Systematic Data Acquisition and Analysis Strategies for Quantitative Proteomics

Lilian Randolph Heil

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee:
Michael J. MacCoss, Chair
Andrew N. Hoofnagle
Devin K. Schweppe

Program Authorized to Offer Degree:
Genome Sciences

University of Washington

Abstract

Systematic Data Acquisition and Analysis Strategies for Quantitative Proteomics

Lilian R. Heil

Chair of supervisory committee:

Michael J. MacCoss

Department of Genome Sciences

Mass spectrometry-based proteomics has emerged as a powerful tool to gain insight into biological systems. As biological studies get more complex, there is a pressing need to increase throughput without sacrificing quality or completeness of the measurements. One limitation of many traditional mass spectrometry methods has been stochastic sampling causing poor reproducibility between runs. Systematic data acquisition methods promise to eliminate these missing values while improving sensitivity but are inherently inefficient and may be limited by instrumentation. Therefore, the field of proteomics could benefit from innovations in systematic data acquisition strategies. Here, we present the fundamentals of quantitative mass spectrometry and systematic data acquisition and analysis methods. We describe a head-to-head comparison of the linear ion trap and Orbitrap for parallel reaction monitoring, followed by a similar comparison between the Astral analyzer and the Orbitrap analyzer for data independent acquisition (DIA). We then present a novel method for improving the efficiency and therefore sensitivity of DIA by shifting acquisition windows across time. Finally, we propose a new search algorithm to build spectral libraries for DIA analysis without prior knowledge of the samples.

Table of Contents

Table of Figures

Table of Tables

Table of Algorithms

Acknowledgements

Although this is nominally my dissertation, this work belongs to the many people who helped me get to this point. Over the past 4 years I have leaned heavily on my mentors, colleagues, and friends and none of this would have been possible without them. Even though I cannot mention everyone by name, I want to thank a few people in particular.

First, I want to thank my thesis mentor, Mike MacCoss, and the rest of the MacCoss and Nunn labs for their unconditional support. As a mentor, Mike has at times believed in me more than I believed in myself, and his refusal to take "no, I don't know how to do that" for an answer is what led to a lot of the work here. I could not have asked for a better lab environment to do my PhD in. I want to thank all of the research scientists who have helped teach me and set a tone for the lab, especially Eric Huang, Gennifer Merrihew, Rich Johnson, Julia Robbins, and the whole Skyline team. Finally, I want to thank the other trainees for their camaraderie, notably Dani Faivre, Deanna Plubell, Chris Hsu, Miranda Mudge, Kristine Tsantilas, and Aaron Maurais.

I want to thank the rest of my committee, Devin Schweppe, Bill Noble, and Andy Hoofnagle and their labs, who were all major parts of my PhD. The proteomics community at UW is phenomenal and I want to thank the other proteomics trainees for creating a truly unique environment. Thank you to Uri Keich for his help and patience with my very rudimentary grasp of pseudo-code. Most of this work was done in collaboration with Thermo Fisher, and I want to thank everyone there for their support and for trusting me with their instruments. In particular, Philip Remes and Jesse Canterbury were a constant source of support (technical, intellectual, and moral).

I would not have even considered getting a PhD if it weren't for my time in the Hicks Lab at UNC. Leslie Hicks has been a mentor and inspiration to me since the day I met her. In addition to introducing me to proteomics, Leslie taught me to be confident and speak up even when I felt like (and often was) the least qualified person in the room. The rest of the Hicks lab has been an amazing support system, especially Emily Werth and Christine Radich who taught me so much about mass spectrometry and life and will be my GalPals4Eva.

I am lucky to have too many friends to thank here. I want to thank the whole GS 2019 cohort, especially Maddy Duran for being my roommate, dog sitter, and co-conspirator. I would not have made it through undergrad without my friends from UNC, particularly my Kappa Kappa Gamma sisters (GKODB) and the 2018-2019 residents of Graham Court Apartment B4. I want to thank my very best friend in the world, Olivia Berkey, for truly being with me from cradle to grave.

Finally, I want to thank my family for their unwavering support. Thank you to my cousins, siblings, nieces, and nephews for always keeping me humble. Especially thank you to by brothers, Mark and Gus, who show me every day that it's ok to be a nerd; my mom, Harrison, who taught me to always be kind and showed me that there are more ways to get things done than brute force; and finally my dad, John Heil, who is my biggest inspiration academically and taught me how to throw a baseball before I could walk.

Dedication

To Charles the dog, who reminds me every day that I need to take a break and get outside.

1 Quantitative proteomics

1.1 Introduction

The central dogma of biology lays out a blueprint for life, and states that information passes from the genome to the transcriptome to the proteome.[1] Of these three levels of information, proteins are the most structurally diverse and functionally relevant and understanding their interactions is critical to understanding biology.[2] Next generation sequencing technologies have allowed for large scale genome and transcriptome-wide studies.[3] However, the dynamic range and complexity of the proteome makes proteome-wide measurements significantly more difficult (**Figure 1-1**).[4,5] The cellular proteome spans 7-orders of magnitude[6] and the plasma proteome spans 10-orders of magnitude.[7] Therefore, analytical method development for proteome-wide measurements is uniquely challenging. Mass spectrometry (MS) is the dominant method for proteome-wide measurements due to its versatility and wide dynamic range.[8]

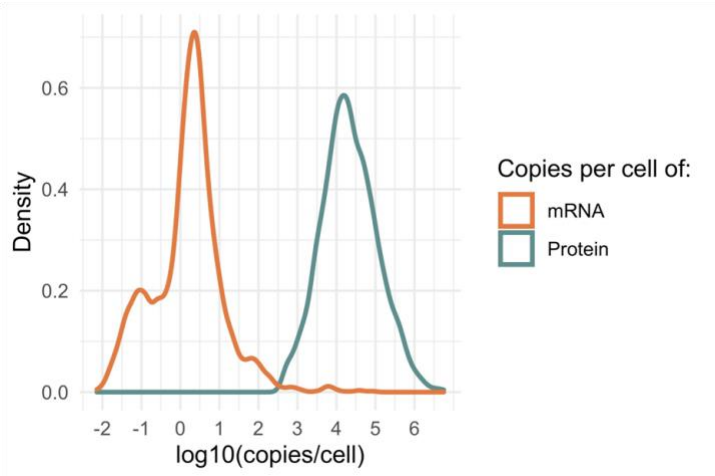

Figure 1-1: Copies per cell of mRNA and protein molecules in yeast
Adapted from Marguerat et al. (2013).[4]

Currently, bottom-up proteomics is the dominant methodology used for large scale quantitative proteomics studies.[9] In bottom-up proteomics, proteins are extracted and enzymatically digested into smaller peptides. Because small peptides are much easier to work with than large proteins, all measurements are done on the peptide level and then protein abundances are inferred. These peptides are separated via liquid chromatography (LC) and analyzed by tandem mass spectrometry (MS/MS).[9,10] The front end separation of liquid chromatography (LC) helps to reduce the complexity of the sample that the mass spectrometer must analyze at any given time, but also imposes a time limit to measure each peptide. Therefore, the MS methods used need to balance sensitivity with efficiency to maximize the quality and quantity of measurements. Mass spectrometers use mass analyzers to separate charged molecules by mass-to-charge ratio. In bottom up proteomics, a peptide is ionized and fragmented in the MS by collision with a gas, and fragments along the peptide backbone can be used to determine the identity and quantity of the peptide. Generally, a peptide identification algorithm is used to identify peptides from these tandem mass spectra based on backbone fragmentation.[11] The mass analyzer, acquisition mode, and analysis method all have major impacts on data quality and selecting the proper tools is necessary to obtain meaningful results.

1.2 Instrumentation

Mass spectrometers are highly complex instruments that fundamentally consist of the following elements: a sample inlet and ionization source where the sample is ionized in the gas phase, a mass analyzer where ions are separated by mass-to-charge ratio (*m/z)*, and a detector which counts the ions.[12] In tandem mass spectrometry, the sample is ionized, ions enter the mass spectrometer and move through ion optics under a vacuum. Ions of a specific *m/z* or range of *m/z'*s can be selected and transferred to a collision cell where they come in contact with a

collision gas that causes fragmentation. The fragment ions are transferred to the mass analyzer where they are separated by *m/z* and the detector reports signal intensity for each *m/z* as a spectrum. The specific components of the instrument vary depending on instrument type and manufacturer and can have drastic impacts on the type of data acquired. Here, we focus primarily on the mass analyzer and detector as they have a significant role in sensitivity and selectivity of measurements.

There are many different types of mass analyzers that are driven by different physical principles to accomplish the same goal of separating ions by *m/z*. The main performance characteristics of mass analyzers are: mass range, spectrum acquisition speed, ion transmission, mass accuracy, and resolution.[12] Spectrum acquisition speed can range from microseconds to seconds, and is often a limiting factor in quantitative proteomics where the front end separation imposes a time limit on peptide measurements.

Resolution or resolving power refers to the ability to separate two ions with a small *m/z* difference. Resolution is also closely tied with mass accuracy and can vary between analyzers (**Figure 1-2**). Resolving power is usually annotated as m/Δm, where m is the mass-to-charge measured, and Δm is the peak full width at half maximum. Resolution plays an important role in measurement selectivity, as lower resolution data means a given observed fragment could be attributed to more possible sources. As resolution and mass accuracy increases, the measurements become generally more selective. The minimum required resolution can range from unit resolution for many clinical assays[13,14] to 50,000 resolving power to distinguish between nitrogen isotopes[15] to over 2,000,000 for some top down proteomics applications.[16] In quantitative bottom-up proteomics, typically high (15-60k) resolution is used, but this is a parameter that should be careful considered in experimental design.

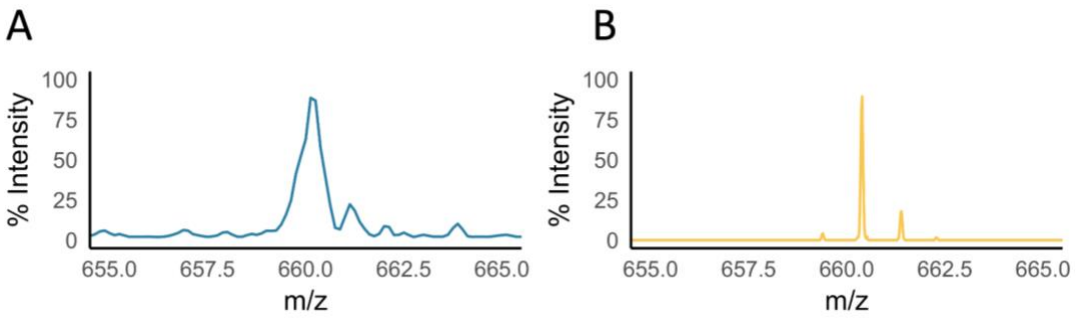

Figure 1-2: Profile spectra acquired with mass analyzers with different resolutions
A spectrum was acquired in linear ion trap at unit resolution (A) and Orbitrap at 30,000 resolving power (B) and the same fragment ion is plotted.

Sensitivity can be impacted by many different variables, including the efficiency of ion transmission to the analyzer and then to the detector, detector gain, and proportion of the ion beam that is actually sampled (duty cycle). The requisite sensitivity for a given experiment depends on the analytes of interest and, generally speaking, quantifying less abundant proteins is more challenging.

Selecting instrumentation that will provide sufficient sensitivity is critical for any experiment and is another factor to be balanced. Ultimately, many factors, including non-performance related considerations such as cost, availability, and familiarity with the instrument platform, must be considered when choosing the right mass spectrometer for an experiment. Nonetheless, understanding the analytical performance of an instrument and how it is impacted by speed, sensitivity, and resolution is beneficial to the community as a whole and is largely lacking in many cases.

1.3 Data acquisition modes

Beyond selecting appropriate instrumentation, it is important to consider how the instrument will acquire data to meet the needs of an experiment. Traditionally, there have been

three main modes of data acquisition: data dependent acquisition (DDA),[17] data independent acquisition (DIA),[18] and targeted data acquisition.[19,20] Despite major advances in instrumentation over the past two decades, many studies still use rudimentary data acquisition strategies, suggesting that there is room for improvement in data acquisition techniques. Acquisition methods must balance sensitivity and reproducibility with the number of peptides measured. Generally, sensitivity and total coverage are in opposition, and therefore maximizing instrument efficiency is critical to allow for high quality measurements of as many peptides as possible.

Because early mass spectrometers were slow (less than 1 Hz acquisition speed), the earliest data acquisition methods relied on manual selection of precursor peaks for fragmentation. DDA was developed as a way to automate the precursor selection process while still selecting precursor ions that were most likely to yield useful spectra. In DDA, a survey spectrum is acquired, and a heuristic is applied to select peaks for MS/MS fragmentation (**Figure 1-3 A**). In the simplest form of DDA, the top N most intense candidate peaks are selected for fragmentation, although in practice there are some exclusion criteria applied to minimize redundancy and maximize the chances of fragmenting a peptide. In DDA, there is only one MS/MS spectrum per peptide so quantification is generally achieved by calculating area under the curve of the precursor signal (label free quantification or LFQ)[21,22] or by using isobaric tandem mass tags (TMT)[23] where quantification is based off reporter ion intensity in a single MS2 or MS3[24] spectrum. The primary limitation of DDA is the stochastic nature of sampling, which can lead to a large portion of missing values in large experimental cohorts.[25] Label-free approaches can use chromatographic alignment and match between runs to integrate MS1 signals in the absence of an identification,[22,26] but TMT approaches have no recourse if a peptide is not

sampled. Recent work has focused on improving reproducibility of MS3 TMT quantification to allow for complete data across larger cohort sizes.[27,28]

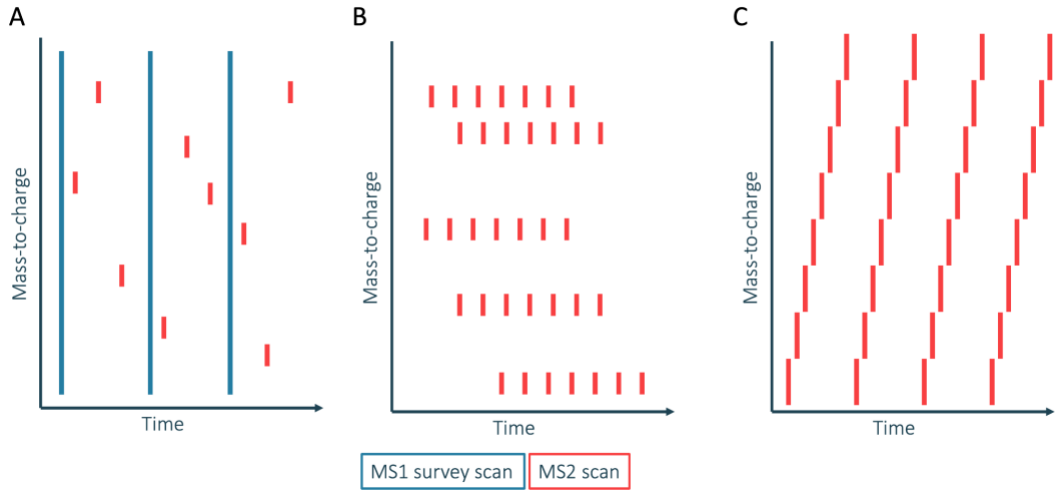

Figure 1-3: Comparison of acquisition strategies

MS/MS spectra acquired in DDA (A), targeted acquisition (B), and DIA (C).

Targeted data acquisition offers an alternative to DDA which solves the missing values problem. In targeted acquisition, the instrument is given a list of precursors to target and it systematically acquires MS2 data for these pre-determined precursors over time (**Figure 1-3 B**). Quantification is achieved by integrating the area under the curve of the fragment ions, which is highly selective and reproducible between runs (**Figure 1-4**). Targeted methods include selected reaction monitoring (SRM)[19] and parallel reaction monitoring (PRM)[20] which apply the same principles on different types of instrument platforms. The primary issue with targeted methods is the lack of efficiency, as many spectra may be acquired for a precursor when it is not detectable. This efficiency has traditionally limited the scope of assays, although real time strategies have been implemented to improve efficiency, increasing assay capacity from a few to thousands of targets in a single run.[29–32]

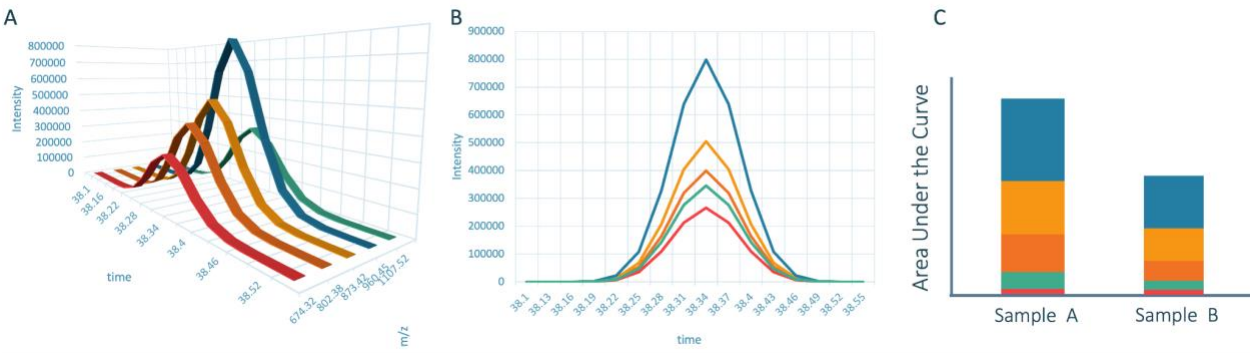

Figure 1-4: MS/MS quantification scheme

Spectra consisting of ion intensity values for different m/z's are acquired across time (A). Intensity of fragment ions in successive spectra are used to generate a peak (B). Area under the curve is used to determine abundance (C).

Data independent acquisition is a powerful acquisition technique that systematically fragments all ions in a specified mass range that has grown in popularity in recent years largely due to improvements in instrument capabilities. In DIA, the instrument systematically acquires MS/MS spectra across a set mass range (**Figure 1-3 C**). DIA fragments peptides in an unbiased manner and allows for MS2 level quantification (**Figure 1-4**),) that is similar to targeted approaches but without any specific peptide targets. Theoretically, DIA gives unbiased coverage across the full proteome and will contain no missing values. However, the downside is many MS/MS spectra may contain no useable information, and valuable instrument time may be wasted on empty spectra. As instruments have become faster and more sensitive, this problem may be less important but there is still room for improvement.

While the limitations of traditional DDA are clear, it is evident that some form of data dependency can help increase efficiency and quality of measurements. Instrument computers are now capable of executing complex algorithms on a millisecond time scale and are no longer limited to using simple acquisition heuristics. An ideal acquisition method should combine the

reproducibility and sensitivity of systematic sampling methods such as DIA and PRM with the efficiency of DDA.

1.4 Data analysis strategies for DIA

Whereas DDA experiments aim to generate simple spectra for a limited number of peptides, DIA experiments aim to collect data in a systematic fashion that balances spanning a significant portion of the m/z range and sampling at a frequency that facilitates sampling a chromatographic peak. To do this, there is often a sacrifice to the precursor isolation efficiency and, thus multiple peptides are often sampled in parallel. Traditionally, search methods were designed for DDA and used a spectrum-centric approach, where each spectrum is queried to identify the best-matching peptide. Spectrum-centric searches assume that each spectrum is generated by a peptide and thus seek to identify the peptide generating each spectrum.

In contrast to DDA data, DIA data is collected in an unbiased manner, and the assumption that each spectrum was generated by a peptide does not hold. A more logical way to search DIA data is using a peptide-centric approach to find and score the best evidence for all candidate peptides independent of the spectrum. One key feature of peptide-centric searches is allowance for the same spectrum to provide evidence for multiple peptides, which is often the case when there are a large number of chimeric spectra.[33] However, this strategy becomes less reliable in cases where different candidate peptides in the same precursor isolation window share a significant number of theoretical ions. In these cases, the same set of observed ions could be used as strong evidence for multiple peptides, leading to an artificially high number of peptide detections.

This problem is exacerbated in DIA data using windows wide enough that post-translationally modified peptides could be considered in the same precursor isolation window as their unmodified counterparts. However, even when very narrow (2 m/z) isolation windows are used, there may be isobaric forms or amino acid substitutions that barely alter the mass. Traditional target/decoy methods fail to control false-discovery rates in these cases as a higher scoring precursor is much more likely to be a target than a random decoy due to sequence conservation and similarities between modified and unmodified peptides.[34] Therefore, additional steps need to be taken in the analysis of DIA data to avoid double counting related ions.

There are several approaches that seek to address the risk of double counting in DIA data. However, many widely used approaches completely ignore the problem and can lead to inflated numbers of peptide detections. Methods to reduce the window size, both experimentally and computationally, can greatly reduce the risk of double counting as more modifications would fall into other isolation windows.[35–37] Other methods may attempt to deconvolve spectra prior to processing (Peckner, DIA-Umpire),[37,38] or to identify cases where multiple candidates share ions and keep only the higher scoring of the two.[39] Currently, the field does not have a rigorous standard to address the problem of double counting, and without any consensus within the field, the results of any DIA experiment should be received with caution.

2 Comparison of unit resolution versus high-resolution accurate mass for parallel reaction monitoring

2.1 Summary

This chapter is based off of the following publication: Lilian R. Heil, Philip M. Remes, Michael J. MacCoss. Comparison of unit resolution versus high-resolution accurate mass for parallel reaction monitoring. J Prot Res. 2021, 20 (9), 4435–4442.

Parallel reaction monitoring (PRM) is an increasingly popular alternative to selected reaction monitoring (SRM) for targeted proteomics. PRM's strengths over SRM are that it monitors all product ions in a single spectrum, thus eliminating the need to select interference-free product ions prior to data acquisition, and that it is most frequently performed on high-resolution instruments, such as quadrupole-orbitrap and quadrupole-time of flight instruments. Here, we show that the primary advantage of PRM is the ability to monitor all transitions in parallel, and that high-resolution data are not necessary to obtain high quality quantitative data. We run the same scheduled PRM assay, measuring 432 peptides from 126 plasma proteins, multiple times on an Orbitrap Eclipse Tribrid mass spectrometer, alternating separate liquid chromatography-tandem mass spectrometry runs between the high resolution Orbitrap and the unit resolution linear ion trap for PRM. We find that both mass analyzers have similar technical precision, and that the linear ion trap's superior sensitivity gives it better lower limits of quantitation on over 62% of peptides in the assay.

2.2 Introduction

Targeted proteomics is a strategy to obtain accurate and precise quantitation by maximizing the time a mass spectrometer (MS) spends measuring a pre-defined subset of analytes in a sample. Historically, targeted assays have been performed using selected reaction monitoring (SRM) on a triple quadrupole mass spectrometer.[40] In an SRM assay, a specific list of precursor and product ion pairs, called transitions, are monitored by setting the quadrupoles to transmit specific gas-phase reactions at specific times during the liquid chromatography (LC) gradient.[41,42] SRM is widely used today due to its reliability and versatility, but it still has limitations. Notably, the need to select interference-free transitions can limit the selectivity and breadth of an experiment; because each precursor/ product ion pair needs to be targeted individually, each additional product ion measured comes at a cost to overall cycle time. Consequently, there is a tradeoff between increasing selectivity and measuring more precursor ions because targeting more precursors means sacrificing selectivity by monitoring fewer transitions per precursor. Additionally, any change to the set of product ions requires rerunning the entire experiment.

Parallel reaction monitoring (PRM) is an alternative to SRM where the entire product ion spectrum is collected for a target precursor instead of selected precursor-product ion reactions. Early versions of PRM showed the potential to benefit from the full scanning capability of a unit-resolution ion trap.[43] However, PRM was popularized on high resolution quadrupole-Orbitrap (q-OT)[20,44] and quadrupole-time of flight[45] mass spectrometers. Since these demonstrations, the performance of PRM was largely attributed to two factors: the ability to monitor more product ions per precursor and the higher confidence in product ions associated with the high resolution product ion spectra.[20,46,47] However, we hypothesize that the primary benefit of PRM is the

ability to measure many product ions in a single scan, and that high resolution spectra are not critical to its success.

To assess the importance of high resolution mass analyzers in PRM, we sought to compare the performance of a unit resolution quadrupole linear ion trap (q-LIT) to a high-resolution q-OT for PRM experiments. We use an Orbitrap Eclipse Tribrid instrument to run a controlled experiment and establish that a q-LIT can be used on its own for targeted quantitative proteomics assays without sacrificing significant precision or quantitative accuracy compared to the q-OT. This instrument has both an Orbitrap and a linear ion trap that can be operated independently, which allows us to inject the same sample multiple times, alternating between the Orbitrap and linear ion trap with each injection, thereby eliminating the variability inherent to comparisons where multiple instruments are used (**Figure 2-1**).

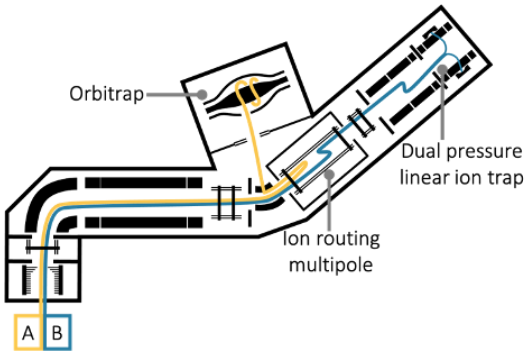

Figure 2-1: Schematic of the Orbitrap Fusion Eclipse Tribrid instrument
Ion paths are highlighted in yellow and blue. To compare Orbitrap to ion trap directly, alternate injections were performed using each mass analyzer. Injections either followed path A, where the instrument operated as a quadrupole-Orbitrap or path B, where the instrument operated as a quadrupole-linear ion trap.

2.3 Methods

Complete methods are discussed in Appendix A

2.4 Results and Discussion

Creation of the PRM assay

To compare mass analyzers, we chose plasma as a sample matrix because it is clinically relevant and its large dynamic range[7,48] is well suited for testing the limits of targeted peptide quantitation. To develop a robust PRM assay, we started with a DIA-based chromatogram library consisting of 3084 peptides mapping to 358 proteins. This library was developed using gas-phase fractionated DIA data that was searched against the Pan-Human library[49] in EncyclopeDIA.[50] The resulting chromatogram library contained information on the identity, fragmentation spectra, and retention times of all detectable peptides. This original library was created using a different instrument, liquid chromatography, and sample than those used in this PRM experiment. Therefore, the library needed to be refined to create a scheduled targeted method that could be acquired with a sufficient number of points across each chromatographic peak.

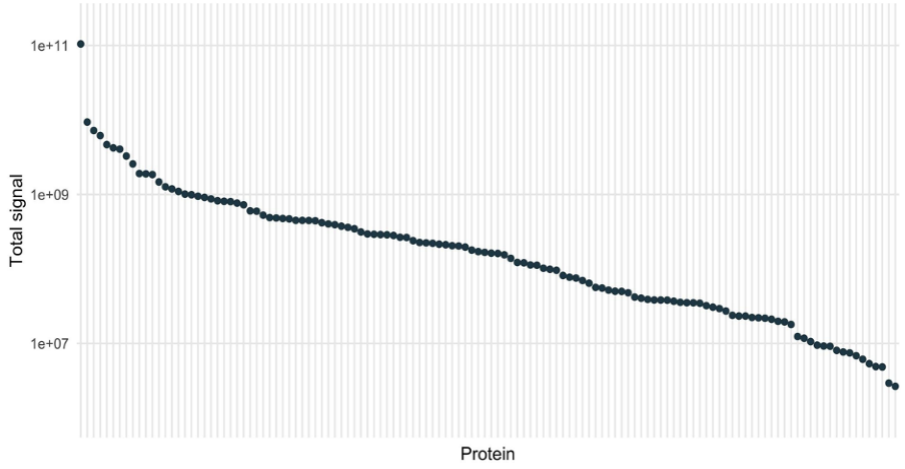

Figure 2-2: Plot of dynamic range of all proteins in the assay

Each point represents the summed area under the curve abundance of all peptides mapping to that protein. The total signal ranges from 1.0×10^{11} for serum albumin down to 2.6×10^{6} for vitamin K-dependent protein C.

To generate a targeted assay from the existing chromatogram library, a DIA run was collected using staggered 8-*m/z* windows with the Orbitrap mass analyzer. The resulting data were used to determine accurate retention times for all peptides. Based on these data, the original peptide library was filtered for peptides that were 1) confidently seen in this sample, 2) biologically relevant (based on a previous desire to focus on proteins useful for assessing general wellness), and 3) came from proteins spanning a wide dynamic range (**Figure 2-2**). To fulfill these criteria, the five peptides per protein with the most intense peak areas from the product ion data were selected, and the list was further refined to ensure all peptides had high dotp with the library chromatograms. Additionally, peptides were preferentially selected if they came from proteins that have been used in other MS-based targeted plasma assays.[51] The final assay consisted of 432 peptides derived from 128 protein coding genes such that upon scheduling no more than 75 peptides would be measured at any given time using a five-minute retention time window (**Figure 2-3**).

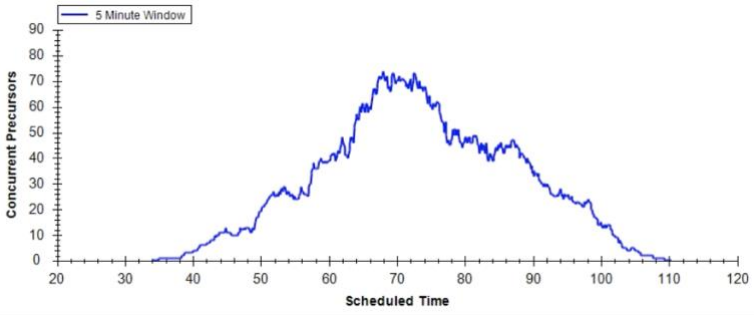

Figure 2-3: Number of precursors monitored at any given time
Scheduling was based on measured retention time with a 5-minute window. Retention times were updated throughout the experiment to account for small changes in chromatography.

The primary consideration in PRM assay development is limiting the number of targets monitored at one time to maintain a short enough cycle time to sample at least eight points across

each peak without sacrificing sensitivity. In the Orbitrap, the length of time required for spectral acquisition does not depend on the range of mass-to-charge, because of the multiplexing property of Fourier transform-based analysis. As a scanning instrument, the linear ion trap pays a time penalty proportional to the range of mass-to-charge acquired, however the total scan time for the range 150–1250 *m/z* is only 8.8 ms with the 125 kTh/s scan speed. On both the Orbitrap and linear ion trap, the analysis takes place while the collision cell is being filled with ions for the next spectrum, so the cost of mass analysis to the overall duty cycle is quite low. The time allotted for ions to accumulate prior to spectral acquisition is typically the limiting factor in cycle time and is crucial for maintaining sensitivity. While increasing the maximum ion injection time could improve sensitivity, it comes at a cost to cycle time. Therefore, we used dynamic maximum injection time to balance speed and sensitivity across the run. Critically, both analyzers maintain a fundamental advantage over triple-quadrupoles because all transitions are collected within a single MS/MS spectrum and therefore can be refined after data acquisition. In contrast, each additional transition measured costs an additional dwell/acquisition time in SRM. This benefit of PRM eliminates the need to perform time consuming experiments to select interference-free transitions prior to acquisition, as is the case when transferring accurate mass data to triple-quadrupole assays.

Although high-resolution mass analyzers have traditionally been favored for PRM, linear ion traps may be ideally suited for PRM experiments due to their speed, sensitivity, and ability to scan whole mass ranges without significantly raising the cycle time.[52] Still, the speed and sensitivity of linear ion traps come at a cost to the resolution, which could present challenges for PRM experiments. First, it may be difficult to select the correct peak for a peptide over the course of a run, there can be multiple peptides with similar precursor masses and shared product

ions which generate multiple possible peaks (**Figure 2-4**). By utilizing a chromatogram library with retention times and fragmentation patterns calibrated on a high resolution instrument, it is possible to build an assay without using synthetic standards to ensure accuracy. A second potential problem with unit resolution is the increased potential for chemical interference, leading to inaccurate quantitation. However, we argue that this problem can be addressed simply by selecting interference-free fragment ions. Furthermore, it is possible that the increased speed and sensitivity of a q-LIT could make it better suited for PRM assays than its high resolution counterparts.

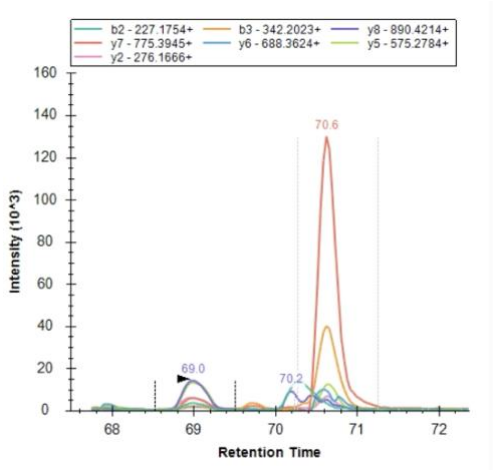

Figure 2-4: Extracted ion chromatograms for all product ions in a 5-minute window for one precursor collected with linear ion trap

Actual peak at 69.0 minutes could be determined based on retention time and accurate product ion fragmentation patterns. Without retention time information, this determination could be difficult.

Evaluation of technical precision

To assess the impact of precursor and product ion selectivity on technical precision, the same sample of human plasma digest was injected and measured using both the linear ion trap and Orbitrap each with two different precursor isolation widths. By collecting multiple replicates

with the same sample on the same instrument, it was possible to compare the technical variance of the mass analyzers directly. Based on five replicate injections, the percent coefficient of variation (% CV) was determined for each peptide in each method (**Figure 2-5 A**). The Orbitrap had lower technical variation than the linear ion trap for most peptides (**Figure 2-5 B**), but the differences in variation would likely make a minor difference in practice where anything below a 10% CV is generally considered high quality. At the 1.7 Th precursor isolation width, the median CV in the linear ion trap of 5.8% was slightly higher than the Orbitrap's 4.1%. Notably, both median CV values are more than adequate for quantitative experiments and at the same 1.7 Th precursor isolation width, the linear ion trap had a higher proportion of peptides, 98.3%, than the Orbitrap, 97.6%, below an arbitrarily selected 20% CV cutoff. The 20% CV cutoff is frequently used as a threshold for quantifiable peptides, and is an important factor in comparing these performances. While the Orbitrap maintains slightly better median precision than the linear ion trap, the practical differences in precision between using these two analyzers are small.

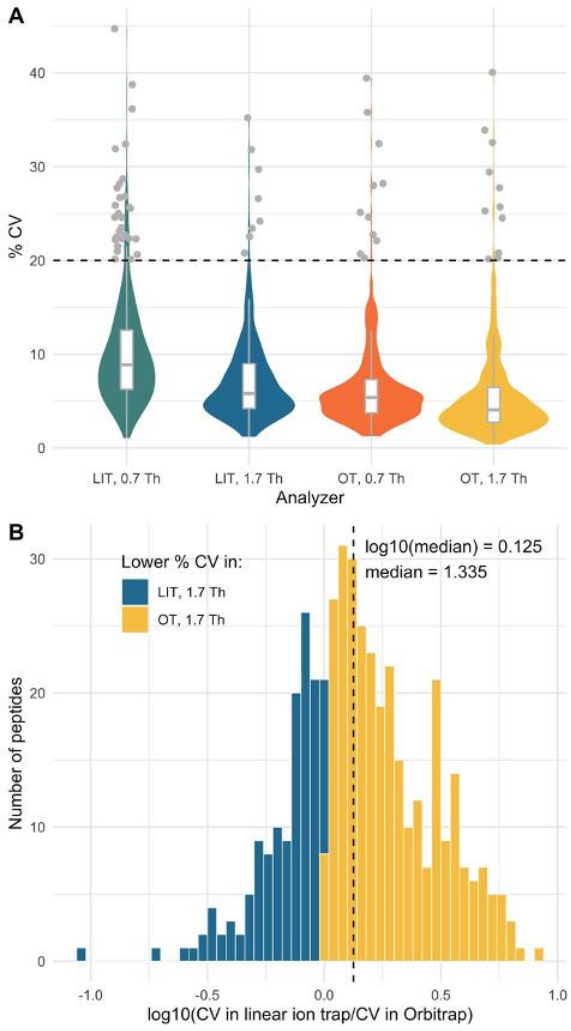

Figure 2-5: Comparison of coefficients of variation between PRM data acquired with the orbitrap (OT) and the linear ion trap (LIT)

CV's for each peptide in the assay when monitored with both mass analyzers at two precursor isolation widths (0.7 and 1.7 Th) (A). The log ratio of every peptide CV in linear ion trap and Orbitrap, where a positive log ratio indicates a better CV in the Orbitrap than linear ion trap (B).

