Journal of The American Society for Mass Spectrometry

Assessing the relationship between mass window width and retention time scheduling on protein coverage for data-independent acquisition

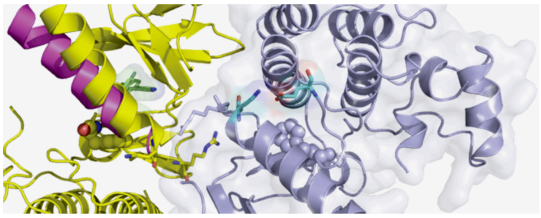
--Manuscript Draft--

Manuscript Number:	ASMS-D-19-00061R2
Full Title:	Assessing the relationship between mass window width and retention time scheduling on protein coverage for data-independent acquisition
Short Title:	Scheduling DIA of small windows along RT
Article Type:	Focus - Full Length Article
Keywords:	Data-independent Acquisition Isolation windows Maxquant pFind Spectronaut
Corresponding Author:	Yansheng of Liu Yale University Yale Cancer Center West Haven, CT UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Yale University Yale Cancer Center
Corresponding Author's Secondary Institution:	
First Author:	Wenxue Li
First Author Secondary Information:	
Order of Authors:	Wenxue Li
	Hao Chi
	Barbora Salovska
	Liangliang Sun
	Yansheng Liu
	Chongde Wu
	George Rosenberger
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Due to the technical advances of mass spectrometers, particularly increased scanning speed and higher MS/MS resolution, the data-independent acquisition mass spectrometry (DIA-MS) became more popular, which enables high reproducibility in both proteomic identification and quantification. The current DIA-MS methods normally cover a wide mass range, with the aim to target and identify as many peptides and proteins as possible and therefore frequently generates MS/MS spectra of high complexity. In this report, we assessed the performance and benefits of using small windows with e.g. 5-m/z width across the peptide elution time. We further devised a new DIA method named RTwinDIA that schedules the small isolation windows in different retention time blocks, taking advantage of the fact that larger peptides are normally eluting later in reversed phase chromatography. We assessed the direct proteomic identification by using shotgun database searching tools such as MaxQuant and pFind, and also Spectronaut with an external comprehensive spectral library of human proteins. We conclude that algorithms like pFind have potential in directly analyzing DIA data acquired with small windows, and that the instrumental time and DIA cycle time, if prioritized to be spent on small windows rather than on covering a

	broad mass range by large windows, will improve the direct proteome coverage for new biological samples and increase the quantitative precision. These results further provide perspectives for the future convergence between DDA and DIA on faster MS analyzers.

Yale University

Cancer Biology Institute



Yansheng Liu, Ph.D.
Assistant Professor of Pharmacology
School of Medicine,
Cancer Biology Institute
Department of Pharmacology
ABC 371A,
840 West Campus Drive,
P.O. Box 27400
West Haven, Connecticut 06516
Telephone: (203) 737-3853
E-Mail: yansheng.liu@yale.edu

April 28, 2019

Gavin E. Reid, PhD.
Professor of Bioanalytical Chemistry
Bio21 Molecular Science and Biotechnology Institute
30 Flemington Rd, room 470
Parkville Victoria 3010, Australia

Dear Gavin,

We are happy that the Reviewers are satisfied with our revision. We addressed the final minor comments accordingly.

I hope we are still in time.

Best regards,
Yansheng

Reviewer #1: The authors were very thorough in addressing all the comments and in providing new data also to perform additional benchmarks. I do appreciate the comprehensive rebuttal. I do agree that the combination of SILAC with DIA is very powerful and will likely be worked on also by other groups.

Author reply: We thank for Reviewer 1's positive feedback.

Reviewer #2: Within this revision Li et al have refined their initial manuscript on a new approach to DIA improving the readability and rigor of the study. I feel the authors have addressed most of my previous concerns and that the manuscript is now suitable for publication. That said the new comparison of the light channel of A2780 SILAC dataset to the Hela dataset doesn't get at my key criticism of the initial work which is does RTwinDIA enable the detection/determination of protein abundance changes between conditions better than the other DIA approach? I concede undertaking an experiment to address this may be beyond the scope of this paper but moving forward this will be a key point that needs to be address before this approach will be widely used.

Author reply: We thank for Reviewer 2's general positive feedback (also for his/her careful proof reading). We agree with the reviewer that there are application works to do to make the approach being widely used. We already fully discussed the robustness of the method in Para. 2, Page 15. We herein add a note to finalize this paragraph: "To make RTwinDIA widely used, the above considerations should be taken in real biological applications." (**Line 28, Page 15**).

Minor corrections
Within the abstract the authors write "the data-independent acquisition mass spectrometry (DIA-MS) became more popular", they may want to reword this to "the use of data-independent acquisition mass spectrometry (DIA-MS) has become more popular"

Author reply: We followed this nice suggestion. Thanks.

Page 4 Line 16: "After digestion, the peptide mixture was acidified with formic acid and then desalted with C18 column (MarocoSpin Columns, NEST Group INC)." Suggested this be changed to "After digestion, the peptide mixture was acidified with formic acid and then desalted with a C18 column (MarocoSpin Columns, NEST Group INC)."

Author reply: We have corrected the grammar as suggested.

Page 7 line 5, suggested to place a space between "AG, Switzerland"

Author reply: We have added the space. Thanks.

Figure S3. It is suggested the authors remake figure S3 and show all the data not just cumulative percentage above an assigned threshold. It would be interesting to see if there are differences in cycle times with the different methods below 3 seconds.

Author reply: We want to communicate that we actually tried different plots before showing the cumulative percentage in Figure S3. The difference between three methods is indeed very small and all the plots we tried (i.e., scatter plot, box plot, and violin plot) showed the same shape (this is because ~70% are exactly 3.2 seconds). We therefore decided to use this plot cumulative percentage so that difference can be visualized, following our previous experience.

Assessing the relationship between mass window width and retention time scheduling on protein coverage for data-independent acquisition

Running title: Scheduling DIA of small windows along RT

Wenxue Li [1,7], Hao Chi [2,7], Barbora Salovska [1,3], Chongde Wu [1], Liangliang Sun [4], George Rosenberger [6], Yansheng Liu [1,5*]

1. Yale Cancer Biology Institute, Yale University, West Haven, Connecticut 06516, USA
2. Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China.
3. Department of Genome Integrity, Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic
4. Department of Chemistry, Michigan State University, Michigan 48824, United States.
5. Department of Pharmacology, Yale University School of Medicine, New Haven, Connecticut 06520, USA
6. Department of Systems Biology, Columbia University, New York, NY, USA.
7. Equal contribution.

Corresponding authors: yansheng.liu@yale.edu

Abstract

Due to the technical advances of mass spectrometers, particularly increased scanning speed and higher MS/MS resolution, the use of data-independent acquisition mass spectrometry (DIA-MS) became more popular, which enables high reproducibility in both proteomic identification and quantification. The current DIA-MS methods normally cover a wide mass range, with the aim to target and identify as many peptides and proteins as possible and therefore frequently generates MS/MS spectra of high complexity. In this report, we assessed the performance and benefits of using small windows with e.g. 5-m/z width across the peptide elution time. We further devised a new DIA method named RTwinDIA that schedules the small isolation windows in different retention time blocks, taking advantage of the fact that larger peptides are normally eluting later in reversed phase chromatography. We assessed the direct proteomic identification by using shotgun database searching tools such as MaxQuant and pFind, and also Spectronaut with an external comprehensive spectral library of human proteins. We conclude that algorithms like pFind have potential in directly analyzing DIA data acquired with small windows, and that the instrumental time and DIA cycle time, if prioritized to be spent on small windows rather than on covering a broad mass range by large windows, will improve the direct proteome coverage for new biological samples and increase the quantitative precision. These results further provide perspectives for the future convergence between DDA and DIA on faster MS analyzers.

Introduction

The last decade has witnessed significant technological development in mass spectrometry (MS). MS-based proteomics has been widely applied to detect and quantify proteins at large scale [1]. In particular, due to the recent advances of fast scanning and high-resolution MS analyzers such as those in Q-TOF or orbitrap type machines, it is possible to scan the entire, primarily populated m/z range for peptide mixtures in a short cycle time (i.e., < 3-5 seconds) with large, sequential mass-to-charge (i.e., m/z) windows. This way, high-resolution MS2 spectra can be acquired multiple times during each peptide's elution peak along the liquid chromatography (LC) gradient. Those methods operate via "data-independent acquisition" (DIA) [2] [3] and can particularly benefit from the high resolution of fragment ions recorded with sufficient data points, enabling both precise identification and quantification of individual proteins [4]. Compared to the traditional data-dependent acquisition (DDA, often referred to as shotgun proteomics) in which peptide fragmentation in the mass spectrometers is guided by the real-time intensity of peptide precursor ions, DIA-MS can record the full MS2 features that are above the detection limit of the mass spectrometer and thus provides consistent sensitivity and high reproducibility for multi-sample measurements.

