Title

2019 Association of Biomolecular Resource Facilities Multi-Laboratory Data-Independent Acquisition Proteomics Study.

Journal

Journal of Biomolecular Techniques, 34(2)

ISSN

1524-0215

Authors

Kirkpatrick, Joanna

Stemmer, Paul M

Searle, Brian C

et al.

Publication Date

2023-07-01

DOI

10.7171/3fc1f5fe.9b78d780

Peer reviewed

Journal of Biomolecular Techniques • Volume 34(2); 2023 Jul

2019 Association of Biomolecular Resource Facilities Multi-Laboratory Data-Independent Acquisition Proteomics Study

Joanna Kirkpatrick[1,2] Paul M. Stemmer[3] Brian C. Searle[4,5] Laura E. Herring[6] LeRoy Martin[7] Mukul K. Midha[8] Brett S. Phinney[9] Baozhen Shan[10] Magnus Palmblad[11] Yan Wang[12] Pratik D. Jagtap[13] Benjamin A. Neely[14]

[1]Leibniz Institute on Aging, Fritz Lipmann Institute, 07745 Jena, Germany,
[2]The Francis Crick Institute, London NW1 1AT, United Kingdom,
[3]Wayne State University, Detroit, Michigan 48202, USA,
[4]Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio 43210, USA,
[5]Pelotonia Institute for Immuno-Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, USA,
[6]UNC Proteomics Core Facility, Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA,
[7]Waters Corporation, Beverly, Massachusetts 01915, USA,
[8]Institute for Systems Biology, Seattle, Washington 98109, USA,
[9]University of California, Davis, Davis, California 95616, USA,

[10]Bioinformatics Solutions Inc., Waterloo, ON N2L 3K8, Canada,
[11]Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands,
[12]National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland 20892, USA,
[13]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA,
[14]National Institute of Standards and Technology, Charleston, South Carolina 29412, USA

Association of Biomolecular Resource Facilities

Published on: Jun 02, 2023

URL: https://jbt.pubpub.org/pub/sapzl1sc

ABSTRACT

Despite the advantages of fewer missing values by collecting fragment ion data on all analytes in the sample as well as the potential for deeper coverage, the adoption of data-independent acquisition (DIA) in proteomics core facility settings has been slow. The Association of Biomolecular Resource Facilities conducted a large interlaboratory study to evaluate DIA performance in proteomics laboratories with various instrumentation. Participants were supplied with generic methods and a uniform set of test samples. The resulting 49 DIA datasets act as benchmarks and have utility in education and tool development. The sample set consisted of a tryptic HeLa digest spiked with high or low levels of 4 exogenous proteins. Data are available in MassIVE MSV000086479. Additionally, we demonstrate how the data can be analyzed by focusing on 2 datasets using different library approaches and show the utility of select summary statistics. These data can be used by DIA newcomers, software developers, or DIA experts evaluating performance with different platforms, acquisition settings, and skill levels.

ADDRESS CORRESPONDENCE TO: Benjamin A. Neely, 331 Fort Johnson Rd., National Institute of Standards and Technology, Hollings Marine Laboratory, Charleston, South Carolina 29412, USA (Phone: 843-460-9841; E-mail: Benjamin.neely@nist.gov).

Conflict of Interest Disclosures: BCS is a founder and shareholder in Proteome Software, which operates in the field of proteomics. BS is the CEO at Bioinformatics Solutions Inc., which creates software for the field of proteomics. The other authors declare no competing interests.

Keywords: data-independent acquisition, label-free quantification, proteomics, spike-in quantification

SUMMARY

Data-independent acquisition (DIA) is an alternative strategy to data-dependent acquisition (DDA) of precursor fragmentation data (MS2) in mass spectrometry. In DDA, the instrument selects and fragments MS1 ions based on signal intensity. In DIA, the mass spectrometer fragments analytes in predefined *m*/*z* windows. The MS2 data contribute to analyte identification and provide relative quantification. Both DDA and DIA approaches rely on sophisticated algorithms, and the interpretation of data is computationally intensive.[1],[2],[3],[4] Benefits of DIA include increased depth of coverage and between-sample uniformity (by avoiding the stochastic nature of DDA acquisition), allowing for unprecedented depth[5] and speed[6] of analysis. Recently, advances in instrumentation and algorithms have resulted in wider adoption.[7],[8],[9] In keeping with the mission of the Association of Biomolecular Resource Facilities (ABRF), the Proteomics Research Group (PRG) developed a multi-laboratory study (Figure 1), providing novice and expert users with samples and generic methods to benchmark their laboratories and empower participants to perform DIA.

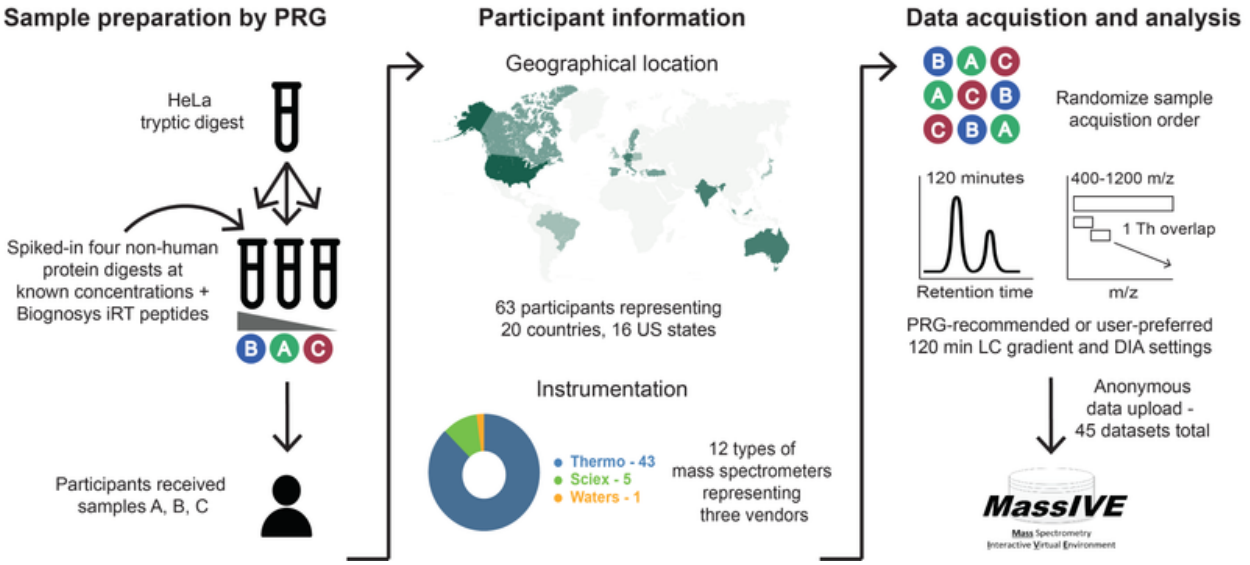

Figure 1

General description of study. The base sample in the study was a tryptic HeLa digest. Each participant received 20 ug of that sample as well as 20 ug of that sample with 2.5 or 10 fmol/µg of 4 nonhuman proteins: beta-galactosidase, lysozyme C, glucoamylase, and protein G. All samples also contained an iRT peptide mix. Participants received dried samples identified only as A, B, or C. Participant identification was confidential, and they could identify their dataset through a randomized identification number. Recommended LC and DIA settings were distributed with the samples, although participants could use their preferred settings. Each laboratory was asked to run the samples in a specific randomized manner. Participants anonymously uploaded data to MassIVE, where it was curated by the PRG (renamed if needed and checked for file integrity) and reuploaded to MassIVE MSV000086479.

Mixtures of proteomes[10],[11] have been used to benchmark proteomic workflows. We selected a set of 4 nonendogenous proteins in a human matrix.[12] The proteins beta-galactosidase, lysozyme C, glucoamylase, and protein G were digested and then spiked into the HeLa digest at 0, 2.5, and 10 fmol/µg (sample A: 2.5 fmol spike, sample B: 10 fmol spike, and sample C was no spike). The 4 added proteins and 2 levels provided a wide range in signal intensities of the peptides such that the depth of spike-in coverage could reflect relative sensitivity between participants.[13],[14] The study announcement was disseminated on the PRG website,[15] at conferences, and via social media. Participants had varying prior experience and included mass spectrometers from different vendors (Figure 2; Table 1; Supplemental Table S1). A generic method (Supplemental File 1) was supplied, and participants were asked to use a standard 2-hour, 2-step liquid chromatography (LC) gradient, a uniform static overlapping windowing strategy, and a cycle time of approximately 3.5 s. Most participants followed these recommendations (Table 2 and Table 3; Supplemental Table S1). Of 63 laboratories that enrolled and received sample sets, 45 returned data. Some users had multiple instruments or compared different DIA methods, resulting in 49 datasets, 43 of which contain data for all replicates.