One factor that can lead to high technical variation is the presence of interfering species. Interference can occur when an interfering precursor ion is co-isolated with the target precursor ion and generates product ions that fall within the integration boundaries of the target product ions, creating a shoulder in one or more product ion chromatograms (**Figure 2-6A and B**).

Between runs, slight changes in chromatography can shift the relative position of the target peptide and interfering species, causing significant changes in quantitation (**Figure 2-6C**). Chemical interference of this nature can be influenced by both the precursor and the product ion selectivity.

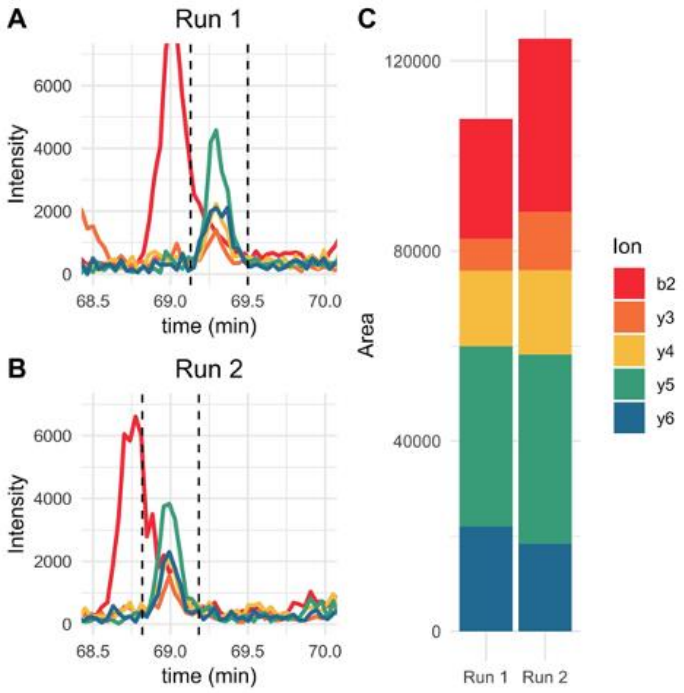

Figure 2-6: Example chromatograms of a peptide with interference
Chromatograms show significant interference impacting the b2 ion (A&B). The variance of the integrated peak area for this peptide is mostly influenced by the b2 ion getting integrated slightly differently in both runs (C).

Precursor ion selectivity is partially determined by precursor isolation width. If a narrower isolation window is used, potentially interfering species may not be co-isolated at all, eliminating that interference. However, reducing the isolation window comes at a cost to sensitivity as the narrow quadrupole transmits fewer ions. Therefore, quadrupole isolation width should be adjusted to strike a balance between sensitivity and selectivity. For both the Orbitrap and the linear ion trap, the 0.7 Th precursor isolation window tended to have higher technical

variance because the traps tended to fill with fewer ions (**Figure 2-7**). When there are fewer ions to measure the relative variance tends to increase[53] which explains the increased precisions with a wider precursor isolation window. Both analyzers performed better with the 1.7 Th isolation width, so those methods were used for the remaining experiments and analyses.

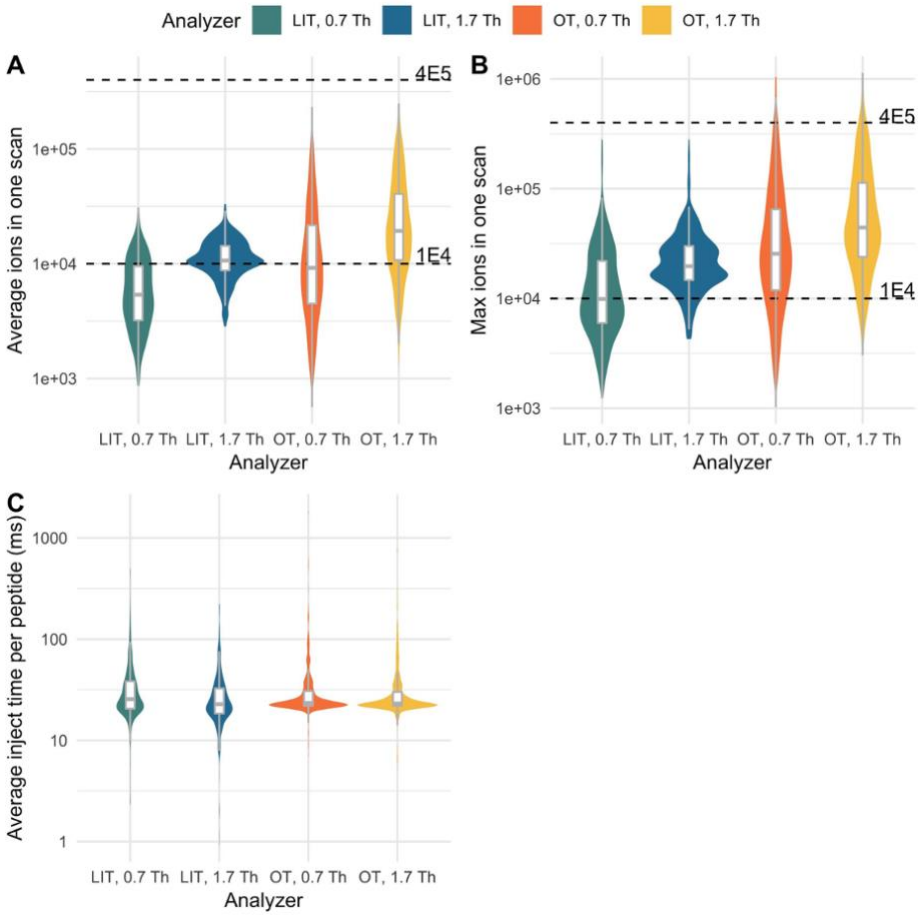

Figure 2-7: Ion fills and inject times for each analyzer

Plots display the average number of ions per scan (A) or maximum number of ions in a scan (B) for all peptides in each mass analyzer. The number of ions per scan was calculated by multiplying the total ion current by the inject time for that scan and dividing by 1000. For the Orbitrap, this value was then multiplied by 0.0917, an instrument and transient length specific conversion factor obtained by measuring the ratio of ion trap peak intensity to Orbitrap peak intensity for a series of calibration ions. When a narrow quadrupole isolation width is used, the mean and max number of ions

gets pretty low, which would explain the higher variance at these isolation widths. Another metric to look at fill is the average inject time per spectrum in each of the analyzers (C). Orbitrap has a narrower distribution of inject times, likely because the maximum inject time was hit more frequently with an AGC target of 1,000,000 ions compared to 50,000 in the Astral analyzer.

Similarly, product ion selectivity is related to the MS/MS resolution. In a lower resolution spectrum, there is a greater chance that another analyte will have an indistinguishable *m/z* at a specific retention time and is therefore more likely to have interference. This difference is reflected in the necessary integration boundaries for each analyzer: in the unit resolution linear ion trap, we must integrate the area +/- 0.7 *m/z* around the theoretical peak, whereas the Orbitrap only requires integration boundaries of 20 ppm. Thus, the unit resolution linear ion trap has the potential to have more interference than the Orbitrap, but with the careful selection of transitions, this problem may be minimized (**Figure 2-8**).

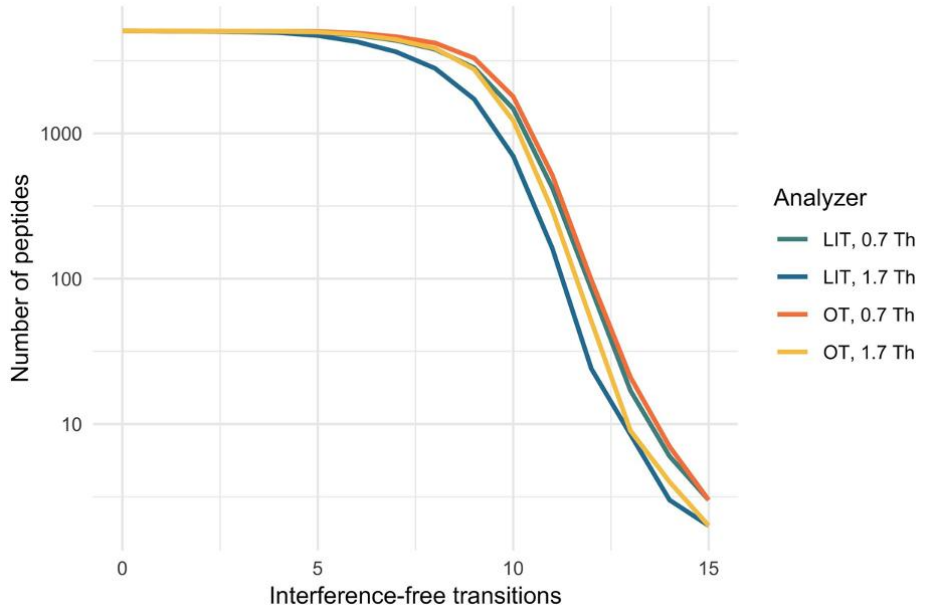

Figure 2-8: Predicted number of interference-free transitions for all peptides in the plasma chromatogram library

The y-axis represents the number of peptides with x transitions or more. Interference-free transitions are defined as precursor/ y-ion pairs that do not overlap within the appropriate m/z tolerance with b- or y- ions generated by precursors within the specified isolation window. At 1.7 Th, the linear ion trap is more susceptible to interference in general, but there is little loss if four or fewer transitions are selected.

The fear of increased variance due to increased interference has been a major reason why the proteomics community has largely trended towards using high resolution mass analyzers for PRM. Indeed, the transition set used in this experiment is likely to have more interference for the linear ion trap than the Orbitrap (**Figure 2-9**), although this prediction does not account for separation of potential interfering species across time. In fact, our results suggest that chemical interference of this nature was minimal, even in the linear ion trap. Thus, the Orbitrap's superior performance can be attributed to several factors, including the fact that it had a higher AGC target and tended to measure more ions per spectrum than the linear ion trap (**Figure 2-7**), a factor that has been demonstrated to improve precision.[31,54,55] Still, the Orbitrap tended to hit its

maximum inject time in most cases (**Figure 2-7C**) and was below its target fill capacity, which is likely why the difference in precision between the analyzers is not more pronounced. Because the Orbitrap was consistently below standard target fill capacity, increasing the AGC target would not have an impact on the number of ions in the trap. Increasing maximum injection time could allow the Orbitrap to fill with more ions, but would slow the cycle time, reducing the number of points across the peak and harming accuracy. Thus, we used the dynamic maximum inject time setting which optimizes the maximum injection time for speed and sensitivity across the run.

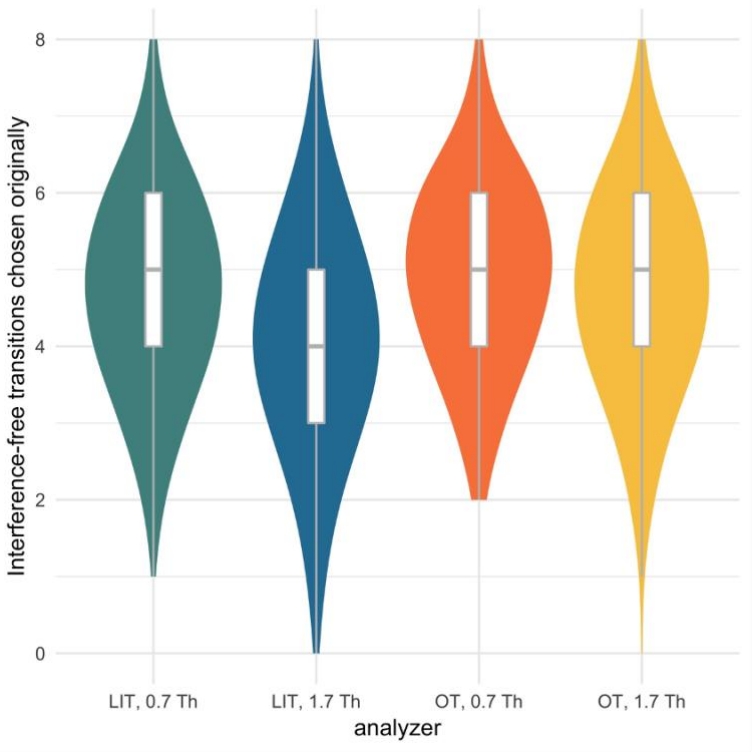

Figure 2-9: Number of theoretically interference-free transitions for each experiment

The original transition set has more interference for linear ion trap with the parameters used in matrix-matched experiment (LIT, 1.7 Th compared to OT, 1.7 Th). However, even

in this set, most peptides have at least four interference-free transitions before refinement.

Assessment of sensitivity with matrix-matched calibration curves

To assess the sensitivity of each analyzer, we determined the lower limit of quantitation for each peptide using a matrix-matched calibration curve.[56] To determine the LLOQ for all peptides in the assay, an 11-point dilution curve was prepared by diluting human plasma in chicken serum (**Appendix Table 1**). This matrix-matched method allows for the dilution of analytes without significantly changing the matrix composition.[56,57] To reduce run-order bias, the complete dilution curve was run with each method from low to high concentration, randomizing the order of analyzers within each point of each replicate, and the process was repeated for a total of three replicates. Data were collected by scheduled PRM using either the Orbitrap or LIT with a 1.7 Th precursor isolation window to acquire the product ion spectra. The PRM data were imported into Skyline where the area under the curve was assessed for each peptide. A report was generated with the peak areas and analyzed using a script modified in house based on the method developed by Pino et al.[56] This method uses a bilinear fit to assess the turning point in the calibration curve and then uses a bootstrapping method to find the lowest concentration where the CV was below 20%. This powerful method has many benefits over alternative methods, including that it is able to estimate LOQs between dilution points, instead of simply finding the lowest measured quantitative point.[56] The ratio of the LOQ for any given peptide in each mass analyzer could be used to compare these results directly.

Peptides tended to have lower limits of quantitation in the linear ion trap compared to the Orbitrap (**Figure 2-10 A**). To obtain a more complete understanding of analyzer performance, the LOQ for each peptide was compared individually by dividing the LOQ in the linear ion trap

by the LOQ in the Orbitrap. A ratio less than one indicates that the linear ion trap has a lower (better) LOQ, whereas the reverse is true for a ratio greater than one. The distribution of these ratios skews in favor of the linear ion trap, with a median ratio of 0.858 (**Figure 2-10 B**). This ratio indicates that LOQs in the linear ion trap were generally lower than those in the Orbitrap, suggesting the linear ion trap more than compensates for the greater potential for interference. The sensitivity of the analyzer can be just as critical as precision in evaluating performance—more sensitive mass analyzers require less analyte to generate meaningful signal. We use the lower limit of quantitation, or the minimum point on a dilution curve where a change in signal reflects a change in quantity, to assess sensitivity quantitatively.[58] Based on this analysis, we determined that the linear ion trap tends to be more sensitive than the Orbitrap for this PRM assay.

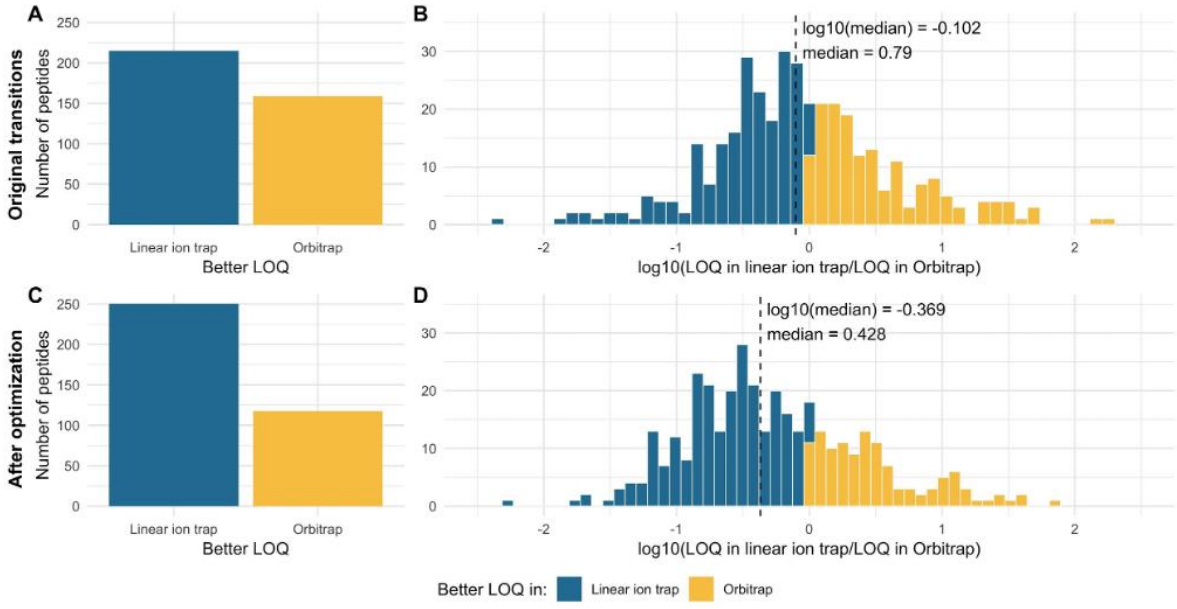

Figure 2-10: LOQ's before (A&B) and after (C&D) transition refinement

A and C show the raw number of peptides with better LOQs in each analyzer. B and D show the log ratio of LOQs, with a log ratio greater than 0 meaning the Orbitrap had a better LOQ than the linear ion trap. Dashed line indicates the median ratio. Before and after LLOQ optimization, 21 and 26 peptides respectively had limits of quantitation

outside the dilution curve in both analyzers and had identical log ratios (not shown). Optimization protocol is defined below.

Theoretically, the sensitivity of each analyzer is limited by the number of ions it requires to produce a quantitative signal. In a targeted experiment on a trapping instrument, the number of ions in each trap is determined by the ion current, which is inherent to the sample and injection setup, and the maximum inject time, a parameter set based on the number of concurrent targets and the desired cycle time. Highly abundant peptides may fill the trap with the desired number of ions in less time, while less abundant peptides may take much longer to fill the trap. In cases where the maximum inject time is too short to fill the trap with the minimum number of ions to obtain sufficient signal, the variation in the measured peptide abundance may be too large to be considered "quantitative". Thus, one crucial difference between the linear ion trap and Orbitrap is the number of ions needed to produce a detectable signal. The linear ion trap uses electron multiplier based detection, with gain sufficient to detect single ions.[52] In contrast, Makarov and Denisov demonstrated Orbitrap detection limits of 20+ charged single myoglobin ions with signal-to-noise 3.5 using 0.76 s transients.[59] With the 0.032 s transient, 15k resolution setting used in our experiments, the minimum detectable number of ions should be ~5x higher, because signal-to-noise ratio varies as the square root of transient length. Therefore, with these settings the Orbitrap needs 1–2 orders of magnitude more ions to create a detectable signal than the linear ion trap. This greater theoretical sensitivity explains why the linear ion trap performs better in this experiment despite the increased potential for interference.

Automated transition refinement

The original transition set was selected to minimize interference in the chromatogram library generated from non-diluted plasma on an Orbitrap and further refinement is required to

optimize the transition set for this experiment. To this end, transitions were automatically selected to produce the lowest limit of quantitation (**Figure 2-11**). For both analyzers, the refined transitions yielded better LOQs for most peptides, and unsurprisingly this change was more pronounced in the linear ion trap than in the Orbitrap (**Figure 2-12**). Consequently, after this transition refinement, the linear ion trap had even better LOQs relative to the Orbitrap than before (**Figure 2-1-C**). The median ratio of LOQs in the linear ion trap compared to the Orbitrap went from 0.790 to 0.428, indicating that the LOQs in the linear ion trap had a larger net improvement, and the linear ion trap outperformed the Orbitrap by an even more significant margin (**Figure 2-10 D**).

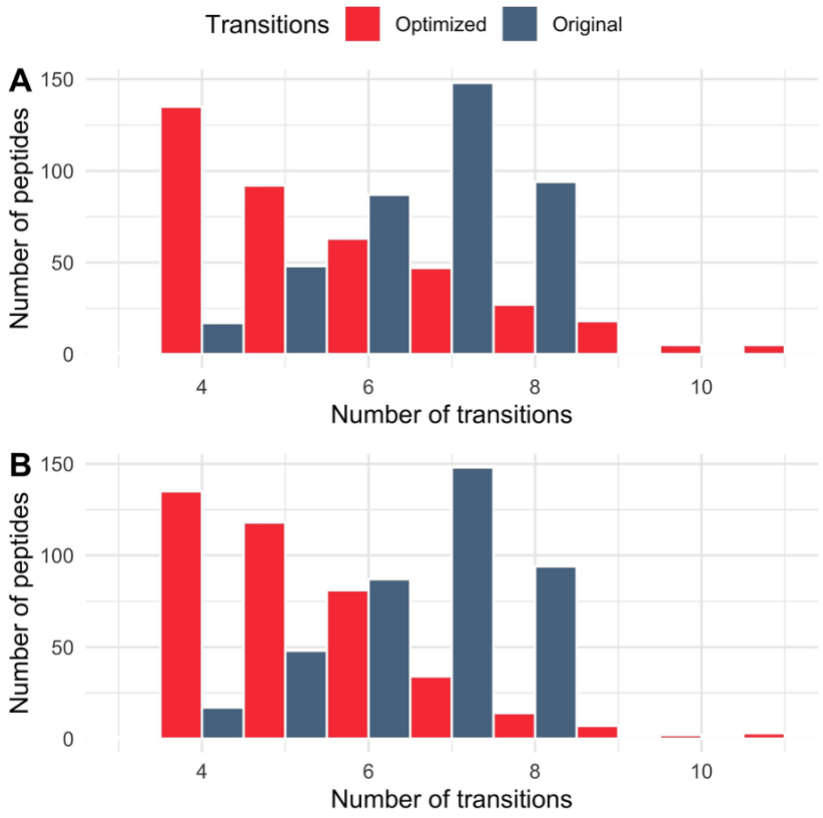

Figure 2-11: Number of transitions selected before and after refinement in (A) linear ion trap and (B) Orbitrap

Before refinement, there were generally more transitions in both analyzers, indicating that this method of manually refining transitions is able to remove interference that crops up at low dilution points. The mode of the optimized distribution is also the minimum number of possible transitions, which is probably due to the fact that many transitions have interference when the signal is low.

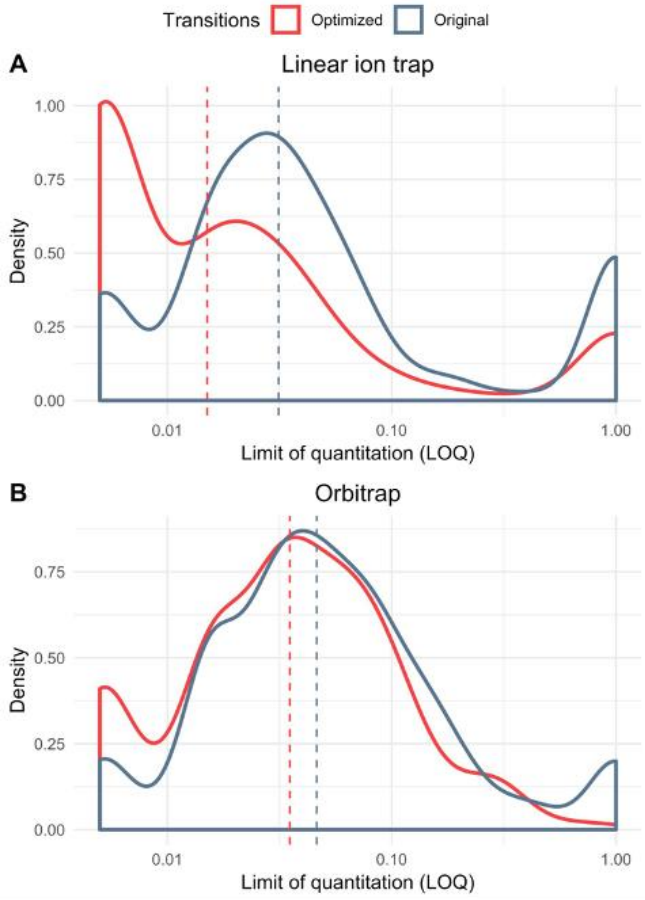

Figure 2-12: Probability density function (PDF) of limit of quantitation before and after transition refinement in (A) linear ion trap and (B) Orbitrap

The x-axis corresponds to the lower limit of quantitation in log scale with the median of each transition set indicated by the appropriately colored dashed line. A lower x-value indicates a lower LOQ, meaning better sensitivity. Both analyzers have better LLOQs after transition selection, but the linear ion trap improves more as a result of optimization.

The observation that transition refinement had a larger impact on the linear ion trap than the Orbitrap may be a result of the assay construction. Because this assay was based on a chromatogram library constructed using interference-free transitions determined using a different sample measured on an Orbitrap, the transitions were never optimized for these specific samples and instrument configuration. The library was generated with an undiluted human plasma digest

and in the case of a similar undiluted sample such as the one used to determine precision; the chosen peptides had minimal interference regardless of the mass analyzer. However, as the volume of the matched matrix in the dilutions increased, interference increased because the transitions were not optimized for these lower dilution points. Additionally, the original transitions selected to be interference-free were selected based on Orbitrap data, and it is likely that many of these transitions may not be optimal in the linear ion trap. Based on the improved mass accuracy and resolution of the Orbitrap, we predicted interference at different precursor/product ion specificities and the original peak picking should have been better for the high resolution Orbitrap than the unit resolution linear ion trap (**Figure 2-9**). Furthermore, peptides that had a lower LOQ in the linear ion trap qualitatively lacked signal at low dilution points using the Orbitrap (**Figure 2-13 A**). Conversely, peptides with a lower LOQ in the Orbitrap subjectively had interference at lower dilution points in the linear ion trap (**Figure 2-13 B**). These subjective observations are supported by the fact that the linear ion trap quantitative data improves significantly more than the Orbitrap after transition refinement (**Figure 2-12**).

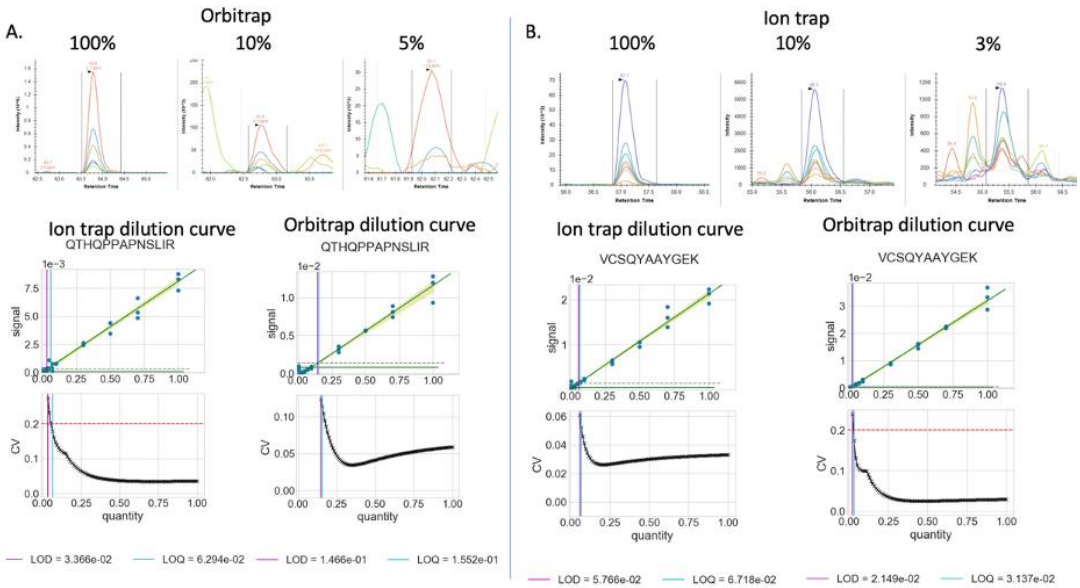

Figure 2-13: Examples of peptides with better LOQ in linear ion trap (A) and Orbitrap (B)

Top panels in each figure show what happens to signal in the analyzer with the worse LOQ. In the case of Orbitrap (A), signal goes to 0 at low points. On the other hand, interference becomes more apparent at lower dilutions when the linear ion trap performs worse (B). The bottom panel shows dilution curves in both analyzers for each peptide.

The ability to refine transitions based on empirical results is a major benefit of using an instrument that acquires full MS2 spectra over a triple quadrupole where transitions must be selected prior to analysis. It is possible to reduce interference post-acquisition simply by selecting a transition set to minimize interference. In both mass analyzers, almost all peptides are predicted to have at least four interference-free transitions (**Figure 2-8**), although the correct transitions may not have been selected in the original assay. When selecting transitions, it is preferable to select as many interference-free transitions as possible to increase the number of ions measured thereby increasing precision.[31,54,55] Thus, the transition refinement method presented here selects at least four transitions because at this threshold, there is very little

difference between analyzers in theoretical interference (**Figure 2-8**). This automated transition refinement heuristic is a simple method to leverage the full power of PRM to accurately quantify peptides.

2.5 Conclusions

Here, we show that the power of PRM lies in its ease of use, sensitivity, and flexibility in selecting transitions post-acquisition—rather than high mass accuracy. We report that unit resolution PRM performs at least as well as high resolution PRM in this targeted plasma assay by demonstrating that of the total set of 432 assayed peptides, over 97% could be quantified precisely (<20% coefficient of variation) with both analyzers, and the linear ion trap produced better lower limits of quantitation compared to the Orbitrap for 62% of all peptides. Here, we operate the Orbitrap at a relatively modest resolving power of 15,000 because increasing the resolution would increase cycle time without an appreciable reduction of interference. While it is possible that increasing the resolving power would improve the signal-to-noise ratio, doing so would come at a cost to the number of targets that can be accurately measured in this assay. In situations where the goal is to measure a small number of peptides with a high degree of precision, increasing the resolving power and maximum injection times in the Orbitrap may still be productive. Nevertheless, we have demonstrated that high resolution spectra and mass accuracy are not a requisite for selectivity, especially when peptide transitions can be refined post-acquisition. While not addressed here, in the rare event where unit resolution tandem mass spectrometry might provide insufficient selectivity, the linear ion trap has the capability to perform multistage MSn, whereas the Orbitrap is limited to MS2 without the LIT. Additionally, it is possible to increase resolving power of the linear ion trap but this comes at a cost to overall speed and still does not get close to Orbitrap-level resolving power. Overall, the superior speed

and sensitivity of the linear ion trap on the Orbitrap Eclipse make it better suited for targeted proteomics than the use of an orbitrap mass analyzer.

3 Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data Independent Acquisition

This chapter is based on the following preprint: Lilian R. Hei[1], Eugen Damoc, Tabiwang N. Arrey, Anna Pashkova, Eduard Denisov, Johannes Petzoldt, Amelia Peterson, Chris Hsu, Brian C. Searle, Nicholas Shulman, Michael Riffle, Brian Connolly, Brendan X. MacLean, Philip M. Remes, Michael Senko, Hamish Stewart, Christian Hock, Alexander Makarov, Daniel Hermanson, Vlad Zabrouskov, Christine C. Wu, Michael J. MacCoss. Evaluating the performance of the Astral mass analyzer for quantitative proteomics using data independent acquisition.

We evaluate the quantitative performance of the newly released Asymmetric Track Lossless (Astral) analyzer. Using data independent acquisition, the Thermo Scientific™ Orbitrap™ Astral™ mass spectrometer quantifies 5 times more peptides per unit time than state-of-the-art Thermo Scientific™ Orbitrap™ mass spectrometers, which have long been the gold standard for high resolution quantitative proteomics. Our results demonstrate that the Orbitrap Astral mass spectrometer can produce high quality quantitative measurements across a wide dynamic range. We also use a newly developed extra-cellular vesicle enrichment protocol to reach new depths of coverage in the plasma proteome, quantifying over 5,000 plasma proteins in a 60-minute gradient with the Orbitrap Astral mass spectrometer.

3.2 Introduction

Quantitative proteomics has advanced significantly over the past decade, becoming a critical tool for biological and biomedical research. However, robust proteome-wide measurements remain challenging due to the complexity of the proteome. Notably, the dynamic range of the proteome is vast, with the cellular proteome spanning 7 orders of magnitude[6] and some biofluids, such as plasma, spanning 10 orders of magnitude.[7] Additionally, ultra-high sensitivity is more critical than ever as the field expands to include single cell measurements.[60–62] Despite advances in technology, there is still room for improvement in quantitative sensitivity, accuracy, and precision across the complete dynamic range. While a large subset of well-characterized proteins are studied frequently, relatively little is known about many lower abundance proteins.[63] Thus, mass spectrometry (MS) has emerged as the dominant method for quantitative proteome-wide measurements,[8] however, further advancing the field will require continued improvement in hardware and data acquisition techniques.

Mass spectrometry offers the promise of comprehensive proteome-wide studies, but quantifying proteins accurately across the full dynamic range remains an unresolved challenge. Recent studies have reported the detection of large numbers of proteins: up to 7700 proteins in a 44-minute run[64], 10,000 proteins in a single 120-minute run[65], and over 12,000 proteins using biochemical fractionation.[66] These results highlight the increasing difficulty in measuring low abundance proteins. Additionally, these studies each implement their own thresholds for false discovery rates and quantifiable peaks, making these results difficult to compare. Even a standard false discovery rate of 1% can have widely different meanings across different tools, instruments, and data sets. Measurements that do not produce any quantitative information inflate numbers while diluting statistical power. Yet, many proteomics studies take quantification

for granted,[57] with very few experiments to benchmark quantification over a significant dynamic range. Therefore, rigorous analytical evaluation of novel technologies, with a primary focus on their quantitative capabilities, is a critical part of the technology development cycle.

The Astral is a novel high resolution accurate mass (HRAM) analyzer that shares some of the operational principles with established mass analyzers, including the Orbitrap, ion trap and time-of-flight.[67] A nearly lossless ion transfer is its most unique characteristic, derived from aligned ion injection and precise modulation of ion motion in three dimensions on a long asymmetric track and resulting in high sensitivity and consequently high analytical acquisition rate.[67] Additionally, due to the configuration of the components (**Figure 1A**), automatic gain control (AGC)[68] may be used to increase sensitivity and inter-spectrum dynamic range without the need for spectrum averaging as is the case with conventional orthogonal acceleration TOFs. The theoretical benefits of such an analyzer are evident, but the practical impacts still need to be assessed. To this end, we evaluated the quantitative performance of the Orbitrap Astral MS for data independent acquisition (DIA) and benchmark it relative to widely used Orbitrap based instruments.

We use a matrix-matched calibration curve[56,57] to compare limits of quantification in both analyzers. This benchmark demonstrates the quantitative gains associated with the Orbitrap Astral mass spectrometer relative to existing state-of-the-art technology, with 22% more quantifiable peptides in a run that is about one-quarter of the length of that used in the Orbitrap. We further demonstrate quantitative gains associated with using dynamic DIA[69] (dDIA) to maximize injection times across the mass range. To highlight the significance of this improved dynamic range, we analyze the plasma proteome. Using a membrane bound particle enrichment method,[70] we quantified 5163 proteins in a single 70-minute LC-MS run with the Orbitrap Astral

mass spectrometer. The benchmarking described here demonstrates the power of the Astral mass analyzer for robust quantitative data independent acquisition.

3.3 Methods

Complete methods are discussed in Appendix B

3.4 Results and Discussion

Peptide detections in a whole cell lysate

To evaluate the performance of the Astral analyzer, we performed a number of comparisons with the Orbitrap. First, we used gas phase fractionation[50,71,72] to generate in-depth chromatogram libraries to compare the total peptide and protein level coverage with each analyzer in the same 24-minute gradient. We use these libraries for data analysis but they are also a way to compare deep proteome coverage when acquisition time is not a limiting factor. The philosophy of our workflow is that the chromatogram library should detect a superset of all observable peptides, while subsequent acquisitions may collect lower quality data, albeit at a faster rate. Here, we created a library from six different injections of the same HeLa digest with each analyzer, each covering a 100 Th mass range with 4 Th isolation windows and a 23 ms maximum injection time. Although the Astral has the ability to acquire spectra at rates up to 200 Hz, for this comparison we reduced the rate significantly and allowed a relatively long maximum injection time so that we could directly compare coverage with the Orbitrap. Even controlling for injection times between analyzers, we can expect to see differences because the Orbitrap fills with significantly more charges than the Astral (**Figure 3-1 A**). While both mass analyzers tended to reach their maximum injection time in every spectrum (**Figure 3-1 B**), the high

sensitivity of the Astral means it has more data points per spectrum (**Figure 3-1 C**) and decreased overhead enabled by additional parallelization (i.e. processing 5 distinct ion packets in parallel) allows the Astral to acquire spectra at a faster rate even at the same injection time (**Figure 3-1 D**). As a result of these differences, there are 70% more detectable peptides in the Astral chromatogram library than the Orbitrap-based chromatogram library (87,257 and 51,603 respectively). Although it should be noted that these Orbitrap parameters may have been suboptimal and longer injection times could have resulted in deeper coverage.