One example of the newly emerging DIA-MS strategies is the sequential window acquisition of all theoretical mass spectra (SWATH-MS) [5]. In its initial implementation, 32 windows of 25 m/z width covering 400– 1200 m/z range were used on a SCIEX TripleTOF with a cycle time of ~3.3 seconds [5]. Later, this window schema was refined to 64 variable windows with half of the scan time per MS/MS, so that the total cycle time is kept the same and sufficient data points per LC peak can be maintained [6]. On orbitrap platforms, the DIA cycle time setting is further associated to the desired MS2 resolution. Recent publications have reported methods using 19- and 24- variable windows implemented on Q-Exactive and Q-Exactive HF at a resolution of 30,000 [7, 8], and 70 windows on Q-Exactive HF-X with fixed width of 9 *m/z* at a resolution of 15,000 [9]. Alternative isolation strategies, such as multiplexed MS/MS (MSX) [10] and "stepping isolation" [11] have been proposed for single-shot DIA, which efficiently increase the DIA selectivity by distributing interferences between scans. However,

they require an additional de-multiplexing step and do not increase the *interscan dynamic range* [12]. To summarize, due to the limit of scanning speed of the mass spectrometers, the current DIA methods still have to use *de facto* rather large isolation windows, resulting in multiplexed MS2 spectra that cannot be directly analyzed by e.g. shotgun database searching engines [13].

In this study, we therefore aim to gauge the benefits and trade-offs of using the factual, much smaller windows (e.g., 5 m/z) for DIA measurement. We focused on single-shot measurement with the same instrument and the same acquisition time, as well as on the direct proteomic identification. *Under these assumptions*, somewhat surprisingly, we found the smaller windows rather than the large m/z windows result in better proteome coverage in DIA workflows. In particular, we report a novel DIA method by distributing different sets of small acquisition windows across the LC retention time (RT) ranges, namely RT windowed DIA (RTwinDIA).

Methods

Material and Reagents

Hela standard peptides were purchased from Thermo Fisher (Pierce™, part no. 88328). Light L-Arginine-HCl (purity >98%, part no. 88427) and L-Lysine-2HCl (purity >98.5%, part no. 88429) were purchased from Thermo Scientific. Heavy L-Arginine-HCl (13C6, 15N4, purity >98%, part noCCN250P1), and L-Lysine-2HCl (13C6, 15N2, purity >98%, part no. CCN1800P1) were purchased from Cortecnet. The human plasma sample (part no. P9523) was purchased from Sigma. RPMI medium 1640 was purchased from Life Technologies (part no. 11875093).

SILAC sample preparation

Ovarian cancer cell A2780 (part no. 93112519-1VL) was purchased from Sigma and cultured in RPMI 1640 media in 10% fetal bovine serum. For SILAC experiment, the SILAC RPMI 1640 media lacking L-Arginine, L-Lysine (Thermo Scientific, part no. 88365) was supplemented

with either light or heavy isotopically labeled lysine and arginine and 10% dialyzed fetal bovine serum (Thermo Fisher, part no. 26400044) as previously described [14].
The cell line was cultured for eight passages with spiked heavy lysine and arginine in SILAC media to reach >99% labeling (checked by mass spectrometry). Heavy and light SILAC cells were collected with three times washing with precooled PBS. The snap frozen cell pellets were stored in -80 °C for proteomics analysis.

Protein extraction and digestion.

Cell pellets were suspended in 10 M urea lysis buffer and complete protease inhibitor cocktail (Roche), ultrasonically lysed by sonication at 4 °C for 2 min using a VialTweeter device (Hielscher-Ultrasound Technology) [14], and then centrifuged at 18,000 × g for 1 h to remove the insoluble material. The supernatant protein mixtures were reduced by 10 mM tris-(2-carboxyethyl)-phosphine (TECP) for 1 h at 37 °C and 20 mM iodoacetamide (IAA) in dark for 45 min at room temperature. All the samples were future diluted by 1:6 (v/v) with 100 mM NH_4HCO_3 and digested with sequencing grade porcine trypsin (Promega) at a protease/protein ratio of 1:20 overnight at 37 °C. After digestion, the peptide mixture was acidified with formic acid and then desalted with a C18 column (MarocoSpin Columns, NEST Group INC). The amount of the purified peptides was determined using Nanodrop One (Thermo Scientific). The A2780 light and heavy peptides were mixed 1:1 as the A280LH mixture for LC-MS run. The purchased human plasma was also dissolved in 10 M urea lysis buffer and complete protease inhibitor cocktail for reduction and alkylation with TECP and IAA, followed by the identical protocol for digestion as described above.

DIA data acquisition on Orbitrap Lumos

Peptide elution was performed on EASY-nLC 1200 systems (Thermo Scientific, San Jose, CA) using a self-packed analytical PicoFrit column (New Objective, Woburn, MA, USA) (75 µm × 50 cm length) using C18 material of ReproSil-Pur 120A C18-Q 1.9 µm (Dr. Maisch GmbH, Ammerbuch, Germany). Peptide separation was conducted by a 1h or 2 h gradient with buffer B (80% acetonitrile containing 0.1% formic acid) from 5% to 37% with flow rate 300

nl/min at 60 °C with column oven (PRSO-V1, Sonation GmbH, Biberach, Germany). Buffer A was composed of 0.1% formic acid in water.

The Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific) instrument coupled to a nanoelectrospray ion source (NanoFlex, Thermo Scientific) was calibrated using Tune (version 3.0) instrument control software. Spray voltage was set to 2,000 V and heating capillary at 275 °C. The mass window settings for BroadDIA, NarrowDIA and RTwinDIA method are described in **Figure 1**, with the exception of the NarrowDIA mass range in plasma samples (**Table S1**). All the DIA-MS methods consisted of one MS1 scan and 40 MS2 scans of isolated windows. The MS1 scan range is 350 - 1650 m/z and the MS1 resolution is 120,000 at m/z 200. The MS1 full scan AGC target value was set to be 2.0E5 and the maximum injection time was 100 ms. The MS2 resolution was set to 30,000 at m/z 200 and the normalized HCD collision energy was 28%. The MS2 AGC was set to be 5.0E5 and the maximum injection time was 50 ms. The default peptide charge state was set to 2. Both MS1 and MS2 spectra were recorded in profile mode.

One microgram of peptides was injected per each MS runs. For technical injections of each MS analysis, two replicates were injected separately in batch blocks (rather than in an adjacent manner). For each biological sample, we have experimental replicates of different LC length (1-hour, 2-hour) with different methods (DDA, NarrowDIA, RTwinDIA, BroadDIA).

DDA data acquisition on Orbitrap Lumos

For DDA-based proteomics, the MS1 singal was recorded by the Orbitrap detector at a resolution of 120,000. The scan range setting was from 350 to 1650 m/z with the RF lens 40%. The AGC value was 5.5E5 and the maximum injection time was 40 ms for MS1. For MS2, the top speed (cycle time 3 s) was used, which means that the maximum dependent scans were performed in each cycle time with desired resolution, AGC and etc. HCD collision energy was 28%. The dynamic exclusion parameters were set to ensure that already sequenced precursors were excluded once from reselection for 30 s. The isolation window was 1.2 m/z and the MS2 resolution was 15,000 for DDA. The AGC value and the maximum injection time

were set to 5e4 and 35 ms, respectively. All the data were collected with 1 h and 2 h gradient as described above.

MS data analysis

MaxQuant: All the shotgun and DIA raw data was directly analyzed by MaxQuant [15] and searched against the human canonical UniProtKB/Swiss-Prot database (downloaded February 2018, 20,258 entries). Oxidation at methionine was set as variable modification, whereas carbamidomethylation at cysteine was set as a fixed modification. Up to two missed cleavages were allowed. The mixed A2780LH sample was searched following standard SILAC setting. Other parameters are kept as default in MaxQuant. Both peptide and protein level were controlled at 1 % FDR [16].The match between run function was disabled and the second peptide search function was enabled (as default).

pFind: All MS/MS data were analyzed using pFind 3.1.5 in this study, in which the Open-pFind workflow was adopted [17]. Open-pFind consists of two search steps—one open and one restricted. First, a sequence tag-based strategy is used to match spectra to a much larger set of possible peptide sequences in the open search, and no modifications are specified initially. After this step, a restricted search is then performed where several key parameters, including modification types and protein sequence entries, are automatically set by semi-supervised machine learning based on the open search results. Finally, the results from both open search and restricted search are merged together and reranked based on a new semi-supervised machine learning model. A standard desktop computer (8-core Intel i7-4910MQ CPU @ 2.90 GHz and 32 GB RAM) was used, with a total of six parallel threads. All datasets were searched against the human database from human human canonical UniProtKB/Swiss-Prot database [18] (released in 2018-12) consisting of 20,408 protein sequences. The target-decoy approach was used in the database search, and decoy proteins were generated by reversing the target protein sequences. Both mass tolerances of precursor ions and fragment ions were set as ±20 ppm. In the procedure of precursor ion extraction of the datasets from DIA mode, the maximum number of precursor ions for each spectrum was not limited, while the corresponding number was set as 6 for the datasets from DDA mode. For

the PSMs results from Open-pFind, the FDR was controlled to be less than 1% at the peptide level, and then protein groups were inferred and the FDR at the protein level was also controlled to be less than 1% based on the target-decoy strategy.