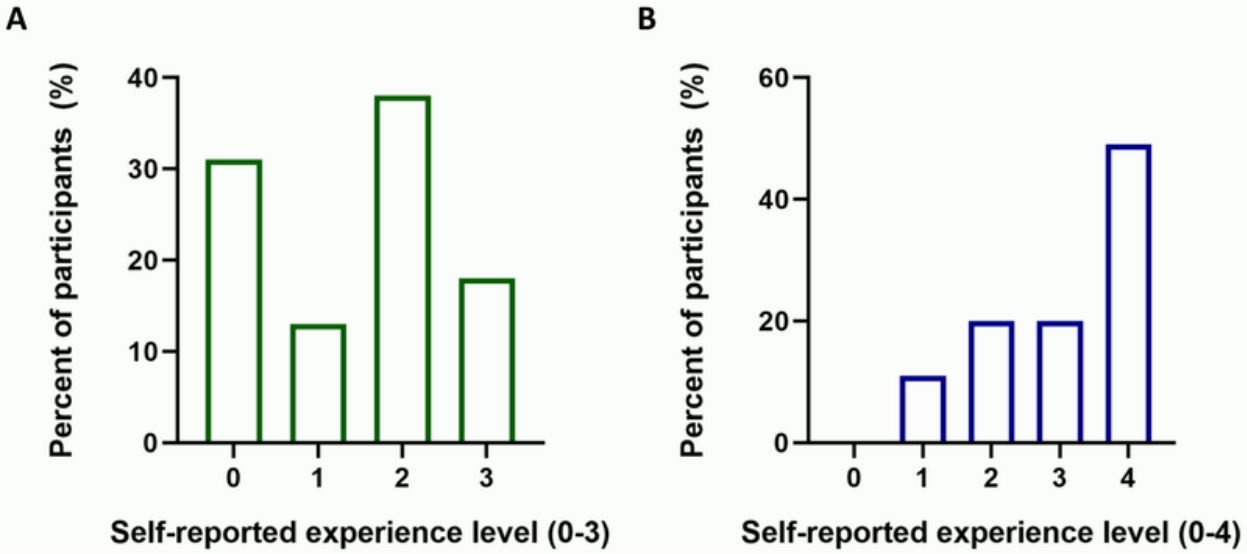

Figure 2

Self-reported experience level of 45 participants. An arbitrary scale was given to participants in a companion survey. **A**. Self-reported experience LC-MS/MS level. When asked about LC-MS/MS experience, participants were given the following choices: 0, “never set up myself”; 1, 0 to 2 years; 2, 3 to 5 years; 3, 6 to 9 years; and 4, >10 years. **B**. Self-reported DIA experience. When asked about DIA experience, participants were given the following choices: 0, Heard about it; 1, Tried it once; 2, Have done it a couple of times; and 3, Expert.

Table 1

Instruments used in the study

Instruments	No. of datasets
Thermo LTQ Orbitrap Elite	1
Thermo LTQ Orbitrap Velos	1
Thermo Orbitrap Fusion	9
Thermo Orbitrap Fusion Lumos	12
Thermo Orbitrap Velos Pro	1
Thermo Q Exactive	3
Thermo Q Exactive HF	6
Thermo Q Exactive HF-X	7
Thermo Q Exactive Plus	3

Sciex TripleTOF 5600	3
Sciex TripleTOF 6600	2
Waters Xevo G2 XS	1

The 45 participants deposited 49 datasets using 12 different instrument platforms from 3 different manufacturers.

Table 2

Window scheme strategies used in the study

Window scheme	No. of datasets
Variable	2
Static, nonoverlap	3
Static, w/ overlap	37
Static, w/ gaps	5
Single window	1
Other*	1

** Used a targeted list.*

The DIA window strategies were divided into groups based on whether the DIA windows were static (ie, the size did not change) and whether the DIA windows overlapped, while 5 datasets had gaps between the DIA windows.

Table 3

Flow rates used in the study

Flow rate	No. of datasets
~300 nL/min	38
>1 µL/min	6
Unknown	5

There were 2 main flow rates employed, either nL/min flow rates (approximately 300 nL/min; 250 to 400 nL/min) or µL/min flow rates (1.5 to 50 µL/min).

Other studies have used multi-laboratory DIA datasets to benchmark software tools.[10],[16],[17],[18],[19] Similarly, this new dataset has many uses, including benchmarking and user training. The range of user experience and instruments contributed to differences in data quality, providing a real-world dataset for evaluating how software normalization strategies are affected by data quality. The incorporation of known spikes facilitates the evaluation of relative quantification using DIA.[20] When the spiked proteins are ignored, each participant performed triplicate injections of experimental replicates, allowing for the generation of useful summary statistics. Overall, this dataset provides opportunities for users to learn about acquisition methods and evaluate computational tools for DIA.

This dataset is also valuable to DIA software developers. The recommended acquisition method was not optimized for any platform, producing datasets conducted on different platforms with a similar acquisition strategy. All samples included internal retention time (iRT) peptides and companion DDA, and gas-phase fractionation data were also generated. Therefore, any software or library approach can use the data to evaluate and improve these approaches. The inclusion of the spike proteins in known amounts creates a unique opportunity to test new DIA strategies, such as MS1-based quantification[21] and in silico–generated libraries.[22],[23],[24],[25],[26]

In an initial analysis detailed herein, we have made a comparative analysis of data from 2 participants that used the same instrument (while a preliminary analysis of the complete dataset can be found here in ref. [27]). Each of these 2 participants acquired additional data for library generation, so it was possible to show how library construction and utilization effects results. Library strategy affected the number of proteins identified (Figure 3A-B), the precision of replicates (Figure 3A-B), and relative abundances of spike-in proteins (Figure 3C-D). Similar to reported observations,[20] these results highlight discrepancies when inferring protein abundance and the need to check relative quantification across the dynamic range. As these data and associated metadata are publicly available, we expect it to be used in benchmarking new tools and library strategies. Ongoing analysis of the dataset will provide more information and best-practice instrument settings for DIA, although the generic method provided performed unexpectedly well. With continued advancement of DIA methods, platforms optimized for DIA and improved computational strategies, this is an exciting time for the field, and we look forward to future multi-laboratory studies, enabling users and developers alike.