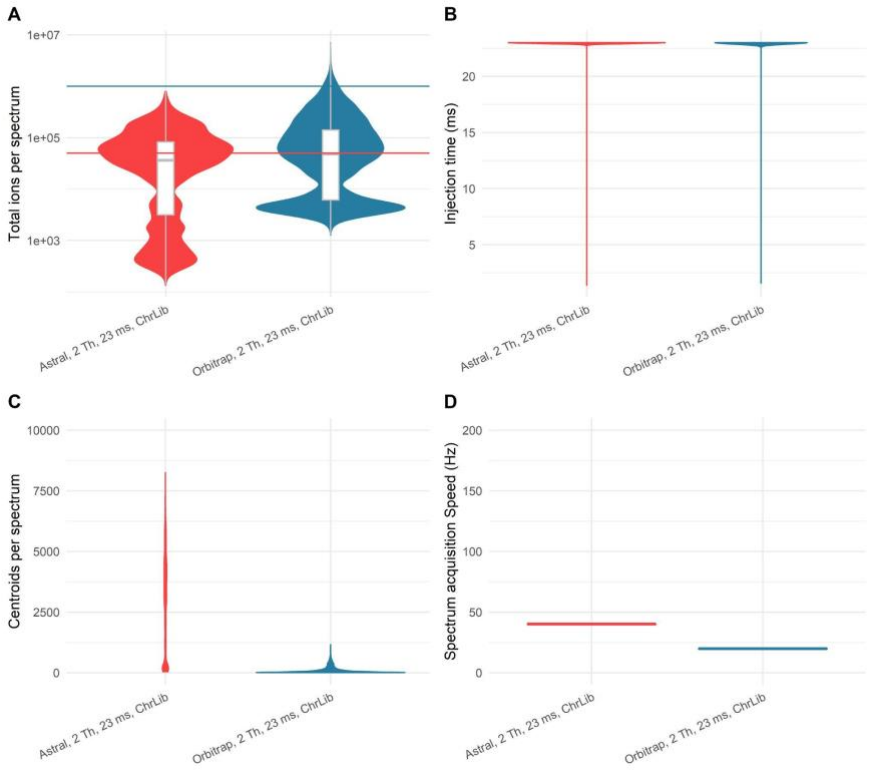

Figure 3-1: Ion statistics for HeLa chromatogram library, 24 minute gradient

Comparison of ions per spectrum (A), injection time (B), centroids per spectrum (C), and overall acquisition speed for both Orbitrap and Astral in a 24-minute chromatogram library run. The Orbitrap detector is less sensitive than the electron multiplier in the Astral and therefore the Orbitrap requires more ions to reach optimal performance.

In contrast to gas-phase fractionation approaches to build a library, the majority of DIA experiments performed today are single runs for quantitative analysis. To evaluate the ability of the Astral to detect peptides across a wide dynamic range, we acquired DIA data for a whole cell HeLa lysate in the Astral analyzer and compared the results to an existing dataset acquired by Heil, et al., on the Orbitrap analyzer in an Orbitrap Fusion Lumos mass spectrometer.[69] There are several differences between these datasets, most notably a 24-minute gradient at 750 nL/minute was used for the Astral data while the Orbitrap data was collected with a 90-minute gradient at 300 nL/minute. The longer gradient better accommodates the Orbitrap acquisition rate and should give it a significant advantage in this head-to-head comparison. Narrow window chromatogram libraries[50] were generated for each method and used for analysis of the quantitative data, using various acquisition schemes in the Astral to compare with the Orbitrap (**Figure 3-2 B**). Although sample specific libraries may not be necessary for the Astral when narrow (2-4 Th) isolation windows can be used across the full mass range, we use the narrow window library here for consistency and to increase search speed.

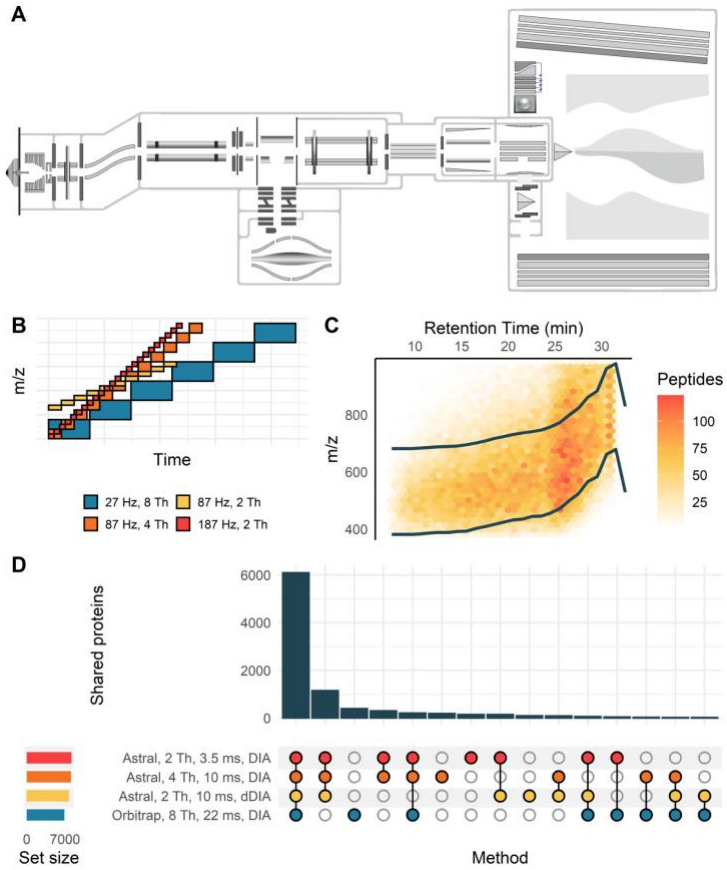

Figure 3-2: Quantitative comparison of the Orbitrap Astral MS and Orbitrap Fusion Lumos performance on a bulk cell digest

Schematic of the Orbitrap Astral mass spectrometer (A) and of isolation windows used for each method, with a limited mass range displayed (B). A dynamic DIA method was created based on peptide feature density with systematic isolation windows moving across the mass range between the two dark colored lines (C). The overlap of protein identifications between each method shows a high level of agreement between different Astral methods, with a subset of these proteins being detectable in the Orbitrap (D). Data were acquired on the Orbitrap Astral MS with a 24-minute gradient compared to a 90-minute gradient on the Orbitrap Fusion Lumos MS.

When creating a quantitative DIA method, it is important to strike a balance between cycle time, spectrum acquisition rate, and isolation window width that fits the specific sample type, experimental goals, and LC parameters. Therefore, we evaluated three different Astral

methods against a standard Orbitrap method. The first Astral method was acquired at 187 Hz using 3.5 ms maximum injection times and 2 Th isolation windows. The second Astral method used longer injection times (10 ms) and wider isolation windows (4 Th) in an effort to increase sensitivity and dynamic range. The third method used a dynamic DIA scheme[69] which adjusts the isolation window range across time (**Figure 3-2C),** allowing for longer injection times (10 ms) with narrow isolation windows (2 Th). Generally, longer injection times allowed the Astral to accumulate more ions (**Figure 3-3 A**) while slowing the overall acquisition rate significantly (**Figure 3-3 B&D**). Additionally, narrower isolation windows reduced the spectral complexity, although spectra generated by the Astral are significantly more complex than those of the Orbitrap (**Figure 3-3 C**). In this 24-minute gradient, all three Astral methods identified 7700-8200 proteins (**Figure 3-2 D**) and 58,000 - 75,000 peptides (**Figure 3-4**). The Orbitrap detected 6885 proteins and 55,981 peptides in a 90-minute gradient, meaning that the Astral analyzer is able to detect more peptides and proteins in a fraction of the time that was needed for Orbitrap analyzer (~1/4).

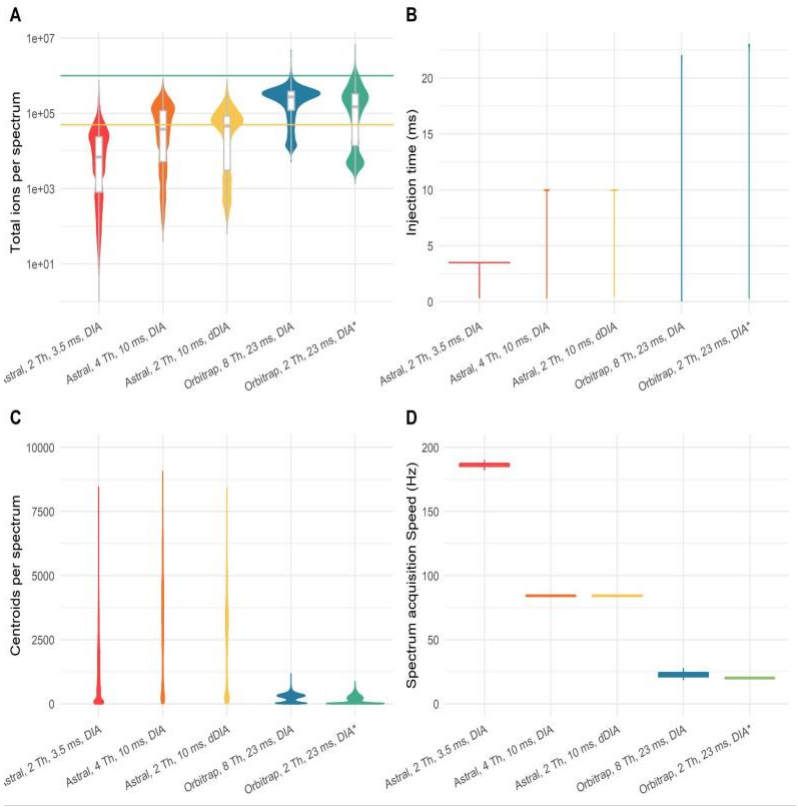

Figure 3-3: Ion statistics for quantitative DIA methods

Comparison of ions per spectrum (A), injection time (B), centroids per spectrum (C), and overall acquisition speed for both 3 Astral methods acquired with a 24-minute gradient, one Orbitrap method acquired with a wide mass range and a 90-minute gradient, and one Orbitrap method acquired with a 75 Th mass range and the same, 24-minute gradient used in the Astral methods.

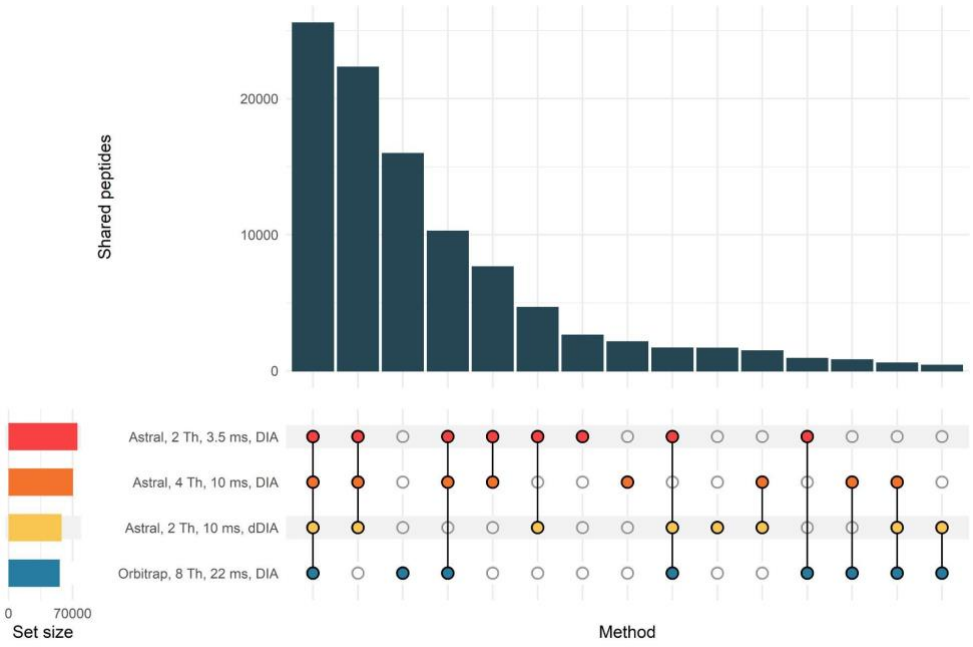

Figure 3-4: UpSet plot of peptide detections across methods

Comparison of peptide-level detections in 1 µg of HeLa digest in the Astral with a 24-minute gradient and the Orbitrap with a 90-minute gradient. Set size indicates total number of peptide detections in each dataset.

To test performance of the Orbitrap while controlling the isolation window and using the same LC setup, we acquired a set of runs in the Orbitrap covering a 75 Th mass range with 2 Th isolation windows. As we observed with the 90-minute gradient, spectra generated in the Orbitrap typically consist of more ion signal distributed across fewer fragment ion peaks, pointing to the increased sensitivity of the Astral analyzer (**Figure 3-3**). In the same mass range, which was a subset of the total mass range covered, the Astral analyzer detected 43% more peptides (**Figure 3-5**). This observation suggests that part, but not all, of the increase in peptide detections seen in the Astral analyzer can be attributed to its high speed and sensitivity allowing it to cover a wider mass range with relatively narrow isolation windows.

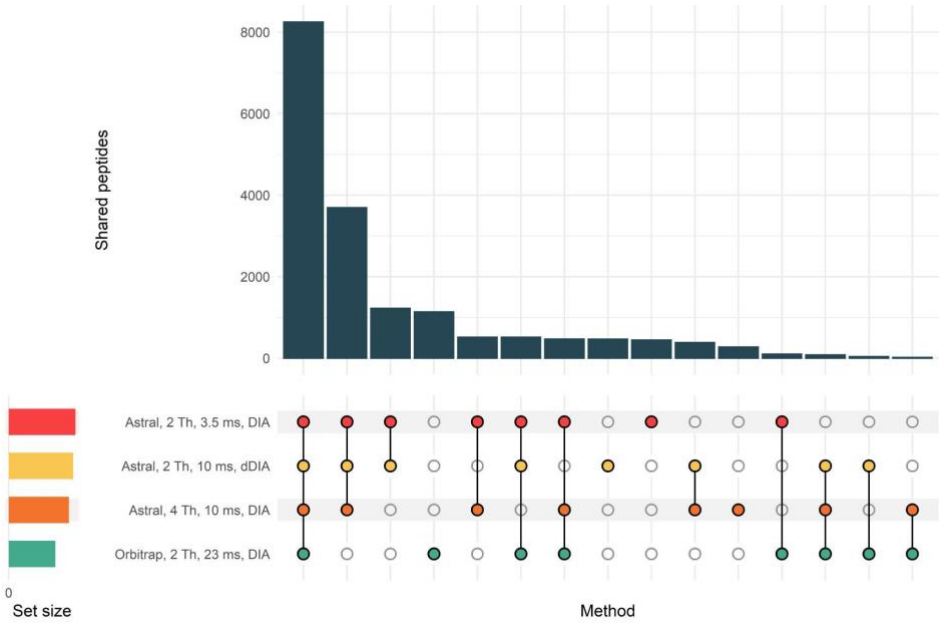

Figure 3-5: UpSet plot of peptide detections across a limited mass range, same LC method

Comparison of peptide-level detections in 1 µg of HeLa digest in the Orbitrap and the Astral with a 24-minute gradient. The Orbitrap method acquired DIA spectra across a limited (75 Th) mass range, and the Astral data was originally acquired across a larger mass range but was filtered down for the analysis to peptides in the same 75 Th mass range.

Quantitative benchmarking of DIA performance

While peptide and protein level identifications can be useful metrics to evaluate instrument performance, these numbers may not fully reflect the quality of the data. Therefore, we evaluated the quantitative precision and accuracy of the Orbitrap Astral MS using a 24-minute gradient relative to the Orbitrap Fusion Lumos using a 90-minute gradient. First, we looked at the technical precision across injection replicates with different inputs (**Figure 3-6**). We found that the coefficient of variation for 1 µg, 500 ng, and 100 ng of HeLa digest tends to be similar if not slightly better for the Astral than the Orbitrap. When compared with the same gradient and reduced mass range in the Orbitrap, a similar trend can be observed, suggesting that

these results were not just due to differences in the chromatography or experimental setup (**Figure 3-7**). Additionally, we observed that using longer maximum injection times in the Astral analyzer improved the technical precision, with the narrow windows and longer injection times afforded by dynamic DIA providing an improvement in technical precision at all input levels.

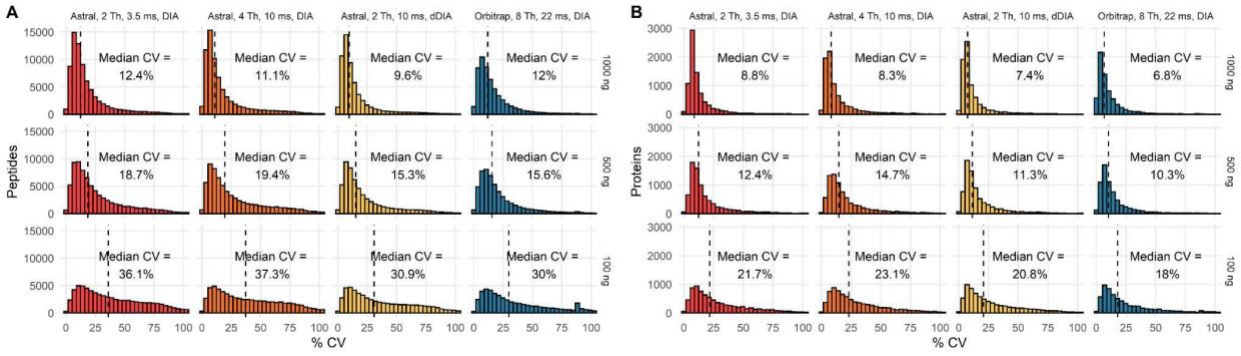

Figure 3-6: Evaluation of the technical precision of the data obtained with the Astral analyzer

Peptide (A) and protein (B) level coefficients of variation across 4 different acquisition methods. CVs are plotted for the same set of peptides with different amounts of HeLa injected with SILAC HeLa added as a background to keep total protein loading level constant (1000 ng). Dashed horizontal line represents the median CV, which is also indicated on the plot.

It is notable that the Astral analyzer is able to measure more peptides than the Orbitrap analyzer and still have similar if not better median CVs. Generally, the variance will increase as analyte abundance decreases. However, these results suggest that the Astral is able to generate very high-quality measurements for a large number of peptides in a short amount of time.

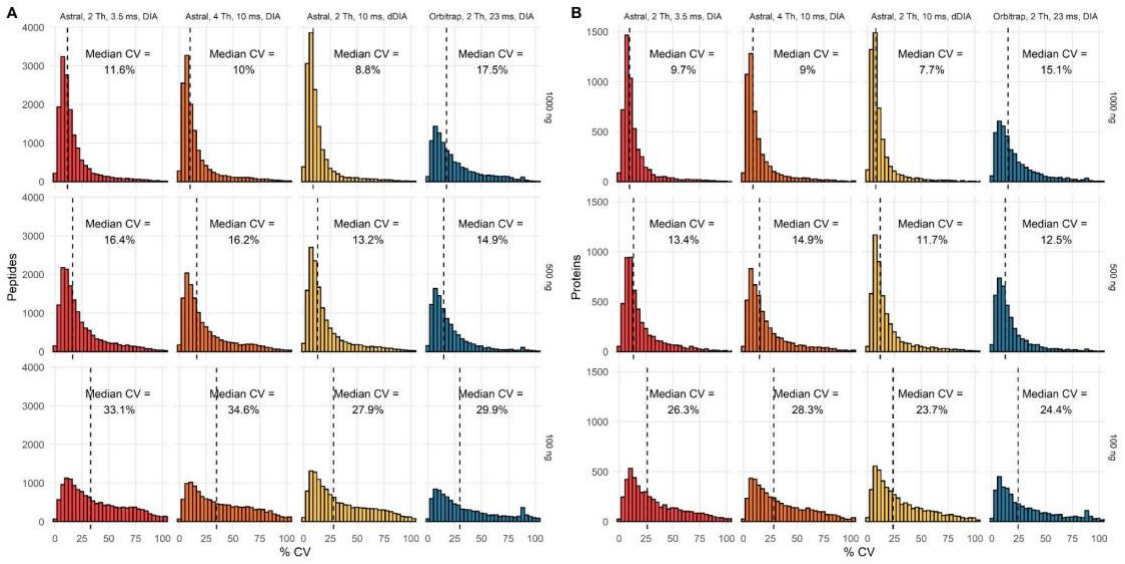

Figure 3-7: Peptide and protein CVs across limited mass range

Peptide (A) and protein (B) level coefficients of variation across 4 different acquisition methods. CVs are plotted for the same set of peptides with different amounts of HeLa injected with SILAC HeLa added as a background to keep total protein loading level constant (1000 ng). Dashed horizontal line represents the median CV, which is also indicated on the plot. The Orbitrap method acquired DIA spectra across a limited (75 Th) mass range, and the Astral data was originally acquired across a larger mass range but was filtered down for the analysis to peptides in the same 75 Th mass range.

Technical precision is an important metric for evaluating performance, but quantitative accuracy is the goal of most proteomics experiments and needs to be evaluated independently. Therefore, we use matrix-matched calibration curves to evaluate proteome-wide quantitative performance.[56,57] By diluting our analyte (HeLa cell digest) in a similar matrix (SILAC labeled HeLa cell digest), we calculate the lower limits of quantification for thousands of peptides at a time and assign a robust analytical figure of merit to sensitivity. We can categorize peptides as quantitative if they can be assigned a lower limit of quantification (LLOQ), meaning that their signal has a linear response to changes in input when the total protein input is less than 1 µg. This measure provides more information than variance-based measurements, which can be

misleading, as the variance of background signals can be quite low. This comparison is especially important here, where the mass analyzers have drastically different levels of background signal.

Using a 9-point matrix-matched calibration curve, we quantified up to 46,038 peptides in the Astral in a 24-minute gradient compared to 40,098 peptides in the Orbitrap in a 90-minute gradient (**Figure 3-8 A**). In addition to quantifying more peptides than the Orbitrap, the Astral tends to produce slightly better LLOQs than the Orbitrap for the peptides that both analyzers can quantify (**Figure 3-8 B**). We can further improve quantification by integrating only the set of fragment ions that yield the best lower limit of quantification per peptide.[73] This procedure improves the lower limit of quantification in all cases but is particularly beneficial for the Astral-based data, likely due to its higher spectral complexity (**Figure 3-8 A&C**). With the refined transition set, the Astral quantifies at most 58,330 peptides (17,454 across at least a 10x linear dynamic range) while the Orbitrap quantifies 47,671 peptides (10,686 across at least a 10x linear range). In all cases, the Astral quantifies more peptides and tends to have a better LLOQ (**Figure 3-8 C**). This quantification is achieved in significantly less time in the Astral than the Orbitrap. In fact, per unit time, the Astral quantifies 5 times as many peptides as the Orbitrap (**Figure 3-9**). Another way to look at the quantitative performance is to observe the ratio of signals between two dilution points. There, the sensitivity of the Astral is evident over a 10x dilution, where accurate quantification is observed for most peptides and proteins (**Figure 3-8 D, Figure 3-10**). To eliminate the possibility that experimental differences were responsible for the increase in sensitivity, we also confirmed that the Orbitrap underperformed relative to the Astral in the reduced mass range 24-minute gradients, with considerably less sensitivity by all metrics (**Figure 3-11**).

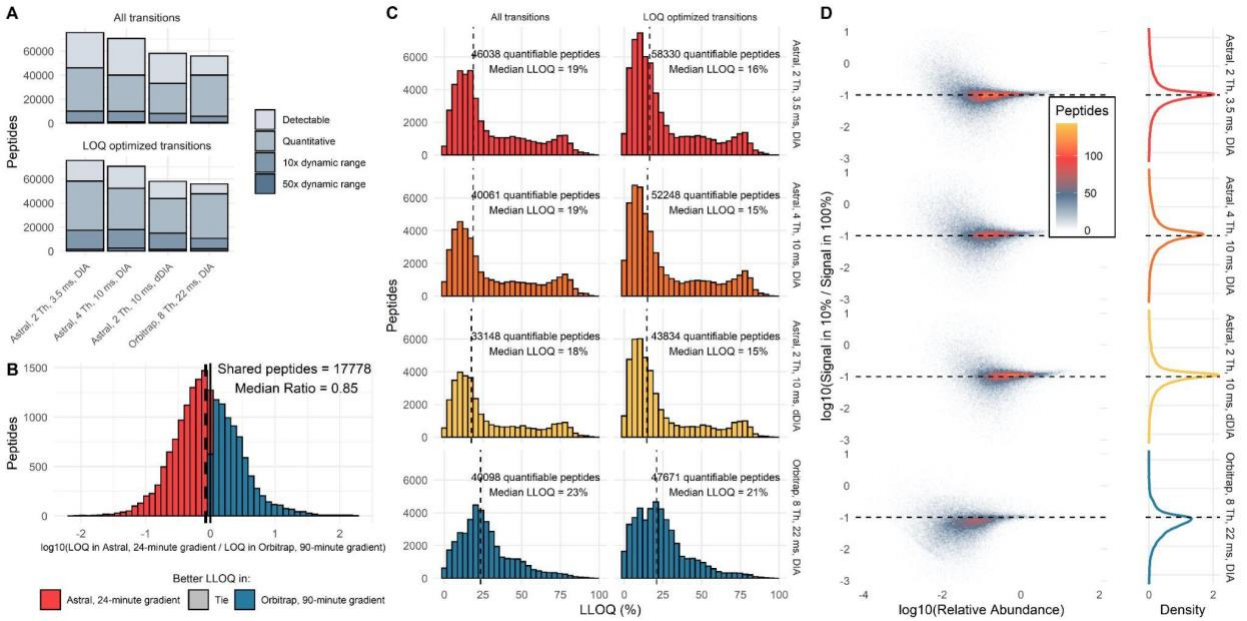

Figure 3-8: Evaluation of the quantitative performance of the Orbitrap Astral MS

Summary of peptides detected in a matrix-matched calibration curve of HeLa into SILAC labeled HeLa, with 1 µg of total protein load (A). Results of A are summarized in Supplemental Table 2. Peptides are considered detectable if they were detected at 1% FDR, quantitative if they could be assigned a lower limit of quantification less than 100%, quantitative over a 10x dynamic range if that LLOQ was less than 10%, and quantitative over a 50x dynamic range if the LLOQ was less than 2%. Transitions were refined in Skyline to optimize LLOQ and metrics were recalculated with this refined transition set. Pairwise comparison of LLOQs for peptides quantified in the Astral and the Orbitrap, black dashed line represents the median LLOQ (B). Histogram of LLOQs for all quantifiable peptides before and after transition refinement, black dashed lines represent the median LLOQ (C). Signal ratio between 2 points on the dilution curve (100% and 10%) with expected ratio shown as dashed black line, signal from the blank was subtracted from each signal to account for carryover and non-zero background (D). Orbitrap data were acquired with a 90-minute gradient, compared to a 24-minute gradient for the Astral data.

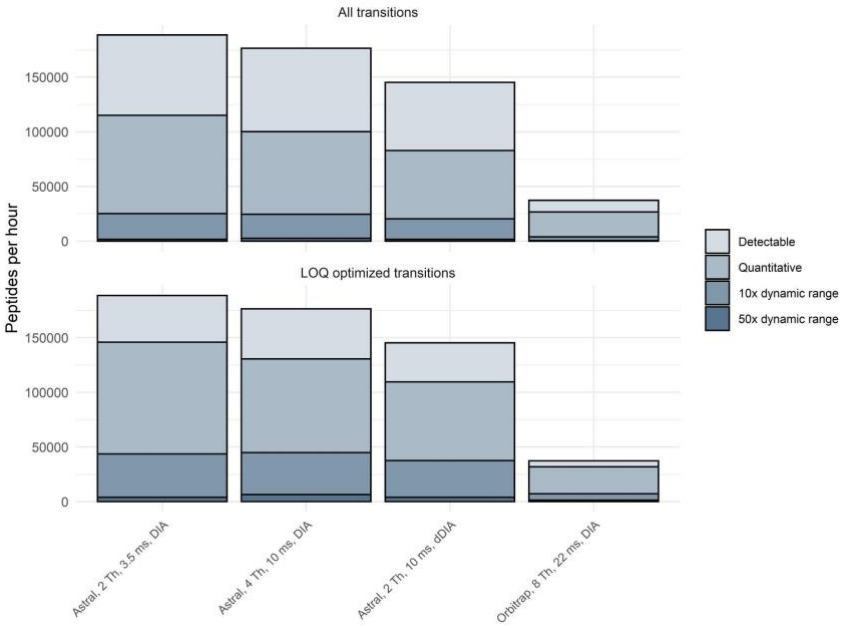

Figure 3-9: Peptides quantified per unit time

Plot of peptides quantified per hour. Same data shown in Figure 3-8 A, scaled to gradient length.

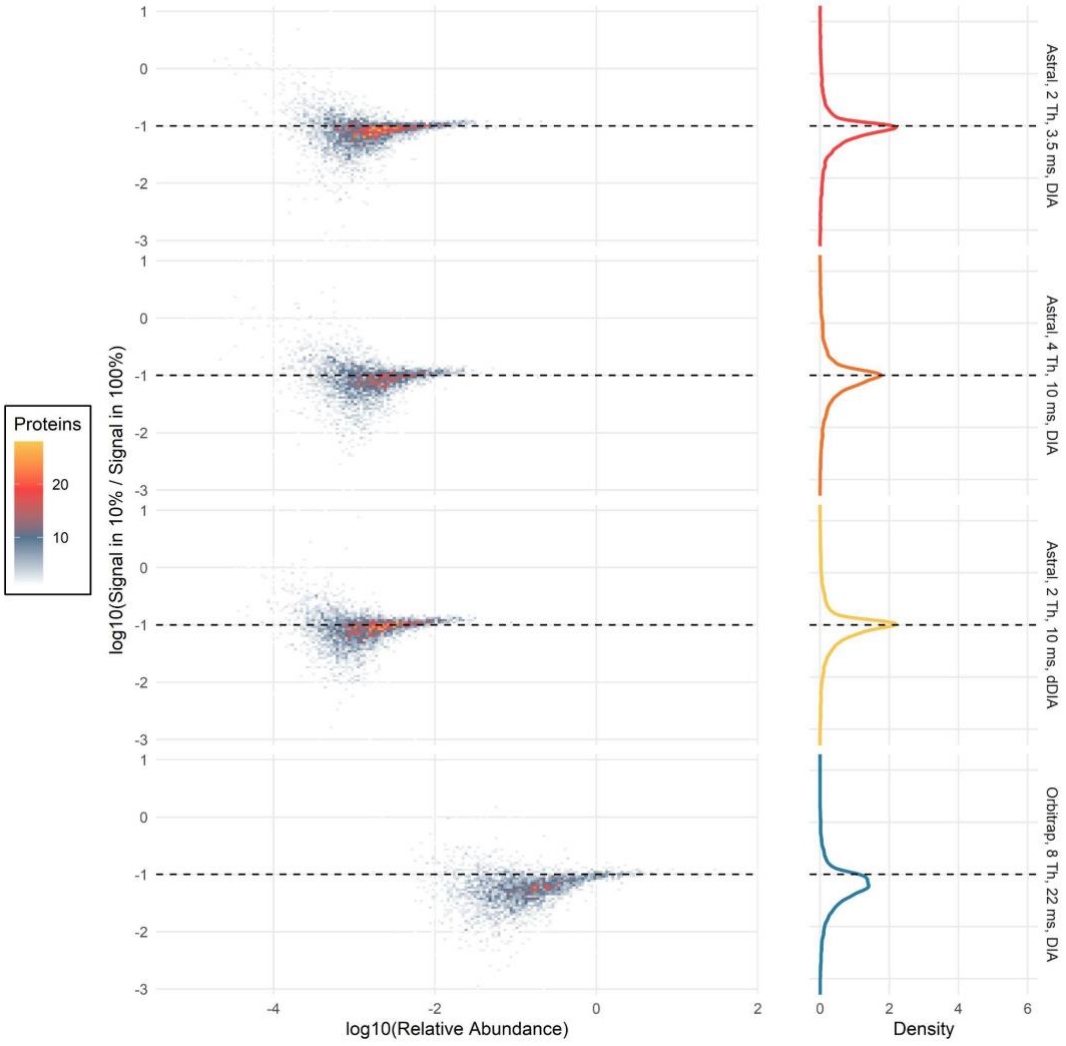

Figure 3-10: Quantification across a 10x dilution using standard protein grouping

Density plot of quantitative ratios across a 10x dilution using protein grouping. Astral data were acquired with a 24-minute gradient, compared to Orbitrap which used a 90-minute gradient.

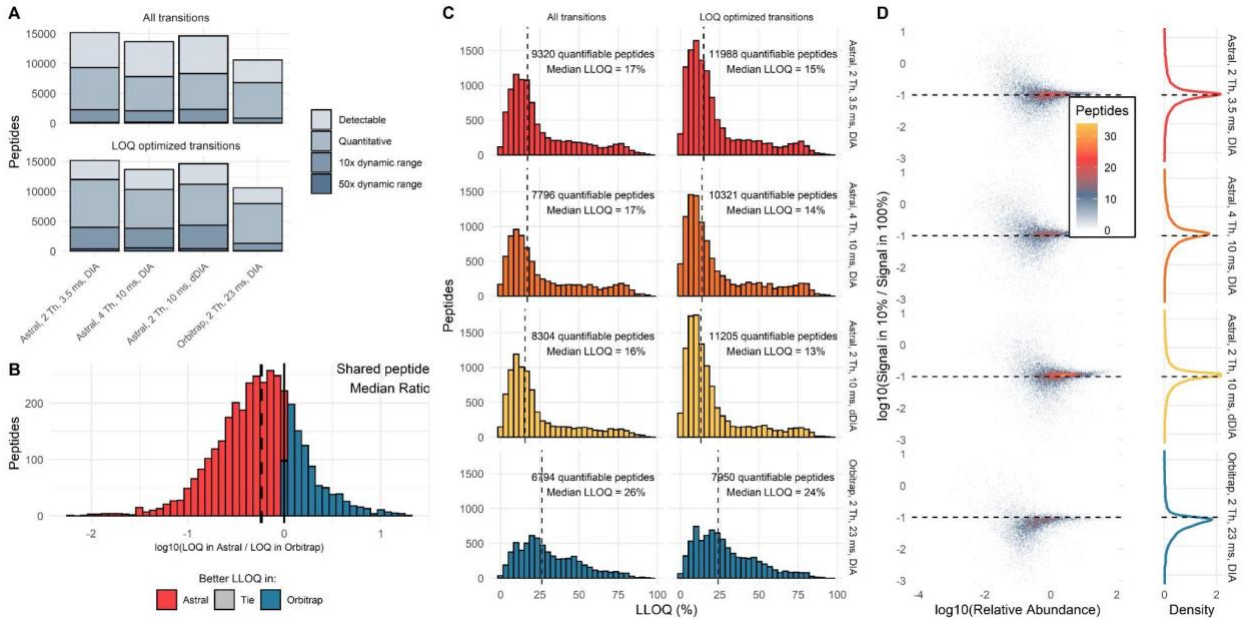

Figure 3-11: Quantification across a shorter mass range

Summary of peptides detected in a matrix-matched calibration curve of HeLa into SILAC labeled HeLa, with 1 µg of total protein load (A). Peptides are considered detectable if they were detected at 1% FDR, quantitative if they could be assigned a lower limit of quantification less than 100%, quantitative over a 10x dynamic range if that LLOQ was less than 10%, and quantitative over a 50x dynamic range if the LLOQ was less than 2%. Transitions were refined in Skyline to optimize LLOQ and metrics were recalculated with this refined transition set. Pairwise comparison of LLOQs for peptides quantified in the Astral and the Orbitrap, black dashed line represents the median LLOQ (B). Histogram of LLOQs for all quantifiable peptides before and after transition refinement, gray dashed lines represent the median LLOQ (C). Signal ratio between 2 points on the dilution curve (100% and 10%) with expected ratio shown as dashed black line (D). The Orbitrap method acquired DIA spectra across a limited (75 Th) mass range, and the Astral data was originally acquired across a larger mass range but was filtered down for the analysis to peptides in the same 75 Th mass range.