Spectronaut: All the DIA data were also analyzed by Spectronaut (Biognosys AG, Switzerland, version 12.0.20491.17.26268) [7, 8]. The spectra library was built with the published external library, referred to as "Pan-Human Library" [19], which has mass spectrometric assays for more than 10,000 human proteins. The optimized non-linear retention time calibration was used and handled by Spectronaut using iRT space [20]. Both peptide precursor and protein FDR were controlled at 1% [16, 21]. As for quantification, interference correction function was enabled, and top 3 peptide precursors were summed for protein quantification. All the other parameters in Spectronaut are kept as default unless mentioned. For SILAC data analysis, the Pan-Human library was labeled using the "Generate Labeled Library" option (Lys8, Arg10) in the Library perspective in Spectronaut. C-terminal peptides were removed from the library. This function ensures a complete labelling of the resulting library to always contain both label-free and labelled version of all arginine and lysine containing peptide precursors and fragment ions. DIA data analysis was performed using the "Labeled" workflow keeping the default Biognosys Factory Settings. For the analysis, both b- and y- ions were kept for optimal peptide identification. For H/L ratio calculation, fragment ion intensities were exported for each light and heavy counterparts of a peptide precursor. The exported results were filtered to remove b-ions prior to further analysis. The H/L ratios were then calculated for all peptide precursor ions.

Figures were made on R Studio (version 1.1.453) and GraphPad Prism (version 8.0.2). The Venn diagrams were generated by Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/). Mann-Whitney test was performed by GraphPad to calculate p value.

Data availability

All the mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [22] partner repository with the dataset identifier PXD013477 (Reviewer account details: **Username:** reviewer13529@ebi.ac.uk **Password:** CpW01z5F)

Results and Discussion

Configuring DIA methods for comparison.

To set up the comparison between large and small windows, we configured three DIA methods on an Orbitrap Fusion Lumos instrument (**Figure 1**). The first method uses 40 sequential windows of 20 m/z width each and covers a broad mass range of 360 – 1160 m/z across the whole LC gradient (hereafter, "BroadDIA"), a setting similar to the original SWATH-MS [5]. The second method (hereafter, "NarrowDIA") uses 40 windows of 5 m/z only and covers a narrow range of 200 m/z where the density of identified MS/MS scans is the highest (e.g., 440 -640 m/z for a cell lysate sample **Figure S1**). To embrace the fact that peptides of larger m/z are generally eluting at a later retention time (RT) (**Figure S2**), we devised the third RT windowed DIA method (hereafter, "RTwinDIA"). In the present implementation of RTwinDIA, the first set of 40 windows (5 m/z each) covering 400-600 m/z was scheduled in the 0-50% RT period, followed by the second 40-window set (600-800 m/z) during 50 -75% RT and the third set (800-1000 m/z) during 75-100% RT. This means, e.g., in a 2-hour RTwinDIA, 400-600 m/z, 600-800 m/z, and 800-1000 m/z are respectively scanned during 0-60, 60-90, and 90-120 min in a single-shot. Please refer to **Table S1** for the exportable window settings in each DIA methods. We further included a routine DDA method with 1.2-m/z isolation for MS/MS sequencing of those highest precursor ions detected in every 3 seconds (i.e., the cycle time directed DDA mode in Lumos, hereafter, "DDA" [23]). All the four MS methods (**Figure 1**) were performed using quadrupole-based isolation, HCD collision, a similar cycle time, and the "high-high" mode in orbitrap analyzer (see **M**ethods). We assessed the performance of these methods with 1- or 2-hour injection of the peptides derived from three biological samples including a) a commercial HeLa cell digest, b) a SILAC lysate of human A2780 cells (heavy to light, 1:1), and c) a human-plasma standard, each in two replicates. All three methods generated nearly identical numbers of MS/MS scans (e.g., about 86,000 in 2-hour run of HeLa samples), suggesting that the same cycle time was achieved. Further analysis suggested that for ~70% of all the cycles, all the three DIA-MS methods finish the data acquisition in 3.2 seconds, and >99% of all the cycles are below 4.2 seconds (**Figure S3**). This translates to an average of >5-7 data points in our LC-

MS settings in all DIA methods used, which was previously deemed to be sufficient [8, 24]. Taken together, our experimental design enabled a simple and fair comparison between large and small windows and between DIA and DDA.

Assessing protein identification performance of data analysis algorithms.

To test if a conventional DDA search engine could be directly used to identify peptides in 5 m/z window DIA, we applied MaxQuant [15] to all data sets generated (without "matching between runs", but with "Second peptide identification" in MaxQuant to allow up to two peptides identified per MS/MS spectrum) (**Figure 2a-b)** [15]. As expected, MaxQuant generated decent results for all DDA runs. Taking HeLa digest as an example, 37,661 peptides (4,259 proteins) were identified in 2-hour measurements and 16,548 peptides (2,572 proteins) in 1-hour (*averaged numbers from two replicates are shown hereafter unless specified*) **(Table S2**). The LC separation and dynamic exclusion worked well in DDA, because merely 7.62 (1-hour) and 9.37% (2-hour) peaks were repeatedly sequenced. However, only 14,242, 13,308, and 15,159 peptides were identified from 2-hour BroadDIA, NarrowDIA and RTwinDIA, fewer peptides than identified within a 1-hour DDA run. More than 80% of peaks were repeatedly sequenced in all DIAs. Interestingly there are ~3,000 proteins identified across all 2-hour DIAs, with RTwinDIA performing the best (n= 3,307 proteins, representing a 7.55% increase from NarrowDIA and a 18.9% increase from BroadDIA). Similar results were obtained from SILAC peptides, despite the fact that the identification numbers and their differences were lower due to the increased sample complexity of SILAC labeling (**Figure 2a-b**). The identifications by MaxQuant were in general low in all plasma DIA runs (about 150 proteins and 1000 peptides), resulting in difficult comparison (**Figure S4**). It should be stressed that MaxQuant was not designed to directly perform the peptide and protein identification for DIA data. It was previously demonstrate that even for a DDA with 2m/z isolation window, the majority of the MS/MS spectra can be assigned to multiple peptides [25], providing a potential explanation for our MaxQuant results.

As stated above, although MaxQuant provided a slightly better identification in RTwinDIA, it is inefficient in analyzing DIA data of 5-m/z windows. An emerging software, pFind (Open-pFind workflow) [17] was shown to significantly increase the MS/MS spectra identification rate by allowing the possibilities of e.g., open modifications and mixed spectra analysis [17]. We herein applied pFind to our data set (**Figure 2c-d**). Particularly, the maximum number of precursor ions for each spectrum was set to be not limited for DIA in pFind, whereas the corresponding number was set to 6 for DDA. We first found that, in DDA of HeLa samples, pFind reported 1.512 times the peptide numbers as reported by MaxQuant. Intriguingly, in HeLa DIA datasets, this ratio increased to 2.082, 2.060, and 2.079 times, for BroadDIA, NarrowDIA, and RTwinDIA respectively **(Table S2**). We further found that pFind facilitates identification in complex samples more significantly. For example, in SILAC samples, the peptide identification ratio between pFind and MaxQuant is 2.239 in 2-hour BroadDIA data (2.498 for 1-hour), 2.768 in 2-hour NarrowDIA (3.242 for 1-hour), and 2.805 for 2-hour RTwinDIA (3.196 for 1-hour). Impressively, in plasma samples, pFind reported 2.758 times the peptide identifications of MaxQuant for DDA runs (similar for both 1-hour and 2-hour injections), and 2.95 times for BroadDIA, and even 3.101 and 3.293 times for NarrowDIA and RTwinDIA. These results highlight the substantially improved ability of pFind in handling complex MS/MS spectra from complex samples, and its potential usage in analyzing DIA datasets. In SILAC and plasma samples, pFind identification numbers are larger in datasets generated with 5-m/z windows than 20-m/z windows, suggesting pFind still has its limitation in handling DIA data of large windows, especially when analyzing proteomes of high complexity (e.g., SILAC) or high dynamic range (e.g., plasma).