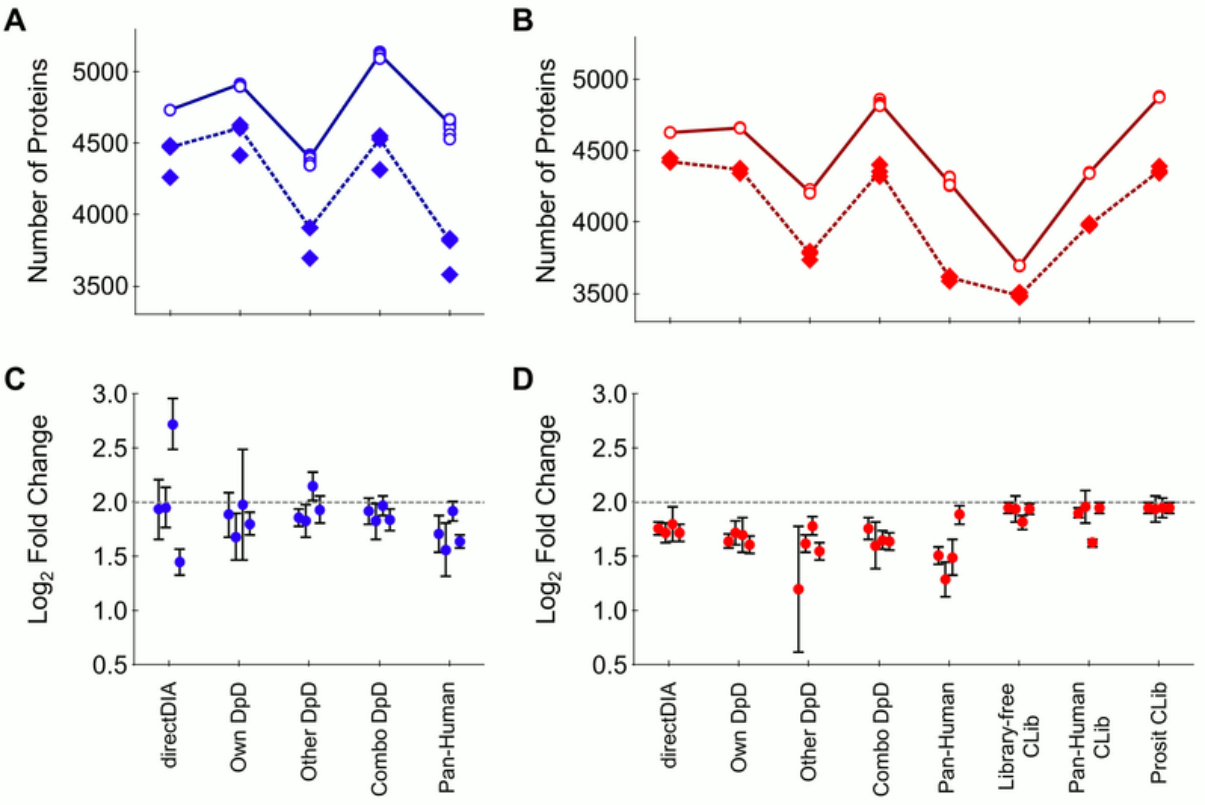

Figure 3

Comparison of library approaches with data from Participants 3 and 48. Different library approaches were used to evaluate protein identification, technical variation, and accurate relative quantification of 4 spike proteins with Participant 48 (**A** and **C**) and Participant 3 (**B** and **D**). Eight library approaches are shown: directDIA, using the DIA files to generate the library; own DpD (DDA plus DIA), using the participant's data to generate a library; other DpD, using the other participant's DDA and DIA-based library; combined DpD, a combined library of the DDA and DIA data from both participants; Pan-Human, the Pan-Human library augmented with empirical evidence of the 4 nonendogenous spike-in proteins; Library-free CLib, chromatogram library; Pan-Human CLib, chromatogram library combined with Pan-human plus spikes; and Prosit CLib, chromatogram library with Prosit-generated spectra. Only Participant 3 generated a chromatogram library. **A-B**. Proteins identified using Participant 48 data (**A**) or Participant 3 data (**B**). Hollow points are total identifications per sample with a solid line being the average identifications. Solid points are the number of proteins below 20% coefficient of variation (CV), with the dotted line being average proteins below 20% CV. **C-D**. Estimated abundance of the 4 spike-in proteins using the different library approaches for Participant 48 (**C**) and 3 (**D**). Log2 fold-change with 95% confidence intervals for the 4 spike-in proteins was determined between the 10 fmol/µg (sample B) and 2.5 fmol/µg (sample A) HeLa digest samples. The expected value was 2 (dotted line). Within each set of 4 points, left to right are ABRF-1, -2, -3, and -4, corresponding to the 4 spike-in proteins (see the Methods section for names). Figures derived from Supplemental Tables S2 to S4.

METHODS

Study samples

Samples were prepared using HeLa cells that were released from cell culture plates using trypsin. The cells were washed with phosphate-buffered saline, and cell pellets were dispersed in MS-grade water and then disrupted by sonication and diluted to a final protein concentration of 1 mg/mL. All digests were carried out using Promega trypsin with overnight incubations at 37 °C in 40 mM triethylammonium bicarbonate buffer and after reducing the sample with 10 mM dithiothreitol (DTT) and alkylating with 30 mM indole-3-acetic acid (IAA). Exogenous proteins were solubilized in MS-grade water and quantified from their absorbance spectra using calculated extinction coefficients.[28] Equimolar amounts of the 4 proteins were combined prior to reduction and alkylation with DTT and IAA and then digestion with trypsin. Digests of HeLa and the exogenous protein mix were desalted using Oasis HLB (Waters) cartridges with a single step elution in 65% (volume fraction) acetonitrile. The 4 exogenous proteins are the following: beta-D-galactosidase from *Escherichia coli* (Sigma, catalog number G8511), protein G from *Streptococcus aureus* (Sigma, catalog number P4689), lysozyme from *Gallus gallus* (Sigma, catalog number L6876), and amyloglucosidase from *Aspergillus niger* (Sigma, catalog number A7420). The digest of the exogenous proteins was added to the HeLa lysate to achieve a concentration of 1 µM for each protein. This stock was diluted with the base HeLa digest to obtain 10, 2.5, or 0 fmol of the added proteins per 1 µg HeLa digest (sample A: 2.5 fmol spike, sample B: 10 fmol spike, and sample C was no spike). Standard iRT peptides (Biognosys) were added to the HeLa plus exogenous protein mixtures. The 3 study samples of HeLa digest with exogenous proteins and iRT peptides were made 1 time and were aliquoted into 10 µg HeLa digest aliquots in 0.5-mL LoBind tubes. These were dried by SpeedVac and stored at -80 °C until shipped. Shipping was at ambient temperature.

Study advertisement, enrollment, and timeline

PRG members designed the study and announced it at the annual ABRF conference in April 2018. The study was also advertised at the annual conference of the American Society of Mass Spectrometry in June 2018. Interested participants' contact details were collected via Google Survey, and the distribution of samples began in September 2018, with the majority of participants receiving the samples by November 2018. Participating laboratories were located in 20 countries and 16 US states. The deadline for data return was extended to June 2019 to accommodate requests by some of the participants. Of the 63 participants who received samples, 45 laboratories returned datasets. Four participants performed multiple methods or acquired data on multiple instruments, resulting in 49 total datasets.

Information given to participants

Each participant received a numerical study ID with the samples. The participants study ID was and is known only to that investigator and to the anonymizer. Documentation with information about study design, sample preparation, data acquisition, and deposition was distributed electronically. This information is included in

supplementary information (Supplemental File 1), although it has been edited from its original form to remove vendor contact information. The study documentation included suggestions for reconstituting the samples, LC gradient conditions, and DIA data acquisition settings for the following platforms: Thermo Fusion and Fusion Lumos, Thermo QE-HFX, Sciex TripleTOF, and Waters Xevo G2 XS platforms. Participants were encouraged to request guidance from members of the PRG if their platform was not included in the original guidelines. For those few investigators, a best attempt was made to design methods with approximately the same DIA cycle time. Finally, there were instructions on how to label the acquired data files and to complete and upload a survey that included self-reported metadata. Throughout the process, participants were encouraged and given the means to remain anonymous even when securing technical assistance.

PRG-suggested sample resuspension and LC conditions

Participants received 3 dried samples that have been described. Participants using microflow received 2 complete sets of the 3 samples. The suggested method was to bring each up in 0.1% formic acid but did not specify the volume. It was expected that nanoflow systems would inject 1 to 2 µg on column, whereas microflow systems might require 4 to 8 µg on column. Participants had discretion to decide the appropriate injection amount for their system and to prepare the samples to allow for replicate injections.

Because of the diversity in LC systems and the latitude for participants to use either nano- or microflow applications, we relied on participants to design appropriate gradients that fit within basic guidelines. The suggestion for the study was a 2-stage linear gradient lasting 110 to 130 minutes that we designated as the PRG gradient. The following was suggested: equilibration (trap or direct load) followed by a step from 5% to 25% acetonitrile over 100 minutes, then 25% to 40% acetonitrile over 20 minutes, and finally 40% to 90% over 10 minutes. The final plateau could be held for 5 minutes before returning to 5% acetonitrile over 1 minute followed by re-equilibration. Roughly half of the datasets reportedly used the PRG gradient (23 of 49 datasets), while 19 of the 49 datasets included specific gradient information that was deposited along with raw files on MassIVE. In general, a multi-step 2-hour separation was performed by all participants. Participants were blinded to the sample identity, so a run order of A, B, C, blank, B, C, A, blank, C, B, A was suggested to minimize systemic bias due to carryover.

PRG-suggested DIA conditions

When constructing a DIA experiment, the MS2 mass range, number of MS2 windows, MS2 window width, and time spent acquiring data all contribute to establishing the instrument's cycle time, which is the time taken to scan all DIA windows 1 time, and therefore how many data points are acquired during a peptide's elution peak. For this study, we recommended static MS2 window widths covering 400 to 1200 *m/z*, with a 1 *m/z* overlap. As an example, this means that 1 window would stop at 420 *m/z*, and the next window would start at 419 *m/z*. The majority of participants followed this recommendation (37 of 49 participants), but other strategies were selected by some participants (Table 2; Supplemental Table S1). The design of the study would produce a

method with a 3.5-second cycle time. We assumed a 30-second peak width at base, and therefore a 3.5-second cycle would produce between 7 and 10 data points per peak. We were aware that for participants with tighter peaks, this would under sample. Overall, the parameters selected by participants largely achieved a 3.5-second cycle (Supplemental Table S1) as confirmed by an evaluation of the window strategies using Skyline 4.2.0.19107 for each SA_R1 raw file for each submission. These are also reported in the windows.txt and windows.png files on MassIVE MSV000086479.