While quantification generally is better in the Astral, the Orbitrap generally performed better across a 100:1 dilution range (**Figure 3-12**). This trend is consistent with the 24-minute gradient and shorter mass range, where the Orbitrap quantification is slightly better, although

fewer peptides are quantified (**Figure 3-12**). Similarly, the Astral quantifies significantly more peptides (17,454 in Astral with 3.5 ms maximum injection time, 10,686 in Orbitrap) across a 10x linear dynamic range compared to the 90-minute gradient in the Orbitrap, but fewer peptides across a 50x dynamic range (1598 in Astral with 3.5 ms injection time, 1948 in Orbitrap). This relative improvement of Orbitrap performance at lower dilution points may be due to the longer injection times allowing for deeper coverage as well as the higher intra-spectrum dynamic range of the Orbitrap analyzer which is critical for quantification in cases where the peptide of interest is a minor species in a chimeric spectrum. The injection time theory is supported by the fact that quantification in the Astral improved by increasing injection time (2563 peptides have 50x linear dynamic range with 10 ms maximum injection time). In both of these comparisons, the performance of the Orbitrap Astral MS was handicapped by either a shorter gradient or a wider mass range. Even the largest maximum injection time assessed (10 ms) was shorter than the Orbitrap (23 ms). Regardless, the Astral is comparable if not better to the Orbitrap in every method assessed here.

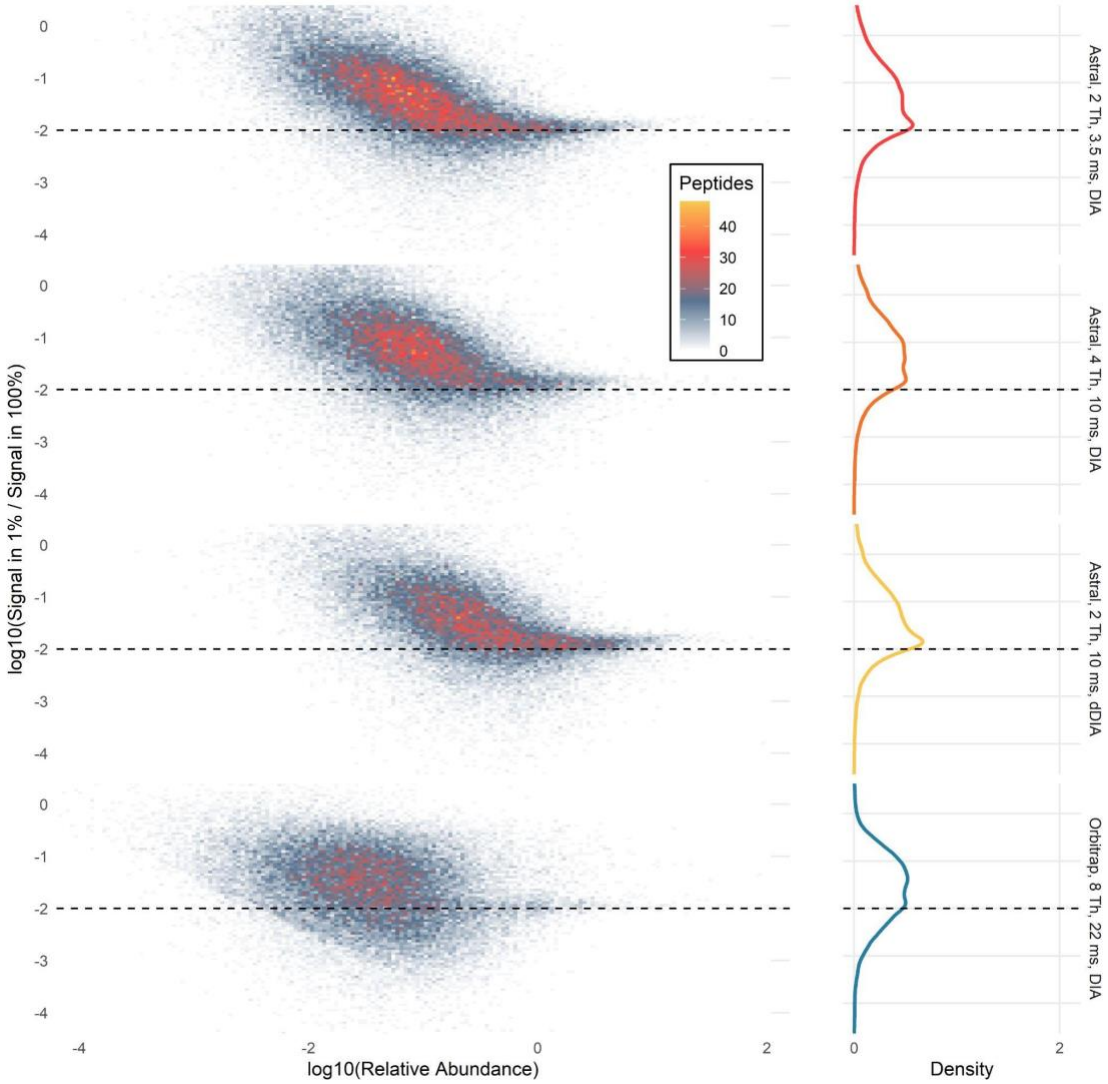

Figure 3-12: Quantification across 2 orders of magnitude

Density plot of peptide quantitative ratios across a 100x dilution. Astral data were acquired with a 24-minute gradient, compared to Orbitrap which used a 90-minute gradient.

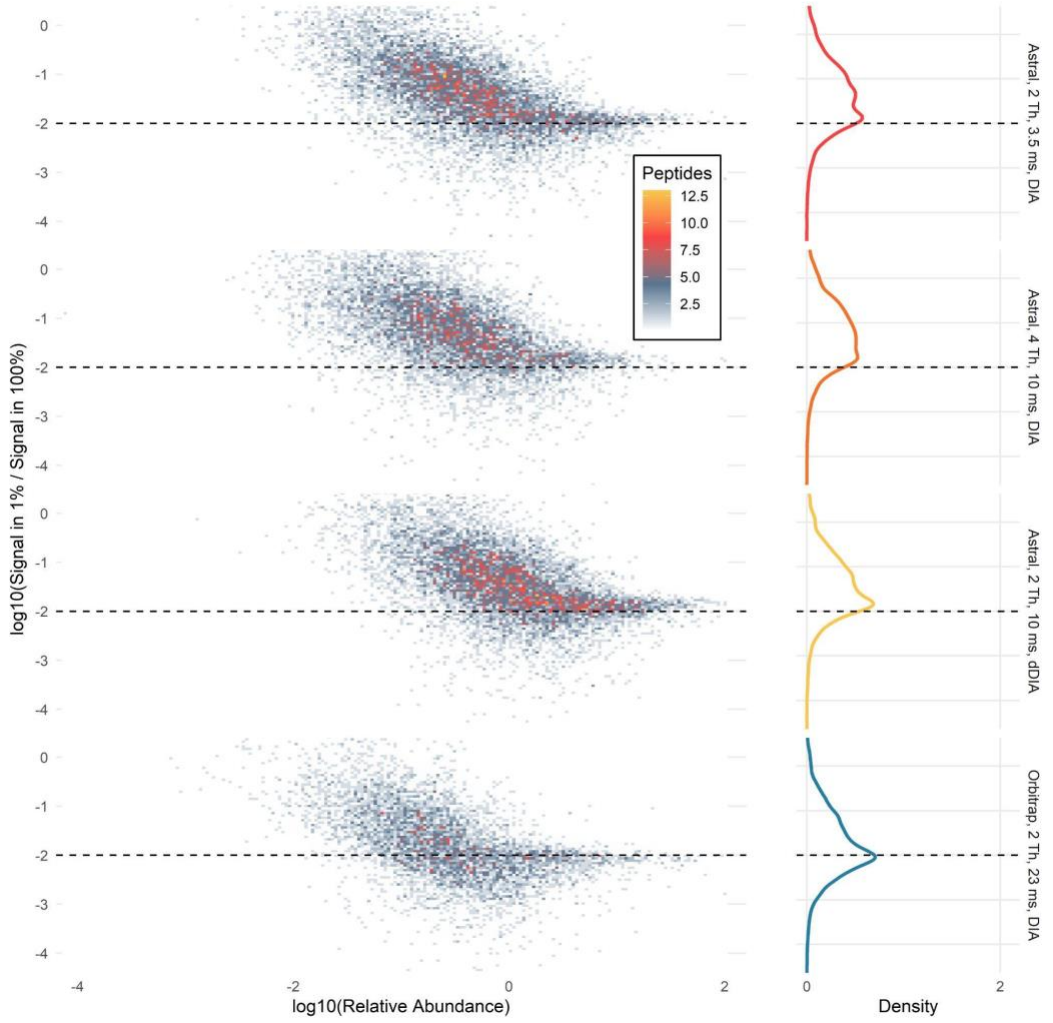

Figure 3-13: Quantification across 2 orders of magnitude, shortened mass range
Density plot of peptide quantitative ratios across a 100x dilution. Astral data were acquired with a 24-minute gradient, compared to Orbitrap which used a 90-minute gradient. The Orbitrap method acquired DIA spectra across a limited (75 Th) mass range, and the Astral data was originally acquired across a larger mass range but was filtered down for the analysis to peptides in the same 75 Th mass range.

Although the Orbitrap Astral MS has the ability to acquire data at rates up to 200 Hz, we predicted that it may be better to use a slower acquisition rate and to increase maximum injection times to increase sensitivity. We found that acquiring spectra at 187 Hz (3.5 ms maximum injection time) and 2 Th isolation windows results in a large number of peptide and protein

detections (**Figure 3-2 D**), but these detections tend to have higher CVs and LLOQs than measurements acquired with longer injection times (**Figures 3-6 &3-8**). However, increasing the maximum injection time to 10 ms required doubling the isolation window to maintain a reasonable cycle time, so we also used a dynamic DIA method which allowed for a 10 ms injection time and 2 Th isolation windows across a variable 300 Th mass range (**Figure 3-2 B&C**). Using dynamic DIA decreases the number of peptide and protein identifications, but results in better technical precision (**Figure 3-6**) and improved limits of quantification (**Figure 3-8 and Figure 3-14**). The longer injection times and narrower isolation windows allowed by dynamic DIA improve quantitative accuracy more as the input gets lower, suggesting that this type of method would be particularly useful in cases where high sensitivity is a priority (**Figure 3-12**).

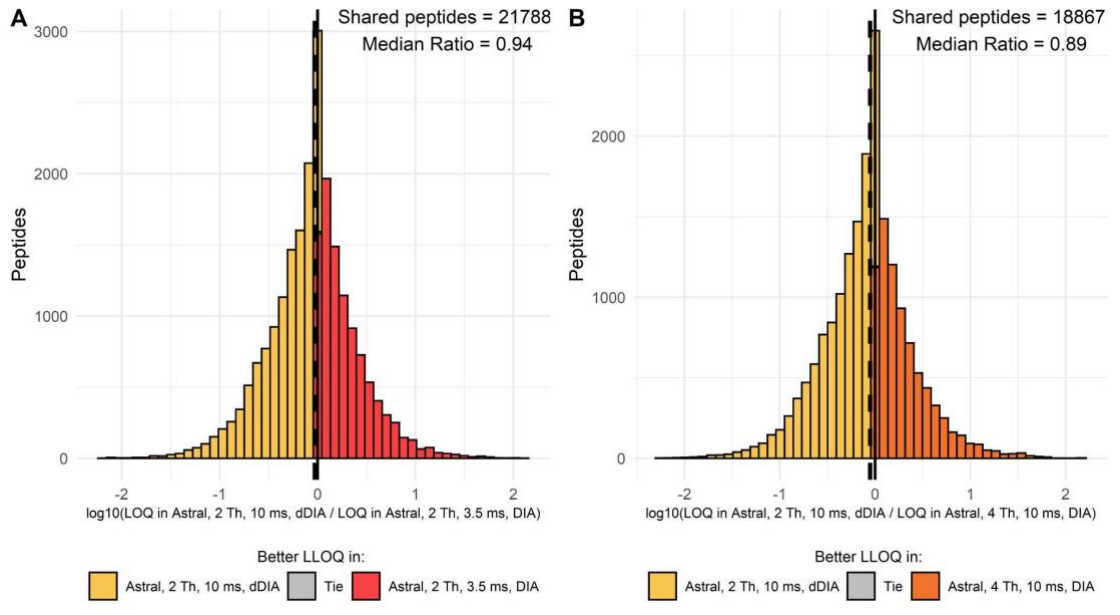

Figure 3-14: Effects of dynamic DIA on limit of quantification

Pairwise comparison of lower limits of quantification obtained by peptides quantified in multiple methods by the Astral. Dashed black line represents the median ratio. A log

ratio greater than 0 means that the dynamic DIA method was more sensitive than the static acquisition method.

We acquired DIA data in the Astral analyzer with a 24-minute gradient at 750 nL/minute as this allows for relatively high throughput. Because the Orbitrap acquires spectra at a much slower rate due to longer detection times and requisite injection times, the dataset we are comparing against was acquired with a 90-minute gradient at 300 nL/minute. There are other differences in the acquisition parameters for the two datasets, such as different columns and a different precursor mass range (380-980 Th in the Astral, 400-1000 Th in the Orbitrap). Therefore, we want to emphasize that this is not a direct comparison, but instead the longer gradient and slower flow rate should put the Orbitrap at a significant advantage. Still, the Astral outperforms the Orbitrap by nearly every quantitative metric while nearly quadrupling the throughput. The Orbitrap has long been the gold-standard in the field for high resolution accurate mass based quantification, and so the improvements in sensitivity and throughput in the Astral analyzer represent a step forward for the field as a whole.

One limitation of the data reported here is that it was acquired on a prototype instrument running a beta version of instrument control software. At the time the data were acquired, calibration for the detector gain was undergoing further optimization, and the AGC target of 500% (50,000 charges) may not have been optimal. On the Astral, the major impact of space charging is localized to a given peak. If any single peak contains more than 1,000 ions, the ions begin to repel each other and slow down slightly. As a result, the peak resolution decreases and the mass shifts to a slightly higher value (**Figure 3-15**). This shift can cause the calibration curve to plateau if it pushes the mass outside of the integration boundary. We believe that the effects of space charging were relatively minimal here, but could be corrected further by decreasing the

AGC target, although this may have a negative impact on LLOQ for some low abundance species that fall in the same isolation window as a more abundant species.

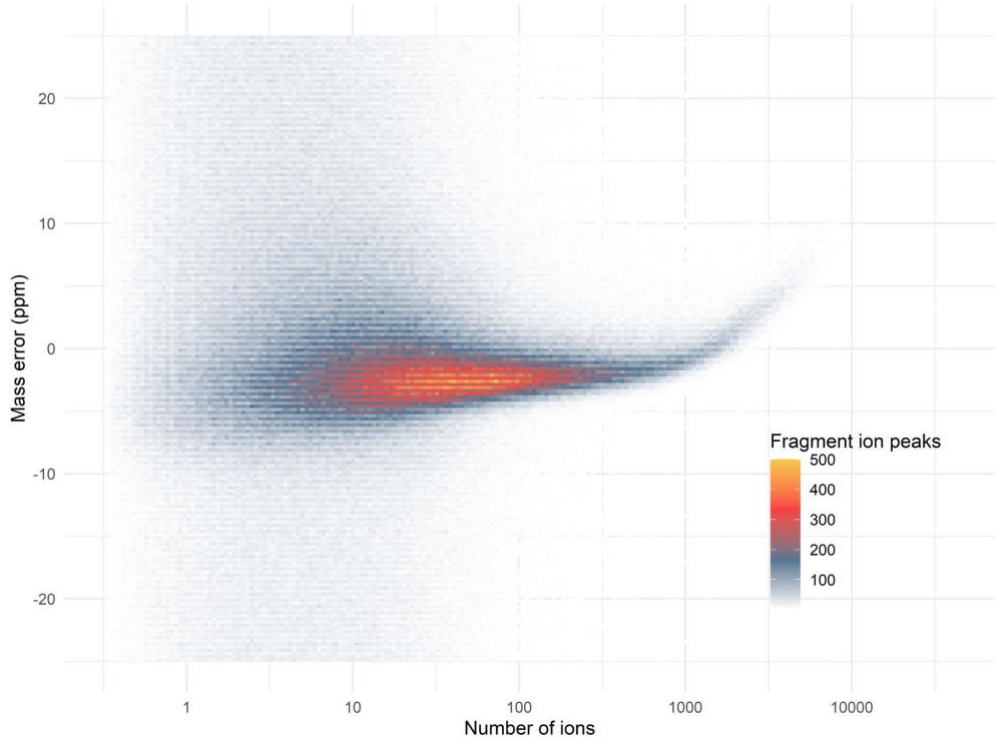

Figure 3-15: Mass error vs number of ions

2-dimensional density plot of fragment mass error vs number of ions. Mass errors are shown for every fragment ion in each point across the peak. A clear inflection point can be seen above 1000 ions which is consistent with localized space charging.

Software tools are used to assess whether there is evidence for a specific peptide in a dataset. The false discovery rate is often assessed using either an analytical or empirical null. These tools have gotten very sensitive, often making use of artificial intelligence for the prediction of retention time and fragment ion intensities. However, some of the peptides that tools assign as being present at a given retention time don't meet the traditional definition of detectability. By definition a signal is detectable if the measured signal, in this case a background subtracted peak area, is statistically different from a blank. Thus, to avoid making multiple

measurements, many use a heuristic requiring the signal to be >3x the standard deviation of the background outside of the peak to be detectable. Using this heuristic is further complicated by the fact that different instruments make use of different ion detection and signal processing strategies and have vastly different background signal making the assessment of whether a signal detectable near impossible to compare between platforms.[74] Because of the differences in sensitivity, ion detection systems, and signal processing of the Orbitrap and Astral mass analyzers we used a matched-matrix calibration curve strategy to assess both the LOD and LLOQ for most peptides in a HeLa digest. We ultimately chose to use the LLOQ because 1) it is what we care about most, 2) is a more conservative measurement than LOD, 3) it would highlight the quantitative performance of each analyzer.

One striking difference between these two high resolution-accurate mass analyzers is the quantity of data produced by the Astral analyzer, as it can acquire spectra much faster than the Orbitrap and each spectrum contains more information, i.e. more centroids per spectrum. While we have demonstrated that this extra data contains quantitative information on more peptides, there is also a lot of additional chemical noise that can complicate analysis. Although an in-depth analysis of the extent of this problem is outside the scope of this study, it was evident when visualizing detections from multiple search engines that a significant portion (upwards of 10% in many cases) of the identifications are matching to background signals with no real peak (**Figure 3-16**). In the majority of these cases, there was extensive peptide backbone coverage generated by low intensity signals, but these signals show little to no correlation across time, meaning that at best they are not providing quantitative information and at worst, they are false matches. We propose that these false identifications are a computational problem that is common to all datasets acquired with an electron multiplier detector.

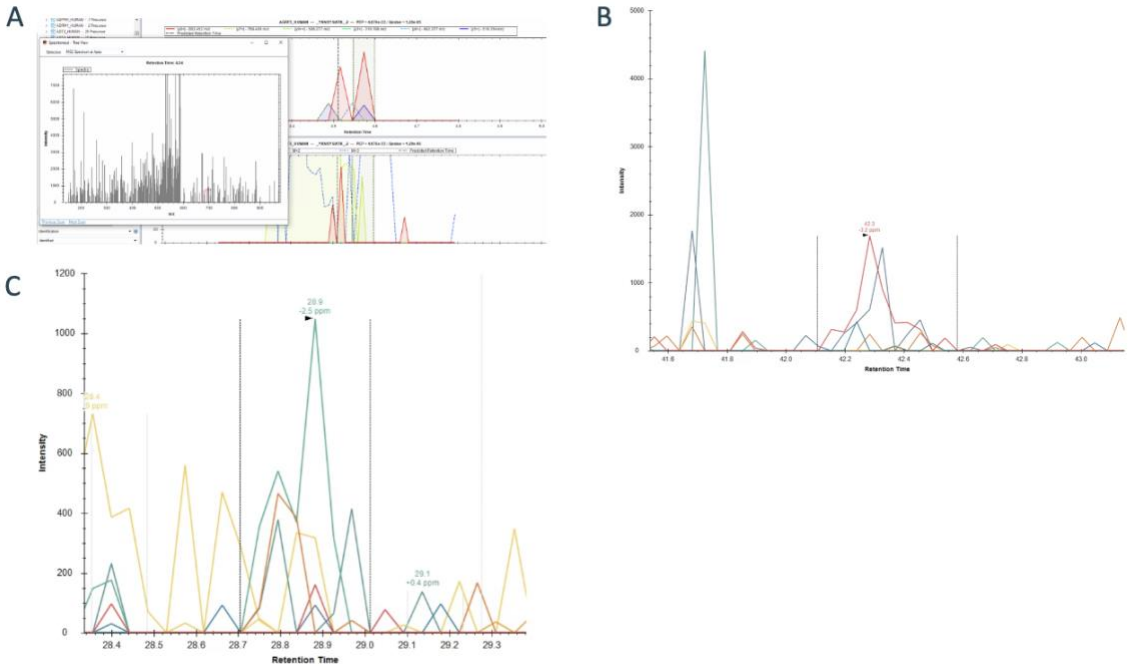

Figure 3-16: Exemplar false discoveries from search results

Examples of peptides detected at 1% FDR that have no real peak in Spectronaut (A), EncyclopeDIA (B), and DIA-NN (C). These are likely false discoveries matched to background signals and accounted for upwards of 10% of all detections regardless of the search tool used. Such false detections are characteristic of analyzers that use electron multiplier detectors, including conventional time-of-flight analyzers, as the high sensitivity can increase chemical background and convolute analysis.

Notably, 64% of mass spectral peaks in the Astral consist of five or fewer ions (**Figure 3-17**), although these low intensity noise signals account for only 6.5% of the total ion current. These extra signals in the Astral can interfere with searches if denoising approaches are not used, and tools that were originally designed for Orbitrap data should be carefully evaluated before being applied to data that is not similarly pre-processed. In this experiment we can largely filter out noise by filtering for peptides that are quantitative, meaning they can be assigned a lower limit of quantification where their change in signal is proportional to the change in input. However, there will be many cases where calibration curves are not readily available for a given

sample and, in many cases, a simpler method would be preferable. We propose looking for peptides that have at least 3 co-eluting transitions, a strategy that has been used before for DIA quality control.[50] In this case, over 76% of quantifiable peptides have at least 3 high quality transitions compared to just 40% of non-quantifiable peptides, suggesting that this type of quality filter may be a good predictor of quantitative behavior. Many of the non-quantitative peptides with high quality chromatographic peaks are contaminants or mis-assigned SILAC labeled peptides that do not respond as expected in the dilution experiment.

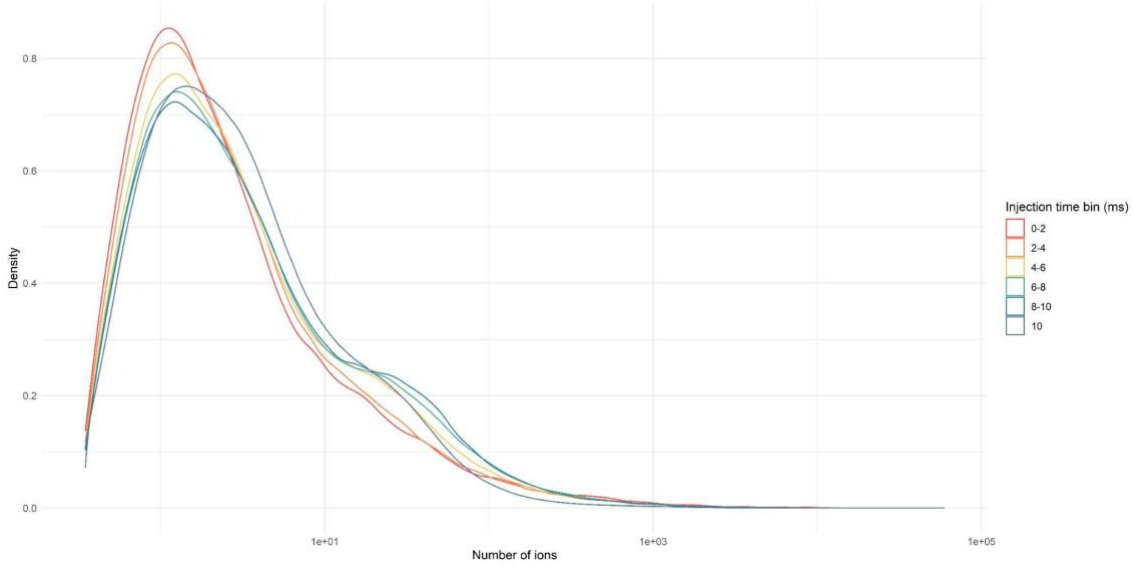

Figure 3-17: Ions per peak

Density plot of ions per spectral peak in the Astral. Spectra were binned by injection time.

While most analyses do not include this type of filtering, we believe that it is critical in reducing noise measurements, and hope that such quality filters become commonplace in the future. We propose that this issue is especially problematic in mass analyzers that are capable of single ion detection, as the spectra are much more complex (**Figure 3-1 C and Figure 3-3 C**) and low-quality identifications are common. In fact, the relative increase in low quality

detections could partially explain the high numbers that have been reported recently from time-of-flight instruments. Unlike other popular methods,[75,76] the isolation windows used on the Astral analyzer are narrow (2-4 Th), which reduces search space and spectral complexity therefore reducing the chemical noise peaks and the frequency of double-counting peptide features. This would suggest that the raw number of peptide and protein identifications alone may be a weak indicator of instrument performance and should be met with some degree of skepticism until quantitative performance can be validated. The Astral analyzer has higher resolution across the full mass range than the Orbitrap analyzer as operated in these experiments (80k vs 15k) and much higher resolving power than most commercially available TOF instruments. The high resolving power and mass accuracy could help decrease the impact of extraneous signals as a smaller mass tolerance can be used for database searching and peak integration.

Exploring the plasma proteome

To assess the potential of the Orbitrap Astral MS to generate new biological insights, we tested its performance on plasma, a matrix that is both common and analytically challenging.[7] While fractionation and enrichment strategies have improved coverage of the plasma proteome, with nanoparticle fractionation detecting 1500-2000 proteins across 5 runs[77] and other methods yielding up to 2700 proteins in a single run,[78] the plasma proteome still remains a formidable analytical challenge. The two main ways to address the dynamic range problem are to increase the dynamic range of the mass spectrometer itself or to use an enrichment or depletion strategy that will reduce the dynamic range of the samples we are measuring. Here, we apply both strategies by using the highly sensitive Orbitrap Astral MS to measure plasma proteins from an extracellular vesicle enriched sample.

The sensitivity and quantitative dynamic range of the Astral analyzer is ideally suited for complex biofluids such as plasma. The dynamic range of plasma is especially challenging for mass spectrometers as electron multiplier detectors may only have a linear dynamic range of 3-4 orders of magnitude, which is much smaller than the dynamic range of plasma. Unlike TOFs, the Orbitrap Astral MS uses automatic gain control of ion injection times, which can expand the dynamic range by two or more orders of magnitude. Therefore, we expected the Astral analyzer to perform especially well on plasma and quantify proteins across a wide dynamic range.

To reduce the dynamic range, we used a simple magnetic bead-based protocol that is capable of enriching for extracellular vesicles (EV) from plasma. Because most of the abundant plasma proteins are not associated with or within vesicles, they are depleted during the EV enrichment process.[70] The EVs are enriched using a combination of the MagResyn hyper-porous polymer matrix that functions as a molecular "net" for membrane bound particles. The polymer matrix of the beads contains a quaternary ammonium surface chemistry resulting in a cationic charge (MagResyn SAX bead). The porous polymer allows vesicles and biomolecules to intercalate within the volume of the beads and positive charge provides a way to distinguish vesicles from lipoprotein particles based on the charge.[79] This protocol results in the enrichment of known EV markers by ~20-fold and the depletion of common plasma proteins by ~95%.

To evaluate plasma proteome coverage in the Astral-based instrument we assessed the quantitative performance of multiple acquisition methods with a short (30-minute) and long (60-minute) gradient in both EV enriched and total plasma. We used multiple isolation window sizes and injection times with each gradient to assess the merits of these parameters. For the 30-minute gradient, we operated the instrument at 187 Hz (3.5 ms maximum injection time, 2 Th isolation windows), 90 Hz (10 ms maximum injection time, 4 Th isolation windows), and 50 Hz (20 ms

maximum injection time, 4 Th dynamic DIA windows). For the 60-minute gradient, we tested two methods both operating at 60 Hz (15 ms maximum injection time), one using 4 Th static isolation windows and one using 2 Th dynamic isolation windows.

Unsurprisingly, we found that the longer gradient produced more peptide and protein level identifications than the shorter gradient (**Figure 3-18 A**). The 60-minute gradient with 4 Th isolation windows and a 15 ms maximum injection time yielded the most peptide and protein-level detections, with 44,668 peptides from 5,163 proteins (**Figure 3-18 C)**. Using a 30-minute gradient with 4 Th isolation windows and a 10 ms maximum injection time, we detected 37,534 peptides from 4,704 proteins. Generally, slowing down the acquisition rate to allow for longer injection times led to better technical precision, although cutting the mass range for dynamic DIA reduced the number of peptide and protein detections. All methods assessed here detected over 4200 proteins with high technical precision, suggesting that the Astral is well-suited for these types of measurements (**Figure 3-18 B**). Further, the Astral is able to detect enrichment and depletion of certain marker proteins that span a wide dynamic range relative to total plasma (**Figure 3-18 D**). [70]

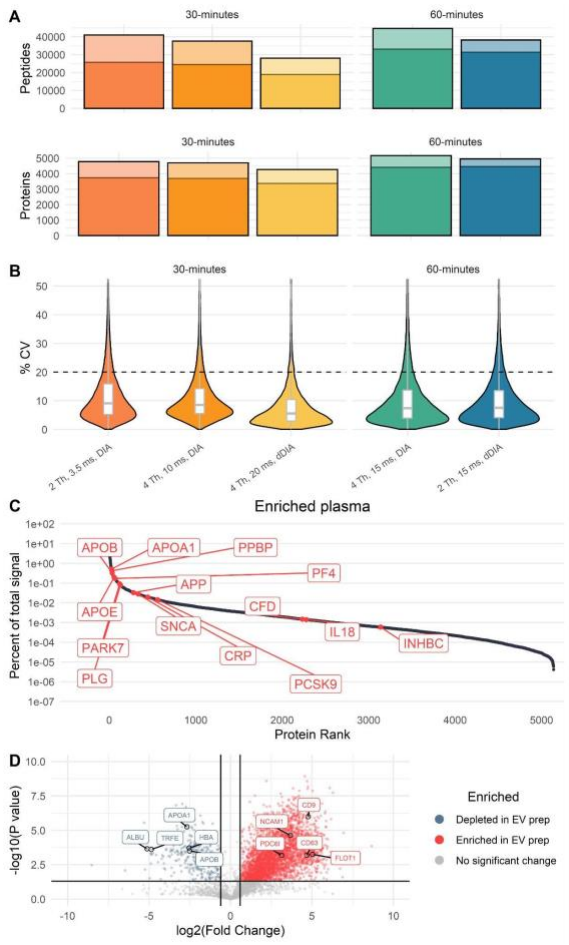

Figure 3-18: Plasma protein quantification

Peptide and protein-level detections from plasma samples (A). Lighter bars represent peptides/ proteins that were detected at 1% FDR but did not have a chromatographic peak with at least 3 co-eluting transitions. Summary of all proteins measured with selected biologically relevant proteins highlighted (B). Coefficient of variation of all peptides detected in EV-enriched plasma with each method (C). The black dashed line is at 20% CV; in all cases more than 83% of all peptides had CVs less than 20%. Volcano plot of proteins enriched and depleted in the EV-enrichment method relative to a total plasma preparation, results are shown for the 60-minute, 4 Th, 15 ms DIA method, and are representative for the results from all five methods (D). Vertical black lines indicate a fold change of +/- 1.5, horizontal black line indicates a significance threshold of 0.95.

While technical precision and concordance with known biology suggest that the measurements produced by the Orbitrap Astral MS are meaningful, it was evident that a certain

proportion of these detections can be attributed to background signals and chemical noise (**Figure 3-16**). Thus, to determine the proportion of high-quality peptides and proteins, we filtered the results for peptides with at least three co-eluting transitions. After applying this filter, the number of peptides and proteins decreased only slightly to a maximum of 31,373 peptides from 4,453 proteins with the dynamic DIA method (**Figure 3-18 A**). One result of filtering for peptides with multiple co-varying transitions is an elimination of peptides with chemical interference leading to an improved coefficient of variation (**Figure 3-19**). Here, increasing the injection time improves the ion statistics which increases both the sensitivity and the precision. Likewise, narrowing the isolation windows improves the measurement selectivity and reduces chemical interference. Both the increased injection time and improved selectivity leads to improved detection limits (**Figure 3-18 A**). Even with a relatively strict quality filter, the plasma proteome coverage presented here is unprecedented.

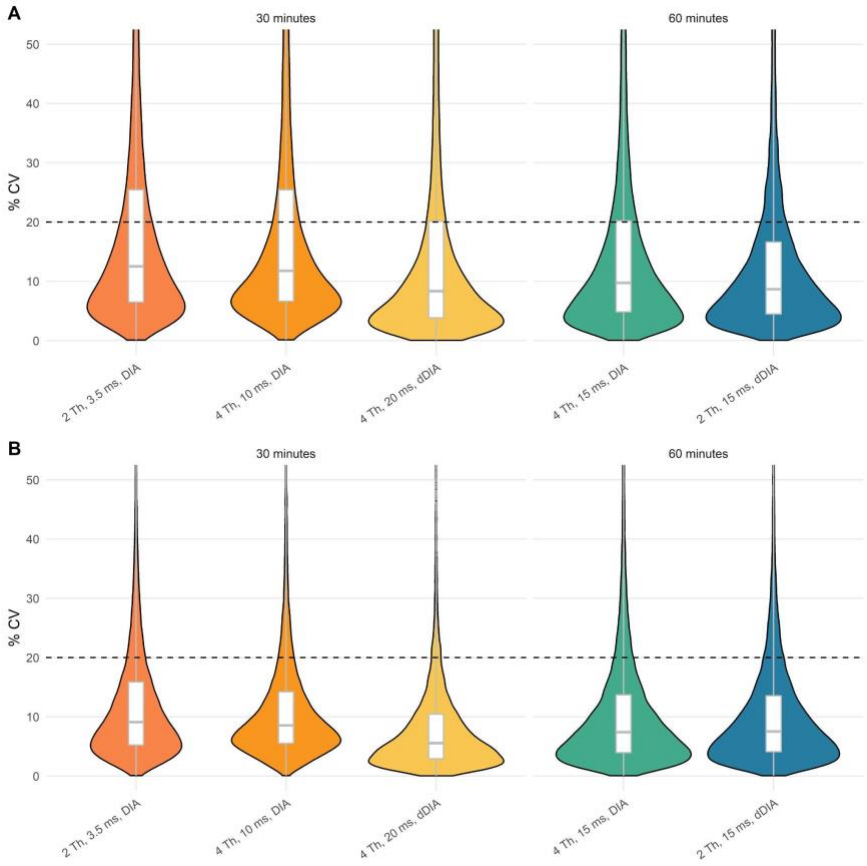

Figure 3-19: Plasma CVs before and after refinement

CVs before (A) and after (B) refinement for peptides with at least 3 co-eluting transitions. Note that the set of transitions was also filtered such that only the transitions that formed a co-eluting peak were integrated in the refined set. Black dashed line is 20% CV cutoff.

Conclusions

Overall, this initial quantitative characterization suggests that the Orbitrap Astral mass spectrometer generates high quality, quantitative data for a large number of peptides. The high acquisition rate, high resolution, sensitivity, and the power of automatic gain control produce a previously unattainable dynamic range across the proteome. In our analysis, the Astral analyzer quantified more peptides and proteins than the Orbitrap could in one fourth of the time, meaning that with the Orbitrap Astral mass spectrometer the analytical throughput can be quadrupled without sacrificing quantitative performance. Additionally, the increased speed of the Astral

analyzer allows for smaller isolation windows while covering the same mass range, producing spectra that are less complex and computationally easier to search. These advances are critical for maintaining precision and accuracy while increasing throughput of measurements.