Next, we applied the widely used DIA software, Spectronaut [7, 8], and a classic peptide-centric data extraction strategy [5] (based on mass spectrometric assays for 10,000 human proteins referred to as "Pan-Human Library" [19]), to further understand the impact of DIA window-size on proteome coverage (**Figure 2e-f**). Herein only DIA runs were analysed. Different than MaxQuant or pFind, Spectronaut using the "Pan-Human Library" identified much more peptides in BroadDIA (n=44,912) than in NarrowDIA (n= 35,316) and RTwinDIA

(n=36,260) for the 2-hour HeLa dataset **(Table S2**). This is expected, because BroadDIA essentially covered a much larger m/z range, where the human peptides were extensively sequenced and included in the "Pan-Human Library" [19]. Somewhat surprisingly, at the protein- level, NarrowDIA and RTwinDIA actually yielded 5,805 and 5,707 protein identifications (protein- level FDR 1%, controlled by Spectronaut), about 25% more than BroadDIA (4,709 proteins). This suggests that, the higher resolution provided by NarrowDIA and RTwinDIA on the ion-dense region of the *m/z vs. RT* space resulted in identification of additional peptide species coming from different proteins. Even when we exclude those proteins identified with only one unique peptide, 5-m/z windows still identified 10.6% more proteins than broad windows. This effect is more extreme for 1-hour runs and for SILAC samples: e.g., the 1-hour NarrowDIA and RTwinDIA increased the protein identification numbers by 114.8% as compared to BroadDIA (and by 78.5% for proteins with >= 2 peptides, **Figure 3a & Figure S5**) and by 170.4% and 112.1% as compared to DDA analysis using MaxQuant and pFind searches (**Figure S6**). Altogether, these results suggest that, when an external comprehensive library is used, the narrow windowed DIA can provide higher proteome coverage than BroadDIA under the same machine- and cycle- time.

Analyzing the high protein level coverage in RTwinDIA result

Because RTwinDIA schedules three 200-m/z windows by three different RT windows, we asked how many peptide and protein identifications would have been missed if the entire RT range is analyzed with the same resolution. Thus, we acquired an additional data set of three 2-hour measurements on the HeLa proteome using 40 5-m/z windows across the entire RT. Each of the three measurements continuously covered the mass range of 400-600 m/z, 600-800 m/z, and 800-1000 m/z. We found that 400-600 m/z, 600-800 m/z and 800-1000 m/z respectively identified very different peptide sequences according to pFind and Spectronaut results (**Figure 4**) because the combined result of the three measurements yielded much more peptides than each of them. Intriguingly, although the combined result exceeded RTwinDIA by 68% more peptide identifications (e.g., 60,918 vs. 36,260 in Spectroaut result, **Figure 4b**), RTwinDIA only had a minimal compromise at the protein level: RTwinDIA yielded

5707 proteins, which is only 3.3% fewer than the 5903 proteins identified by three 200-m/z runs combined (**Figure 4a**). This is likely because that most of the additional peptides are derived from the same set of proteins that are relatively more abundant than others in the sample. The identical result can be obtained by pFind. Considering only one-third of the machine time used, RTwinDIA is indeed time-efficient in covering more proteins.

The acquisition window setting for DIA has been often optimized to have variable sizes recently [6], i.e., smaller windows are used for m/z region of higher precursor ion density and intensity, and vice versa. We thus used a DIA of 40 variable windowed schema (VariableDIA) we recently published [24] to analyze the HeLa proteome in a 2-hour measurement. As shown in **Figure S7**, according to Spectronaut, VariableDIA generated better results than BroadDIA at both protein and peptide levels by 7.20% and 8.60% of number increase respectively (i.e., optimized variable window setting is indeed useful), but still had lower protein coverage than NarrowDIA and RTwinDIA. Interestingly, pFind generated the least peptide- and protein- identification numbers for variable windows compared to all the other MS methods, likely due to the fact that some windows have to be much larger 20 m/z in the variable setting that further complicated MS2 spectra.

Labeling and label-free based quantification benefits for DIAs using small windows

Besides the direct protein identification gain demonstrated above, we further assessed the quantification performance of DIA when small m/z windows are used. Interestingly, we found that the smaller windows essentially increased the selectivity of DIA and the signal-to-noise ratio for the same peptide space targeted (**Figure S8**). Accordingly, we observed a 5-7% increase in the quantitative precision in regard to the heavy-to-light SILAC ratios, when all the MS2 level y-ions were summarized (**Figure 3b & Figure S5**). DIA has a great potential in dealing with complex samples such as proteomes labeled by SILAC. Previously, we have applied DIA in pulse-chase SILAC (pSILAC) experiments to quantify the protein specific turnover rate, for understanding post-transcriptional regulations in complicated biological systems such as human aneuploidy [14] and cell line heterogeneity [26]. The increased

quantitative precision in SILAC data by using small DIA windows therefore has immediate implication for similar studies in the future.

As for label-free quantification, we first performed correlation analysis between replicates for different DIA methods. We found that all the three DIA methods achieved nice reproducibility (R=0.947, 0.979, and 0.979 for BroadDIA, NarrowDIA, and RTwinDIA) at the absolute scale (**Figure S9**). To quickly estimate the relative quantification performance for label-free experiment, we have compared the light channel of A2780 SILAC DIA data to the HeLa DIA data. Such a comparison has the advantages to represent a biological comparison between two human cell lines and essentially covers a wide quantitative range depending on the difference between two cell lines. As shown in **Figure S10**, all the three DIA methods had a high and comparable accuracy for relative label-free quantification between HeLa and A2780 cell proteome (R=0.8919, 0.9163, 0.9151, **Figure S10a-c**). In all DIA methods, the effective ratio has a good linearity that goes beyond 32:1 to 1:32. Interestingly, similar to SILAC data, we could also observe that NarrowDIA and RTwinDIA have a slight but significantly better quantitative precision than BroadDIA (P<0.0001, **Figure S10d**). Previously, it was reported that there is a compromise between DIA window size and the percentage of MS events reaching AGC level. In our data, for example, about 45% and 17% of MS2 scans in HeLa 1-hour and 2-hour runs triggered AGC in our settings (See **Table S3** for more), which seem to be able to provide decent and comparable quantification as shown above.

Taken together, our data suggests that all the three DIA methods similarly achieved decent relative quantification reproducibility, while DIAs of narrow windows (NarrowDIA and RTwinDIA) achieved a slight and significantly higher precision than BroadDIA, which is consistent for both labeling and label-free experiments.

Considerations for the application of RTwinDIA and NarrowDIA for proteomic analysis.

Herein, as a pilot study we assessed both identification and quantification with variable DIA method settings and algorithms. The usage of small windows is not new in DIA methods [5,

6, 13]. One example is the PAcIFIC approach [3, 27]. Recently small window DIA schemes have been used to generate DIA libraries [28]. Herein, we proposed a new method, RTwinDIA, which takes the advantage of the correlation between peptide mass and elution RT, and uniquely schedules different small windows across different retention time bins as a direct DIA measurement. Considering the peptide distribution along RT and the practical LC analytical robustness, we simply used only three blocks of m/z range vs. RT and did not design more sophisticated smaller blocks overlapping along the RT in current version of method. The 200 m/z range was selected based on the same number of windows (N=40) and a reasonably small size per window (5 m/z), so that the cycle time is similar for all comparisons. Although RTwinDIA achieved slightly better results than NarrowDIA in MaxQuant and pFind analyses, it did not provide more peptide and protein identifications than NarrowDIA in Spectronaut analysis. However, due to the different peptide sets targeted, RTwinDIA and NarrowDIA indeed covered many different peptides (~45% being different) and proteins (~10% being different) (see Venn diagrams in **Figure S11**).

Our results have implications on the potential usage of small windowed DIA methods such as RTwinDIA. *First,* we found that small windowed DIA methods improved the protein- level coverage as compared to BroadDIA in both HeLa and A2780 SILAC data sets according to Spectronaut. Therefore, if the experiment purpose is to quickly characterize protein-level changes to as many proteins as possible in human cell samples by directly using e.g., human proteome sequence file or combined assay libraries such as "Pan-Human Library", NarrowDIA and RTwinDIA can be powerful. In such cases, a sample-specific, comprehensive library is often not needed or cannot be acquired when there are limits of sample amount or machine time. *Second,* we found that RTwinDIA seems to provide significant benefit over BroadDIA and even NarrowDIA in 1-hour measurement in MaxQuant, pFind, and especially Spectronaut analysis. Also, RTwinDIA provided more protein identifications than NarrowDIA for plasma samples (Table S1) and increased quantification precision in the complex SILAC samples. These results therefore support the usage of NarrowDIA and RTwinDIA in short gradient (e.g., 1-hour) based protein identification tasks, especially for analyzing high

dynamic range, or multiplexed samples. Third, post-translational modifications (PTM) of peptides may also complicate the data matrix. Using pFind we have found that small windowed methods have identified oxidation with higher frequency than BroadDIA (e.g., 7.59% vs. 4.10% for 1-hour methods), but similar frequency of carbamidomethylation for all three DIAs (**Figure S12**). Because we did not perform any PTM enrichment, future experiments are needed to confirm the advantage of small windowed DIA in tracking small modifications. Last but not the least, we suggest RTwinDIA could potentially extend its usability in the future. For example, we are currently applying RTwinDIA on the separation system provided by capillary zone electrophoresis (CZE) in which the RT showed a much stronger correlation to peptide mass [29].