Specifying the instrument data acquisition time was difficult because of the diversity of platforms and instrument types. For example, for trap-based instruments, the acquisition time is related to the transient time, automatic gain control (AGC) target, and maximum injection time. We provided general recommendations for QE-HFX, Fusion, and Fusion Lumos and personalized recommendations for others, in which given resolutions, with known transient times and maximum injection times, could be suggested to achieve a 3.5-second cycle. For non-trap–based instruments (such as the triple TOF line), it was much easier to specify instrument time because it was part of the method. In general, though, we suggested the following: For the Fusion and Fusion Lumos, we suggested 40 windows 21 *m/z* wide at 30 000 resolution or 62 DIA windows 14 *m/z* wide at 15 000 resolution. For the QE-HFX, we suggested 40 windows 21 *m/z* wide at 30 000 resolution. For tripleTOFs, we suggested 80 DIA windows 11 *m/z* wide. For specific recommended settings, such as maximum injection time and AGC (for trap based) or collision energy, please see Supplemental File 1 or consult the actual settings of users in Supplemental Table S1.

Participant actions

The actions and results of participants 3 and 48 will be discussed in detail.

Participant 3 LC and DIA conditions

Participant 3 self-reported 6 to 9 years of liquid chromatography tandem mass spectrometry (LC-MS/MS) experience, had performed DIA a couple of times, and used a Thermo Fusion Lumos. The 3 samples were brought up in 20 µL 0.1% (volume fraction) formic acid to approximately 0.5 µg/µL. Peptide mixtures (2 µL injection; approximately 1 µg) were run in the order specified: A, B, C, blank, B, C, A, blank, C, B, A. The analysis was performed using an UltiMate 3000 Nano LC coupled to a Fusion Lumos mass spectrometer (Thermo Fisher Scientific) with a nano-ESI source. A trap/elute setup was used by trapping with a PepMap 100 C18 trap column (75 µm id × 2 cm length; Thermo Fisher Scientific) at 3 µL/min for 10 minutes with 2% acetonitrile (volume fraction) and 0.05% trifluoroacetic acid (volume fraction) followed by separation on an Acclaim PepMap RSLC 2 µm C18 column (75 µm id × 25 cm length; Thermo Fisher Scientific) at 40 °C. Peptides were separated along the suggested PRG LC gradient, except that the suggested 90% acetonitrile (volume fraction) was not possible with the mobile phase setup used. Specifically, a 130-minute gradient of 5% to 32% mobile phase B (80% acetonitrile [volume fraction], 0.08% formic acid [volume fraction]) over 100

minutes, followed by a ramp to 50% mobile phase B over 20 minutes, and lastly to 95% mobile phase B over 10 minutes at a flow rate of 300 nL/min.

Instrument acquisition settings for DIA were exactly those suggested for the 62 windows 14 *m/z* width at 15 000 fragment ion scan resolution (Supplemental File 1). Specifically, a default charge of 4 was used, no internal mass calibration was used, and the ion funnel radio frequency (RF) was 30%, a full-scan resolution of 120 000 (determined at 200 *m/z*), with an ion target value of 1.0×10^6 and maximum injection of 20 ms. Full-scan data were acquired from 393 to 1200 *m/z* in profile mode. For DIA settings, quad isolation was set at 14 *m/z*, and a list of 62 mass centers was used to accomplish the suggested DIA window scheme, starting at 400 *m/z* and ending at 1193 *m/z*. This resulted in 62 DIA windows of 14 *m/z* width with 1 *m/z* overlap on edge of each window (eg, one window would stop at 420 *m/z* and the next would begin at 419 *m/z*). Fragmentation was performed using higher-energy collisional dissociation (HCD) at a normalized collision energy of 32. Fragmentation profile data were collected from 200 to 2000 *m/z* at 15 000 resolution. The maximum injection time was 30 ms with an ion target value of 1.0×10^6, and inject parallelizable ions was set to off. Data were acquired under Tune version 2.1 in XCalibur 4.0.

Participant 48 LC and DIA conditions

Participant 48 self-reported >10 years LC-MS/MS experience, was an expert in DIA, and had used a Thermo Fusion Lumos. The 3 samples were brought up in 20 µL 0.1% (volume fraction) formic acid to approximately 0.5 µg/µL. Peptide mixtures (approximately 1 µg) were run in the order specified: A, B, C, blank, B, C, A, blank, C, B, A. The analysis was performed using a Nano Acquity (Waters) coupled to a Fusion Lumos mass spectrometer (Thermo Fisher Scientific). A trap/elute setup was used by trapping with a trap column (nanoAcquity Symmetry C18, 5 µm, 180 µm × 20 mm) and an analytical column (nanoAcquity BEH C18, 1.7 µm, 75 µm × 250 mm). The outlet of the analytical column was coupled directly to the MS using a Proxeon nanospray source. The peptides were introduced into the mass spectrometer via a PicoTip Emitter (360 µm OD × 20 µm ID; 10 µm tip [New Objective]), and a spray voltage of 2.2 kV was applied. The capillary temperature was set at 300 °C. Mobile phase A was water with 0.1% formic acid (volume fraction), and mobile phase B was acetonitrile with 0.1% formic acid (volume fraction). The samples were loaded with a constant flow of mobile phase A (5 µL/min) onto the trapping column. Trapping time was 6 minutes. Peptides were eluted via the analytical column with a constant flow of 300 nL/min with the analytical column held at 40 °C. Peptides were separated along the suggested PRG LC gradient (Supplemental File 1).

Instrument acquisition settings for DIA were exactly those suggested for the 40 windows 21 *m/z* width at 30 000 fragment ion scan resolution (Supplemental File 1). Specifically, a default charge of 4 was used, internal mass calibration was used, and the ion funnel RF was 30%, a full-scan resolution of 120 000 (determined at 200 *m/z*), with an ion target value of 1.0×10^6 and a maximum injection of 20 ms. Full-scan data were acquired from 399 to 1200 *m/z* in profile mode. For DIA settings, quad isolation was set at 21 *m/z*, and a list of 40 mass centers were used to accomplish the suggested DIA window scheme, starting at 409.5 *m/z* (center

mass) and ending at 1189.5 *m/z* (center mass). This resulted in 40 DIA windows of 21 *m/z* width with 1 *m/z* overlap on the edge of each window. Fragmentation was performed using HCD at a normalized collision energy of 30. Profile data were collected from 200 to 2000 *m/z* at 30 000 resolution. The maximum injection time was 60 ms with an ion target value of 1.0×10^6, and inject parallelizable ions was set to True. Data were acquired under Tune version 2.1 in XCalibur 4.0.

Participant 3 DDA and gas-phase fractionation runs

Participant 3 also performed additional analyses in order to provide data used for constructing spectral and chromatogram libraries. The remaining amounts (approximately 12 µL) of samples A (2.5 fmol spike) and B (10 fmol spike) were combined to obtain a solution that contained the spiked-in proteins at approximately 6 fmol spike per µg HeLa digest. The same conditions were used as specified for DIA, including the amount of sample injected and the gradient used. Data-acquisition settings were changed to standard DDA. For the DDA runs, the Fusion Lumos was operated in positive polarity and data-dependent mode (topN, 3-s cycle time) with a dynamic exclusion of 60 seconds (with 10 ppm error). Full-scan resolution using the orbitrap was set at 120 000, and the mass range was set to 375 to 1500 *m/z* collected in profile mode. A full-scan ion target value was 4.0×10^5, allowing for a maximum injection time of 50 ms. Monoisotopic peak determination was used, specifying peptides, and an intensity threshold of 1.0×10^4 was used for precursor selection. Data-dependent fragmentation was performed using HCD at a normalized collision energy of 32 with quadrupole isolation at 0.7 *m/z* width. The fragment-scan resolution using the orbitrap was set at 30 000, with 110 *m/z* as the first mass, an ion target value of 2.0×10^5, and a 60-ms maximum injection time, and the data type was set to centroid.

To enable chromatogram library construction, “gas-phase fractionation” was performed.[29] The same injection volume and gradient were used. Five successive runs were performed using a staggered window approach described in detail by Searle et al.[29] Briefly, a series of nonoverlapping 4-*m/z*-wide DIA windows are collected over a short enough mass range to maintain a reasonable DIA cycle. Then, the cycle repeats but offset by 2 *m/z*. This is repeated multiple times so that the full desired precursor mass range is covered. In the case of Participant 3, there were 5 runs, each with 2 cycles of 40 windows that were 4 *m/z* wide (detailed in Searle et al).[29] The first run went 400 to 560 *m/z* and then 398 to 558 *m/z*. The next 4 runs were 560 to 720 *m/z*, 720 to 880 *m/z*, 880 to 1040 *m/z*, and 1040 to 1200 *m/z*. The raw file names were *TW1, *TW2, *TW3, *TW4, and *TW5, respectively, shorthand for tight window. For each run, the instrument specifics were as follows: The Fusion Lumos was operated in positive polarity, and no full-scan data were acquired. Fragmentation was performed using HCD at a normalized collision energy of 32 with quadrupole isolation at 4 *m/z* width in conjunction with the 2 lists of 40 window centers. Fragment-scan resolution using the orbitrap was set at 30 000, and the mass range was set to 200 to 2000 *m/z* collected in profile mode. The default charge was 4, the RF lens was 30%, and the ion target value was 1.0×10^6, allowing for a maximum injection time of 60 ms.