Our results also highlight the need for quantitative characterization and careful consideration of an instrument's strengths when selecting the best instrument for the project. Despite all of the strengths of the Astral, the intra-spectrum dynamic range of the Orbitrap is superior, albeit takes a longer acquisition time to achieve, and it is important to select the appropriate analyzer for a given measurement based on the specific goals and practical limitations. Here, we use dilution curves spanning a wide dynamic range to determine the lower limit of quantification for tens of thousands of peptides in addition to limits of detection and coefficients of variation. These figures of merit provide much more context and meaning than simply evaluating protein-level identifications or even looking at a single three-proteome mixture, as has become the common practice in the field.

Not only is the quantitative performance of the Orbitrap Astral mass spectrometer analytically robust, but we have demonstrated the potential to obtain improved detection of low abundance proteins in the plasma proteome when combined with a novel, yet simple, protocol.[70] We can now quantify over 5000 plasma proteins in a single one-hour LC-MS/MS run (**Figure 3-18 C**), which was previously unattainable. This depth and quality of coverage is just one application where speed, sensitivity, and dynamic range of the new analyzer enable improved quantitative coverage in a shorter time span.

4 Dynamic Data Independent Acquisition Mass Spectrometry with Real-Time Retrospective Alignment

4.1 Summary

This chapter is based off of the following manuscript: Lilian R Heil, Philip M Remes, Jesse D Canterbury, Ping Yip, William D Barshop, Christine C Wu, Michael J MacCoss. Dynamic Data Independent Acquisition Mass Spectrometry with Real-Time Retrospective Alignment. BioRXiv. https://doi.org/10.1101/2022.11.29.518428.

Data independent acquisition (DIA) mass spectrometry has grown in popularity in recent years due to the reproducibility and quantitative rigor of a systematic tandem mass spectrometry (MS/MS) sampling method. However, traditional DIA methods may spend valuable instrument time acquiring MS/MS spectra with no usable information in them, hurting sensitivity and quantitative performance. We have developed a DIA strategy that dynamically adjusts the MS/MS windows during the chromatographic separation. The method focuses MS/MS acquisition on the most relevant mass range at each point in time – improving the quantitative sensitivity by increasing the time spent on each DIA window. We demonstrate an improved lower limit of quantification, on average, without sacrificing the number of peptides detected.

4.2 Introduction

As mass spectrometry (MS) gains popularity for the detection and quantification of proteins, a variety of data acquisition schemes have been developed to accommodate the diverse range of applications. For untargeted experiments, acquisition methods can be characterized as

data dependent or data independent. Data independent acquisition (DIA) systematically collects MS/MS spectra given mass-to-charge (*m/z*) range.[18] While systematic sampling greatly improves reproducibility between runs, there is an inherent trade-off between the total mass range covered and the quality of the results which must be considered when creating new methods. For example, a wider mass range can theoretically sample a greater range of peptides, but comes at a cost to selectivity, sensitivity, and/or quantitative accuracy, depending on the adjustment made (**Figure 4-1**). An ideal MS method strikes a balance between maintaining sensitivity by increasing the time per scan, selectivity by minimizing isolation width, coverage by maximizing mass range, and speed by minimizing cycle time. Therefore, we aim to have a method that has the sensitivity and specificity of a narrow DIA precursor isolation window while having the peptide coverage of a wide isolation window while maintaining the same cycle time.

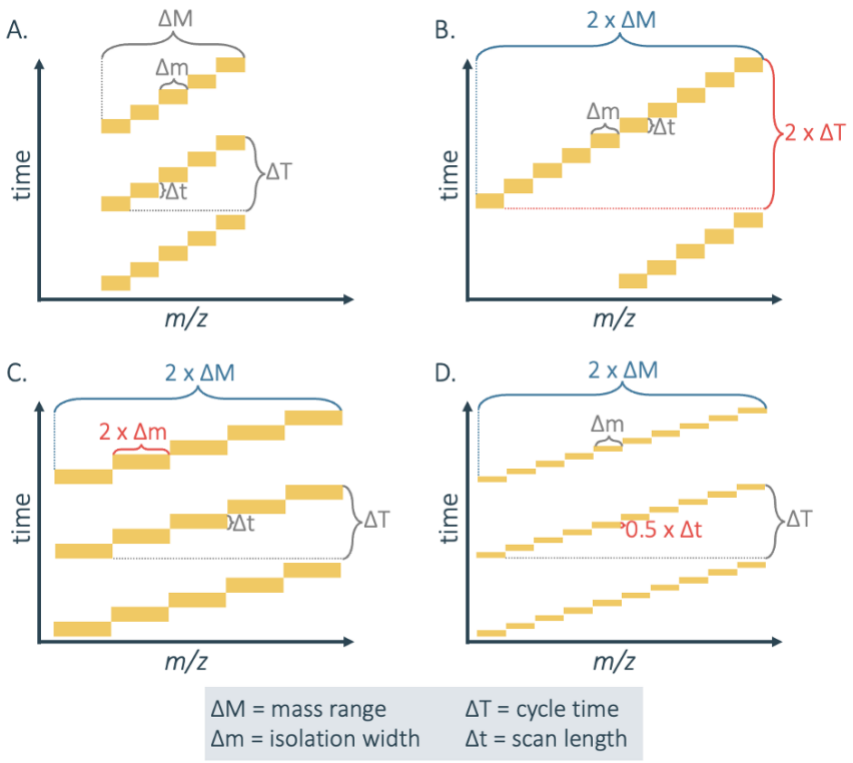

Figure 4-1: Tradeoffs associated with DIA methods

A wider mass range can theoretically sample a greater range of peptides, but comes at a cost to selectivity, sensitivity, and/or quantitative accuracy, the exact nature of which depends on the isolation width selected. A generic DIA method is shown in (A), with (B-D) illustrating possible alterations to compensate for doubling the mass range. The parameter negatively impacted by doubling the mass range is shown in red. If the isolation width is kept narrow, the amount of time that can be spent acquiring each spectrum and the number of data points acquired per LC peak is limited, decreasing sensitivity, accuracy, and reproducibility (B). If the isolation width is increased, selectivity, dynamic range, and sensitivity may suffer (C). On the other hand, a narrower mass range improves data quality by allowing a combination of either improved selectivity or sensitivity, at the cost of reduced peptide coverage (D).

In a typical proteomics experiment, peptides are separated prior to analysis by liquid chromatography (LC). Reversed-phase liquid chromatography is the most widely used separation in proteomics and separates peptides by hydrophobicity.[80] Larger peptides follow a trend of

increasing hydrophobicity.[81] Previously, Bern et al. (2010)[82] and Li et al. (2019)[83] have capitalized on the relationship between hydrophobicity and m/z by dynamically adjusting the DIA isolation windows during the separation. However, these methods have been limited by the robustness of the liquid chromatography. Therefore, we aim to build on this concept to dynamically adjust the DIA window bounds to accommodate shifts in retention time. As a proof-of-concept, we chose to use half as many isolation windows acquired for double the time (and resolution) although a number of different parameters could be adjusted, including cutting the cycle time or isolation window width in half (**Figure 4-2**). To minimize the effect of shifts in retention time,[82,83] we align the current acquisition to a run acquired previously so that the optimal DIA windows are adjusted dynamically.[31]

4.3 Methods

Complete methods are discussed in Appendix C

4.4 Results and Discussion

To create a dynamic DIA method, we start by creating a chromatogram library using gas phase fractionation DIA.[72] This chromatogram library serves two purposes: it provides a reduced list of peptides detectable in the quantitative runs[50] and is used to optimize the isolation window placements for scheduled DIA. The optimal DIA windows at different time points across the gradient are selected to maximize the number of library peptides covered while achieving the desired cycle time (**Figure 4-2A**).

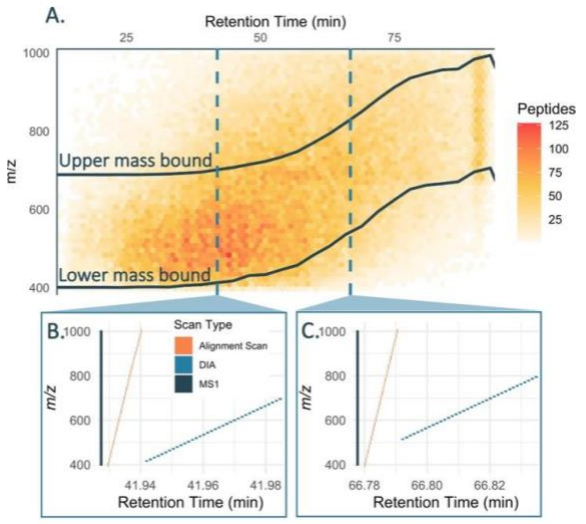

Figure 4-2: Overview of dynamic DIA method

All features detected in the chromatogram library are binned by retention time and m/z with more features corresponding to a lighter color, and the optimal m/z boundaries are shown in dark gray (A). The spectra acquired at 41.9 minutes (B) differs from the spectra acquired at 66.8 minutes (C) only in the m/z range of precursor windows that were isolated.

To use dynamic scheduling, we adapt the method described by Remes *et al.* (2020).[31] Briefly, each instrument cycle consists of an MS1 scan, a set of fast DIA alignment scans covering the entire mass range acquired in the linear ion trap, and a set of high-resolution scheduled DIA spectra (**Figure 4-2 B&C**). We acquire an alignment run using the scheduling from the original library acquisition. If this run is acquired right after the library acquisition, shifts in retention time will be minimal and dynamic scheduling is not necessary (**Figure 4-3A)**. Additionally, if a chromatogram library has not been acquired, any single MS run can be used as the alignment run and to generate a method directly. Any acquisition mode could be used for an alignment run as long as the requisite alignment scans are included and there is information that can be used to determine feature density and therefore optimal scheduling bounds (**Figure 4-3B)**. The alignment run is preprocessed using Haar compression and subsequent, quantitative runs are

acquired using the same sets of DIA alignment scans to determine retention time shifts relative to the alignment run, updating the scheduling accordingly. We have contemplated and tested several improvements to the methodology that don't change its essence as presented here, for example by including the alignment scans as part of the chromatogram library acquisition to avoid the extra alignment run; or by using "self-alignment" of the DIA scans to eliminate extra alignment acquisitions all-together.

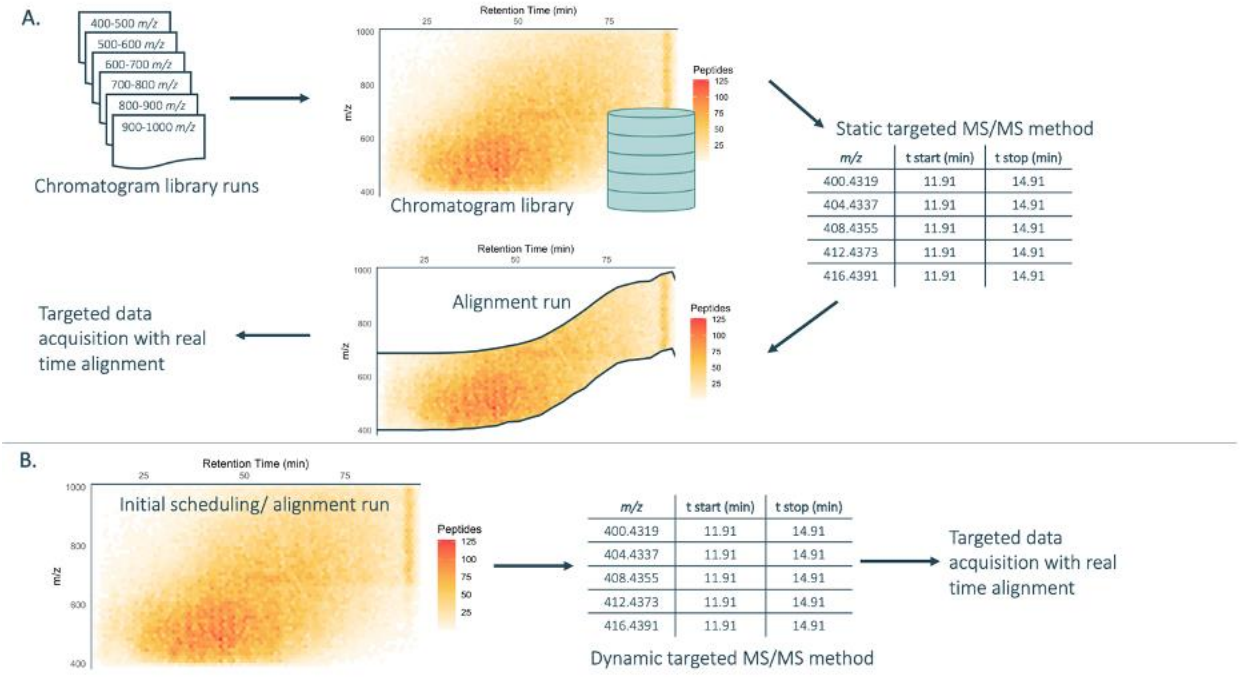

Figure 4-3: Workflows for generating dynamic DIA methods

Two alternative workflows for generating dynamic DIA methods. In (A), as demonstrated in this paper, a chromatogram library is generated and based on peptide ID density, a dynamic DIA method is created. An alignment run is acquired using these optimal window boundaries without dynamic scheduling, and the subsequent runs are based on the alignment scans from this run. In (B), an alternative workflow is shown where a single run may be used for method creation and alignment runs. Any acquisition method may be used as long as the required alignment scans are included, and based on feature density, a method can be made and immediately applied with real time alignment.

As proof of concept, we applied this dynamic DIA method to a HeLa cell lysate, generating a chromatogram library with 115,917 peptides spanning an *m/z* range of 400-1000. We elected to keep the isolation window width the same as our traditional DIA method, while doubling fill time and resolution. Therefore, we were theoretically able to measure 81,833 of the library peptides (70.6%) (**Figure 4-2 A**). It is important to note that because we use gas phase fractionation to acquire chromatogram libraries, we expect the quantitative data to contain fewer detectable peptides than the theoretical limit.

We used a matrix-matched calibration curve to compare the quantitative performance of dynamic DIA relative to our gold standard DIA method.[56] We generated calibration curves and determined the lower limit of quantitation (LoQ) (**Appendix C)**.[56] Here, we define peptides as quantifiable if they were able to be assigned a lower limit of quantitation less than the undiluted sample.

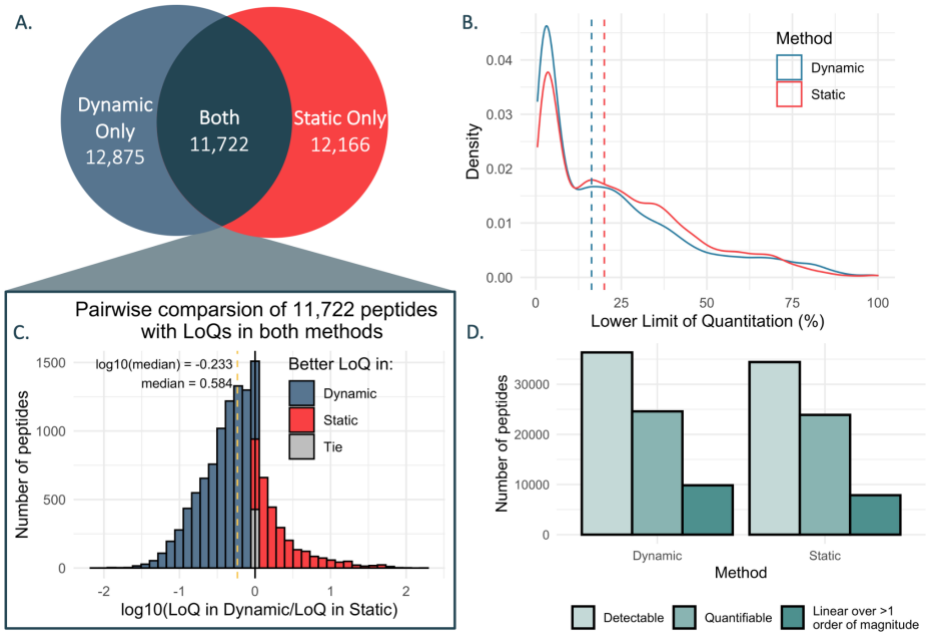

Figure 4-4: Comparison of quantifiable peptides using dynamic and traditional DIA

The total number of peptides that are deemed quantifiable in each method are shown with overlaps between the methods indicated (A). The relative density of limits of quantitation for all peptides in each method suggests that the dynamic DIA method typically leads to better LOQs (B). In cases where peptides are quantifiable in both methods, there is over a two-fold improvement (reduction) in lower limit of quantitation when using the dynamic method (C). Dynamic DIA detects and quantifies more peptides than traditional DIA (D).

With traditional DIA, we were able to detect 34,443 peptides in the matrix-matched calibration curve experiment, of which 23,888 were deemed quantifiable based on their LoQs. With dynamic DIA, we detected 36,355 peptides, and quantified 24,597 (**Figure 4-4**). Further, the median LoQ is a relative concentration of 20% with traditional DIA compared to 16% for dynamic DIA (**Figure 4-4B)**. We observe almost a 2 fold improvement in LoQ with dynamic DIA compared to the traditional static method in peptides that can be quantified from both datasets (**Figure 4-4C**). Because dynamic DIA focuses on a smaller *m/z* range, the instrument spends more time measuring peptides in this range and more ions are measured per spectrum (**Figure 4-5**).

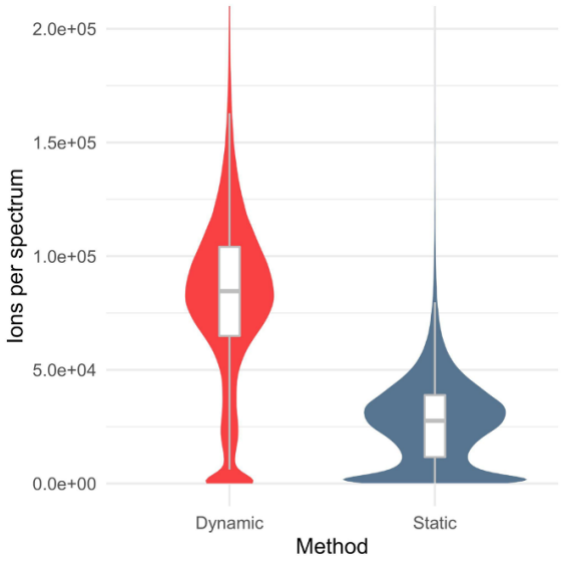

Figure 4-5: Number of ions per spectrum

By leveraging the dynamic DIA method to increase maximum injection times allowed by automatic gain control, each spectrum contains more ions which accounts for the increased sensitivity.

4.5 Conclusions

In this work, we have demonstrated the potential of scheduled DIA to improve quantitative sensitivity and precision while maintaining the number of peptide detections. Here, we present one way to use the extra time generated by shrinking the mass range, which is increasing the maximum injection time. However, we anticipate that this is not the only way to use this method and the optimal parameters will vary based on the experiment. This method is ideally suited to shorter chromatographic gradients with narrow elution peaks as demonstrated by Li et al. (2019)[83], cases where peptides of interest span a large m/z range, and complex mixtures where narrow isolation windows are important for selectivity. Narrow isolation windows could be especially useful when using unit resolution mass analyzers, where interference is a concern.

Narrowing acquisition windows is especially important given recent work showing that the unit resolution linear ion trap is more sensitive than Orbitrap analyzers in both targeted and DIA methods.[73,84] Such sensitivity improvements are critical advancements for advancing any mass spectrometry-based measurements, including the still nascent field of single cell and low input proteomics, an area that has generated widespread interest in recent years.

5 Building spectral libraries from narrow window data independent acquisition mass spectrometry data

5.1 Summary

This chapter is based off of the following publication: Lilian R. Heil, William E. Fondrie, Christopher D. McGann, Alexander J. Federation, William S. Noble, Michael J. MacCoss, and Uri Keich. Building spectral libraries from narrow window data independent acquisition mass spectrometry data. J Prot Res. 2022, 21 (6), 1382–1391.

Advances in library-based methods for peptide detection from data independent acquisition (DIA) mass spectrometry have made it possible to detect and quantify tens of thousands of peptides in a single mass spectrometry run. However, many of these methods rely on a comprehensive, high quality spectral library containing information about the expected retention time and fragmentation patterns of peptides in the sample. Empirical spectral libraries are often generated through data-dependent acquisition and may suffer from biases as a result. Spectral libraries can be generated *in silico* but these models are not trained to handle all possible post-translational modifications. Here, we propose a false discovery rate controlled spectrum-

centric search workflow to generate spectral libraries directly from gas-phase fractionated DIA tandem mass spectrometry data. We demonstrate that this strategy is able to detect phosphorylated peptides and can be used to generate a spectral library for accurate peptide detection and quantitation in wide window DIA data. We compare the results of this search workflow to other library-free approaches and demonstrate that our search is competitive in terms of accuracy and sensitivity. These results demonstrate that the proposed workflow has the capacity to generate spectral libraries while avoiding the limitations of other methods.

5.2 Introduction

Data independent acquisition (DIA) mass spectrometry (MS) has emerged as a powerful method to simultaneously quantify tens of thousands of peptides in a single MS run.[85,86] Unlike data dependent acquisition (DDA) tandem mass spectrometry methods which select precursor ions to fragment in real time, in DIA the instrument cycles through a series of pre-defined isolation windows to fragment all ions in a specific mass-to-charge (*m/z*) range.[18] Decreasing the precursor isolation width comes at a cost to either total mass range covered or the number of points across the peak, and therefore may not be a viable option. Consequently, DIA tandem mass spectra may contain ions drawn from mass ranges of 20 Da or more and can be highly complex due to the presence of product ions from multiple co-isolated peptides. As DIA rises in popularity, the increasing volume and complexity of data presents computational challenges for peptide detection.

Due to the massive amounts of data generated by modern mass spectrometers, database searches are essential to determine which precursor peptide generated a spectrum at a large scale. Traditional search methods in proteomics use a "spectrum-centric" approach[33] to identify the

peptide sequence that best explains a given tandem mass spectrum (MS2). These searches are optimized for situations where each spectrum represents exactly one peptide, an assumption which originates from typical DDA mass spectrometry data, but is often invalid for unprocessed DIA mass spectrometry data. Therefore "peptide-centric" methods were developed, which aim to detect the best evidence for each possible peptide sequences in a set of spectra. Methods such as PECAN[87] employ a peptide-centric method to detect peptides in DIA data based on a protein sequence database. Unfortunately, the large search space limits the statistical power of these searches, and it is not possible to detect modified peptides without sacrificing statistical power due to multiple hypothesis testing.[87] This limitation is inherent to peptide-centric searches as the candidate peptide search space increases with added search complexity, while spectrum-centric searches largely avoid this problem due to a limited number of spectra.

One DIA search strategy is to group co-eluting fragments and precursor ion signals and assign them to groups of pseudo-spectra that can be searched using spectrum-centric DDA database search methods.[37,88–90] While this method enables the detection of post-translational modifications without a spectral library, it relies on the presence of a precursor MS1 signal, which may not be present for all peptides that are detectable at the product ion MS2 level. This problem is common in trapping instruments such as Orbitraps where automatic gain control modulates the fill time to maximize the number of ions in the mass analyzer while minimizing negative impacts of overcrowding such as space charging and ion coalescence.[91] Automatic gain control enables the MS2 spectra to be substantially more sensitive than an MS1 spectrum by filling the trap with a narrow *m/z* range and excluding ions outside the window.[92]

Another way to increase statistical power of DIA searches is to use spectral libraries which contain information about the retention times and fragmentation patterns of peptides that

might be in a sample. Using this information, spectral libraries can be used to perform targeted signal extraction on DIA tandem mass spectra to detect and quantify peptides with great sensitivity and precision.[36,50,85,93–95] By incorporating additional features into scoring models, the use of libraries significantly boosts the performance of peptide-centric searches.[96] Because analyses using spectral libraries are limited by the depth and quality of the library, an incomplete or low quality spectral library can derail analysis of even the best data.[97] Thus, the method of library generation is of utmost importance in DIA mass spectrometry experiments.

Traditionally, spectral libraries have been generated by offline fractionation via liquid chromatography followed by DDA tandem mass spectrometry. These experiments are time consuming, sample intensive, and yield libraries which may not represent the behavior of peptides in the sample used for quantitative DIA mass spectrometry analysis.[50] The inaccuracy of DDA-based spectral libraries arises because offline fractionation changes the composition of the sample matrix relative to the unfractionated sample, causing shifts in retention time.[50,72] Additionally, DDA methods typically use different, charge state-dependent collision energies as opposed to DIA methods that typically assume the same charge state for every fragmentation spectrum, creating differences in peptide fragmentation patterns.[50,72,98] Further, it has been shown that larger DDA based libraries created with offline fractionation yield fewer peptide detections.[99] To conserve sample and time, publicly available spectral libraries generated from DDA methods can be used in place of creating a DDA fractionated library in-house. However, even high quality publicly available libraries, such as the Pan-Human library,[49] may lack crucial peptides specific to a sample and are not available for every possible experimental condition and organism. This problem may be exacerbated when specific post-translational modifications

(PTMs) are of interest that may not be detectable in a library without enrichment or may be unique to a specific set of experimental conditions.

An alternative method to overcome the problems with DDA-based spectral library generation is to use a computational model to predict entire spectral libraries. Recently, tools to predict peptide fragmentation and retention time have improved such that it is possible to accurately predict these features for all possible unmodified peptides in a proteome, producing spectral libraries that are entirely predicted *in silico*.[100–102] Programs such as EncyclopeDIA can integrate predicted spectral libraries to detect and quantify peptides in quantitative DIA mass spectrometry data.[50,100] Notably, DIA-NN has used both predicted and DIA-based spectral libraries for highly sensitive searches.[39] However, robust methods to predict which peptides are likely to be detectable in a sample are not readily available and these searches are hindered by the large search space which imposes a limit on sensitivity.[50,103] Additionally, these methods need to be trained specifically for individual post-translational modifications which can greatly limit their use.

To reduce the search-space and calibrate spectral libraries to a specific sample, gas-phase fractionation of a pooled sample can be used to obtain deep coverage of the whole sample space. In this method, a single sample is injected six times, and DIA mass spectrometry is performed using narrow (4-*m/z*) staggered isolation windows that can be computationally demultiplexed to 2-*m/z* effective isolation width.[35] This acquisition method, similar to the PAcIFIC method,[71] can be used to generate sample-specific chromatogram libraries based on the predicted library which increase the power of the search.[50,72] While this approach is not ideal for large scale quantitative experiments due to the required number of injections, a benefit of using this approach is that the effective isolation windows are similar to DDA experiments. Therefore, the data combine the

systematic nature of DIA with the relatively low-complexity of DDA data. Because these searches are limited by the scope of the predictable library, which typically only covers unmodified and fully tryptic peptides, a method to generate libraries that cover more peptides is essential. Thus, limitations in spectral library generation have complicated the widespread adoption of DIA in many circumstances.

To address the current obstacles in library construction for DIA data, we propose a workflow to detect peptides directly from gas-phase fractionated, narrow-window DIA data using a modified spectrum-centric approach outlined in **Appendix D Algorithm 1** (sketched overview in **Figure 5-1**). The method involves a spectrum-centric database search to score all possible spectrum/peptide combinations – where each peptide entry consists of a peptide sequence with a specified charge state. Then, a series of steps for score calibration and a novel false-discovery rate estimation method are used to generate a list of peptides and their corresponding spectra. We use the results of this search to build spectral libraries that include PTMs, thereby enabling us to detect and quantify these peptides in wide-window DIA mass spectrometry data. In contrast to other spectrum-centric approaches, this search workflow utilizes information from the time domain to generate a spectral library. We demonstrate that this search method outperforms DIA-Umpire,[37] a pseudo-spectrum based method, to detect phosphorylated peptides without the use of a spectral library. Additionally, we show that the peptides detected with our method can be accurately quantified in DIA mass spectrometry data. The DIA-based, empirically generated spectral library produced by our proposed workflow overcomes the limitations of many DIA search methods and simplifies the generation of high-quality libraries for many experiments that were previously unable to use DIA mass spectrometry.

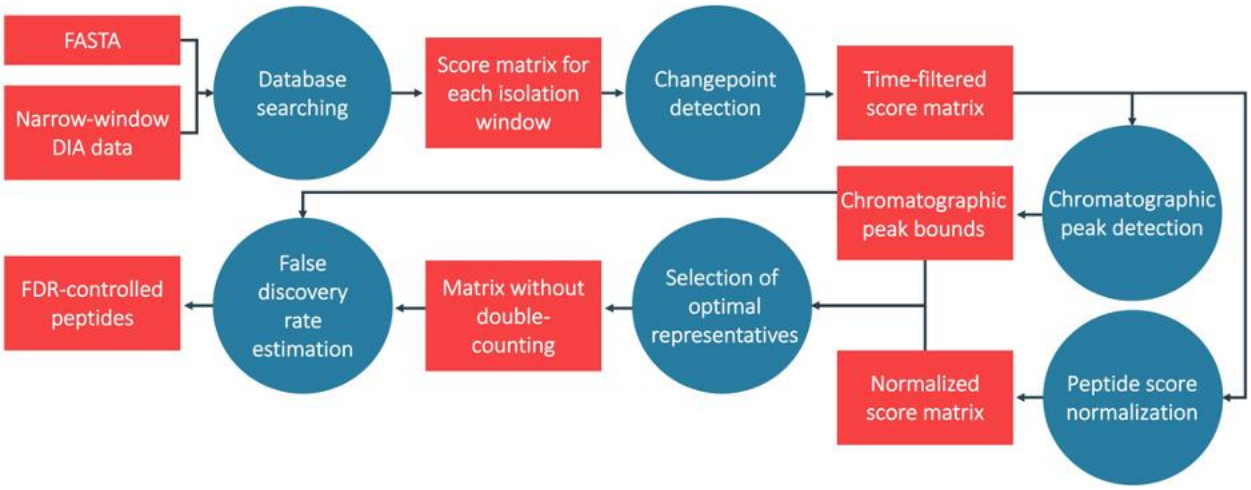

Figure 5-1: Search workflow

Workflow for direct searching of narrow-window DIA data. Each blue circle corresponds to a subheading in the Approach section. Raw data is searched against a FASTA file and the search results are converted into a matrix of scores. Then, spectra at the beginning and end of the run are removed from the matrix using changepoint detection, the chromatographic peak width is determined, scores are normalized, and highly correlated peptides (at their respective charge-states) are competed against one another directly. Finally, this set of peptides is subjected to FDR control where peptides are considered at multiple precursor charge-states.

5.3 Approach

Peptide detection in DIA data presents unique challenges and opportunities that should be considered when designing a search method. The proposed workflow aims to address these challenges and capitalize on the unique information provided by DIA tandem mass spectrometry data. As with any peptide database search, it is important to select a sensitive and powerful score metric to quantify the quality of each peptide spectrum match (PSM). Additionally, traditional spectrum-centric type searches fail to utilize the time-domain information present within DIA data and may struggle to account for highly chimeric spectra. However, simply retaining the top n best-scoring peptide candidates per spectrum with no restrictions can lead to artificially large

numbers of peptide detections because multiple peptide sequences sharing predicted fragments may score highly while only one is present. More generally, due to the complexity of DIA data, traditional target/decoy competition methods can be insufficient for DIA searches.

Here, we present a workflow to detect peptides in narrow-window DIA data that is designed to address the challenges of DIA searches with a multi-step procedure, which we describe in detail in the following sections. Note that we use the term peptide to denote a peptide sequence and charge state combination. Our notation is summarized in **Appendix Table 2** and the overall procedure is summarized in (**Figure 5-1**). All processing scripts are publicly available at https://github.com/uw-maccosslab/Narrow-DIA-Search.

The workflow presented here generates a list of peptide-spectrum matches that can be easily converted to a spectral library. The resulting spectral library contains each matched spectrum and its retention time. Downstream library searches can be performed using this library. We do not apply any processing to the library spectra, and instead select the relevant fragments during the library search step itself, which is a standard input parameter for these types of searches. The specific search parameters may vary based on application and user preference; therefore, our aim is to generate a high-quality library that can be used for any library-based search.

Database searching

Our procedure begins by conducting a Tide search[104] using the XCorr score function[11] with Tailor calibration[105] and a sufficiently large top-match parameter so that every possible mass spectrum/peptide combination is assigned a score. We use data acquired with gas-phase fractionation, wherein a pooled sample is injected six separate times, each injection spanning a

100-m/z mass range.[71] The acquisition scheme uses 4-m/z isolation windows that are acquired using a staggered isolation scheme so that they can be computationally demultiplexed to 2-m/z (**Figure 5-2**).[35,72] The Tide search results are stored in matrices, each covering a single 2-m/z precursor isolation window, where each row represents a theoretical peptide at a specific charge state, each column represents a tandem mass spectrum, and each entry is the Tailor calibrated XCorr score of the match between the corresponding peptide and spectrum (**Figure 5-3**, **Algorithm 2**). Each matrix M consists of m peptides P and n mass spectra S, where P and S are specific to each 2-m/z isolation window, and it contains no missing values. In a typical experiment, there are 50 matrices for each of the six gas phase fractions. For a 90-minute gradient searched against all unmodified fully tryptic yeast peptides, each of these 300 matrices is approximately 2000 peptides by 1500 spectra. For every target peptide in the database, there is a corresponding decoy, generated by Tide by shuffling the peptide sequence while preserving the first and last amino acids.[104] Here, a "target/decoy pair" refers to a target and the decoy generated from its sequence.

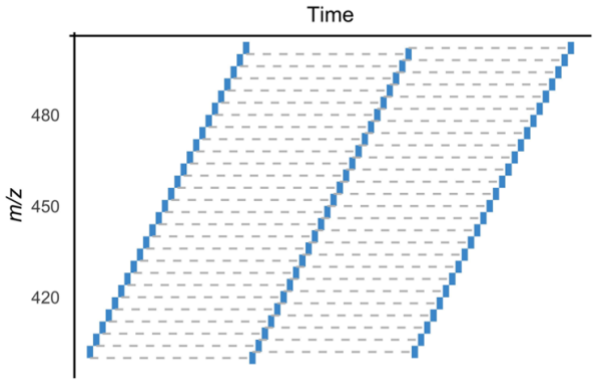

Figure 5-2: Precursor isolation scheme for fraction 400-500

Library data is acquired as described by Pino et al. (2020), using a staggered isolation scheme. Isolation windows are originally 4-m/z in width but are later demultiplexed to 2-m/z.

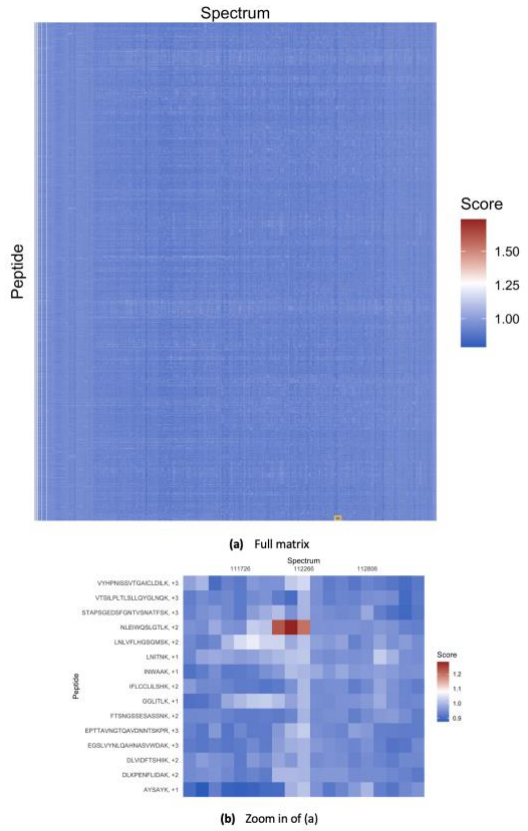

Figure 5-3: Representation of a single 2-m/z matrix produced after Tide search
(A) The full matrix has 1733 peptides (rows) and 1439 spectra (columns). (B) A zoom in on the area of (A) marked by the yellow box. The peptide NLEIQQSLGTLK is accepted at 1% FDR based on this example.

Changepoint detection

DIA data includes uninteresting mass spectra at the beginning and end of the run before and after peptides are eluting. The matrix columns (**Figure 5-3**) corresponding to these mass spectra have score distributions that are distinct from the rest of the run, and are identified using changepoint detection (**Algorithm 3**). Specifically, we considered (in this order): two, three, or four changepoints in the mean and variance of the median of the score distribution for each spectrum. The changepoints were identified using binary segmentation and if the number of spectra between the first and last changepoints was at least 50% of the total number of spectra,

then we stopped at that number of changepoints and continued the rest of the analysis using only the spectra between those changepoints (**Figure 5-4**).[106,107] This process focuses our analysis on mass spectra where peptides are likely to be present, thus ensuring more accurate score calibration.