A critical element for broader application of RTwinDIA is its robustness. We have checked the retention time variation in our LC system during ~ 2 weeks of measurement. The real time deviation was less than +/- 1 min for a 120 min measurement, and +/- 0.5 min for a 60 min measurement **(Figure S13).** Similar performance can be expected in current LC systems. Considering that Spectronaut is using a non-linear regression retention time calibration, that our implementation of RTwinDIA has only three blocks, and that DIA analysis tools such as Spectronaut will automatically discard most partial peak groups shaped by the boundary region of RT windows during scoring, the impact of LC stability might be minimal for RTwinDIA results. For large-scale applications using long-term measurement, retention time drift has to be recalibrated regularly (to first injections), which is essentially similar to scheduled SRM measurements [30]. Except for the RT variation between windows, RTwinDIA is still a DIA method, which is known to have the particular advantage of reproducibility over DDA, and has been recently tested in hundreds to thousands injections of different biological and clinical samples [31, 32]. Also, RTwinDIA should be feasible and easily transferable to other MS platforms if they provide the fast scanning speed and the high resolution needed (such as QE-HF and later Orbitrap series from Thermo Scientific, as well as Q-TOF instruments from different vendors such as SCIEX , Bruker, and etc.). To make

RTwinDIA widely used, the above considerations should be taken in real biological applications.

We assessed three algorithms for DIA data analysis. The MaxQuant software was not designed to analyze mixed MS2 spectrum and therefore will miss the identification for many DIA scans. The pFind software, although powerful for shotgun analysis and mixed spectrum identification, does not consider the co-elution of peptide fragments along RT as peak groups and does not assume the relative abundance of fragment ions for a given peptide [33], and therefore in general still has less peptide and protein identifications compared to Spectronaut. It is interesting that pFind identified more plasma peptides than Spectronaut, suggesting this software handles complex MS2 scans well. Another promising example is that, from those 1-hour injections of HeLa and SILAC cell digests, pFind actually identified more proteins in the 5-m/z DIA runs, as compared to both 20-m/z DIAs and even DDA results (**Figure 2c**). These results highlight that, a small window like 5-m/z or similar sizes would enable those shotgun database searching tools which can handle mixed spectra analysis, such as pFind or DeMix [25], to be directly applied in DIA-MS, providing immediate, alternative data analysis options for DIA. Spectronaut identification results could be missed if the corresponding peptides/ proteins are not included in reference library used. Therefore, extensive sample specific libraries can be used in Spectronaut analysis to improve the identification. Furthermore, it should be noted that both MaxQuant and pFind still use MS1 intensity for quantification, which may not be ideal for DIA-MS especially when the MS1 resolution is not sufficient [34]. Future developments may be needed for traditional shotgun searching tools to incorporate MS2 level quantitative features.

In this report our assessment has certain limitations. We did not yet analyze other MS parameter effects and other relevant topics. For example, the MS2 resolution was set to be 30,000 for DIA, and 15,000 for DDA, based on our experience and other publications [7, 8]. The lower MS2 resolution for DIA will reduce the cycle time and change the noise levels in both large and narrow windows [9]. The AGC setting and the number of windows (or the

use of windows with variable widths) may also further impact the comparison results between DIA of different windows. Despite of these limitations, the identification benefit of using smaller DIA windows may largely remain due to the use of target-decoy strategy in all algorithms such as pFind and Spectronaut for separating signals from noise. Also our study here did not analyze a scenario where an optimal, sample-specific spectral library was already generated for the particular sample by e.g., peptide fractionation [13] or even discovery DIA runs of small windows [28]. Although using such optimal libraries may significantly increase the identification in BroadDIA, they are less likely to change the quantitative precision difference (**Figure 3b**), which is determined by the method selectivity and noise levels. Last but not the least, we also did not test other DIA algorithms yet such as OpenSWATH [35] or Skyline [36], although they have been demonstrated to provide similar identification and quantification results in a previous benchmark study [37].

Finally, our study, together with previous studies such as applications of PAcIFIC and gas-phase fractionation [3, 10, 27, 28, 38], provided insights about the future convergence between DDA and DIA analysis, and how proteome coverage can be benefited from hypothetical future mass spectrometers with increased scanning speed. If one could apply 5-m/z windows for the whole ca. 800 m/z range using a cycle time of 2-3 seconds (**Figure 1**), the DDA and DIA boundaries (both analytical methods and algorithms) will eventually vanish. This means, the MS analyzers would have to be 4-10 times faster than the current solutions while other parameters, such as ion transmission (related to AGC settings in Orbitrap-type platforms) and MS2 resolution should not be compromised.

Conclusions

In conclusion, we proposed and assessed a new DIA executing method, RTwinDIA, which uniquely schedules different small windows of increasing mass along the peptide elution time. These results support the direct usage of RTwinDIA and other DIA methods of small windows in the short gradient (e.g., 1-hour) based protein identification tasks, especially for analyzing high dynamic range, or multiplexed samples. Our results highlighted advantages

of performing DIA-MS using smaller windows, such as the increased proteome coverage, the direct identification by shotgun searching engines like pFind, higher signal-to-noise ratio, and the increase of quantitative precision, providing hints for future DIA method options.

Acknowledgements.

We thank Lukas Reiter and Oliver Bernhardt from Biognosys AG and Daoyang Chen from Michigan State University for the helpful discussions. We thank Semin He from Institute of Computing Technology CAS Beijing for the resource support in pFind analysis. This research was supported in part by a Pilot Grant from Yale Cancer Center.

Figure legends.

Figure 1. The DDA and three DIA methods with their distinctive ion isolation schema. All the three DIA methods, i.e., BroadDIA, NarrowDIA, and RTwinDIA use 40 sequential isolation windows and have the same desired MS2 resolution. BroadDIA covers a wide range of 360-1160 m/z with a fix window width of 20 m/z. NarrowDIA covers a 200-m/z range of the most precursor ion density, e.g., 440-64 0m/z depending on the sample measured. RTwinDIA uniquely schedules three ranges of 200-m/z in different retention time periods.

Figure 2. Peptide and protein numbers identified by DDA, BroadDIA, NarrowDIA, and RTwinDIA using MaxQuant (a-b), pFind (c-d), and Spectronaut (e-f) analysis. The error bar presents s.d. based on experimental replicates. Both peptide- level and protein-level FDR were controlled at 1% by respective softwares.

Figure 3. Advantages of identification and quantitative precision for DIA-MS with narrow winidows. (a) Distribution of proteins with different number of unique peptides compared between methods. The numbers of those proteins identified by two or more than two unique peptides were shown for the SILAC sample of 1hr- measurement time. (b) The box plot of the log-transformed heavy-to-light ratios in the absolute scale, which are calculated from all the y-ion intensities between for the SILAC sample. Note the median value was shown for the absolute value of log2 ratios. This means, e.g., the value of 0.2842 indicates a ratio of 2^0.2842 = 1.218, whereas a value of 0.3673 indicates a ratio of 2^0.3673=1.2890.

Figure 4. Comparison of peptide and protein identifications between RTwinDIA and three narrow DIA measurements. These three measurements continuously covered the mass range of 400-600 m/z, 600-800 m/z, and 800-1000 m/z over the entire RT range.

"Combined" results are shown for which three data sets are searched together using pFind and Spectronaut.

References

1. Aebersold, R., Mann, M.: Mass-spectrometric exploration of proteome structure and function. Nature. **537**, 347-355 (2016)
2. Venable, J.D., Dong, M.Q., Wohlschlegel, J., Dillin, A., Yates, J.R.: Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nature methods. **1**, 39-45 (2004)
3. Panchaud, A., Scherl, A., Shaffer, S.A., von Haller, P.D., Kulasekara, H.D., Miller, S.I., et al.: Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. Analytical chemistry. **81**, 6481-6488 (2009)
4. Ting, Y.S., Egertson, J.D., Payne, S.H., Kim, S., MacLean, B., Kall, L., et al.: Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. Molecular & cellular proteomics : MCP. **14**, 2301-2307 (2015)
5. Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., et al.: Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Molecular & cellular proteomics : MCP. **11**, O111 016717 (2012)
6. Sajic, T., Liu, Y., Aebersold, R.: Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications. Proteomics Clin Appl. **9**, 307-321 (2015)
7. Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinovic, S.M., Cheng, L.Y., Messner, S., et al.: Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Molecular & cellular proteomics : MCP. **14**, 1400-1410 (2015)
8. Bruderer, R., Bernhardt, O.M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., et al.: Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. Molecular & cellular proteomics : MCP. **16**, 2296-2309 (2017)
9. Kelstrup, C.D., Bekker-Jensen, D.B., Arrey, T.N., Hogrebe, A., Harder, A., Olsen, J.V.: Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. Journal of proteome research. **17**, 727-738 (2018)
10. Egertson, J.D., Kuehn, A., Merrihew, G.E., Bateman, N.W., MacLean, B.X., Ting, Y.S., et al.: Multiplexed MS/MS for improved data-independent acquisition. Nature methods. **10**, 744-746 (2013)
11. Moseley, M.A., Hughes, C.J., Juvvadi, P.R., Soderblom, E.J., Lennon, S., Perkins, S.R., et al.: Scanning Quadrupole Data-Independent Acquisition, Part A: Qualitative and Quantitative Characterization. Journal of proteome research. **17**, 770-779 (2018)

