- Zhu Y, et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat Commun.* 2018;9:882.
- Cong Y, et al. Improved single-cell proteome coverage using narrow-bore packed nanoLC columns and ultrasensitive mass spectrometry. *Anal Chem.* 2020;92:2665–71.
- Cong Y, et al. Ultrasensitive single-cell proteomics workflow identifies > 1000 protein groups per mammalian cell. *Chem Sci.* 2021;12:1001–6.
- Li Y, et al. An integrated strategy for mass spectrometry-based multiomics analysis of single cells. *Anal Chem.* 2021;93:14059–67.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