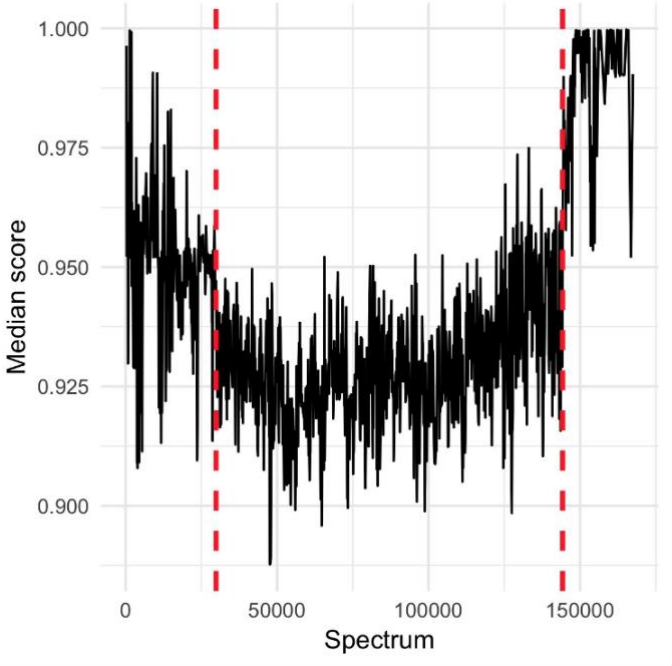

Figure 5-4: Plot of median XCorr scores for each spectrum across a DIA window
Changepoint detection automatically removes spectra with median scores distinct from the rest of the run. To account for shifts in score distributions early in the chromatographic gradient, only spectra between red dashed lines are retained for analysis. We first identify two changepoints and retain only the n' mass spectra between these two changepoints provided $n' \geq 0.5n$, where n is the original number of spectra. If $n' < 0.5n$ then we repeat the changepoint detection while increasing the number of changepoints, this time considering the n' spectra between the first and the last changepoints.

Chromatographic peak detection

As peptides elute from the column, we expect to observe them in multiple consecutive mass spectra, a feature that could potentially help to distinguish between true and false detections, but using this information is not straightforward. Notably, the associated chromatographic peak widths may vary due to many factors. Therefore, we propose a computational method to deduce the chromatographic peak width of each peptide (**Figure 5-5**).

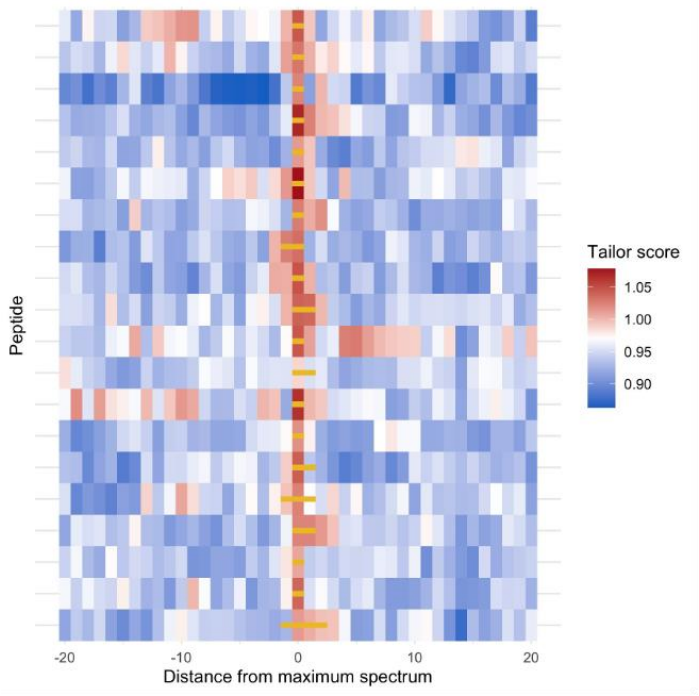

Figure 5-5: Peak boundaries assigned with our method

*The figure graphically depicts the +/- 20 Xcorr scores surrounding the top scoring spectrum for a number of target peptides. Peak boundaries (assigned by **Algorithm 4**) are shown with yellow lines.*

In the case of DIA spectra, determining the chromatographic peak boundaries is equivalent to assigning each peptide p_i a time-consecutive sequence of MS2 spectra which contain fragment ions that can be explained by p_i. Computationally this amounts to scanning the i^{th} row of the PSM matrix M for a stretch of column-consecutive high scoring entries anchored at

the maximal entry for that row. While this process is conceptually straightforward, the baseline Xcorr score distributions vary substantially between peptides. Therefore, we found that further normalization applied to each row significantly improves our ability to assign correct peak widths. Specifically, we use robust z-scores that provide normalization that is robust to outliers that appear when peptides have a few high scoring spectra.

Hence, matching the consecutive sequence of MS2 spectra to a peptide to assign peak width amounts to applying row-wise robust z-score normalization, and then identifying the maximal entry for each row (peptide). Then, we extend the sequence by adding as many entries to the left and right of the maximal entry with a score ≥ 70% of the maximal entry. When executing target/decoy competition-based FDR control, it is essential that each target/decoy peptide pair shares the same width group. Therefore, we assign each target/decoy pair the maximum of the two assigned widths. Without assigning target/decoy pairs to the same group, we are unable to perform direct target decoy competition during FDR control. In **Algorithm 4** we provide more details on the approach.

Following this last step, we compete each target peptide with its corresponding decoy so that only the larger scoring one is retained while the other is removed from consideration (**Algorithm 6**). This direct competition is essential to our false-discovery rate control and it is performed early in the procedure here to reduce the computational overhead of subsequent steps. Even with accurate chromatographic peak width boundaries, it is not obvious how to use this information. After exploring several options, including methods to reward peptides with wider peaks, we ended up exploiting this information to split the peptides into groups with similar chromatographic peak widths. Subsequently controlling the FDR separately in each group boosts power relative to an ungrouped approach.

Peptide score normalization

Many search algorithms implement spectrum-level score calibration to normalize scores and boost detection power.[105,108] This normalization helps boost power because it can account for biases unique to specific spectra (e.g., more sparse spectra might tend to have lower scores). It has been noted that similar biases are present for each peptide, where properties such as length and charge state systematically influence their scores.[109] Therefore, we aimed to implement peptide-level score normalization to increase detection power. This normalization would not be possible in DDA-type scenarios where sampling for a given peptide will be inherently sparse. However, in the case of DIA data, each peptide's m/z range is systematically sampled across the run, making this type of normalization meaningful. This calibration helps eliminate biases for or against specific subsets of peptides. For example, singly charged peptides tend to have lower scores because they have shorter sequences and therefore yield fewer ions.

Recall that when determining the peak width above we used robust z-score normalization to normalize the rows of the matrix scores M. However, the goal here is different: for width-detection the focus of the normalization is on effectively normalizing the relative differences within a row. In contrast, the purpose of peptide normalization is to be able to compare maximal PSM scores across different peptides. Accordingly, we found that it is better to use a different, Tailor-like, normalization here. Specifically, analogously to Tailor we normalize scores for a given peptide to the 99th quantile for the peptide. This normalization boosts our power to detect target peptides because the scores are now calibrated between the peptide candidates for each mass spectrum (**Algorithm 5**). Note that this normalization is not optimal for peak-width detection because the additive constant that is added to each score tends to overly-smooth the

differences within each row making it difficult to find a uniform cutoff that will define the peak width as effectively as when using robust z-scores.

Selection of optimal representatives

One potential danger in searching DIA data is assigning the same set of fragment ions in a tandem mass spectrum to multiple peptide sequences that happen to share a number of theoretical fragment ions. Although this risk is mitigated by using the narrower 2-m/z isolation windows, there are still peptides that fall in the same m/z isolation window and share fragment ions. This issue is particularly relevant when searching for post-translationally modified peptides with multiple possible localization sites. In these cases, many possible localization isoforms may score highly when only one is truly present. To address this problem, we developed a greedy procedure designed to select the optimal representatives from groups of highly similar (potentially) co-eluting peptides. In the above scenario this means that similar, modified peptides are competing against each other and only the highest scoring modified peptide is retained.

While one can determine the degree of similarity between two peptides by directly comparing their theoretical spectra, this would be a very time consuming job. Therefore, as an alternative we postulate that peptides with shared fragment ions will have highly correlated scores when matched against all tandem mass spectra, or equivalently across all retention times (**Figure 5-6**). Thus we can assess the similarity between two peptides p_i and p_j by determining the similarity between their corresponding rows of scores M_i and M_j. More specifically, we look at the normalized dot product, or the cosine of the angle, between these two vectors.

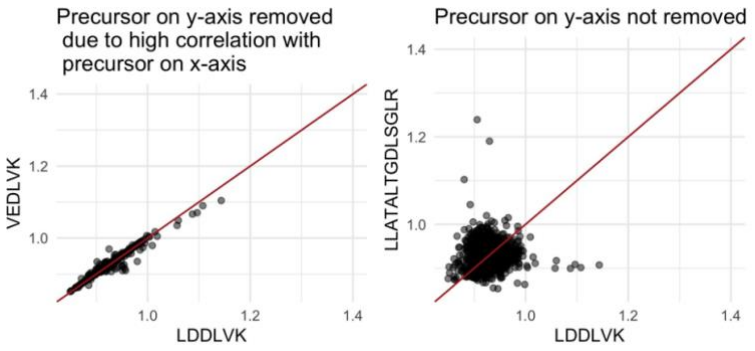

Figure 5-6: Pairwise scatterplot with each point representing a single spectrum
The peptides are identified as significantly correlated (left). On the right, the peptides are not significantly correlated. If the peptides on the left share more than 25% of their ions, the lower scoring peptide (y-axis) will be removed.

In practice our procedure goes through the list of peptides sorted in decreasing order and for each considered peptide p_i, it removes all lower scoring peptides p_j that (i) have overlapping elution profiles, as determined from the chromatographic peak width and boundary selection, (ii) have a statistically significant angle-cosine correlation relative to random peptides, and (iii) share at least 25% of theoretical fragment ions.[110] This filtering process is applied separately to the target and decoy peptides and is described in detail in **Algorithm 7**.

False discovery rate estimation

Accurate FDR estimation is challenging for DIA data due to the presence of potentially chimeric tandem mass spectra. We aimed to create an FDR control method that will leverage the information provided by DIA data to increase power. Therefore, to control the FDR, we employ a multistep procedure that includes a novel target/decoy peptide competition (TDC) step.

First, we select the maximum score and corresponding spectrum for each peptide. Then, we retain only the highest remaining scoring peptide for each tandem mass spectrum. Although DIA data contains chimeric spectra, it is unlikely that peptides in the same isolation window

share the same maximally matched spectrum, especially after representative selection. In a two hour gradient, we may collect over 150,000 spectra, but detect fewer than 10,000 peptides in most gas-phase fractionated DIA runs. Indeed, we find that using more than one peptide per tandem mass spectrum for FDR control actually reduces power at low false discovery rates possibly due to less-than-perfect score calibration (**Figure 5-7**).

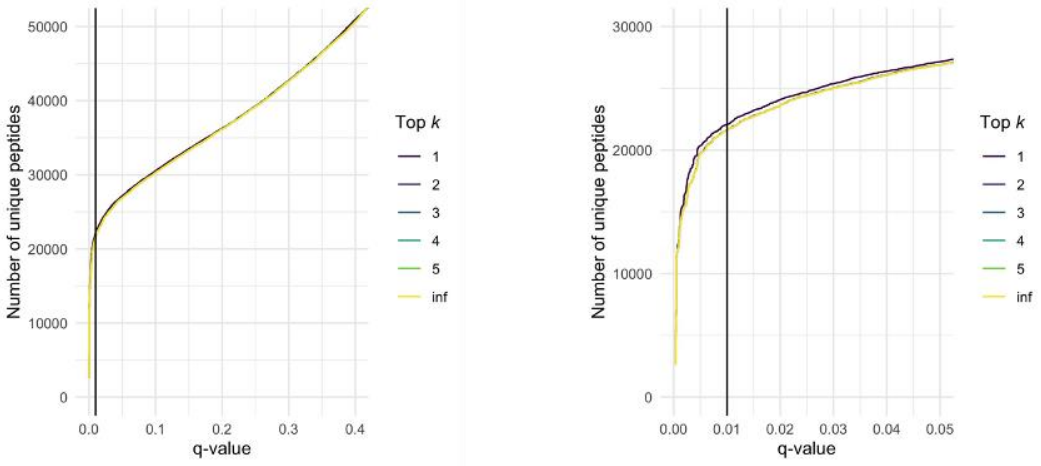

Figure 5-7: Plot of q-values when the top k peptides are retained

Here, the top k scoring peptides for each spectrum are retained for FDR control. Although many DIA spectra are chimeric, it is rare that multiple "true" peptides would share the same optimally matching spectrum. Therefore, using a k value of 1 does not seem to eliminate a significant number of peptide matches, rather it increases power presumably by culling some high scoring fake matches. The vertical line represents 1% FDR threshold. Left is full plot, right is zoomed in from 0-5% FDR.

Each of the peptides remaining after the last step is associated with a score as well as a target/decoy label. While we can pool all of these remaining peptides together to do a standard TDC wherein at a given score threshold the ratio of decoys to targets is calculated,[111] grouping peptides according to their chromatographic peak widths and applying TDC separately to each group yielded more discoveries in our analysis. Moreover, when considering PTMs, we can possibly further divide the groups according to the post-translational modification status of the

peptides, again followed by FDR control within each group using standard target/decoy procedures described in **Algorithm 1**. Using grouping in this way may allow for making more discoveries,[112–114] but it can also be detrimental if some groups do not contain a sufficient number of peptides with high scores.[115] Therefore, we advise using caution when a small number of high scoring targets are present in the data. In those cases, the use of groups may reduce discriminatory power by dividing an already low number of high scoring targets into even smaller groups.

5.4 Methods

Complete methods are discussed in Appendix E

5.5 Results and discussion

The proposed workflow increases sensitivity for detecting phosphorylated peptides relative to pseudo-spectrum approach

To assess the ability of our workflow to detect post-translationally modified peptides by generating libraries in cases where spectral libraries are not readily available, we sought to detect phosphorylated peptides in DIA data from HEK 293T cells. While there are several methods for the detection of phosphorylated peptides in DIA data, most rely on the use of existing libraries.[116] Although there are high-quality spectral libraries available for phosphorylated peptides,[117–119] that is not the case for many other post-translational modifications.

As a reference point, we used DIA-Umpire[37] as an alternative library-free approach. Specifically, pseudospectra were extracted using DIA-Umpire and searched using Tide (with Percolator used only for false discovery rate estimation). These results provide a proof-of-

concept that our workflow offers a powerful alternative method for detecting PTMs directly from DIA mass spectrometry data.

Overall, our direct search workflow detects more peptides at every confidence level relative to the DIAUmpire approach (**Figure 5-8 A**). At an FDR threshold of 1%, our workflow detects 1300 unique peptides while the DIA-Umpire approach detects 851 peptides. We observe overlap between the phosphorylated peptides detected by both searches at a 1% FDR (**Figure 5-8 B**) although our workflow is able to detect many more unique peptides. These results suggest that our workflow is a viable alternative to other library-free methods for detecting PTMs in DIA data.

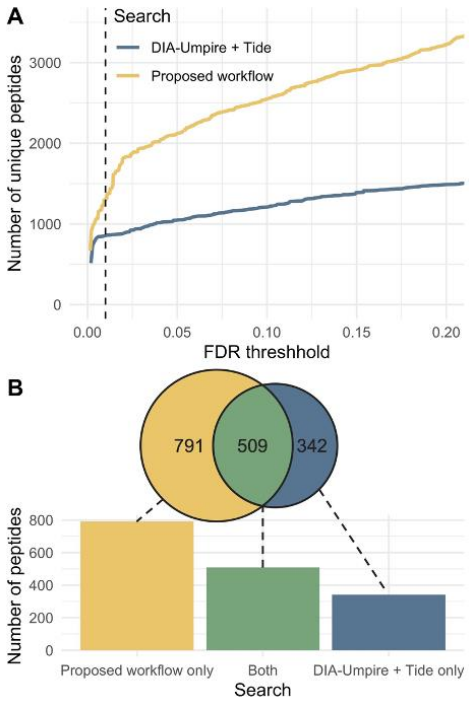

Figure 5-8: Comparison of searches for phosphorylated peptides

(A) Plot of number of peptides detected at a given FDR threshold obtained by searching for doubly charged phosphorylated peptides using our workflow compared to a search with DIA-Umpire results with Tide and adjusting FDR with Percolator. (B) The 1% FDR threshold is indicated with a dashed vertical line. The number of peptides unique to each

search as well as those shared by both methods are also displayed at a 1% FDR threshold.

One reason our workflow is more sensitive in detecting phosphorylated peptides than the pseudo-spectrum based approach (**Figure 5-8 A**) is that it is independent of signal in the precursor (MS1) spectra. Due to automatic gain control in trapping-type instruments, MS1 signal can be less sensitive than MS2.[91,92] DIAUmpire and similar methods are optimized to detect peptides in wide-window DIA data where the MS2 signal may be less sensitive than it is with narrow isolation windows due to limits on total ion signal. Therefore, our workflow is more sensitive due to its ability to detect target peptides without an MS1 signal, while DIA-Umpire cannot. As further evidence of the differences in these methods, both methods detect a large number of unique peptides at 1% FDR (**Figure 5-8 B**). This finding suggests that the two methods use orthogonal methods to detect peptides. While the DIA-Umpire approach may be more sensitive in some cases, the fact that our workflow does not rely on MS1 signal helps it perform better for narrow-window DIA mass spectrometry data.

Peptides detected with the proposed workflow are quantitatively accurate

To ensure that our workflow is producing accurate results, we assessed the quantitative accuracy of peptides detected with our workflow compared to EncyclopeDIA, a peptide-centric method for peptide detection in DIA data.[50] We created a library of all peptides detected with our workflow at 1% FDR in narrow window DIA data. We compared these results to a library generated by searching the narrow-window data in EncylopeDIA using a Prosit predicted library.[100] We then used both libraries to search quantitative wide-window DIA data of 100% and 50% yeast in YEPD media diluted with yeast grown in ^{15}N media and compared quantitative accuracy of all peptides.

In this analysis, we use the ratio of signal between the two datasets as "ground truth" to evaluate search accuracy. If a peptide is really present, we would expect to see a two-fold decrease in signal in the 50% sample relative to the 100% sample, which corresponds to a quantitative ratio of two. Peptides detected in both searches were highly accurate (**Figure 5-9 B**) with a median quantitative ratio of 2.14. Peptides unique to the library generated with the proposed workflow and the predicted library had median quantitative ratios of 2.02 and 2.31, respectively.

The library created by our direct search detects similar, but slightly lower, numbers of quantitative peptides relative to a Prosit predicted library in the *S. cerevisiae* dataset (**Figure 5-9**) At 1% FDR, both searches have a high degree of overlap in quantitative peptides as determined by EncyclopeDIA (**Figure 5-9 A**). These two searches have much higher overlap than the previous analysis (**Figure 5-8 B**) because neither rely on MS1 signal. Although the Prosit library is able to detect 9% more quantitative peptides, the peptides unique to that search are less accurate than the peptides found in both searches and those unique to the direct search (**Figure 5-9 B**). Therefore, the extra detections in the Prosit search may in part be due to a higher false-positive rate. These results suggest that our search workflow compares favorably to using another popular library-free DIA search method, using EncyclopeDIA as a peptide-centric DIA approach to search a predicted spectral library.

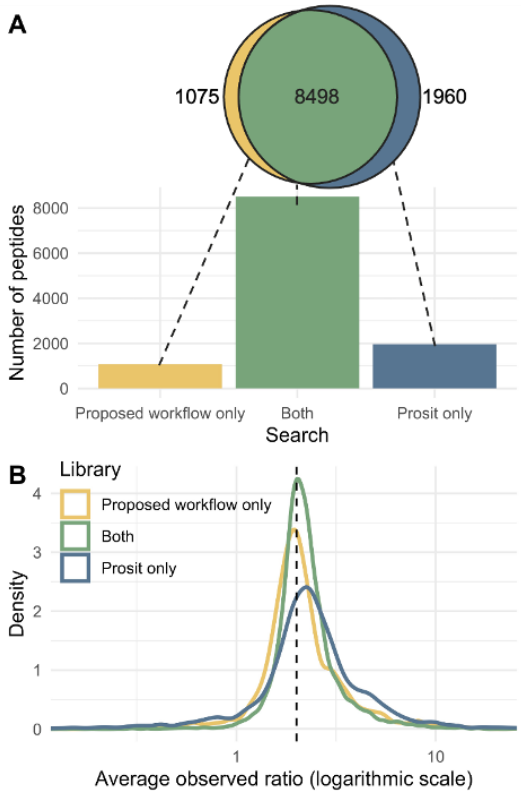

Figure 5-9: Comparison of peptides detected in quantitative data

Quantitative peptides were detected in the EncyclopeDIA search of wide-window DIA data using both a library generated from our workflow and a narrow-window calibrated Prosit predicted library. The overlap between peptides detected with both methods at a 1% FDR threshold (A) is shown along with a normalized density plot of the observed ratio of the peptides in each group between an undiluted sample and a sample diluted 50% (B). For the quantitative comparison, the expected ratio of 2 is shown as a dashed vertical line.

We used a 1% FDR threshold to generate our library here, but one could also use a higher FDR threshold (**Figure 5-10**). However, the higher cutoff comes with an increased risk of false detections, and we therefore suggest using the conservative 1% threshold. We note that the 1% FDR cutoff library is highly efficient in the sense that we are detecting a much larger fraction in the EncyclopeDIA quantitative search than the narrow-window calibrated Prosit library (**Figure 5-10**). At the same time this gives further evidence of the accuracy of our search. Based on this

comparison with a more conventional peptide-centric search approach, we have validated the performance of our proposed method. Not only are we able to detect similar numbers of peptides, but the peptides we detect are highly quantitatively accurate when spiked in at known ratios. Therefore, our search compares favorably to a library-based peptide-centric approach.

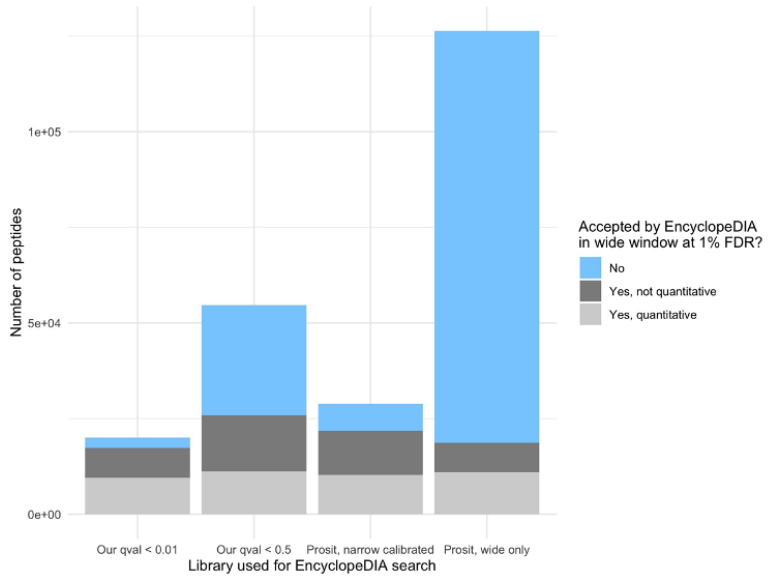

Figure 5-10: Summary of the number of peptides in EncyclopeDIA runs using alternative libraries

The figure summarizes the results of running EncyclopeDIA on the wide-window yeast DIA data using four different libraries: the first two runs used narrow-window yeast DIA data processed by our novel tool using a q-value cutoff of 0.01 and 0.5 to generate the library, and the third run used EncyclopeDIA's ability to take advantage of the same narrow-window data to significantly reduce the initial Prosit-generated library. The last run used the full-size Prosit library of all possible tryptic peptides.

5.6 Conclusions

Here, we propose a library-free workflow to detect peptides in narrow-window DIA data without relying on MS1 signal. We address many of the issues in searching for PTMs in DIA data, including the reliance on a spectral library. Existing spectral libraries may not contain all

relevant peptides and may not be available for all PTMs. Additionally, spectral and retention time prediction methods are not trained for many PTMs. Our method is able to detect PTMs from narrow-window DIA mass spectrometry data to create libraries that may be used for quantitative DIA mass spectrometry analysis.

The gas-phase fractionated DIA data used here could easily be collected in almost any experiment with minimal sample use and instrument time compared to deep-fractionation DDA-based libraries. These results show that this type of search is a promising avenue for future development, but it will require computational adjustments to become readily useable for many applications. While DIA has many advantages over DDA methods, its uses can be limited by computational hurdles. Researchers wishing to study non-model organisms and/or specific post-translational modifications may be unable to rely on many existing tools and spectral libraries. Therefore, this type of search workflow is crucial for moving the field forward and making DIA accessible to a broader range of applications.

6 Future Directions

Mass spectrometry-based proteomics has become mainstream over the past several decades due to its versatility and power. However, poor reproducibility and an emphasis on number of identifications rather than quantitative performance still limit the field. Systematic data acquisition strategies have the potential to improve the quality of measurements and ultimately lead to previously unattainable biological discoveries. This dissertation explores several such systematic data acquisition and analysis strategies and demonstrates their quantitative power. However, for such systematic strategies to be more widely implemented, future work should focus on making these methods more accessible to average users. Easily

accessible, open source user interfaces to create, run, and analyze methods are critical for the future of the field. We hope that the results presented here provide compelling evidence for the benefits of investing in infrastructure to implement these systematic acquisition and analysis methods on a wider scale.

Appendix A Supplementary methods for Chapter 2

Sample preparation and mass spectrometry

Sample preparation

A 50:50 mix of female and male pooled human plasma from Golden West Biologicals was created (Lot # MSG11500F and MSG11500M). Both the pooled human plasma and chicken serum (Cat #SC30 Lot #190509-0130, Equitech-Bio, Inc.) were split into aliquots containing approximately 300 μg of total protein. Both aliquots were reduced, alkylated, and digested with trypsin using S-Trap Mini sample preparation kit (ProtiFi) according to manufacturer guidelines.[120] The digested samples were diluted to a concentration of approximately 0.1 mg/mL with 0.1% formic acid in water with Pierce Peptide Retention Time Standard (PRTC) (Thermo Fisher Scientific) at a concentration of 0.4 nmol/mL. A dilution series was prepared by diluting the human plasma digest in the chicken serum digest at 11 different dilution points of 0, 0.5, 1, 3, 5, 7, 10, 30, 50, 70 and 100% human plasma (**Appendix Table 1).**

Appendix Table 1: List of points on the dilution curve

List of points on the dilution curve. Dilution curve was made up of 11 points, with more points at low concentration to get more accurate lower limits of quantitation

Label	% Human	% Chicken

A	0	100
B	0.5	99.5
C	1	99
D	3	97
E	5	95
F	7	93
G	10	90
H	30	70
I	50	50
J	70	30
K	100	0

Liquid Chromatography-Mass Spectrometry

LC-MS/MS analysis was performed with a Thermo Easy-nanoLC coupled to a Thermo Orbitrap Eclipse Tribrid Mass spectrometer. Peptides were separated using an in-house manufactured column created by pulling 75 μm inner diameter fused silica capillary (TSP075375, Molex) to an approximately 5 μm ID tip using a Sutter Instruments CO2 laser based micropipette puller. The pulled tip capillary was packed with 30 cm of 3 μm ReproSil-Pur C18 beads (Dr. Maisch GmbH) using an in-house developed pressure bomb with high pressure helium gas. The trap column consisted 150 μm inner diameter fused silica capillary (TSP150375, Molex) fritted on one end with Kasil and packed with 4 cm of 3 μm ReproSil-Pur C18 beads (Dr. Maisch GmbH). Mobile-phase A was 0.1% formic acid in water and mobile-phase B was 0.1% formic acid in 80% acetonitrile. Peptides were separated with reversed-phase liquid chromatography over a 90 minute gradient from 0 to 40% B followed by a 5 minute ramp to 75% B, a 5 minute hold at B, and an additional 15 minute wash at 100% B.

Data for the chromatogram library were acquired previously using the same chromatography setup coupled to a Thermo Lumos Tribrid Mass spectrometer. Tandem mass spectrometry data were acquired in 6 gas-phase fractions with 4-*m/z* staggered data independent acquisition (DIA) windows as described in detail by Pino et al. (2020).[72] Retention time calibration was performed on the Eclipse with the same LC method using 8-*m/z* staggered DIA window isolation scheme described previously.[72]

Parallel Reaction monitoring

All parallel reaction monitoring experiments were performed using one of four methods. The four methods can be distinguished based on the following characteristics (analyzer, quadrupole precursor isolation width): linear ion trap, 0.7 Th; linear ion trap, 1.7 Th; Orbitrap, 0.7 Th; and Orbitrap, 1.7 Th. Each method featured a standard MS1 scan and a set of targeted MS2 scans, all acquired with the same mass analyzer. The same target list with scheduled acquisition was used in all methods, and scheduling was adjusted throughout the experiment to account for retention time drift. For MS2 acquisition, all methods used standard AGC target, dynamic maximum inject time with 20 desired points across the expected peak width of 40 s. Fragmentation occurred in the ion routing multipole using HCD with normalized collision energy of 30%. When the linear ion trap was used, MS2 were acquired using turbo scans covering a mass range of 150–1250 *m/z*. The Orbitrap methods collected MS2 spectra at 15,000 resolution covering a range of 150–2000 *m/z*.

Replicate injections of human plasma digest

To assess technical precision, 3 μL of the same vial of 100% human plasma digest was injected in five separate replicates. Each replicate consisted of four total injections, one injection using each of the four methods, run in a random order.

Matrix-matched calibration curve

The calibration curve was created by acquiring three replicates of each dilution point (**Appendix Table 1**) in both the linear ion trap and Orbitrap both using only the 1.7 Th precursor isolation width. Data were acquired starting with the lowest dilution point, running one injection each in the linear ion trap and Orbitrap, then moving to the next lowest dilution point until the full curve was run. Following a quality control run, the process was repeated for a total of 3 replicates. The order of linear ion trap and Orbitrap acquisition was randomly assigned for each replicate of each dilution point.

Data availability

The raw MS files, EncyclopeDIA .elib file, and all Skyline documents are available have been deposited in the ProteomeXchange consortium[121] (identifier PXD023334) via Panorama Public[122] (https://panoramaweb.org/qLITvsqOT.url).

Data processing and analysis

Library creation and refinement

A chromatogram library was created by searching gas-phase fractionation DIA data against the Pan-Human library[49] in EncyclopeDIA (v 0.9.0),[50] using a method that has been described in detail previously.[72] The chromatogram library was loaded into Skyline (v 20.2.1.315)[123,124] along with the 8-*m/z* DIA data collected for this experiment demultiplexed and

converted to mzML format in MSConvert (Proteowizard, V 3.0.20239).[125] From these data, the five most abundant peptides were selected for each protein, and the remaining peptides were filtered for a high dotP, high retention time correlation with the chromatogram library, and minimal interference. Preference was given to peptides from plasma proteins that have been previously identified for their role in monitoring human wellness (https://panoramaweb.org/Passport).[51] The isolation list of all 432 peptides plus 14 PRTC peptides was scheduled with a 5 minute retention time window (**Figure 2-3**). As PRM data were acquired, results were loaded into Skyline and used to adjust scheduling windows as needed.

Data parsing

Raw files from the PRM run were imported into Skyline Daily (v 20.2.1.315)[123,124] for peak detection, integration, and visualization. Runs using the linear ion trap were integrated using a tolerance of +/- 0.7 *m/z*, while the Orbitrap was processed with a tolerance of 20 ppm. Peak picking was manually refined for all data files, and area under the curve abundance was obtained for each peptide. For the replicate injection experiment, all peptide abundances were normalized to the median signal of the run. For the matrix-matched dilution curve runs, peptide abundances were normalized to the global standards (PRTC peptides). Thirty-seven peptides that were observed to have truncated peaks in one or more of the matrix-matched dilution curve runs or to be present in the chicken digest were removed from this analysis. Area under the curve data were exported from Skyline, and analyzed in R (v 3.6.1) and Python (v 3.8.2). Lower limits of quantitation (LLOQs) were calculated for each peptide in each analyzer from the dilution curve data using the method described by Pino et al. (2020).[56] The code used for analysis and figure generation is available on Github (https://github.com/uw-maccosslab/iontrap_vs_orbitrap).

Transition refinement to optimize limit of quantitation

A list of areas under the curve for all singly charged b- and y- ions beyond the second residue were exported from Skyline. For each peptide and each mass analyzer, the optimal set of transitions was determined independently as follows. First, the LLOQ was calculated for all transitions separately using the method developed and described in detail by Pino et al.[56] The LLOQ was then calculated for the summed area of the four transitions that had the lowest LLOQ, which constituted the initial accepted transition set. Then, the LLOQ was calculated for the accepted transition set and each of the remaining transitions. Each additional transition that lowered the LOQ was added to the accepted transition set. Finally, the refined LLOQ was recalculated with the full set of transitions.

Appendix B Supplementary methods for Chapter 3

Cell Culture

HeLa S3 cells were cultured in DMEM/10% FBS, with stable isotope labeling performed according to manufacturer instructions using the Thermo Scientific™ SILAC™ Protein Quantitation Kit (Catalog A33972, Thermo Fisher Scientific). Briefly, cells were exchanged into DMEM (supplemented with Penicillin-Streptomycin and Thermo Scientific™ GlutaMAX ™ containing either unlabeled L-lysine and L-arginine, or the same supplemented DMEM containing 13C6 15N2 L-lysine and 13C6 15N4 L-arginine for SILAC labeling. Labeling was performed over 8 doublings and cells were then harvested at ~85% confluence and pelleted by centrifugation.

Matrix-matched calibration curve sample preparation

Both SILAC labeled and unlabeled HeLa cell lysates were prepared using protein aggregation capture.[126,127] Briefly, cell pellets were lysed with probe sonication in a lysis buffer containing 2% SDS. A Pierce BCA assay (Thermo Fisher Scientific) was used to estimate protein concentration and dilute lysates to a final concentration of 2 µg/µL in 1% SDS. Following reduction in 20 mM dithiothreitol and alkylation in 40 mM iodoacetamide, both samples were diluted to 70% acetonitrile and bound to MagReSyn Hydroxyl particles (ReSyn Biosciences) at a ratio of 100:1 (beads to protein). Subsequent washing was performed on a magnetic rack, with three washes in 95% acetonitrile followed by two washes in ethanol. Following the final wash, trypsin (in 50 mM ammonium bicarbonate) was added at an enzyme to protein ratio of 1:33. Samples were digested at 47˚ C for 4 hours, eluted from the beads and dried down via vacuum centrifugation. The digests were resuspended at a concentration of 200 ng/µL in 0.1% formic acid, Pierce peptide retention time calibration mix (PRTC) was spiked in to a final concentration of 5 pmol/µL, and digests were combined to form the following dilutions (X% unlabeled in (100-X)% heavy-labeled): 0, 0.5, 1, 5, 10, 25, 50, 70, 100%.

Plasma preparation and membrane particle enrichment

An enriched membrane particle fraction was prepared from plasma using Mag-Net, a magnetic bead-based protocol developed by Wu et al. (forthcoming). Plasma particle enrichment and protein aggregation capture[126,127] steps were performed on a Thermo Scientific™ KingFisher Flex. Briefly, HALT cocktail (protease and phosphatase inhibitors, Thermo Fisher Scientific) was added to 100 µL plasma and then mixed 1:1 by volume with Binding Buffer (BB, 100 mM Bis-Tris Propane, pH 6.3, 150 mM NaCl). MagReSyn® strong anion exchange beads (ReSyn

Biosciences) were first equilibrated 2 times in Equilibration/Wash Buffer (WB, 50 mM Bis Tris Propane, pH 6.5, 150 mM NaCl) with gentle agitation and then combined in a 1:4 ratio (volume beads:volume starting plasma) with the plasma:BB sample for 45 minutes at room temperature. The beads were washed with WB 3 times 5 minutes with gentle agitation. The enriched membrane particles on the beads were then solubilized and reduced in 50 mM Tris, pH 8.5/1% SDS/10 mM Tris (2-carboxyethyl) phosphine (TCEP) with 800 ng enolase standard added as a process control. Following reduction, the plate was removed from the Kingfisher Flex. Samples were alkylated with 15 mM iodoaceamide in the dark for 30 minutes and then quenched with 10 mM DTT for 15 minutes. Total unfractionated plasma (1 µL) was prepared in parallel (reduction, alkylation, and quenched) and added to the Kingfisher plate as a control for enrichment. The samples were processed using protein aggregation capture with minor modifications.[126,127] Briefly, the samples were adjusted to 70% acetonitrile, mixed, and then incubated for 10 minutes at room temperature to precipitate proteins onto the bead surface. The beads were washed 3 times in 95% acetonitrile and 2 times in 70% ethanol for 2.5 minutes each on magnet. Samples were digested for 1 hour at 47 ˚C in 100 mM ammonium bicarbonate with Thermo Scientific™ Pierce™ porcine trypsin at a ratio of 20:1 trypsin to protein. The digestion was quenched in 0.5% formic acid and spiked with Pierce Retention Time Calibrant peptide cocktail (Thermo Fisher Scientific) to a final concentration of 50 fmol/µL. Peptide digests were lyophilized and stored at -80°C.