12. Kaufmann, A., Walker, S.: Comparison of linear intrascan and interscan dynamic ranges of Orbitrap and ion-mobility time-of-flight mass spectrometers. Rapid Commun Mass Spectrom. **31**, 1915-1926 (2017)
13. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., Aebersold, R.: Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Molecular systems biology. **14**, e8126 (2018)
14. Liu, Y., Borel, C., Li, L., Muller, T., Williams, E.G., Germain, P.L., et al.: Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. Nature communications. **8**, 1212 (2017)
15. Cox, J., Mann, M.: MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nature biotechnology. **26**, 1367-1372 (2008)
16. Elias, J.E., Gygi, S.P.: Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature methods. **4**, 207-214 (2007)
17. Chi, H., Liu, C., Yang, H., Zeng, W.-F., Wu, L., Zhou, W.-J., et al.: Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. Nature Biotechnology. **36**, 1059 (2018)
18. Leinonen, R., Diez, F.G., Binns, D., Fleischmann, W., Lopez, R., Apweiler, R.: UniProt archive. Bioinformatics (Oxford, England). **20**, 3236-3237 (2004)
19. Rosenberger, G., Koh, C.C., Guo, T., Rost, H.L., Kouvonen, P., Collins, B.C., et al.: A repository of assays to quantify 10,000 human proteins by SWATH-MS. Scientific data. **1**, 140031 (2014)
20. Bruderer, R., Bernhardt, O.M., Gandhi, T., Reiter, L.: High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. Proteomics. **16**, 2246-2256 (2016)
21. Rosenberger, G., Bludau, I., Schmitt, U., Heusel, M., Hunter, C.L., Liu, Y., et al.: Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses. Nature methods. **14**, 921-927 (2017)
22. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., et al.: The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. **47**, D442-D450 (2019)
23. Espadas, G., Borras, E., Chiva, C., Sabido, E.: Evaluation of different peptide fragmentation types and mass analyzers in data-dependent methods using an Orbitrap Fusion Lumos Tribrid mass spectrometer. Proteomics. **17**, (2017)
24. Mehnert, M., Li, W., Wu, C., Salovska, B., Liu, Y.: Combining Rapid Data Independent Acquisition and CRISPR Gene Deletion for Studying Potential Protein Functions: A Case of HMGN1. Proteomics. e1800438 (2019)
25. Zhang, B., Pirmoradian, M., Chernobrovkin, A., Zubarev, R.A.: DeMix workflow for efficient identification of cofragmented peptides in high resolution data-dependent tandem mass spectrometry. Molecular & cellular proteomics : MCP. **13**, 3211-3223 (2014)
26. Liu, Y., Mi, Y., Mueller, T., Kreibich, S., Williams, E.G., Van Drogen, A., et al.: Multi-omic measurements of heterogeneity in HeLa cells across laboratories. Nature biotechnology. (2019)

27. Panchaud, A., Jung, S., Shaffer, S.A., Aitchison, J.D., Goodlett, D.R.: Faster, quantitative, and accurate precursor acquisition independent from ion count. Analytical chemistry. **83**, 2250-2257 (2011)
28. Searle, B.C., Pino, L.K., Egertson, J.D., Ting, Y.S., Lawrence, R.T., MacLean, B.X., et al.: Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Nature communications. **9**, 5128 (2018)
29. Chen, D., Ludwig, K.R., Krokhin, O.V., Spicer, V., Yang, Z., Shen, X., et al.: Capillary Zone Electrophoresis-Tandem Mass Spectrometry for Large-Scale Phosphoproteomics with the Production of over 11,000 Phosphopeptides from the Colon Carcinoma HCT116 Cell Line. Analytical chemistry. **91**, 2201-2208 (2019)
30. Picotti, P., Aebersold, R.: Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nature methods. **9**, 555-566 (2012)
31. Liu, Y., Buil, A., Collins, B.C., Gillet, L.C., Blum, L.C., Cheng, L.Y., et al.: Quantitative variability of 342 plasma proteins in a human twin population. Molecular systems biology. **11**, 786 (2015)
32. Bruderer, R., Muntel, J., Muller, S., Bernhardt, O.M., Gandhi, T., Cominetti, O., et al.: Analysis of 1508 plasma samples by capillary flow data-independent acquisition profiles proteomics of weight loss and maintenance. Molecular & cellular proteomics : MCP. (2019)
33. Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., et al.: OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotech. **32**, 219-223 (2014)
34. Collins, B.C., Hunter, C.L., Liu, Y., Schilling, B., Rosenberger, G., Bader, S.L., et al.: Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. Nature communications. **8**, 291 (2017)
35. Rost, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S.M., Schubert, O.T., et al.: OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nature biotechnology. **32**, 219-223 (2014)
36. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., et al.: Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. **26**, 966-968 (2010)
37. Navarro, P., Kuharev, J., Gillet, L.C., Bernhardt, O.M., MacLean, B., Rost, H.L., et al.: A multicenter study benchmarks software tools for label-free proteome quantification. Nature biotechnology. **34**, 1130-1136 (2016)
38. de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Frohlich, F., et al.: Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature. **455**, 1251-1254 (2008)

Figure 1

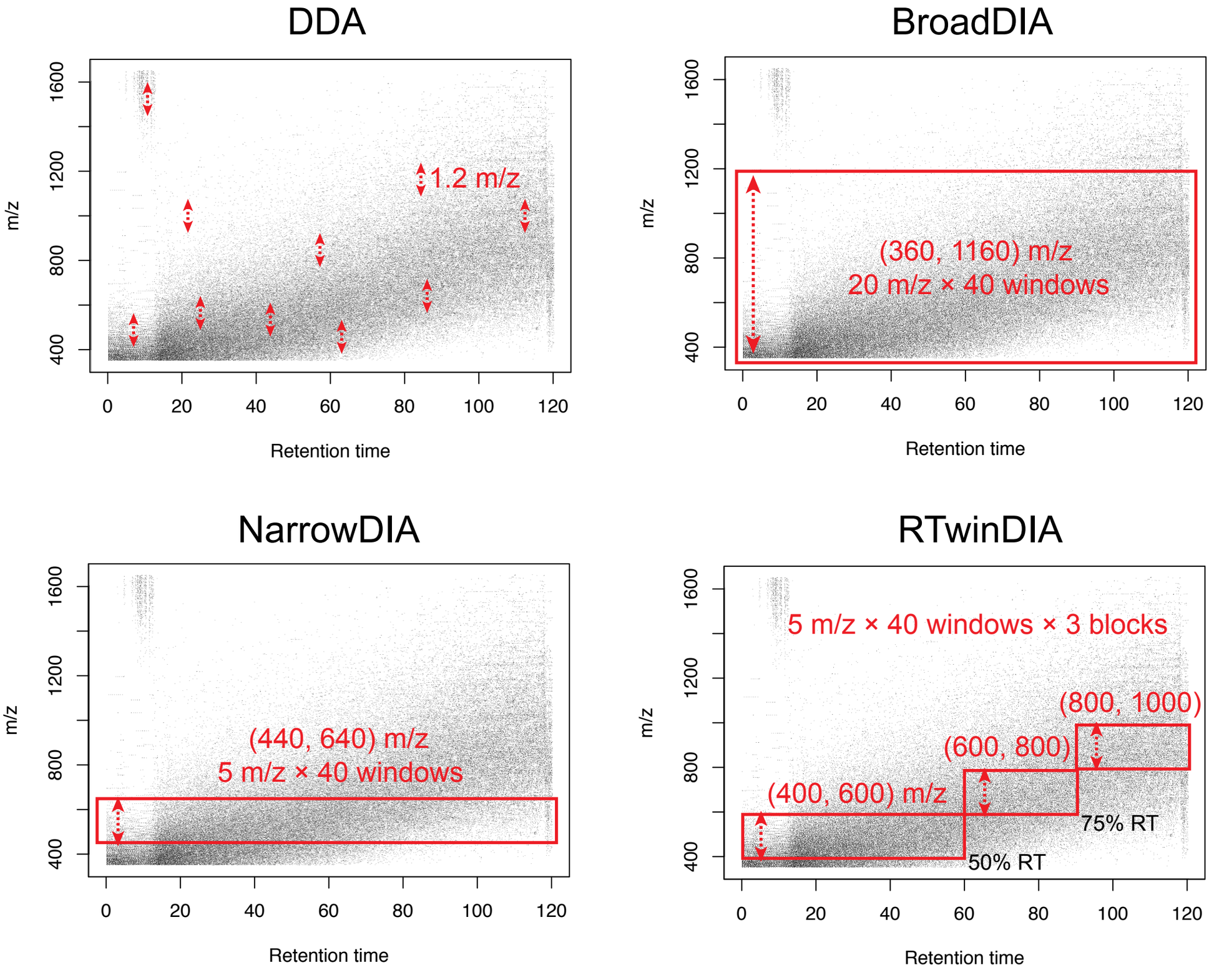

DDA
1.2 m/z
m/z
Retention time
BroadDIA
(360, 1160) m/z
20 m/z × 40 windows
m/z
Retention time
NarrowDIA
(440, 640) m/z
5 m/z × 40 windows
m/z
Retention time
RTwinDIA
5 m/z × 40 windows × 3 blocks
(800, 1000)
(600, 800)
(400, 600) m/z
75% RT
50% RT
m/z
Retention time