Participant 48 DDA runs (Lumos)

Participant 48 also acquired additional data for library construction. The remaining parts of samples A (2.5 fmol spike) and B (10 fmol spike) were combined to obtain a solution that was approximately 6 fmol spike per µg HeLa digest. The same conditions were used as specified for DIA, including 1 µg injection and the same gradient, with the main change being data acquisition settings. For the data-dependent acquisition runs, the Fusion Lumos was operated in positive polarity and data-dependent mode (topN, 3-second cycle time) with a dynamic exclusion of 15 seconds (with 10 ppm error). Full-scan resolution using the orbitrap was set at 60 000, and the mass range was set to 375 to 1500 *m/z* collected in profile mode. The full-scan ion target value was 2.0×10^5, allowing for a maximum injection time of 50 ms. Monoisotopic peak determination was used, specifying peptides, and an intensity threshold of 5.0×10^4 was used for precursor selection. Only multiply charged (2+ to 7+) precursor ions were selected for fragmentation. Isotopes were excluded. Data-dependent fragmentation was performed using HCD at a normalized collision energy of 30 with quadrupole isolation at 1.4 *m/z* width. The fragment-scan resolution using the orbitrap was set at 15 000, with 120 *m/z* as the first mass, an ion target value of 2.0×10^5, and a 22-ms maximum injection time, and the data type was set to centroid.

Participant 48 DDA/DIA on spike-in proteins (Lumos and QE)

Participant 48 also performed additional analyses of the spike-in proteins alone to be included as a library when the data was searched against the Pan-Human library.[30] Participant 48 was supplied with a tryptic digest of approximately 16.9 pmol of each protein. The sample was resuspended in 170 µL (approximately 100 fmol/µL), the iRT kit was added, and 4 injections were made in DDA and DIA, respectively, with the amount on the column ranging from 100 to 800 fmol (100, 200, 400, 800). These DDA and DIA runs of the spike-in proteins are described below.

The same conditions were used for the LC, with the exception that the nanoUPLC hardware was an M-Class NanoAcquity from Waters. The same gradient was applied (2-step PRG recommended), with the main change being the MS instrument settings for the QE-HFX (Thermo). For the data-dependent acquisition runs, the QE-HFX was operated in positive polarity and data-dependent mode (top15) with a dynamic exclusion of 20 seconds (with 10 ppm error). The full-scan resolution using the orbitrap was set at 60 000, and the mass range was set to 350 to 1650 *m/z* collected in profile mode. The full-scan ion target value was 3.0×10^6, allowing for a maximum injection time of 20 ms. The peptide setting was set to “preferred,” and an intensity threshold of 1.0×10^4 was used for precursor selection and an AGC of 1.0×10^3. Only multiply charged (2+ to 5+) precursor ions were selected for fragmentation. Isotopes were excluded. Data-dependent fragmentation was performed using HCD at a normalized collision energy of 27 with quadrupole isolation at 1.6 *m/z* width. The fragment-scan resolution using the orbitrap was set at 15 000, with 120 *m/z* as the first mass, an ion target value of 2.0×10^5, and a 25-ms maximum injection time, and the data type was set to profile. The default charge state was set to 2+. Data were acquired under Tune version 2.9 in XCalibur 4.0.

For the data-independent acquisition runs, the same conditions as for the DDA were applied to the LC gradient. The following parameters were different for the QE-HFX acquisition: MS1 full scans were acquired using the orbitrap resolution set at 120 000, the mass range was set to 400 to 1200 *m/z*, and data were collected in profile mode. The full-scan ion target value was 3.0×10^6, allowing for a maximum injection time of 20 ms. Data-independent scans were set to 40 fixed windows, each of width 21 *m/z* (the same as for the Lumos DIA by Participant 48). A maximum injection time of 60 ms was set with an ion target value of 3.0×10^6. Fragmentation (HCD) for the DIA scans in MS2 was carried out with a normalized collision energy of 30, the first MS2 mass was set to 200 *m/z*, and the data type was set to profile. The default charge state was set to 3+.

Analysis of Participant 3 and 48 data

A preliminary analysis of the majority of participants' data was presented at the ABRF 2019 meeting and is available for reference.[27] Herein, we describe the analysis of 2 participants using Spectronaut (v13.6.190905.43655; Biognosys AG)[31] and Scaffold DIA (v1.3.1; Proteome Software). Spectronaut and Scaffold DIA are 2 of the many available software packages capable of DIA analysis. They were selected for this project because of the expertise of the authors. We recommend similar analysis in other programs (see the section Data Usage), although the settings listed may not necessarily translate. Participants 3 and 48 both performed the replicate analysis of 3 samples using an Orbitrap Fusion Lumos. They both collected DDA in replicate of a combined sample, while data of just the digested spike proteins were only collected on a QE-HFX by Participant 48. This allowed for the comparison of different library generation techniques within Spectronaut: directDIA, where only the DIA data is used; DpD, directDIA plus DDA where separate search archives (ie, libraries) are constructed of the DIA data and the DDA and then combined; and Pan-Human plus spikes, where a search archive of the spikes alone was combined with the Pan-Human library.[30] Since only Participant 3 collected data using gas-phase fractionation, this was used for generating a chromatogram library in Scaffold DIA and was not utilized for Participant 48's data.

The following settings were used in Spectronaut for directDIA libraries (setting tabs are bold), which can be retrieved as .xls and .kit files on MassIVE MSV000086479 as 03_lumos_directDIA and 48_lumos_directDIA. **Sequences:** Trypsin/P selected, maximum pep length 52, minimum pep length 7, 2 missed cleavages, KR as special amino acids in decoy generation, and toggle N-terminal M set to true. **Labeling:** no labeling settings were used. **Applied modifications:** maximum of 5 variable modifications using fixed carbamidomethyl (C), variable acetyl (protein N-term), and oxidation (M). **Identification:** per run machine learning, Q-value cutoff of 0.01 for precursors and proteins, single hits defined by stripped sequence and do not exclude single hit proteins, PTM localization set to true with a probability cutoff of 0.75, and kernel density p-value estimator. **Quantification**: interference correction was used with excluding all multichannel interferences with a minimum of 2 and 3 for MS1 and MS2, respectively. Proteotypicity filter was set to none, major protein grouping by protein group ID, minor peptide grouping by stripped sequence, major group quantity set to mean peptide quantity, a Major Group Top N was used (minimum of 1, maximum of 3), minor group quantity set to

mean precursor quantity, a Minor Group Top N was used (minimum of 1, maximum of 3), quantity MS-level used MS2 area, data filtering by q-value, cross run normalization was used with global median normalization and automatic row selection, no modifications or amino acids were specified, and best N fragments per peptide was set to between 3 and 6, with ion charge and type not used. **Workflow:** no workflow was used. **Post analysis**: no calculated explained total ion current (TIC) or sample correlation matrix, differential abundance grouping using major group (from quantification settings) and smallest quantitative unit defined by precursor ion (from quantification settings), differential abundance was not used for conclusions, but the following settings were used in the attached files: Student's *t* test, no groupwise testing correction, run clustering was set using the Manhattan distance metric and Ward's method for linkage strategy, and runs were ordered by clustering without *z*-score transformation. The fasta files used are included in MassIVE MSV000086479, but, briefly, the UniProtKB Swiss-Prot 2018_06 human database (taxonomy:9606), canonical only was concatenated with the 4 spike protein entries (ABRF-1 P00722 beta-galactosidase [*Escherichia coli*], ABRF-2 P00698 lysozyme C [*Gallus gallus*], ABRF-3 P69328 glucoamylase [*Aspergillus niger*], and ABRF-4 Q54181 protein G' [*Streptococcus* sp. group G]), the iRT Fusion sequence supplied by Biognosys (LGGNEQVTRYILAGVENSKGTFIIDPGGVIRGTFIIDPAAVIRGAGSSEPVTGLDAKTPVISGGPYEYRVE ATFGVDESNAKTPVITGAPYEYRDGLDAASYYAPVRADVTPADFSEWSKLFLQFGAQGSPFLK), and a contaminants database of 247 entries. These 2 .fasta files are on MassIVE MSV000086479 as sp_human_180620_plus_PRG_ABRF_4_prot.fasta and contaminants_ 20120713.fasta, respectively.

For DpD approaches, the DDA data was used to generate a search archive with Pulsar using the same settings described for directDIA, and the resulting search archive was combined with a library made directly from the DIA data using the settings described above. There were 3 DpD libraries created: each participant individually and then a combined library. These are included as 03_lumos_DpD, 48_lumos_DpD, and 03_lumos_48_lumos_DpD as .xls and .kit on MassIVE MSV000086479. The Pan-Human search archive was downloaded within Spectronaut and is also available as Pan Human Library – ETH .xls and .kit on MassIVE MSV000086479. This was combined with a directDIA plus DDA library of digested spike proteins and is on MassIVE MSV000086479 as 48_qehfx_spikes_DpD .xls and .kit.