Liquid Chromatography-Mass spectrometry

All data were acquired using a Thermo Scientific™ Vanquish™ Neo UHPLC system coupled to an Orbitrap Astral mass spectrometer with a Thermo Scientific™ Easy-spray™ source. Data were acquired in data independent acquisition mode with a normalized collision

energy of 25% and a default charge state of 2. A 110 cm Thermo Scientific™ µPac™ Neo HPLC column was used for the quantitative experiments with a 24-minute gradient of from 4 to 45% B and a flow rate of 750 nL/min. For plasma experiments, a two step 30 or 60 minute gradient at 750 nL/min from 4 to 45% B was used on a 110 cm µPac Neo HPLC column. Mobile phase A was 0.1% formic acid in water, and mobile phase B was 80% acetonitrile and 0.1% formic acid in water. For Orbitrap experiments, MS1 spectra were acquired in the Orbitrap every 0.6 s at a resolution of 30,000, and MS/MS spectra were acquired in the Orbitrap at 15,000 resolving power with a maximum injection time of 23 ms. For the Orbitrap Astral MS experiments, MS1 spectra were acquired in the Orbitrap at a resolving power of 240,000 every 0.6 s, and MS/MS spectra were acquired in the Astral analyzer with varying injection times. In all MS/MS experiments, DIA precursor isolation windows spanned 380-980 Th unless noted otherwise. The MS1 mass range was the same as the MS/MS precursor range. All experiments besides chromatogram libraries were acquired in triplicate.

For the chromatogram library runs, the same parameters were used in both the Orbitrap and the Astral, with 6 gas phase fractions each spanning a 100-Da range with 4 Da isolation windows and a 23 ms maximum injection time. For the dilution curve, data were acquired from low concentration to high concentration, completing one replicate from each method before running a blank and starting the second replicate at the lowest concentration. Raw data files, EncyclopeDIA search results, and all Skyline documents have been deposited in the ProteomeXChange Consortium (identifier PXD042704) via Panorama Public (https://panoramaweb.org/AstralBenchmarking.url).

A reference dataset from Heil et al, 2022[69] acquired with the Orbitrap was used. Briefly, the dataset was acquired using the Orbitrap mass analyzer on a Thermo Scientific™ Orbitrap

Fusion Lumos™. Separation was performed with a 30 cm packed column attached to a Thermo Scientific™ EASY-nano™ LC across a two-step 90-minute gradient from 0 to 75% B at a flow rate of 300 nL/min. Mobile phase A was 0.1% formic acid in water, and mobile phase B was 80% acetonitrile and 0.1% formic acid in water. MS/MS spectra were acquired in the Orbitrap at 15k resolving power and a 22 ms maximum injection time using 8 Th isolation windows spanning a mass range from 400-1000 Th.

Dynamic data independent acquisition method development

A single injection on Astral was analyzed with EncyclopeDIA using the Pan-Human library[49] to create a preliminary list of peptides in the sample. Then, dynamic DIA boundaries were selected as described by Heil et al. (2022)[69] to maximize the number of peptides covered in a 300 Th mass range with either 2 or 4 Th isolation windows. In dynamic DIA, the 300 Th mass range that contains the highest peptide feature density is divided into equal isolation windows. This mass range is adjusted across time while the number and size of isolation windows remains the same (**Figure 3-2 C).**

Data analysis

Prior to searching, raw files were converted to mzMLs using Proteowizard MSConvert (from ProteoWizard release 3.0.22335).[125] For each experiment, a chromatogram library was generated by searching the 6-fraction chromatogram library runs against a Prosit[100] predicted spectral library in EncyclopeDIA (v 3.0.0).[50] Then, the quantitative DIA runs were searched against the appropriate chromatogram library in EncyclopeDIA with V2 scoring enabled and adjust inferred retention time boundaries set to true, and the quantitative results were loaded into Skyline (version 22.2.1.335)[123] for protein grouping and subsequent analysis. Orbitrap data were

searched and integrated with a 10 ppm mass tolerance, Astral data were searched with 10 ppm mass tolerance and integrated with 25 ppm mass tolerance to account for space charging at the apex of the peak. Conversion to mzML and importing into Skyline were done using the ProteoWizard Docker image available on Docker Hub at " proteowizard/pwiz-skyline-i-agree-to-the-vendor-licenses:3.0.22335-b595b19". Encyclopedia was run in a Docker image available Docker Hub at "mriffle/encyclopedia:3.0.0-rc.3".

For the HeLa dilution curve, all 27 quantitative runs for a given method (3x 9 dilution points) were searched together in EncyclopeDIA and integrated in Skyline without any normalization. A version of Skyline that can refine transitions to give the best lower limit of quantification. This feature will be available in future Skyline releases. Peak areas and lower limits of quantification were exported from Skyline for further analysis.

For plasma data, all six (3x total plasma and 3x enriched plasma) were searched together in EncyclopeDIA and imported into Skyline for quantification. TIC normalized peak areas were exported from Skyline and used for subsequent analysis.

Appendix C Supplementary methods for Chapter 4

Cell Culture

HeLa S3 cells were stable isotope labeled with SILAC using the Thermo Scientific SILAC Protein Quantitation Kit with DMEM (Catalog A33972, ThermoFisher Scientific) as recommended by the manufacturer. Briefly, cells were cultured in DMEM/10% FBS and then exchanged into DMEM (for SILAC) containing either $^{13}C_6$ $^{15}N_2$ L-lysine and $^{13}C_6$ $^{15}N_4$ L-arginine or unlabeled L-lysine and L-arginine and 10% dialyzed FBS in 80 cm^2 flasks. Media

was also supplemented with Penicillin-Streptomycin and GlutaMAX (Thermo Scientific). Cells were labeled for 8 cell doublings by maintaining log phase growth by exchanging labeled media and splitting as necessary. After washing with PBS, cells from each flask were harvested at ~85% confluency and pelleted by centrifugation at 1500 x g for 10 min. Cell pellets were stored at -80°C.

Sample Preparation

The cell lysates were prepared using an automated single pot solid phase sample preparation (SP3) protocol.[126] Briefly, cell pellets were lysed in 2% SDS with 10 seconds of probe sonication. Total protein concentration was estimated using a Pierce BCA assay (Thermo Fisher Scientific) and 1% SDS solution was added to bring the final concentration to approximately 4 μg/μL. Proteins were reduced in 20 mM dithiothreitol and alkylated in 40 mM iodoacetamide. Following reduction and alkylation, each sample was diluted to 70% acetonitrile and bound to MagResyn Hydroxyl particles (Resyn Biosciences) at a ratio of 100:1 (beads to protein) for a total of 10 minutes.

Subsequent washing and digestion steps were performed on a Kingfisher Flex (Thermo Fisher Scientific). The samples were washed three times in 95% acetonitrile and twice in ethanol. Then, samples were digested with trypsin in 50 mM ammonium bicarbonate at an enzyme to protein ratio of 1:25. The peptide samples were dried down via vacuum centrifugation and resuspended in 0.1% formic acid. A dilution curve was made by diluting the normal digest in the SILAC labeled digest with an equal amount of Pierce Peptide Retention Time Calibration standard (Thermo Fisher Scientific) spiked into each sample. A total of 11 dilution points were

made, consisting of the following fractions of normal digest: 0, 0.5, 1, 3, 5, 7, 10, 30, 50, 70, and 100%.

LC-MS/MS analysis

LC-MS/MS analysis was performed with a Thermo Easy-nanoLC coupled to a Thermo Orbitrap Lumos Tribrid mass spectrometer run in development mode. Peptide separation was performed with an in-house manufactured column consisting of 75 μm inner diameter fused silica capillary (TSP075375, Molex) pulled to an approximately 5 μm tip using a CO2 laser-based micropipette puller (Sutter Instruments). The capillary was packed to a final length of 30 cm with 3 μm ReproSil-Pur C18 beads (Dr. Maisch GmbH) using an in house-built pressure bomb with high pressure helium gas. A trap column was created by packing a 150 μm inner diameter fused silica capillary (TSP150375, Molex) fritted on one end with Kasil with approximately 3 cm of 3 μm ReproSil-Pur C18 beads (Dr. Maisch GmbH). Mobile phase A was 0.1% formic acid in water and mobile phase B was 0.1% formic acid in 80% acetonitrile. Peptide separation was performed via reversed-phase liquid chromatography over a 90 minute gradient from 0 to 40% B followed by a 5 minute ramp to 75% B and a 5 minute hold at 75% B.

Data for the chromatogram library were acquired on the instrument setup prior to quantitative runs. Tandem mass spectra were acquired in 6 gas-phase fractions with 4-*m/z* staggered DIA windows as described in detail by Pino et al. (2020).[72] After the library was acquired, a retention time alignment run was performed, using the dynamic DIA method described below without any retention time alignment. This run was preprocessed and used to align all future dynamic runs.

For quantitative analysis, two separate DIA methods were used: a traditional static method and a dynamic DIA method. The traditional static DIA method consisted of one MS1 spectrum followed by a set of MS2 spectra acquired at 15k resolving power with 8-*m/z* isolation windows covering a span of 400-1000 *m/z*. The dynamic DIA method began with the same MS1 spectrum covering the full 400-1000 *m/z* range. In addition, it included a set of alignment DIA spectra acquired in the linear ion trap with 20-*m/z* isolation windows. These spectra were used to align the next set of high-resolution spectra to a reference run that was acquired immediately after chromatogram library acquisition. This third set of spectra consisted of 8-*m/z* isolation windows at 30k resolving power, covering a variable *m/z* range of approximately 300 *m/z*. Based on the retention time alignment, the exact m/z positioning of these spectra was adjusted over time. The bounds of these spectra were optimized as described below.

The matrix-matched calibration curve was acquired by injecting each dilution point twice, acquiring once with the static DIA method and a second time with the dynamic method. The curve was run from low concentration to high concentration, switching the order of the DIA methods for each point. In total, three replicates of the curve were acquired for both DIA methods.

Raw data files, EncyclopeDIA search results, and all Skyline documents have been deposited in the ProteomeXChange Consortium (identifier PXD038508) via Panorama Public (https://panoramaweb.org/dynamicDIA.url).

Chromatogram library creation

A chromatogram library was created by analyzing the gas phase fractionated DIA data in EncyclopeDIA (v 1.4.10).[50] Each fraction was searched against a Prosit predicted library[100]

generated from the human Uniprot database. A detailed description of this process is provided by Pino et al. (2020).[72] The resulting chromatogram library was exported for downstream analyses.

Selection of dynamic DIA windows

To optimize the window boundaries for the dynamic DIA window, a maximization method was used. The code to do so is available on GitHub, and the process is described conceptually here. First, the window span was determined based on the instrument acquisition rate, desired isolation width, and desired points across the peak. At 30k resolving power, the instrument acquisition rate is 14.6 Hz. To acquire 10 points across LC peaks that are on average 25 seconds wide at the base requires a cycle time of 2500 msec; therefore, there is time for at most 2.5 s * 14.6 Hz = 36 acquisitions per cycle. Using 8-*m/z* isolation windows means that a 300-*m/z* span can be measured with these 36 acquisitions. Similarly, at 15k resolving power, the acquisition rate is 27 Hz, allowing for an optimal 2.5 s*27 Hz = 68 acquisitions per cycle. With 8-*m/z* static windows we can cover the 600-*m/z* span in 75 windows, meaning some peptides may have fewer points across the peak but we do not anticipate this impacting quantification. To create the dynamic DIA method, all of the features in the chromatogram library were binned by retention time and *m/z*. For each time bin, the *m/z* range that covers the most detectable features from the chromatogram library was selected. These window boundaries with scheduling were uploaded directly into the instrument method for the dynamic DIA setting.

Real time alignment

Real time retrospective retention time alignment was performed by the instrument as described by Remes et al. (2020).[31] Briefly, the retention time shift relative to a reference run is determined based on the average cross correlation of a set of high-speed DIA spectra acquired in

the linear ion trap in a reference run relative to the current run. The retention time shift is used to determine the active targets and acquire data for these targets. In this case, rather than specific scheduled peptide targets, scheduled DIA windows were used. The DIA spectra were listed as targeted MS2 spectra with specific retention time windows, thus the execution of this method is the same as described in the original publication.

Data processing

Data were searched with EncyclopeDIA (v 1.2.2)[50] using the protocol described by Pino et al. (2020).[72] All quantitative files were searched against the sample-specific chromatogram library and filtered for peptides at 1% FDR with at least three quantitative transitions to generate a new quantitative chromatogram library. Using the quantitative library for each method respectively, quantitation was performed in Skyline,[123] and the areas under the curve for each peptide were used for quantitation. Lower limits of detection and quantitation were determined using a bootstrapping method.[56] Further analysis of the results was performed in R. Analysis code is available on GitHub,https://github.com/uw-maccosslab/dynamic-DIA.

Appendix D Supplementary algorithms for Chapter 5

Appendix Table 2: Notation used for search algorithm equations

$allS$	set of MS2 spectra structured as a list of arrays, one for each isolation window
$allD$	database of all possible peptides (where each peptide is a specific peptide/charge state combination) structured as a list of arrays, each corresponding to one isolation window
D	database of all possible peptides within a single precursor isolation window
S	all MS2 spectra associated with a single precursor isolation window
P	all possible peptides associated with a single isolation window
m	number of peptides
n	number of MS2 spectra

p_i	a single peptide
s_j	a single MS2 spectrum
$M_{i,j}$	score assigned to the match between peptide p_i and spectrum s_j
A	vector of changepoints
$peak$	an m-dimensional structure array which stores the following information for each peptide: the maximally matching spectrum ($peak.m$), the corresponding score ($peak.s$), the width of the peak ($peak.w$), and its left and right boundaries ($peak.l$, $peak.r$)
$TDpairs$	a data structure containing target/decoy pairing information
$groups$	desired grouping designations for FDR control
$isTarget$	a logical vector indicating the target/decoy status of each peptide in the isolation window
Ip	indices of considered peptides
τ	acceptance threshold
$allGroups$	desired grouping aggregated across all precursor isolation windows
$allIsTarget$	target/decoy identification aggregated across all precursor isolation windows
$allScores$	assigned peptide scores aggregated across all precursor isolation windows
$allPs$	IDs of all considered peptides aggregated across all precursor isolation windows
$qvalue$	q-values of all considered target peptides aggregated across all precursor isolation windows

Algorithm 1: DIA analysis

procedure DIASEARCH($allS, allD, groups$) ▷ Proposed search procedure
  $allGroups[1 : length(groups)] = [\,]$
  $allIsTarget = [\,]$
  $allScores = [\,]$
  $allPs = [\,]$
  for $i_w = 1 :$ LENGTH($allS$) **do** ▷ loop on isolation windows
    $D_T := allD[i_w]$; $S := allS[i_w]$
    $[M, TDpairs, isTarget] :=$ DATABASESEARCHING(D_T, S)
    $D_{ext} = [D_T, D_T]$ ▷ D_{ext} includes precursor ID of targets and corresponding decoys
    $M :=$ CHANGEPOINTDETECTION(M) ▷ Remove leading and trailing "junk" spectra from M
    $peak :=$ CHROMATOGRAPHICPEAKDETECTION(M, $TDpairs$, $isTarget$) ▷ Detect chromatographic peaks
    $M :=$ PEPTIDESCORENORMALIZATION(M) ▷ Peptide-level score normalization
    $Ip :=$ TDC(M, $TDpairs$, $peak$, $isTarget$) ▷ keep the higher scoring peptide from each target-decoy pair
    $Ipt :=$ OPTIMALREPRESENTATIVESELECTION(M, Ip, $isTarget$, $peak$) ▷ Select optimal target representatives
    $Ipd :=$ OPTIMALREPRESENTATIVESELECTION(M, Ip, NOT($isTarget$), $peak$) ▷ Select optimal decoy representatives
    $Ip := [Ipt, Ipd]$
    $mAll := 0$ ▷ $mAll$ is the windows-aggregated number of peptides
    for i_p in Ip **do**
      if $i_p \neq$ WHICH.MAX($M[:, peak[i_p].m]$) **then**
        $Ip = Ip[-i_p]$ ▷ Retain only the maximum peptide per spectrum
      end if
      $mAll := mAll + 1$
      $i_g =$ WHICHGROUP($D_{ext}[i_p]$, $isTarget[i_p]$, $peak[i_p]$, $groups$) ▷ Determine the group of i_p based on its features
      $allGroups[i_g] = [allGroups[i_g], mAll]$
    end for
    $allPs = [allPs, D_{ext}[Ip]]$
    $allScores = [allScores, peak[Ip].s]$
    $allIsTarget = [allIsTarget, isTarget[Ip]]$
  end for
  for $i_g = 1 :$ LENGTH($groups$) **do**
    $Ip = allGroups[i_g]$
    $qvalues[Ip] =$ QVALUESVIATDC($allScores[Ip]$, $allIsTarget[Ip]$)
  end for
  return ($allPs[allIsTarget]$, $qvalue[allIsTarget]$, $allScores[allIsTarget]$)
end procedure
procedure QVALUESVIATDC($scores$, $isTarget$) ▷ Estimate FDR using target decoy competition
  $n :=$ LENGTH($scores$)
  $sortPerm :=$ ORDER($scores$)
  $scores := scores[sortPerm]$ ▷ sort scores in decreasing order
  $isTarget := isTarget[sortPerm]$
  $nTargetWins :=$ CUMSUM($isTarget$)
  $nDecoyWins := [1 : n] - nTargetWins$
  $estFDR := \min(1, (nDecoyWins + 1)/\max(1, nTargetWins))$
  $qvalues[n] := estFDR[n]$
  for $i = n - 1 : 1$ by -1 **do**
    $qvalues[i] := \min(estFDR[i], qvalues[i + 1])$
  end for
  return ($qvalues$[INVERSEPERMUTATION($sortPerm$)])
end procedure

Algorithm 2: Database searching

procedure DATABASESEARCHING($D_T = (p_i)_1^m,\ S = (s_j)_1^n$) ▷ Perform database search to obtain score matrix
  D_D = CREATEDECOYDB(D_T)
  for $1 \leq j \leq n$ **do**
    $M_T[i, :]$:= SCOREALLPEPTIDES(s_j, D_T) ▷ SCOREALLPEPTIDES returns the scores of the matches between the spectrum s_j and every peptide $p_i \in D$ (here we used Tailor-normalized XCorr)
    $M_D[i, :]$:= SCOREALLPEPTIDES(s_j, D_D)
  end for
  $isTarget[1 : m]$ = TRUE
  $isTarget[m + 1 : 2m]$ = FALSE
  $TDpairs[1 : m] = [m + 1 : 2m]$
  $TDpairs[m + 1 : 2m] = [1 : m]$
  return [M := CONCAT(M_T, M_D), $TDpairs$, $isTarget$] ▷ Return concatenated target/decoy scores
end procedure

Algorithm 3: Changepoint detection

procedure CHANGEPOINTDETECTION(M) ▷ Use changepoint detection to remove empty spectra
  n = NROWS(M)
  for $j = 1 : n$ **do**
    $meds[j]$:= MEDIAN($M[:, j]$) ▷ Determine median score for a given spectrum
  end for
  for $ncp = 2 : 4$ **do** ▷ ncp is total number of changepoints
    A := CHANGEPOINT($meds$, ncp) ▷ Inputs for changepoint function (described by Killick et al., 2016) are vector of scores and number of changepoints, l
    if $A_{ncp} - A_1 \geq 0.5 \times n$ **then**
      $M := M[:, A_1 : A_{ncp}]$ ▷ Filter score matrix for spectra between the calculated changepoints
      return M
    end if
  end for
  return `ERROR`
end procedure

Algorithm 4: Chromatographic peak detection

procedure CHROMATOGRAPHICPEAKDETECTION($M, TDpairs, isTarget$) ▷ Assign chromatographic peak boundaries to each peptide in the matrix
  $m = \text{NROWS}(M)$
  for $i = 1 : m$ **do**
    $\text{med} = \text{MEDIAN}(M[i, :])$
    $\text{MAD} = \text{MEDIAN}(|M[i, :] - \text{med}|)$
    $M^{RZ}[i, :] := (M[i, :] - \text{med})/MAD$ ▷ Determine normalize scores for a given peptide using the robust z-score
  end for
  for $i_t = 1 : m$ **do**
    if $isTarget[i_t]$ **then**
      $i_d = TDpairs[i_t]$
    else
      continue
    end if
    $[peak[i_t].m, l_t, r_t] := \text{FINDPEAKINROW}(M^{RZ}[i_t, :])$ ▷ Calculate peak boundaries for targets
    $[peak[i_d].m, l_d, r_d] := \text{FINDPEAKINROW}(M^{RZ}[i_d, :])$ ▷ Calculate peak boundaries for decoys
    if $l_t + r_t < l_d + r_d$ **then** ▷ For each target/decoy pair, use the wider of the two assigned peak widths
      $l = l_d;\ r = r_d$
    else
      $l = l_t;\ r = r_t$
    end if
    $peak[i_t].l = \max(peak[i_t].m - l,\ 1)$
    $peak[i_t].r = \min(peak[i_t].m + r,\ n)$
    $peak[i_t].w = l + r + 1$
    $peak[i_t].s = M[i_t,\ peak[i_t].m]$
    $peak[i_d].l = \max(peak[i_d].m - l,\ 1)$
    $peak[i_d].r = \min(peak[i_d].m + r,\ n)$
    $peak[i_d].w = l + r + 1$
    $peak[i_d].s = M[i_d,\ peak[i_d].m]$
  end for
  return $peak$ ▷ an m-dimensional structure array storing peak data for each peptide
end procedure
procedure FINDPEAKINROW(M_r) ▷ Find the consecutive spectra about the maximum score with a robust z-score greater than 75% of the maximum robust z-score
  $m = \text{WHICH.MAX}(M_r)$
  $M_r = M_r / M_r[m]$
  for $l = 0 : m - 1$ **do**
    if $M_r[m - l] < 0.75$ **then**
      $l = l - 1$
      break
    end if
  end for
  for $r = 0 : n - m$ **do**
    if $M_r[m + r] < 0.75$ **then**
      $r = r - 1$
      break
    end if
  end for
  return $[m,\ l,\ r]$
end procedure

Algorithm 5: Peptide score normalization

procedure PEPTIDESCORENORMALIZATION(M)
    $m = \text{NROWS}(M)$
    for $i = 1 : m$ **do**
        $q^{0.99} := \text{QUANTILE}(M[i, :], 0.99)$ ▷ Normalize scores for a given peptide by dividing by the 99th quantile score for that peptide
        $M[i, :] = M[i, :] \,/\, q^{0.99}$
    end for
    return M ▷ return the peptide normalized scores
end procedure

Algorithm 6: Target/decoy competition

procedure TDC(M, $TDpairs$, $peak$, $isTarget$) ▷ Perform target/decoy competition, keeping only the higher scoring of the pair
    $m = \text{NROWS}(M)$
    $Ip := \emptyset$
    for $i_t = 1 : m$ **do**
        if $isTarget[i_t]$ **then**
            $i_d = TDpairs[i_t]$
            if $peak.s[i_t] > peak.s[i_d]$ **then**
                $Ip = [Ip, i_t]$
            else
                $Ip = [Ip, i_d]$
            end if
        end if
    end for
    return Ip
end procedure

Algorithm 7: Selection of optimal representatives

procedure OPTIMALREPRESENTATIVESELECTION(M, Ip, $isTarget$, $peak$)
  $Ipt := Ip[isTarget]$
  $Ipd := Ip[\text{NOT}(isTarget)]$
  $Ipt := IIpt[\text{ORDER}(peak.s[Ipt])]$ ▷ Sort scores in decreasing order
  for i_p in Ipt **do**
    $pRow := M[i_p, :]$
    $overlaps = \{i \in Ipt : peak[i_p] \cap peak[i] \neq \emptyset \text{ AND } peak.s[i] < peak.s[i_p]\}$ ▷ Identify co-eluting peptides
    $acosNull = 0$
    $draws := \text{PERMUTE}(Ipd)$
    for $l = 1 : \min(1000, \text{LENGTH}(Ipd))$ **do** ▷ Calculate a null distribution of correlations
      $pRow2 = M[draws[l], :]$
      $acosNull := \max(acosNull, (|pRow \cdot pRow2|)/(\|pRow\|\|pRow2\|))$
    end for
    for i in $overlaps$ **do** ▷ Determine if which of the co-eluting peptides are significantly correlated
      $pRow2 = M[i, :]$
      $acosTest = (|pRow \cdot pRow2|)/(\|pRow\|\|pRow2\|)$
      if $acosTest > acosNull$ AND $[(pRow2 > 0) \cdot (pRow > 0)/n] > 0.25$ **then** ▷ Check if acos distribution and predicted ion overlap are significant
        $Ipt = Ipt[-which(Ipt == i)]$ ▷ Remove the lower scoring peptide from the final matrix
      end if
    end for
  end for
  return Ipt
end procedure

Appendix E Supplementary methods for Chapter 5

Data acquisition

Here, we analyze a dataset from Pino et al. 2020,[56] which is available through the ProteomeXChange Consortium[121] with the dataset identifier PXD014815 and Panorama Public[122] at https://panoramaweb.org/ matrix-matched_calcurves.url. A second dataset enriched for phosphorylated peptides was generated for this project. All original data have been deposited to PRIDE Archive (http://www.ebi.ac.uk/pride/ archive) with the dataset identifier PXD029733.

Phosphorylated peptide enriched dataset

The phosphorylated peptide dataset was acquired in house. HEK 293T cells were lysed by resuspending in lysis buffer (8 M Urea, 100 mM Tris pH 8.0, 100 mM NaCl) with Pierce protease and phosphatase inhibitors (Thermo Fisher Scientific). Lysates were incubated on ice for 30 minutes. Lysate protein concentration was measured with a Pierce BCA assay kit (Thermo Fisher Scientific). Lysates were reduced with 5 mM dithiothreitol (DTT) for 30 minutes at 55° C and alkylated with 15 mM iodoacetamide for 30 minutes in the dark, before being quenched with additional 5 mM DTT for 15 minutes.

Lysates were digested on KingFisher Flex robot (Thermo Fisher Scientific) according to R2-P1 protocol described in Leutert et al., 2019.[128] In brief, magnetic beads (MagReSyn Amine, ReSyn Biosciences) were combined with lysates in a 96-well plate at a 10:1 w/w ratio of bead to protein. Ethanol was added to bead/lysate mixture to a final concentration of 70%. Samples were then washed by transferring beads to a wash plate containing 80% ethanol. This wash was repeated two more times. Proteins were digested by transferring beads to plate containing trypsin (1:50) and ammonium bicarbonate pH 8.2. Samples were incubated for 6 hours at 37° C. Beads were then removed and the remaining solvent was evaporated using vacuum centrifugation. Samples were enriched for phosphorylated peptides as described in Leutert et al.[128]

In brief, peptides were enriched on KingFisher Flex using 25 μL of 5% Fe^{3+} NTA magnetic beads (Cube Biotech) per 100 μg of peptide. Samples were reconstituted in 80% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) and incubated with beads for 30 minutes before being washed three times with 50% ACN, 0.1% TFA. Phosphorylated peptides were then eluted from beads with 50% ACN, 2.5% NH_4OH. Phosphorylated peptides were then dried in a speed vacuum resuspended in 3% ACN 5% formic acid for LC-MS/MS analysis.

Data were acquired using on an Orbitrap Exploris 480 (Thermo Fisher Scientific) coupled to an Easy-nLC 1200 (Thermo Fisher Scientific). Peptides were separated on an in-house pulled 100 μm inner diameter C18 column packed with 30 cm of 3 μm beads (Dr. Maisch GmbH) using a 90-min gradient from 3 to 32% ACN with 0.1% formic acid. Mass spectrometry was performed as DIA gas-phase fractions with six fractions. Each fraction covered a full MS scan range of 100 m/z from 400-1000 in total (400-500, 500-600, 600-700, 700-800, 800-900, 900-1000). Full MS scans were acquired at 30,000 resolution with an AGC target of 100% and a maximum injection time of 100 ms. MS/MS spectra were acquired with staggered 4 m/z windows at 30,000 resolution, AGC target of 1000%, maximum inject time of 25 ms, and a HCD collision energy of 27%.

Yeast dataset

The *S. cerevisiae* dataset comes from Pino et al. 2020.[56] From this dataset, a chromatogram library generated from data collected in 6 gas-phase fractions was used along with quantitative DIA data on three 100% and three 50% yeast samples. Briefly, yeast strains BY4741 and S288C were cultured in both YEPD and ^{15}N minimal media and harvested at mid-log phase. The cells were lysed, digested, and cleaned up using MCX . Dilutions were achieved by diluting cultures from the YEPD media in cultures grown in the ^{15}N minimal media.

For LC-MS/MS analysis, peptides were separated via reversed-phase liquid chromatography over a 90-minute gradient. Tandem mass spectrometry data were acquired using DIA performed on a Thermo QExactive HF Orbitrap mass spectrometer monitoring a mass range of 388.43190–1012.70480 m/z using normalized collision energy of 27 with assumed charge state of +2. Gas-phase fractionated “narrow-window” DIA data were acquired on a sample of yeast grown in YEPD media with the following parameters: 4 m/z staggered DIA windows,

30,000 resolution, 55 ms maximum inject time, and 1e6 AGC target. For the quantitative data, 24 m/z overlapping DIA windows were acquired using the same resolution, maximum inject time, and AGC as the narrow-window DIA data.

Data processing

All raw files are preprocessed to demultiplex overlapping windows using ProteoWizard's MSConvert (v3.0.1908) using vendor peak picking, overlap-only demultiplexing with a mass error of 10 ppm, and the "SIM as spectra" option turned on.[125] All spectra are assigned possible charge states +1, +2, and +3. Database searches are performed in Crux Tide (v 3.2)[104] using a precursor tolerance of 1.007 m/z and Tailor calibration.[105] All searches include carbadomethylation of cysteine as a fixed modification and appropriate post-translational modifications or missed cleavages. All processing and analysis were performed in Python (v 3.8.2) and R (v 3.6.1).

Analysis of phosphorylated peptide-enriched dataset

For the phosphorylated peptide search, the human reference proteome FASTA database was downloaded from Uniprot (UP000005640, https://www.uniprot.org/) in July of 2019. Prior to database searching, a Tide index was constructed by performing an in silico digestion of the FASTA with trypsin. The index contained fully tryptic peptides with no missed cleavages and exactly one variable modification of 79.966331 on serine, threonine, or tyrosine. Using this index, a Tide search of the phosphorylated peptide enriched DIA data was performed in Crux as described in the previous section. The results from this Tide search were processed according to the procedure described in Section 5-3. No groups were used for FDR estimation due to the small number of high scoring peptides in possible groups.

For DIA-Umpire analysis, pseudo-spectrum extraction was performed on demultiplexed mzXML files with DIA-Umpire (V2.0)[37] using window-type MSX. A Tide search was performed on the resulting pseudo-spectra using the same index as the direct search with a precursor tolerance of 20 parts per million and a top-match of three. FDR estimation was performed using Crux Percolator[129] with all default settings. Peptide-level q-values from this search were used for further analysis.

Analysis of *S. cerevisiae* dataset

The FASTA database for *S. cerevisiae* strain ATCC 204508 reference proteome was downloaded off of Uniprot (UP000002311, https://www.uniprot.org/) in September of 2020. A Tide index was constructed from the S. cerevisiae proteome using default settings for tryptic peptides. A Tide search was performed on the gas-phase fractionated yeast dataset[56] with a top match of 10,000. All peptides with q-values below 1% were converted to SSL format and built into a spectral library using Bibliospec[130] in Skyline. The spectral library .blib file was converted to an EncyclopeDIA-compatible format in EncyclopeDIA for later use to search quantitative DIA data.

The predicted spectral library was made by generating a list of all fully tryptic unmodified peptides from *S. cerevisiae* in EncyclopeDIA and predicting their spectra using the Prosit pre-trained spectral prediction model accessed through the online server (https://www.proteomicsdb.org/prosit/).[100] A chromatogram library was generated from the narrow window DIA data using EncyclopeDIA (v 0.9.0)[50] with a normal target/decoy approach requiring at least three quantitative ions. All six gas-phase fractionated DIA runs were searched individually and combined in EncyclopeDIA to generate a chromatogram library .elib file.

Both the spectral library from our workflow and chromatogram library from the Prosit search were used to search quantitative wide-window DIA data in EncyclopeDIA.[50] All six quantitative DIA mass spectrometry runs were searched against each library and the results were combined using the "Save Quant Reports" option in EncyclopeDIA. Peptides accepted at 1% EncyclopeDIA FDR with at least five quantitative transitions were used for quantitative analysis. Area under the curve data calculated by EncyclopeDIA was analyzed in R to assess fold-change abundances.

References

(1) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227* (5258), 561–563. https://doi.org/10.1038/227561a0.

(2) Brown, T. A. Transcriptomes and Proteomes. In *Genomes. 2nd edition*; Wiley-Liss, 2002.

(3) Shendure, J.; Ji, H. Next-Generation DNA Sequencing. *Nat. Biotechnol.* **2008**, *26* (10), 1135–1145. https://doi.org/10.1038/nbt1486.

(4) Marguerat, S.; Schmidt, A.; Codlin, S.; Chen, W.; Aebersold, R.; Bähler, J. Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells. *Cell* **2012**, *151* (3), 671–683. https://doi.org/10.1016/j.cell.2012.09.019.

(5) Milo, R. What Is the Total Number of Protein Molecules per Cell Volume? A Call to Rethink Some Published Values. *BioEssays* **2013**, *35* (12), 1050–1055. https://doi.org/10.1002/bies.201300066.

(6) Zubarev, R. A. The Challenge of the Proteome Dynamic Range and Its Implications for In-Depth Proteomics. *PROTEOMICS* **2013**, *13* (5), 723–726. https://doi.org/10.1002/pmic.201200451.

(7) Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **2002**, *1* (11), 845–867. https://doi.org/10.1074/mcp.R200007-MCP200.

(8) MacCoss, M. J.; Alfaro, J. A.; Faivre, D. A.; Wu, C. C.; Wanunu, M.; Slavov, N. Sampling the Proteome by Emerging Single-Molecule and Mass Spectrometry Methods. *Nat. Methods* **2023**, *20* (3), 339–346. https://doi.org/10.1038/s41592-023-01802-5.

(9) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, *113* (4), 2343–2394. https://doi.org/10.1021/cr3003533.

(10) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, *422* (6928), 198–207. https://doi.org/10.1038/nature01511.

(11) Eng, J. K.; Mccormack, A. L.; Yates, J. R. *An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database*; 1994. https://pubs.acs.org/sharingguidelines.

(12) Hoffmann, E. de; Stroobant, V. *Mass Spectrometry: Principles and Applications*, 3rd ed.; John Wiley & Sons Ltd, 2007.

(13) Hoofnagle, A. N.; Becker, J. O.; Wener, M. H.; Heinecke, J. W. Quantification of Thyroglobulin, a Low-Abundance Serum Protein, by Immunoaffinity Peptide Enrichment and Tandem Mass Spectrometry. *Clin. Chem.* **2008**, *54* (11), 1796–1804. https://doi.org/10.1373/clinchem.2008.109652.

(14) Carr, S. A.; Abbatiello, S. E.; Ackermann, B. L.; Borchers, C.; Domon, B.; Deutsch, E. W.; Grant, R. P.; Hoofnagle, A. N.; Hüttenhain, R.; Koomen, J. M.; Liebler, D. C.; Liu, T.; MacLean, B.; Mani, D. R.; Mansfield, E.; Neubert, H.; Paulovich, A. G.; Reiter, L.; Vitek, O.; Aebersold, R.; Anderson, L.; Bethem, R.; Blonder, J.; Boja, E.; Botelho, J.; Boyne, M.; Bradshaw, R. A.; Burlingame, A. L.; Chan, D.; Keshishian, H.; Kuhn, E.; Kinsinger, C.; Lee, J. S. H.; Lee, S.-W.; Moritz, R.; Oses-Prieto, J.; Rifai, N.; Ritchie, J.; Rodriguez, H.; Srinivas, P. R.; Townsend, R. R.; Eyk, J. V.; Whiteley, G.; Wiita, A.; Weintraub, S. Targeted Peptide Measurements in Biology and Medicine: Best Practices

for Mass Spectrometry-Based Assay Development Using a Fit-for-Purpose Approach *. *Mol. Cell. Proteomics* **2014**, *13* (3), 907–917. https://doi.org/10.1074/mcp.M113.036095.