Figure 2

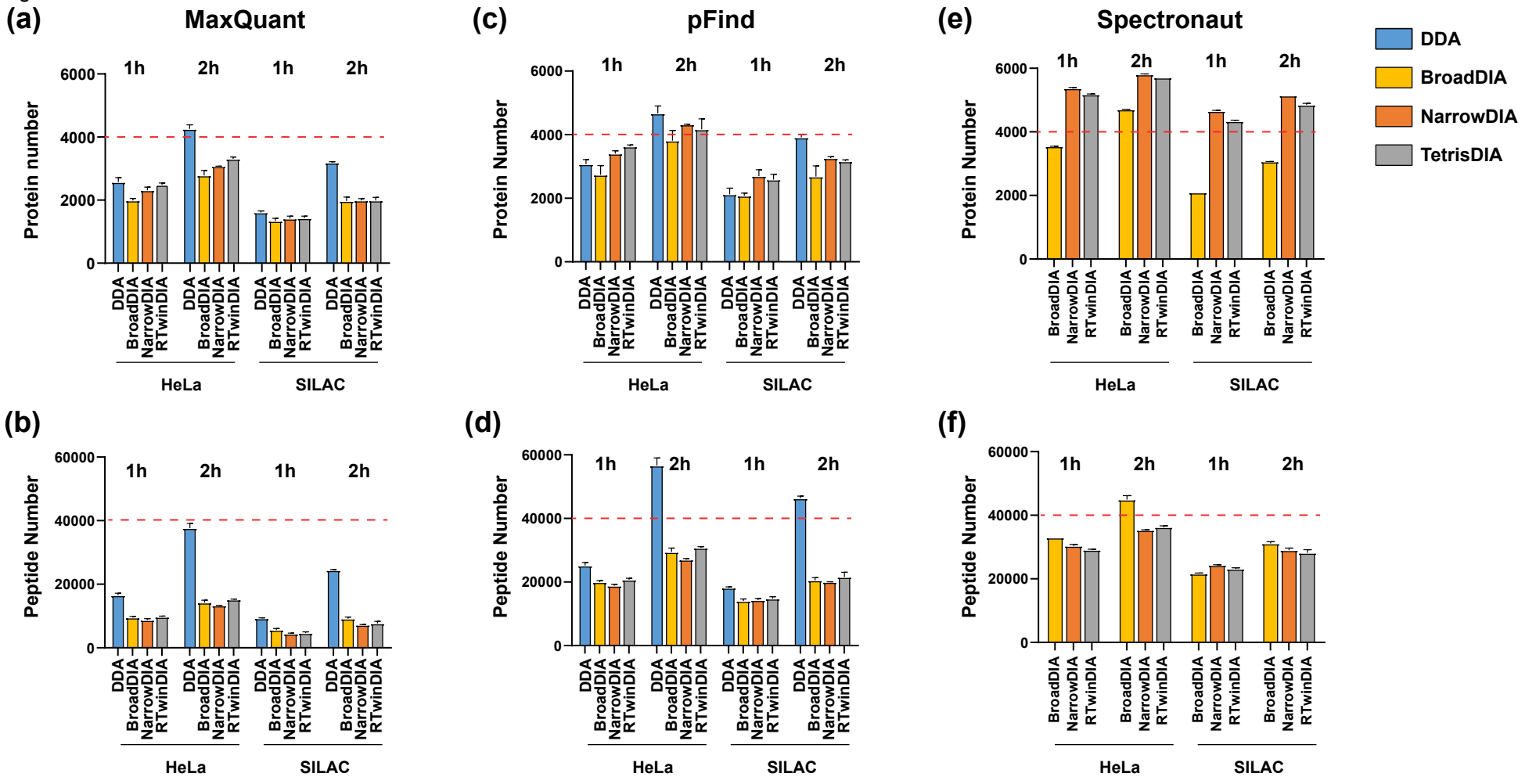

(a)
MaxQuant
(c)
pFind
(e)
Spectronaut
DDA
BroadDIA
NarrowDIA
TetrisDIA
Protein number
Protein Number
1h
2h
HeLa
SILAC
(b)
(d)
(f)
Peptide Number
RTwinDIA

Figure 3

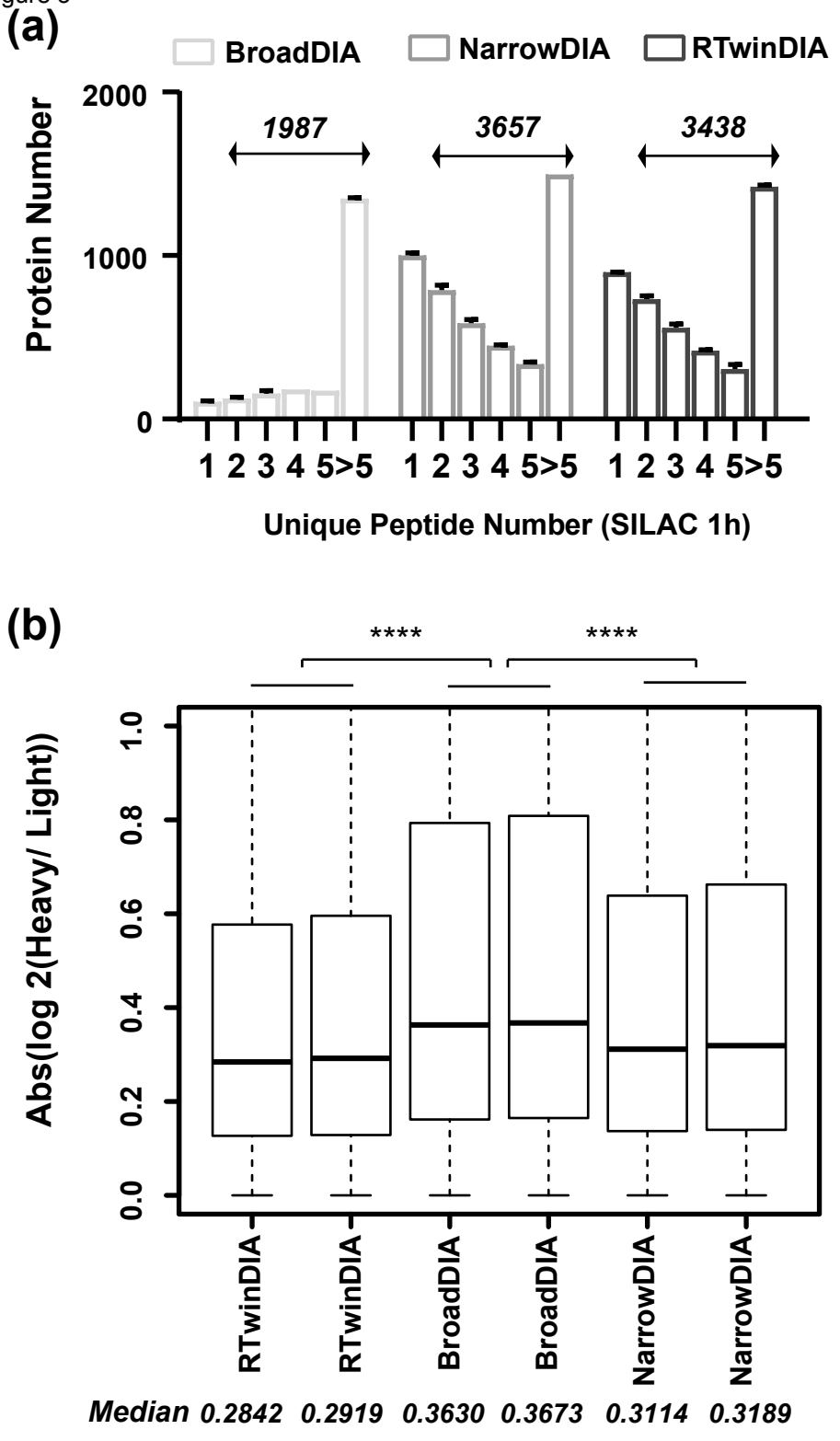
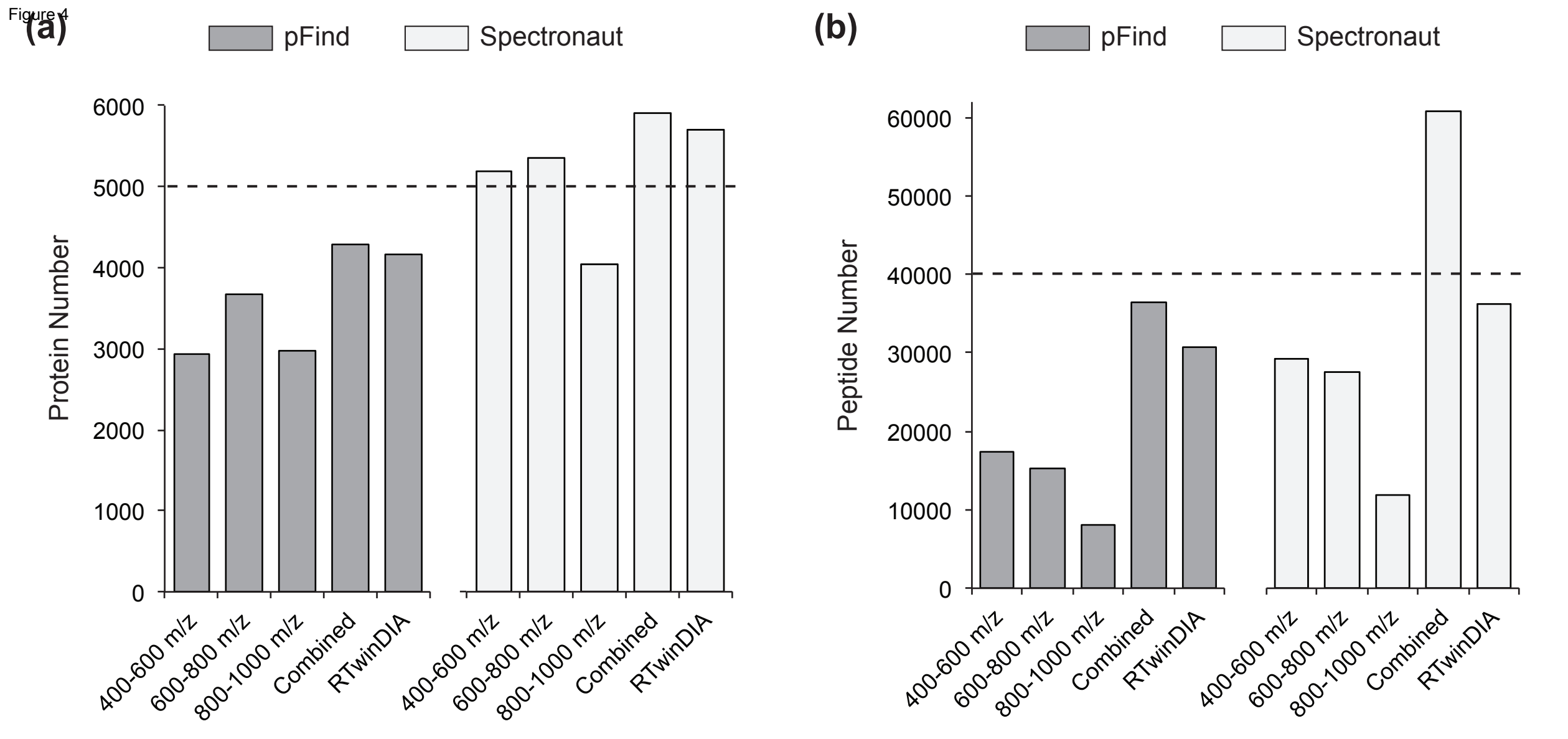