When searching with the DpD search archives or the Pan-Human–derived archive, the following settings were used: **Data extraction:** maximum intensity extraction for MS1 and MS2, dynamic MS1 mass tolerance strategy with a correlation factor of 1, and a dynamic MS2 mass tolerance strategy with a correction of 1. **XIC extraction:** a dynamic XIC RT extraction window was used with a correction factor of 1. **Calibration:** allowed source-specific iRT calibration with an automatic calibration mode, used a maximum intensity MZ extraction strategy, precision iRT was set to true with excluded deamidated peptides and a local (nonlinear) iRT<->RT regression type, used Biognosys iRT kit, and no calibration carryover. **Identification:** same settings as used in directDIA. **Quantification:** same as in directDIA. **Workflow:** no in silico library optimization, multichannel workflow definition from library annotation with a fallback option as labeled, and no profiling or

unify peptide peaks strategy was used. **Protein inference:** automatic. **Post analysis:** same settings as used in directDIA.

Because Participant 3 also collected gas-phase fractionation data (described above), this DIA data was also processed in Scaffold DIA (v1.3.1) 3 different ways: (1) by creating a chromatogram library using only the gas-phase fractionation data, (2) using these data combined with the Pan-Human library, and (3) using these data combined with a Prosit in silico library. The Pan-Human library was converted directly using EncyclopeDIA (v0.8.1)[29] from phl004_canonical_sall_osw.csv, downloaded from the SwathAtlas repository (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetDIALibs). The Prosit predictions used the EncyclopeDIA library generation defaults.[23] These defaults were predictions for +2H/+3H peptides between 396.4 and 1002.7 *m/z* with up to 1 missed cleavage. The normalized collision energy (NCE) setting was 33, assuming all peptides were fragmented in DIA as +2H peptides. The additional 4 ABRF peptides were predicted using the same pipeline but for all +2H/+3H/+4H/+5H peptides between 396.4 and 1002.7 *m/z* with up to 2 missed cleavage. Again, the NCE setting was 33, assuming all peptides were fragmented in DIA as +2H peptides. The resulting library files for Pan-Human and Prosit human plus PRG spike proteins can be found on MassIVE MSV000086479 as combined_prg_pan_human.dlib and combined_prg_sprothuman.dlib, respectively.

All raw data files were converted to mzML format (within Scaffold DIA) using ProteoWizard (v3.0.18342).[32] In the first case in which an external library was not used, the reference spectral library was created by EncyclopeDIA (v0.8.1).[29] Reference samples were individually searched against the same fasta described above, with a peptide mass tolerance of 10.0 ppm and a fragment mass tolerance of 10.0 ppm. The fixed modifications considered were the following: Carbamidomethylation C. In the second case, when combining the data with the Pan-Human library, the reference spectral library files were individually searched against a combined fasta and the dlib with a peptide mass tolerance of 10.0 ppm and a fragment mass tolerance of 10.0 ppm. The variable modifications considered were the following: Oxidation M and Carbamidomethylation C (since these are used in the Pan-Human library). In the third case, when combining the data with Prosit predictions, the reference spectral library files were individually searched against a combined fasta and the dlib with a peptide mass tolerance of 10.0 ppm and a fragment mass tolerance of 10.0 ppm. The variable modifications considered were the following: Carbamidomethylation C.

For all 3 search approaches, the digestion enzyme was assumed to be trypsin with a maximum of 1 missed cleavage site allowed. Only peptides with charges in the range +2 to +3 and length in the range 6 to 30 were considered. Peptides identified in each search were filtered by Percolator (v3.01.nightly-13-655e4c7-dirty)[33], [34],[35] to achieve a maximum false discovery rate (FDR) of 0.01. Individual search results were combined, and peptides were again filtered to an FDR threshold of 0.01 for inclusion in the reference library.

Analytical samples (ie, the replicate injections of the 3 ABRF samples Participant 3 analyzed) were aligned based on retention times and individually searched against 03 Chromatogram Library.elib, 03 - PH plus CL

library.elib, or 03 - Prosit plus CL library.elib (created as described above and available on MassIVE MSV000086479) with search settings identical to those used to create the reference library. Peptide quantification was performed by EncyclopeDIA (v0.8.1).[29] For each peptide, the 5 highest quality fragment ions were selected for quantitation. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis were grouped to satisfy the principles of parsimony. Proteins with a minimum of 2 identified peptides were thresholded to achieve a protein FDR threshold of 1.0%. These files are available as 03 - CL only.sdia, 03 - PH plus CL.sdia, and 03 - Prosit plus CL.sdia.

For all approaches, intensity values of the 4 spike-in proteins were used to compare relative quantification between the different approaches. Specifically, Sample A versus Sample B was used to evaluate how well each approach measured the predicted 4-fold difference in protein concentration. To easily calculate the 95% confidence interval of each fold-change, the topTable function within the limma package (v3.40.6)[36] in R (v3.6.0; 64-bit) was used with the argument “confint=TRUE.” To accomplish this, first exported intensity values were transformed to log2 values, and this matrix was used with limma. A summary figure outlining the workflow for the data from Participants 3 and 48 is shown in Figure 4.

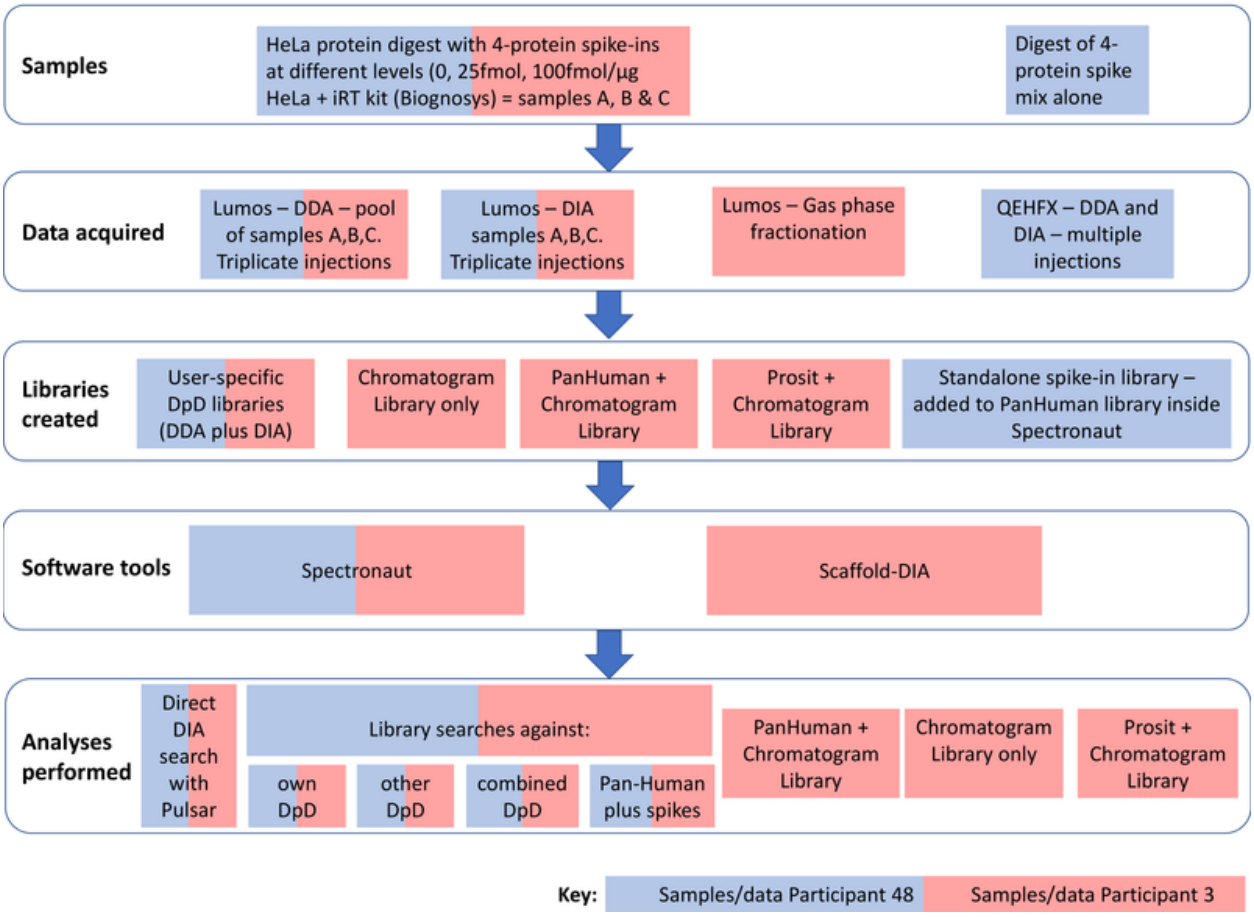

Figure 4
Summary flowchart showing steps taken from samples assigned to Participants 3 and 48.