(15) Li, J.; Cai, Z.; Bomgarden, R. D.; Pike, I.; Kuhn, K.; Rogers, J. C.; Roberts, T. M.; Gygi, S. P.; Paulo, J. A. TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *J. Proteome Res.* **2021**, *20* (5), 2964–2972. https://doi.org/10.1021/acs.jproteome.1c00168.

(16) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1626–1632. https://doi.org/10.1007/s13361-015-1182-2.

(17) Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D. Data-Controlled Automation of Liquid Chromatography/Tandem Mass Spectrometry Analysis of Peptide Mixtures. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (6), 532–540. https://doi.org/10.1016/1044-0305(96)00057-8.

(18) 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. *Nat. Methods* **2004**, *1* (1), 39–45. https://doi.org/10.1038/nmeth705.

(19) Kondrat, R. W.; McClusky, G. A.; Cooks, R. G. Multiple Reaction Monitoring in Mass Spectrometry/Mass Spectrometry for Direct Analysis of Complex Mixtures. *Anal. Chem.* **1978**, *50* (14), 2017–2021. https://doi.org/10.1021/ac50036a020.

(20) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics MCP* **2012**, *11* (11), 1475–1488. https://doi.org/10.1074/mcp.O112.020131.

(21) Chelius, D.; Bondarenko, P. V. Quantitative Profiling of Proteins in Complex Mixtures Using Liquid Chromatography and Mass Spectrometry. *J. Proteome Res.* **2002**, *1* (4), 317–323. https://doi.org/10.1021/pr025517j.

(22) Jaitly, N.; Monroe, M. E.; Petyuk, V. A.; Clauss, T. R. W.; Adkins, J. N.; Smith, R. D. Robust Algorithm for Alignment of Liquid Chromatography−Mass Spectrometry Analyses in an Accurate Mass and Time Tag Data Analysis Pipeline. *Anal. Chem.* **2006**, *78* (21), 7397–7409. https://doi.org/10.1021/ac052197p.

(23) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904. https://doi.org/10.1021/ac0262560.

(24) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 Eliminates Ratio Distortion in Isobaric Multiplexed Quantitative Proteomics. *Nat. Methods* **2011**, *8* (11), 937–940. https://doi.org/10.1038/nmeth.1714.

(25) Liu, M.; Dongre, A. Proper Imputation of Missing Values in Proteomics Datasets for Differential Expression Analysis. *Brief. Bioinform.* **2021**, *22* (3), bbaa112. https://doi.org/10.1093/bib/bbaa112.

(26) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805. https://doi.org/10.1021/pr101065j.

(27) Schweppe, D. K.; Eng, J. K.; Yu, Q.; Bailey, D.; Rad, R.; Navarrete-Perea, J.; Huttlin, E. L.; Erickson, B. K.; Paulo, J. A.; Gygi, S. P. Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J. Proteome Res.* **2020**, *19* (5), 2026–2034. https://doi.org/10.1021/acs.jproteome.9b00860.

(28) Yu, Q.; Liu, X.; Keller, M. P.; Navarrete-Perea, J.; Zhang, T.; Fu, S.; Vaites, L. P.; Shuken, S. R.; Schmid, E.; Keele, G. R.; Li, J.; Huttlin, E. L.; Rashan, E. H.; Simcox, J.; Churchill, G. A.; Schweppe, D. K.; Attie, A. D.; Paulo, J. A.; Gygi, S. P. Sample Multiplexing-Based Targeted Pathway Proteomics with Real-Time Analytics Reveals the Impact of Genetic Variation on Protein Expression. *Nat. Commun.* **2023**, *14* (1), 555. https://doi.org/10.1038/s41467-023-36269-7.

(29) Gallien, S.; Kim, S. Y.; Domon, B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Mol. Cell. Proteomics MCP* **2015**, *14* (6), 1630–1644. https://doi.org/10.1074/mcp.O114.043968.

(30) Rougemont, B.; Bontemps Gallo, S.; Ayciriex, S.; Carrière, R.; Hondermarck, H.; Lacroix, J. M.; Le Blanc, J. C. Y.; Lemoine, J. Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by Dickeya Dadantii Proteomic Analysis during Plant Infection. *Anal. Chem.* **2017**, *89* (3), 1421–1426. https://doi.org/10.1021/acs.analchem.6b03201.

(31) Remes, P. M.; Yip, P.; MacCoss, M. J. Highly Multiplex Targeted Proteomics Enabled by Real-Time Chromatographic Alignment. *Anal. Chem.* **2020**, *92* (17), 11809–11817. https://doi.org/10.1021/acs.analchem.0c02075.

(32) Zhu, H.; Ficarro, S. B.; Alexander, W. M.; Fleming, L. E.; Adelmant, G.; Zhang, T.; Willetts, M.; Decker, J.; Brehmer, S.; Krause, M.; East, M. P.; Gray, N. S.; Johnson, G. L.; Kruppa, G.; Marto, J. A. PRM-LIVE with Trapped Ion Mobility Spectrometry and Its Application in Selectivity Profiling of Kinase Inhibitors. *Anal. Chem.* **2021**, *93* (41), 13791–13799. https://doi.org/10.1021/acs.analchem.1c02349.

(33) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell. Proteomics MCP* **2015**, *14* (9), 2301–2307. https://doi.org/10.1074/mcp.O114.047035.

(34) Moosa, J. M.; Guan, S.; Moran, M. F.; Ma, B. Repeat-Preserving Decoy Database for False Discovery Rate Estimation in Peptide Identification. *J. Proteome Res.* **2020**, *19* (3), 1029–1036. https://doi.org/10.1021/acs.jproteome.9b00555.

(35) Egertson, J. D.; Kuehn, A.; Merrihew, G. E.; Bateman, N. W.; MacLean, B. X.; Ting, Y. S.; Canterbury, J. D.; Marsh, D. M.; Kellmann, M.; Zabrouskov, V.; Wu, C. C.; MacCoss, M. J. Multiplexed MS/MS for Improved Data-Independent Acquisition. *Nat. Methods* **2013**, *10* (8), 744–746. https://doi.org/10.1038/nmeth.2528.

(36) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-Independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis*. *Mol. Cell. Proteomics* **2012**, *11* (6), O111.016717. https://doi.org/10.1074/mcp.O111.016717.

(37) Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nat Methods* **2015**, *12* (3), 258–264. https://doi.org/10.1038/nmeth.3255.

(38) Peckner, R.; Myers, S. A.; Jacome, A. S. V.; Egertson, J. D.; Abelin, J. G.; MacCoss, M. J.; Carr, S. A.; Jaffe, J. D. Specter: Linear Deconvolution for Targeted Analysis of Data-Independent Acquisition Mass Spectrometry Proteomics. *Nat. Methods* **2018**, *15* (5), 371–378. https://doi.org/10.1038/nmeth.4643.

(39) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17* (1), 41–44. https://doi.org/10.1038/s41592-019-0638-x.

(40) Yost, R. A.; Enke, C. G. Selected Ion Fragmentation with a Tandem Quadrupole Mass Spectrometer. *J. Am. Chem. Soc.* **1978**, *100* (7), 2274–2275. https://doi.org/10.1021/ja00475a072.

(41) Picotti, P.; Aebersold, R. Selected Reaction Monitoring–Based Proteomics: Workflows, Potential, Pitfalls and Future Directions. *Nat. Methods* **2012**, *9* (6), 555–566. https://doi.org/10.1038/nmeth.2015.

(42) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222. https://doi.org/10.1038/msb.2008.61.

(43) Sherrod, S. D.; Myers, M. V.; Li, M.; Myers, J. S.; Carpenter, K. L.; MacLean, B.; MacCoss, M. J.; Liebler, D. C.; Ham, A.-J. L. Label-Free Quantitation of Protein

Modifications by Pseudo Selected Reaction Monitoring with Internal Reference Peptides. *J. Proteome Res.* **2012**, *11* (6), 3467–3479. https://doi.org/10.1021/pr201240a.

(44) Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics MCP* **2012**, *11* (12), 1709–1723. https://doi.org/10.1074/mcp.O112.019802.

(45) Schilling, B.; MacLean, B.; Held, J. M.; Sahu, A. K.; Rardin, M. J.; Sorensen, D. J.; Peters, T.; Wolfe, A. J.; Hunter, C. L.; MacCoss, M. J.; Gibson, B. W. Multiplexed, Scheduled, High-Resolution Parallel Reaction Monitoring on a Full Scan QqTOF Instrument with Integrated Data-Dependent and Targeted Mass Spectrometric Workflows. *Anal. Chem.* **2015**, *87* (20), 10222–10229. https://doi.org/10.1021/acs.analchem.5b02983.

(46) Gallien, S.; Bourmaud, A.; Kim, S. Y.; Domon, B. Technical Considerations for Large-Scale Parallel Reaction Monitoring Analysis. *J. Proteomics* **2014**, *100*, 147–159. https://doi.org/10.1016/j.jprot.2013.10.029.

(47) Gallien, S.; Duriez, E.; Demeure, K.; Domon, B. Selectivity of LC-MS/MS Analysis: Implication for Proteomics Experiments. *J. Proteomics* **2013**, *81*, 148–158. https://doi.org/10.1016/j.jprot.2012.11.005.

(48) Ignjatovic, V.; Geyer, P. E.; Palaniappan, K. K.; Chaaban, J. E.; Omenn, G. S.; Baker, M. S.; Deutsch, E. W.; Schwenk, J. M. Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res.* **2019**, *18* (12), 4085–4097. https://doi.org/10.1021/acs.jproteome.9b00503.

(49) Rosenberger, G.; Koh, C. C.; Guo, T.; Röst, H. L.; Kouvonen, P.; Collins, B. C.; Heusel, M.; Liu, Y.; Caron, E.; Vichalkovski, A.; Faini, M.; Schubert, O. T.; Faridi, P.; Ebhardt, H. A.; Matondo, M.; Lam, H.; Bader, S. L.; Campbell, D. S.; Deutsch, E. W.; Moritz, R. L.; Tate, S.; Aebersold, R. A Repository of Assays to Quantify 10,000 Human Proteins by SWATH-MS. *Sci. Data* **2014**, *1* (1), 140031. https://doi.org/10.1038/sdata.2014.31.

(50) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and

Quantification by Data Independent Acquisition Mass Spectrometry. *Nat. Commun.* **2018**, *9* (1), 5128. https://doi.org/10.1038/s41467-018-07454-w.

(51) Bollinger, J. G.; Stergachis, A. B.; Johnson, R. S.; Egertson, J. D.; MacCoss, M. J. Selecting Optimal Peptides for Targeted Proteomic Experiments in Human Plasma Using in Vitro Synthesized Proteins as Analytical Standards. *Methods Mol. Biol. Clifton NJ* **2016**, *1410*, 207–221. https://doi.org/10.1007/978-1-4939-3524-6_12.

(52) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. A Two-Dimensional Quadrupole Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (6), 659–669. https://doi.org/10.1021/jasms.8b01797.

(53) MacCoss, M. J.; Toth, M. J.; Matthews, D. E. Evaluation and Optimization of Ion-Current Ratio Measurements by Selected-Ion-Monitoring Mass Spectrometry. *Anal. Chem.* **2001**, *73* (13), 2976–2984. https://doi.org/10.1021/ac010041t.

(54) Peterson, D. W.; Hayes, J. M. Signal-to-Noise Ratios in Mass Spectroscopic Ion-Current-Measurement Systems. In *Contemporary Topics in Analytical and Clinical Chemistry: Volume 3*; Hercules, D. M., Hieftje, G. M., Snyder, L. R., Evenson, M. A., Eds.; Springer US: Boston, MA, 1978; pp 217–252. https://doi.org/10.1007/978-1-4615-6734-9_5.

(55) Blackler, A. R.; Klammer, A. A.; MacCoss, M. J.; Wu, C. C. Quantitative Comparison of Proteomic Data Quality between a 2D and 3D Quadrupole Ion Trap. *Anal. Chem.* **2006**, *78* (4), 1337–1344. https://doi.org/10.1021/ac051486a.

(56) Pino, L. K.; Searle, B. C.; Yang, H.-Y.; Hoofnagle, A. N.; Noble, W. S.; MacCoss, M. J. Matrix-Matched Calibration Curves for Assessing Analytical Figures of Merit in Quantitative Proteomics. *J. Proteome Res.* **2020**, *19* (3), 1147–1153. https://doi.org/10.1021/acs.jproteome.9b00666.

(57) Grant, R. P.; Hoofnagle, A. N. From Lost in Translation to Paradise Found: Enabling Protein Biomarker Method Transfer Using Mass Spectrometry. *Clin. Chem.* **2014**, *60* (7), 941–944. https://doi.org/10.1373/clinchem.2014.224840.

(58) *IUPAC Compendium of Chemical Terminology: Gold Book*, 2.1.0.; Nič, M., Jirát, J., Košata, B., Jenkins, A., McNaught, A., Eds.; IUPAC: Research Triagle Park, NC, 2009. https://doi.org/10.1351/goldbook.

(59) Makarov, A.; Denisov, E. Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1486–1495. https://doi.org/10.1016/j.jasms.2009.03.024.

(60) Slavov, N. Single-Cell Protein Analysis by Mass-Spectrometry. *Curr. Opin. Chem. Biol.* **2021**, *60*, 1–9. https://doi.org/10.1016/j.cbpa.2020.04.018.

(61) Brunner, A.-D.; Thielert, M.; Vasilopoulou, C. G.; Ammar, C.; Coscia, F.; Mund, A.; Hoerning, O. B.; Bache, N.; Apalategui, A.; Lubeck, M.; Richter, S.; Fischer, D. S.; Raether, O.; Park, M. A.; Meier, F.; Theis, F. J.; Mann, M. *Ultra-High Sensitivity Mass Spectrometry Quantifies Single-Cell Proteome Changes upon Perturbation*; preprint; Systems Biology, 2020. https://doi.org/10.1101/2020.12.22.423933.

(62) Schoof, E. M.; Furtwängler, B.; Üresin, N.; Rapin, N.; Savickas, S.; Gentil, C.; Lechman, E.; Keller, U. auf dem; Dick, J. E.; Porse, B. T. Quantitative Single-Cell Proteomics as a Tool to Characterize Cellular Hierarchies. *Nat. Commun.* **2021**, *12* (1), 3341. https://doi.org/10.1038/s41467-021-23667-y.

(63) Kustatscher, G.; Collins, T.; Gingras, A.-C.; Guo, T.; Hermjakob, H.; Ideker, T.; Lilley, K. S.; Lundberg, E.; Marcotte, E. M.; Ralser, M.; Rappsilber, J. Understudied Proteins: Opportunities and Challenges for Functional Proteomics. *Nat. Methods* **2022**, *19* (7), 774–779. https://doi.org/10.1038/s41592-022-01454-x.

(64) Skowronek, P.; Thielert, M.; Voytik, E.; Tanzer, M. C.; Hansen, F. M.; Willems, S.; Karayel, O.; Brunner, A.-D.; Meier, F.; Mann, M. Rapid and In-Depth Coverage of the (Phospho-)Proteome With Deep Libraries and Optimal Window Design for Dia-PASEF. *Mol. Cell. Proteomics MCP* **2022**, *21* (9), 100279. https://doi.org/10.1016/j.mcpro.2022.100279.

(65) Kawashima, Y.; Nagai, H.; Konno, R.; Ishikawa, M.; Nakajima, D.; Sato, H.; Nakamura, R.; Furuyashiki, T.; Ohara, O. Single-Shot 10K Proteome Approach: Over 10,000 Protein

Identifications by Data-Independent Acquisition-Based Single-Shot Proteomics with Ion Mobility Spectrometry. *J. Proteome Res.* **2022**, *21* (6), 1418–1427. https://doi.org/10.1021/acs.jproteome.2c00023.

(66) Bekker-Jensen, D. B.; Kelstrup, C. D.; Batth, T. S.; Larsen, S. C.; Haldrup, C.; Bramsen, J. B.; Sørensen, K. D.; Høyer, S.; Ørntoft, T. F.; Andersen, C. L.; Nielsen, M. L.; Olsen, J. V. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **2017**, *4* (6), 587-599.e4. https://doi.org/10.1016/j.cels.2017.05.009.

(67) Stewart, H.; Grinfeld, D.; Giannakopulos, A.; Petzoldt, J.; Shanley, T.; Garland, M.; Denisov, E.; Peterson, A.; Damoc, E.; Zeller, M.; Arrey, T. N.; Pashkova, A.; Renuse, S.; Hakimi, A.; Kühn, A.; Biel, M.; Kreutzmann, A.; Hagedorn, B.; Colonius, I.; Schütz, A.; Stefes, A.; Dwivedi, A.; Mourad, D.; Hoek, M.; Reitemeier, B.; Cochems, P.; Kholomeev, A.; Ostermann, R.; Quiring, G.; Ochmann, M.; Möhring, S.; Wagner, A.; Petker, A.; Kanngiesser, S.; Wiedemeyer, M.; Balschun, W.; Hermanson, D.; Zabrouskov, V.; Makarov, A.; Hock, C. Parallelized Acquisition of Orbitrap and Astral Analyzers Enables High-Throughput Quantitative Analysis. bioRxiv June 5, 2023, p 2023.06.02.543408. https://doi.org/10.1101/2023.06.02.543408.

(68) Schwartz, J. C.; Zhou, X.-G.; Bier, M. E. Method and Apparatus of Increasing Dynamic Range and Sensitivity of a Mass Spectrometer. US5572022A, November 5, 1996. https://patents.google.com/patent/US5572022A/en (accessed 2023-03-20).

(69) Heil, L. R.; Remes, P. M.; Canterbury, J. D.; Yip, P.; Barshop, W. D.; Wu, C. C.; MacCoss, M. J. Dynamic Data Independent Acquisition Mass Spectrometry with Real-Time Retrospective Alignment. bioRxiv December 2, 2022, p 2022.11.29.518428. https://doi.org/10.1101/2022.11.29.518428.

(70) Wu, C. C.; Tsantilas, K. A.; Park, J.; Plubell, D. L.; Naicker, P.; Govender, I.; Buthelezi, S.; Stoychev, S.; Jordaan, J.; Merrihew, G. E.; Huang, E.; Parker, E. D.; Riffle, M.; Hoofnagle, A. N.; MacCoss, M. J. Mag-Net: Rapid Enrichment of Membrane-Bound Particles Enables High Coverage Quantitative Analysis of the Plasma Proteome. bioRxiv June 11, 2023, p 2023.06.10.544439. https://doi.org/10.1101/2023.06.10.544439.

(71) Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R. Precursor Acquisition Independent from Ion Count: How to Dive Deeper into the Proteomics Ocean. *Anal Chem* **2009**, *81* (15), 6481–6488. https://doi.org/10.1021/ac900888s.

(72) Pino, L. K.; Just, S. C.; MacCoss, M. J.; Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **2020**, *19* (7), 1088–1103. https://doi.org/10.1074/mcp.P119.001913.

(73) Heil, L. R.; Remes, P. M.; MacCoss, M. J. Comparison of Unit Resolution Versus High-Resolution Accurate Mass for Parallel Reaction Monitoring. *J. Proteome Res.* **2021**, *20* (9), 4435–4442. https://doi.org/10.1021/acs.jproteome.1c00377.

(74) Wells, G.; Prest, H.; IV, C. W. R. *Signal, Noise, and Detection Limits in Mass Spectrometry*; Application Note DE16430639; Agilent, 2023.

(75) Meier, F.; Brunner, A.-D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; Aebersold, R.; Collins, B. C.; Röst, H. L.; Mann, M. DiaPASEF: Parallel Accumulation–Serial Fragmentation Combined with Data-Independent Acquisition. *Nat. Methods* **2020**, *17* (12), 1229–1236. https://doi.org/10.1038/s41592-020-00998-0.

(76) Skowronek, P.; Krohs, F.; Lubeck, M.; Wallmann, G.; Itang, E. C. M.; Koval, P.; Wahle, M.; Thielert, M.; Meier, F.; Willems, S.; Raether, O.; Mann, M. Synchro-PASEF Allows Precursor-Specific Fragment Ion Extraction and Interference Removal in Data-Independent Acquisition. *Mol. Cell. Proteomics MCP* **2022**, *22* (2), 100489. https://doi.org/10.1016/j.mcpro.2022.100489.

(77) Donovan, M. K. R.; Huang, Y.; Blume, J. E.; Wang, J.; Hornburg, D.; Ferdosi, S.; Mohtashemi, I.; Kim, S.; Ko, M.; Benz, R. W.; Platt, T. L.; Batzoglou, S.; Diaz, L. A.; Farokhzad, O. C.; Siddiqui, A. Functionally Distinct BMP1 Isoforms Show an Opposite Pattern of Abundance in Plasma from Non-Small Cell Lung Cancer Subjects and Controls. bioRxiv January 8, 2023, p 2022.01.07.475393. https://doi.org/10.1101/2022.01.07.475393.

(78) Tognetti, M.; Sklodowski, K.; Müller, S.; Kamber, D.; Muntel, J.; Bruderer, R.; Reiter, L. Biomarker Candidates for Tumors Identified from Deep-Profiled Plasma Stem Predominantly from the Low Abundant Area. *J. Proteome Res.* **2022**, *21* (7), 1718–1735. https://doi.org/10.1021/acs.jproteome.2c00122.

(79) Woo, H.-K.; Cho, Y. K.; Lee, C. Y.; Lee, H.; Castro, C. M.; Lee, H. Characterization and Modulation of Surface Charges to Enhance Extracellular Vesicle Isolation in Plasma. *Theranostics* **2022**, *12* (5), 1988–1998. https://doi.org/10.7150/thno.69094.

(80) Xie, F.; Smith, R. D.; Shen, Y. Advanced Proteomic Liquid Chromatography. *J. Chromatogr. A* **2012**, *1261*, 78–90. https://doi.org/10.1016/j.chroma.2012.06.098.

(81) Mant, C. T.; Burke, T. W. L.; Black, J. A.; Hodges, R. S. Effect of Peptide Chain Length on Peptide Retention Behaviour in Reversed-Phase Chromatogrphy. *J. Chromatogr. A* **1988**, *458*, 193–205. https://doi.org/10.1016/S0021-9673(00)90564-8.

(82) Bern, M.; Finney, G.; Hoopmann, M. R.; Merrihew, G.; Toth, M. J.; MacCoss, M. J. Deconvolution of Mixture Spectra from Ion-Trap Data-Independent-Acquisition Tandem Mass Spectrometry. *Anal. Chem.* **2010**, *82* (3), 833. https://doi.org/10.1021/ac901801b.

(83) Li, W.; Chi, H.; Salovska, B.; Wu, C.; Sun, L.; Rosenberger, G.; Liu, Y. Assessing the Relationship Between Mass Window Width and Retention Time Scheduling on Protein Coverage for Data-Independent Acquisition. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (8), 1396–1405. https://doi.org/10.1007/s13361-019-02243-1.

(84) Borràs, E.; Pastor, O.; Sabidó, E. Use of Linear Ion Traps in Data-Independent Acquisition Methods Benefits Low-Input Proteomics. *Anal. Chem.* **2021**, *93* (34), 11649–11653. https://doi.org/10.1021/acs.analchem.1c01885.

(85) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical Advances in Proteomics: New Developments in Data-Independent Acquisition. *F1000Research* **2016**, *5*, F1000 Faculty Rev-419. https://doi.org/10.12688/f1000research.7042.1.

(86) Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Müller, M.; Lisacek, F. Processing Strategies and Software Solutions for Data-Independent

Acquisition in Mass Spectrometry. *Proteomics* **2015**, *15* (5–6), 964–980. https://doi.org/10.1002/pmic.201400323.

(87) Ting, Y. S.; Egertson, J. D.; Bollinger, J. G.; Searle, B. C.; Payne, S. H.; Noble, W. S.; MacCoss, M. J. PECAN: Library-Free Peptide Detection for Data-Independent Acquisition Tandem Mass Spectrometry Data. *Nat Methods* **2017**, *14* (9), 903–908. https://doi.org/10.1038/nmeth.4390.

(88) Silva, J. C.; Denny, R.; Dorschel, C. A.; Gorenstein, M.; Kass, I. J.; Li, G.-Z.; McKenna, T.; Nold, M. J.; Richardson, K.; Young, P.; Geromanos, S. Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. *Anal. Chem.* **2005**, *77* (7), 2187–2200. https://doi.org/10.1021/ac048455k.

(89) Purvine, S.; Eppel*, J.-T.; Yi, E. C.; Goodlett, D. R. Shotgun Collision-Induced Dissociation of Peptides Using a Time of Flight Mass Analyzer. *PROTEOMICS* **2003**, *3* (6), 847–850. https://doi.org/10.1002/pmic.200300362.

(90) Li, Y.; Zhong, C.-Q.; Xu, X.; Cai, S.; Wu, X.; Zhang, Y.; Chen, J.; Shi, J.; Lin, S.; Han, J. Group-DIA: Analyzing Multiple Data-Independent Acquisition Mass Spectrometry Data Files. *Nat. Methods* **2015**, *12* (12), 1105–1106. https://doi.org/10.1038/nmeth.3593.

(91) Schwartz, J. C.; Kovtoun, V. V. Automatic Gain Control (AGC) Method for an Ion Trap and a Temporally Non-Uniform Ion Beam. US7960690B2, June 14, 2011. https://patents.google.com/patent/US7960690B2/en (accessed 2023-06-01).

(92) Amodei, D.; Egertson, J.; MacLean, B. X.; Johnson, R.; Merrihew, G. E.; Keller, A.; Marsh, D.; Vitek, O.; Mallick, P.; MacCoss, M. J. Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (4), 669–684. https://doi.org/10.1007/s13361-018-2122-8.

(93) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat. Biotechnol.* **2014**, *32* (3), 219–223. https://doi.org/10.1038/nbt.2841.

(94) Keller, A.; Bader, S. L.; Shteynberg, D.; Hood, L.; Moritz, R. L. Automated Validation of Results and Removal of Fragment Ion Interferences in Targeted Analysis of Data-Independent Acquisition Mass Spectrometry (MS) Using SWATHProphet. *Mol. Cell. Proteomics MCP* **2015**, *14* (5), 1411–1418. https://doi.org/10.1074/mcp.O114.044917.

(95) Bruderer, R.; Bernhardt, O. M.; Gandhi, T.; Miladinović, S. M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; Vitek, O.; Rinner, O.; Reiter, L. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteomics MCP* **2015**, *14* (5), 1400–1410. https://doi.org/10.1074/mcp.M114.044305.

(96) Searle, B. C.; Swearingen, K. E.; Barnes, C. A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating High Quality Libraries for DIA MS with Empirically Corrected Peptide Predictions. *Nat. Commun.* **2020**, *11* (1), 1548. https://doi.org/10.1038/s41467-020-15346-1.

(97) Midha, M. K.; Campbell, D. S.; Kapil, C.; Kusebauch, U.; Hoopmann, M. R.; Bader, S. L.; Moritz, R. L. DIALib-QC an Assessment Tool for Spectral Libraries in Data-Independent Acquisition Proteomics. *Nat. Commun.* **2020**, *11* (1), 5251. https://doi.org/10.1038/s41467-020-18901-y.

(98) Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B. C.; Aebersold, R. Data-Independent Acquisition-Based SWATH-MS for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2018**, *14* (8), e8126. https://doi.org/10.15252/msb.20178126.

(99) Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-Based Data-Independent Acquisition. *Mol. Cell. Proteomics MCP* **2020**, *19* (1), 181–197. https://doi.org/10.1074/mcp.RA119.001714.

(100) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; Reimer, U.; Ehrlich, H.-C.; Aiche, S.; Kuster, B.; Wilhelm, M. Prosit: Proteome-Wide Prediction of Peptide Tandem Mass

Spectra by Deep Learning. *Nat Methods* **2019**, *16* (6), 509–518. https://doi.org/10.1038/s41592-019-0426-7.

(101) Yang, Y.; Liu, X.; Shen, C.; Lin, Y.; Yang, P.; Qiao, L. In Silico Spectral Libraries by Deep Learning Facilitate Data-Independent Acquisition Proteomics. *Nat Commun* **2020**, *11* (1), 146. https://doi.org/10.1038/s41467-019-13866-z.

(102) Van Puyvelde, B.; Willems, S.; Gabriels, R.; Daled, S.; De Clerck, L.; Vande Casteele, S.; Staes, A.; Impens, F.; Deforce, D.; Martens, L.; Degroeve, S.; Dhaenens, M. Removing the Hidden Data Dependency of DIA with Predicted Spectral Libraries. *Proteomics* **2020**, *20* (3–4), 1900306. https://doi.org/10.1002/pmic.201900306.

(103) Noble, W. S. Mass Spectrometrists Should Search Only for Peptides They Care About. *Nat. Methods* **2015**, *12* (7), 605–608. https://doi.org/10.1038/nmeth.3450.

(104) Diament, B. J.; Noble, W. S. Faster SEQUEST Searching for Peptide Identification from Tandem Mass Spectra. *J Proteome Res* **2011**, *10* (9), 3871–3879. https://doi.org/10.1021/pr101196n.

(105) Sulimov, P.; Kertész-Farkas, A. Tailor: A Nonparametric and Rapid Score Calibration Method for Database Search-Based Peptide Identification in Shotgun Proteomics. *J. Proteome Res.* **2020**, *19* (4), 1481–1490. https://doi.org/10.1021/acs.jproteome.9b00736.

(106) Killick, R.; Eckley, I. A. Changepoint: An R Package for Changepoint Analysis. *J. Stat. Softw.* **2014**, *58*, 1–19. https://doi.org/10.18637/jss.v058.i03.

(107) Scott, A. J.; Knott, M. A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics* **1974**, *30* (3), 507. https://doi.org/10.2307/2529204.

(108) Eng, J. K.; Hoopmann, M. R.; Jahan, T. A.; Egertson, J. D.; Noble, W. S.; MacCoss, M. J. A Deeper Look into Comet – Implementation and Features. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (11), 1865–1874. https://doi.org/10.1007/s13361-015-1179-x.

(109) Martiánez-Bartolomé, S.; Navarro, P.; Martián-Maroto, F.; López-Ferrer, D.; Ramos-Fernández, A.; Villar, M.; Garciáa-Ruiz, J. P.; Vázquez, J. Properties of Average Score Distributions of SEQUEST: The Probability Ratio Method. *Mol. Cell. Proteomics* **2008**, *7* (6), 1135–1145. https://doi.org/10.1074/mcp.M700239-MCP200.

(110) Lin, A.; Plubell, D. L.; Keich, U.; Noble, W. S. Accurately Assigning Peptides to Spectra When Only a Subset of Peptides Are Relevant. *J. Proteome Res.* **2021**, *20* (8), 4153–4164. https://doi.org/10.1021/acs.jproteome.1c00483.

(111) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat. Methods* **2007**, *4* (3), 207–214. https://doi.org/10.1038/nmeth1019.

(112) Fu, Y.; Qian, X. Transferred Subgroup False Discovery Rate for Rare Post-Translational Modifications Detected by Mass Spectrometry. *Mol. Cell. Proteomics MCP* **2014**, *13* (5), 1359–1368. https://doi.org/10.1074/mcp.O113.030189.

(113) Yi, X.; Gong, F.; Fu, Y. Transfer Posterior Error Probability Estimation for Peptide Identification. *BMC Bioinformatics* **2020**, *21* (1), 173. https://doi.org/10.1186/s12859-020-3485-y.

(114) Baker, P. R.; Medzihradszky, K. F.; Chalkley, R. J. Improving Software Performance for Peptide Electron Transfer Dissociation Data Analysis by Implementation of Charge State- and Sequence-Dependent Scoring. *Mol. Cell. Proteomics MCP* **2010**, *9* (9), 1795–1803. https://doi.org/10.1074/mcp.M110.000422.

(115) Efron, B. Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis. *J. Am. Stat. Assoc.* **2004**, *99* (465), 96–104.

(116) Searle, B. C.; Lawrence, R. T.; MacCoss, M. J.; Villén, J. Thesaurus: Quantifying Phosphopeptide Positional Isomers. *Nat Methods* **2019**, *16* (8), 703–706. https://doi.org/10.1038/s41592-019-0498-4.

(117) Lawrence, R. T.; Searle, B. C.; Llovet, A.; Villén, J. Plug-and-Play Analysis of the Human Phosphoproteome by Targeted High-Resolution Mass Spectrometry. *Nat. Methods* **2016**, *13* (5), 431–434. https://doi.org/10.1038/nmeth.3811.

(118) Sharma, K.; D'Souza, R. C. J.; Tyanova, S.; Schaab, C.; Wiśniewski, J. R.; Cox, J.; Mann, M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep.* **2014**, *8* (5), 1583–1594. https://doi.org/10.1016/j.celrep.2014.07.036.

(119) de Graaf, E. L.; Giansanti, P.; Altelaar, A. F. M.; Heck, A. J. R. Single-Step Enrichment by Ti4+-IMAC and Label-Free Quantitation Enables In-Depth Monitoring of Phosphorylation Dynamics with High Reproducibility and Temporal Resolution. *Mol. Cell. Proteomics MCP* **2014**, *13* (9), 2426–2434. https://doi.org/10.1074/mcp.O113.036608.

(120) HaileMariam, M.; Eguez, R. V.; Singh, H.; Bekele, S.; Ameni, G.; Pieper, R.; Yu, Y. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteome Res.* **2018**, *17* (9), 2917–2924. https://doi.org/10.1021/acs.jproteome.8b00505.

(121) Vizcaíno, J. A.; Deutsch, E. W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J. A.; Sun, Z.; Farrah, T.; Bandeira, N.; Binz, P.-A.; Xenarios, I.; Eisenacher, M.; Mayer, G.; Gatto, L.; Campos, A.; Chalkley, R. J.; Kraus, H.-J.; Albar, J. P.; Martinez-Bartolomé, S.; Apweiler, R.; Omenn, G. S.; Martens, L.; Jones, A. R.; Hermjakob, H. ProteomeXchange Provides Globally Coordinated Proteomics Data Submission and Dissemination. *Nat. Biotechnol.* **2014**, *32* (3), 223–226. https://doi.org/10.1038/nbt.2839.

(122) Sharma, V.; Eckels, J.; Taylor, G. K.; Shulman, N. J.; Stergachis, A. B.; Joyner, S. A.; Yan, P.; Whiteaker, J. R.; Halusa, G. N.; Schilling, B.; Gibson, B. W.; Colangelo, C. M.; Paulovich, A. G.; Carr, S. A.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B. Panorama: A Targeted Proteomics Knowledge Base. *J. Proteome Res.* **2014**, *13* (9), 4205–4210. https://doi.org/10.1021/pr5006636.

(123) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinforma. Oxf. Engl.* **2010**, *26* (7), 966–968. https://doi.org/10.1093/bioinformatics/btq054.

(124) Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J. The Skyline Ecosystem: Informatics for Quantitative Mass Spectrometry Proteomics. *Mass Spectrom. Rev.* **2020**, *39* (3), 229–244. https://doi.org/10.1002/mas.21540.

(125) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani,

K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* **2012**, *30* (10), 918–920. https://doi.org/10.1038/nbt.2377.

(126) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments. *Nat. Protoc.* **2019**, *14* (1), 68–85. https://doi.org/10.1038/s41596-018-0082-x.

(127) Batth, T. S.; Tollenaere, M. A. X.; Rüther, P.; Gonzalez-Franquesa, A.; Prabhakar, B. S.; Bekker-Jensen, S.; Deshmukh, A. S.; Olsen, J. V. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol. Cell. Proteomics MCP* **2019**, *18* (5), 1027–1035. https://doi.org/10.1074/mcp.TIR118.001270.

(128) Leutert, M.; Rodríguez-Mias, R. A.; Fukuda, N. K.; Villén, J. R2-P2 Rapid-Robotic Phosphoproteomics Enables Multidimensional Cell Signaling Studies. *Mol. Syst. Biol.* **2019**, *15* (12), e9021. https://doi.org/10.15252/msb.20199021.

(129) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets. *Nat. Methods* **2007**, *4* (11), 923–925. https://doi.org/10.1038/nmeth1113.

(130) Frewen, B.; MacCoss, M. J. Using BiblioSpec for Creating and Searching Tandem MS Peptide Libraries. *Curr. Protoc. Bioinforma.* **2007**, *20* (1), 13.7.1-13.7.12. https://doi.org/10.1002/0471250953.bi1307s20.