(a)
BroadDIA
NarrowDIA
RTwinDIA
Protein Number
2000
1000
0
1987
3657
3438
1 2 3 4 5>5 1 2 3 4 5>5 1 2 3 4 5>5
Unique Peptide Number (SILAC 1h)
(b)
****
****
Abs(log 2(Heavy/ Light))
0.0
0.2
0.4
0.6
0.8
1.0
RTwinDIA
RTwinDIA
BroadDIA
BroadDIA
NarrowDIA
NarrowDIA
Median 0.2842 0.2919 0.3630 0.3673 0.3114 0.3189

Figure 4


(a)
pFind
Spectronaut
Protein Number
6000
5000
4000
3000
2000
1000
0
400-600 m/z
600-800 m/z
800-1000 m/z
Combined
RTwinDIA
400-600 m/z
600-800 m/z
800-1000 m/z
Combined
RTwinDIA
(b)
pFind
Spectronaut
Peptide Number
60000
50000
40000
30000
20000
10000
0
400-600 m/z
600-800 m/z
800-1000 m/z
Combined
RTwinDIA
400-600 m/z
600-800 m/z
800-1000 m/z
Combined
RTwinDIA

Supplementary text and figures

Supplemental Table 1

Click here to access/download

Supplemental Information

Supplementary Table 1.xlsx

Supplemental Table 3

Click here to access/download

Supplemental Information

Supplementary Table 3.xlsx

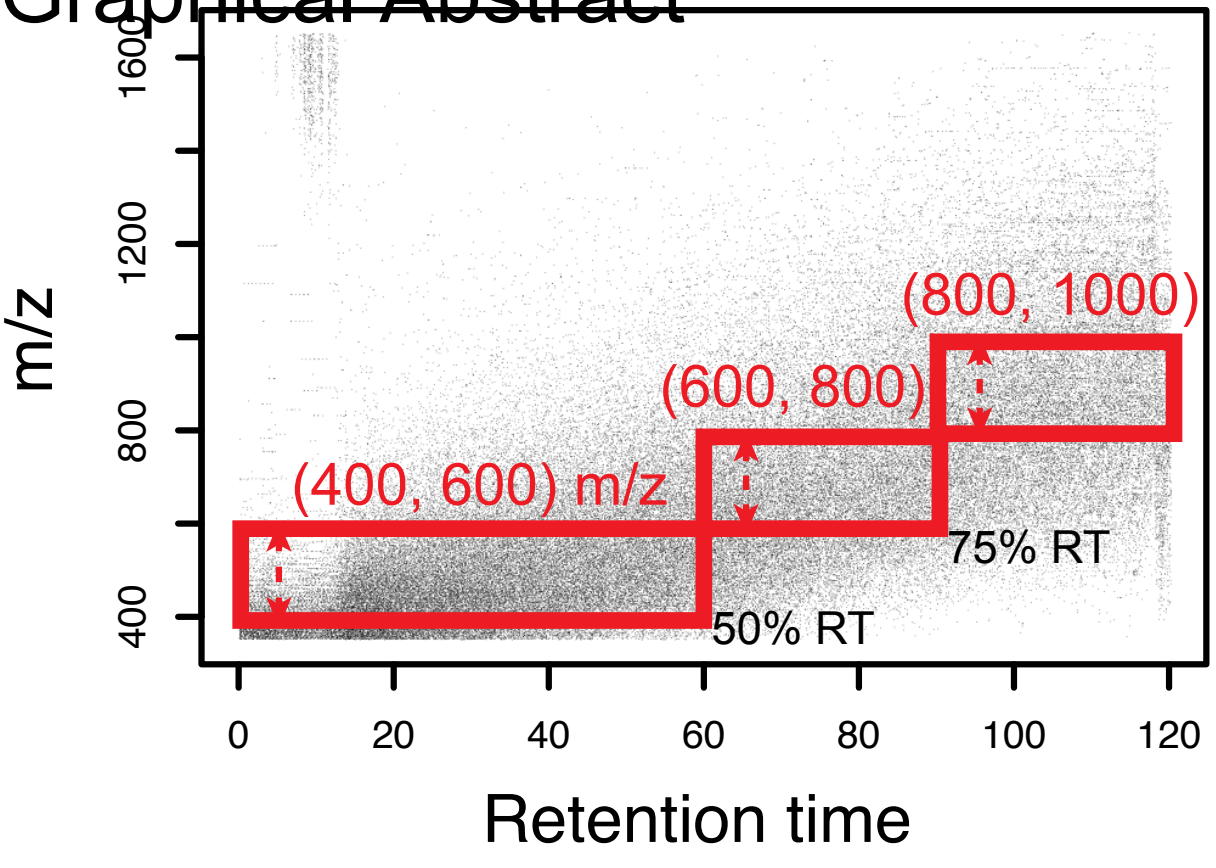
Supplemental Table 4

Click here to access/download

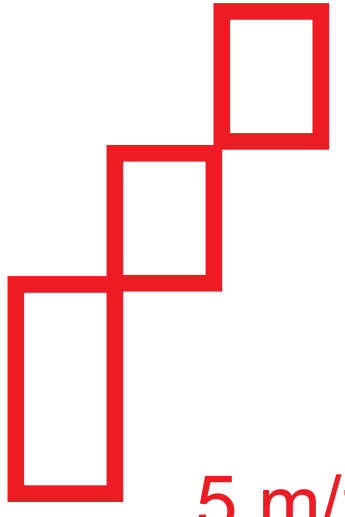
Supplemental Information

Supplementary Table 4.xlsx

Graphical Abstract



“RTwinDIA”



5 m/z ×
40 windows ×
3 RT blocks