TECHNICAL VALIDATION

Data return and curation

Participants uploaded their data to a private FTP server hosted by MassIVE. Each participant was given a folder designated by their participant number, and the file naming scheme was described in Supplemental File 1. The instrument-naming scheme was changed for Supplemental Table S1 to reflect instrument names following the PSI-MS recommended names (https://github.com/HUPO-PSI/psi-ms-CV/blob/master/psi-ms.obo). Following the end of the study, file integrity was confirmed by opening each file. In some cases, the file was corrupt, and the participant was requested by the anonymizer to reupload the files. In some cases, the original file was also corrupt, and those data were not available. In the case of missing files, we made every effort with the participant to find and upload the missing data. Despite these efforts, not all participants were able to provide the requested 9 raw data files.

Once the data was curated, files were manually inspected using the TIC to look for any noticeable qualities, such as TIC without peaks. Notes were made in Supplemental Table S1. The majority of replicates were consistent, although it should be noted that this does not imply a measure of data quality. Next, the embedded parameters in the raw files were used to determine and/or confirm MS-acquisition settings. Though participants were encouraged to self-report MS settings, there was missing information and discrepancies. To avoid reliance on the participant, the first replicate of sample A was used to determine DIA window scheme, MS1 and MS2 resolution (and injection time if applicable), and DIA cycle time. All other files from that participant were assumed to have the same acquisition settings. In the case of Sciex TripleTOFs, participants reported MS1 and MS2 settings. The DIA windowing strategy was determined using Skyline,[37] while the scan header provided MS1 and MS2 information (in the case of Thermo instruments). The DIA cycle time was determined using Spectronaut (v13.6.190905.43655; Biognosys AG).[31] This information is provided in Supplemental Table S1.

Although LC conditions were not available from data files, many participants submitted LC gradient specifics along with raw data. These are included in Supplemental Table S1 as well as in the MassIVE MSV000086479. Finally, in cases in which those metadata could not be surmised and were not self-reported, we contacted the participant directly to request that information. After these efforts, there are still some participants with missing information. This is noted in Supplemental Table S1.

Survey

Participants were given the option to self-report the specific LC columns they used, the LC parameters, the MS instrument settings, and any attempt to identify the amount of spike in proteins in the labeled samples. These survey questions are provided in Supplemental File 2.

DATA AVAILABILITY

All data including raw files, fasta, search result files, and data keys are available on the online public data repository MassiVE (https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000086479). Details explaining the nature of each file available for download are also shown in Supplemental Table S5.

USAGE NOTES

All raw files are available on MassIVE MSV000086479, and many programs can use these files directly. The software to analyze DIA includes, but is not limited to, the following: DIA-NN,[38] DIA-Umpire,[39] EncyclopeDIA,[29] OpenSWATH,[40] PEAKS Studio X (Bioinformatics Solutions Inc.), PECAN,[41] Protalizer (Vulcan Analytical), Scaffold DIA (Proteome Software), Skyline,[37] and Spectronaut (Biognosys AG)[31] as well as DIA-specific statistical packages (ex. iq: Protein Quantification in Mass Spectrometry-Based Proteomics; https://CRAN.R-project.org/package=iq). Many of these programs have excellent online resources, including tutorials of the analysis pipeline. The relevant information, such as DIA window placement or instrument settings, can be found in Supplemental Table S1 and in supplemental files on

MassIVE MSV000086479. Specifically to the analysis performed in this paper, we have included .sne (Spectronaut) and .sdia (Scaffold DIA) files, which are located on MassIVE MSV000086479 and can be opened with free viewers for these programs. The libraries used in the analysis can be found on MassIVE MSV000086479 as .kit, .xls, .dlib, or .elib files and can be used directly by some of the software listed. Alternatively, the DDA files available on MassIVE MSV000086479 can be used to create libraries. It should be noted that all samples included the iRT peptides, which can be used, if needed, to map the elution patterns into iRT space. Finally, in the case of in silico libraries such as Prosit[22] and MS2PIP,[24],[26] the original publication or tutorials should be consulted for instructions for how to combine with empirical data.[9],[23]

CODE AVAILABILITY

An analysis of raw data was performed using Spectronaut (v13.6.190905.43655; Biognosys AG)[31] and Scaffold DIA (v1.3.1; Proteome Software). No other code was used for this data generation or example analysis.

AUTHOR CONTRIBUTIONS

JK assisted in study design and execution, analyzed data, and wrote the manuscript with contributions from all authors. PS assisted in study design and execution, feedback, and edits on the manuscript and prepared experimental samples. BCS assisted with data analysis, the design and generation of figures, feedback, and edits on the manuscript. LH assisted in study design and execution, the design and generation of figures, feedback, and edits on the manuscript. LM assisted in study design and execution, feedback, and edits on the manuscript. MM assisted in study design and execution, feedback, and edits on the manuscript. BP assisted in study design and execution, feedback, and edits on the manuscript. BS assisted in study design and execution, feedback, and edits on the manuscript. MP assisted in study design and execution, feedback, and edits on the manuscript. YW assisted in study design and execution, feedback, and edits on the manuscript and developed the participant survey. PJ assisted in study design and execution, feedback, and edits on the manuscript. BN assisted in study design and execution, analyzed data, curated study data, and wrote the manuscript with contributions from all authors.

ACKNOWLEDGMENTS

The authors would like to thank all the participants who took the time to measure and return data for this study. Of the 45 participating laboratories, 9 wished to remain anonymous, and 5 are authors (BN, JK, MM, LM, and YW). We wish to thank and acknowledge these 31 participants (listed in no particular order): Alex Campos, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA; Andrea Petretto and Chiara Lavarello, IRCCS Istituto G. Gaslini, Genoa, Italy; Bianka Świderska, Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics Polish Academy of Sciences, Warsaw, Poland; Bin Fang, Proteomics and Metabolomics Core Facility, Moffitt Cancer Center, Tampa, FL, USA (supported by National Cancer Institute [NCI] cancer center supporting Grant P30-CA076292); Büşra Aytül Akarlar and Nurhan Özlü, Koç University

Proteomics Facility, Istanbul, Turkey; Jasjot Singh and Dominic Winter, Institute for Biochemistry and Molecular Biology, University of Bonn, Bonn, Germany; Medicharla V. Jagannadham, CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India; Sadr ul Shaheed and Chris Sutton, University of Oxford United Kingdom and Institute of Cancer Therapeutics, University of Bradford, Bradford, West Yorkshire, United Kingdom; Renu Goel, Translational Health Science and Technology Institute, Faridabad, India; Eric Spooner, Whitehead Institute for Biomedical Research, Cambridge, MA, USA; Gabriela Grigorean, Mass Spectrometry Facility, University of Michigan, Ann Arbor, MI, USA; Marcel P. de Vries and Justina C. Wolters, University Medical Center Groningen, University of Groningen, Groningen, Netherlands; Jeremy L. Balsbaugh, UConn Proteomics & Metabolomics Facility, University of Connecticut, Storrs, CT, USA; Juraj Lenčo, Charles University, Prague, Czechia; Kristyna Pimkova and Jenny Hansson, Stem Cell Centre, Lund University, Lund, Sweden; Liisa Arike, Institute of Biomedicine, University of Gothenburg, Sweden; Liyan Chen and Radoslaw M. Sobota, Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Singapore; Matt Padula, Proteomics Core Facility, University of Technology Sydney, Ultimo, Australia; Michael Ford, MS Bioworks, Ann Arbor, MI, USA; Nazlı Ezgi Özkan-Küçük and Nurhan Özlü, Omics Laboratory, Koc University Research Center for Translational Medicine, Istanbul, Turkey; Peter Hains, Children's Medical Research Institute, Australia; Yingchun Zhao and Peter Villalta, Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA; Anuli C. Uzozie and Philipp F. Lange, University of British Columbia, Vancouver, Canada; Tejan Lodhiya and Raju Mukherjee, Indian Institute of Science Education and Research Tirupati, India; Romina Belli and Daniele Peroni, Facility of Mass Spectrometry and Proteomics, University of Trento, Trento, Italy; Sanjeeva Srivastava and Vipin Kumar, IIT Bombay, Mumbai, India; Sophie Braga-Lagache, Proteomics Mass Spectrometry Core Facility Bern, Switzerland; Sylvie Bourassa and Arnaud Droit, CHU de Quebec and Laval University, Quebec, Canada; Tobias Kockmann, Functional Genomics Center Zurich, ETH Zurich/University of Zurich, Zurich, Switzerland; Valdemir Melechco Carvalho, Fleury Group, São Paulo, Brazil; and Zach Rolfs and Lloyd Smith, University of Wisconsin–Madison, Madison, WI, USA.

The authors also wish to thank Matt Herring for graphical advice when designing figures. PDJ was supported by NCI Informatics Technology for Cancer Research Grant 1U24CA199347 and National Science Foundation (U.S.) Grant 1458524. This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001999), the United Kingdom Medical Research Council (FC001999), and the Wellcome Trust (FC001999). The Fritz Lipmann Institute is a member of the Leibniz Association and is financially supported by the Federal Government of Germany and the State of Thuringia. We would also like to thank the software companies (Biognosys AG, Bioinformatics Solutions Inc., and Proteome Software) for providing extended trial licenses of their software to study participants and Biognosys AG for the provision of the iRT peptides added to all samples. The identification of certain commercial equipment, instruments, software, or materials does not imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

SUPPLEMENTAL MATERIALS

pdf	Supplemental File 1.pdf	158 KB
pdf	Supplemental File 2.pdf	384 KB
xlsx	Supplemental Tables.xlsx	53 KB

References

- ABRF. Proteomics Research Group (PRG). https://abrf.memberclicks.net/proteomics-prg- ↩
- Bache N, Geyer PE, Bekker-Jensen DB, et al. A novel LC system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol Cell Proteomics.* 2018;17(11):2284-2296. doi:10.1074/mcp.TIR118.000853. ↩
- Barkovits K, Pacharra S, Pfeiffer K, et al. Reproducibility, specificity and accuracy of relative quantification using spectral library-based data-independent acquisition. *Mol Cell Proteomics.* 2019;19(1):181-197. doi:10.1074/mcp.RA119.001714. ↩
- Bilbao A, Varesio E, Luban J, et al. Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics.* 2015;15(5-6):964-980. doi:10.1002/pmic.201400323. ↩
- Bruderer R, Bernhardt OM, Gandhi T, 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(5):1400-1410. doi:10.1074/mcp.M114.044305. ↩
- Bruderer R, Bernhardt OM, Gandhi T, et al. Optimization of experimental parameters in data-independent mass spectrometry significantly increases depth and reproducibility of results. *Mol Cell Proteomics.* 2017;16(12):2296-2309. doi:10.1074/mcp.RA117.000314. ↩
- Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol.* 2012;30(10):918-920. doi:10.1038/nbt.2377.

↩

- Collins BC, Hunter CL, Liu Y, et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat Commun.* 2017;8(1):291. doi:10.1038/s41467-017-00249-5.

↩

- Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics.* 2014;13(9):2513-2526. doi:10.1074/mcp.M113.031591.

↩

- Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: Neural networks and interference correction enable deep coverage in high-throughput proteomics. *Nat Methods.* 2020;17(1):41-44. doi:10.1038/s41592-019-0638-x.

↩

- Fröhlich K, Brombacher E, Fahrner M, et al. Benchmarking of analysis strategies for data-independent acquisition proteomics using a large-scale dataset comprising inter-patient heterogeneity. *Nat Commun.* 2022;13(1):2622. doi:10.1038/s41467-022-30094-0.

↩

- Gabriels R, Martens L, Degroeve S. Updated MS(2)PIP web server delivers fast and accurate MS(2) peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. *Nucleic Acids Res.* 2019;47(W1):W295-W299. doi:10.1093/nar/gkz299.

↩

- Gessulat S, Schmidt T, Zolg DP, et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods.* 2019;16(6):509-518. doi:10.1038/s41592-019-0426-7.

↩

- Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem.* 1989;182(2):319-326. doi:10.1016/0003-2697(89)90602-7.

↩

- Gotti C, Roux-Dalvai F, Joly-Beauparlant C, Mangnier L, Leclercq M, Droit A. Extensive and accurate benchmarking of DIA acquisition methods and software tools using a complex proteomic standard. *J Proteome Res.* 2021;20(10):4801-4814. doi:10.1021/acs.jproteome.1c00490.

↩

-

↩

- Huang T, Bruderer R, Muntel J, Xuan Y, Vitek O, Reiter L. Combining precursor and fragment information for improved detection of differential abundance in data independent acquisition. *Mol Cell Proteomics.* 2020;19(2):421-430. doi:10.1074/mcp.RA119.001705.

↩

- Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods.* 2007;4(11):923-925. doi:10.1038/nmeth1113.

↩

- Käll L, Storey JD, MacCoss MJ, Noble WS. Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res.* 2008;7(1):29-34. doi:10.1021/pr700600n.

↩

- Käll L, Storey JD, Noble WS. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics.* 2008;24(16):i42-48. doi:10.1093/bioinformatics/btn294.

↩

- Liu K, Li S, Wang L, Ye Y, Tang H. Full-spectrum prediction of peptides tandem mass spectra using deep neural network. *Anal Chem.* 2020;92(6):4275-4283. doi:10.1021/acs.analchem.9b04867.

↩

- Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol.* 2018;14(8):e8126. doi:10.15252/msb.20178126.

↩

- MacLean B, Tomazela DM, Shulman N, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics.* 2010;26(7):966-968. doi:10.1093/bioinformatics/btq054.

↩

- Muntel J, Gandhi T, Verbeke L, et al. 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(5):348-360. doi:10.1039/c9mo00082h.

↩

- Navarro P, Kuharev J, Gillet LC, et al. A multicenter study benchmarks software tools for label-free proteome quantification. *Nat Biotechnol.* 2016;34(11):1130-1136. doi:10.1038/nbt.3685.

↩

- Pino LK, Just SC, MacCoss MJ, Searle BC. Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries. *Mol Cell Proteomics.* 2020;19(7):1088-1103. doi:10.1074/mcp.P119.001913.

↩

- Pino LK, Searle BC, Huang EL, Noble WS, Hoofnagle AN, MacCoss MJ. Calibration using a single-point external reference material harmonizes quantitative mass spectrometry proteomics data between platforms and laboratories. *Analytical chemistry.* 2018;90(21):13112-13117. doi:10.1021/acs.analchem.8b04581.

↩

- Reubsaet L, Sweredoski MJ, Moradian A. Data-independent acquisition for the Orbitrap Q Exactive HF: a tutorial. *Proteome Res.* 2019;18(3):803-813. doi:10.1021/acs.jproteome.8b00845.

↩

- Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. doi:10.1093/nar/gkv007.

↩

- Rosenberger G, Koh CC, Guo T, et al. A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci Data.* 2014;1:140031. doi:10.1038/sdata.2014.31.

↩

- Rost HL, Aebersold R, Schubert OT. Automated SWATH data analysis using targeted extraction of ion chromatograms. *Methods Mol Biol.* 2017;1550:289-307. doi:10.1007/978-1-4939-6747-6_20.

↩

- Searle BC, Egertson JD, Bollinger JG, Stergachis AB, MacCoss MJ. Using data independent acquisition (DIA) to model high-responding peptides for targeted proteomics experiments. *Mol Cell Proteomics.* 2015;14(9):2331-2340. doi:10.1074/mcp.M115.051300.

↩

- Searle BC, Pino LK, Egertson JD, et al. Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat Commun.* 2018;9(1):5128. doi:10.1038/s41467-018-07454-w.

↩

- Searle BC, Swearingen KE, Barnes CA, et al. Generating high-quality libraries for DIA-MS with empirically-corrected peptide predictions. *Nat Commun.* 2020;11(1):1548. doi:10.1038/s41467-020-15346-1.

↩

↩

- Ting YS, Egertson JD, Payne SH, et al. Peptide-centric proteome analysis: an alternative strategy for the analysis of tandem mass spectrometry data. *Mol Cell Proteomics.* 2015;14(9):2301-2307. doi:10.1074/mcp.O114.047035.

↩

- Tsou CC, Avtonomov D, Larsen B, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods.* 2015;12(3):258-264. doi:10.1038/nmeth.3255.

↩

- Van Puyvelde B, Daled S, Willems S, et al. A comprehensive LFQ benchmark dataset on modern day acquisition strategies in proteomics. *Sci Data.* 2022;9(1):126. doi:10.1038/s41597-022-01216-6.

↩

- Van Puyvelde B, Willems S, Gabriels R, et al. Removing the hidden data dependency of DIA with predicted spectral libraries. *Proteomics.* 2020;20(3-4):e1900306. doi:10.1002/pmic.201900306.

↩

- van Riper S, Chen E, Remmer H, et al. Identification of low abundance proteins in a highly complex protein mixture. *Zenodo.* 2015. doi:10.5281/zenodo.3563207.

↩

- Wang Y, Kirkpatrick J, Neely BA, et al. ABRF PRG study to evaluate data-independent acquisition for protein quantification in core facility settings. *Zenodo.* 2019. doi:10.5281/zenodo.3946819.

↩