MS.1

Improving the robustness, throughput and comprehensiveness of quantitative proteomics

Michael J. MacCoss[1], Brian Searle[1,2], Lindsay Pino[1], Deanna Plubell[1], Danielle Faivre[1], Gennifer Merrihew[1], Jarrett Egertson[1], Sonia Ting[1], Brendan MacLean[1]

[1]University of Washington, [2]Institute for Systems Biology

Our goal is to develop high throughput method for sampling peptides with a mass spectrometer that can be used as a quantitative measure of the phenotype. To do this we would like a tandem mass spectrometry (MS/MS) method that can comprehensively sample all peptides in a sample continuously throughout the chromatographic elution. MS/MS acquired using data independent acquisition (DIA) offers significant advantages in terms of selectivity, sensitivity, and dynamic range over a single stage of mass analysis. Quantitative analysis using MS/MS has significant technical advantages over MS1 analysis and we are just now in a situation where we can get high selectivity (<4 m/z) across a majority of the mass range (i.e. 400–1000 m/z) using a rapid duty cycle (<3 sec). That said, this does not mean that there are not substantial challenges to overcome. For example, we need methods to assess whether the peptide measurements are quantitative versus just qualitative. Additionally, global methods like proteomics struggle significantly with signal calibration – making it difficult to compare quantitative measurements between batches, labs, and instrument platforms. Given the prevalence of complex proteoforms we need to think carefully about what the desired outcome is of a quantitative proteomics experiment using bottom-up methodologies. Finally, while most labs feel it is important to measure as many proteins and peptides as possible, the complications associated with doing this is non-trivial – ultimately with an increase in the number of analytes measured increases the multiple testing burden and the number of samples required to have the same statistical power. The talk will present the current state of the art of performing quantitative proteomics using DIA. I will present use cases of what we can do, where we think the limitations are, and what work is being done to improve the methods further.

MS.2

Complex-centric proteome profiling by SEC-SWATH-MS

Isabell Bludau[1], Moritz Heusel[1], George Rosenberger[1,2], Robin Hafen[1], Max Frank[1,3], Amir Banaei-Esfahani[1], Claudia Martelli[1], Charlotte Nicod[1], Peng Xue[1], Yujia Cai[3], Yansheng Liu[4], Ashok Venkitaraman[5], Vihandha Wickramasinghe[6], Hannes Roest[3], Ben Collins[1], Matthias Gstaiger[1], Ruedi Aebersold[1]

[1]Institute of Molecular Systems Biology, ETH Zurich, Switzerland, [2]Columbia University, New York, United States, [3]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Canada, [4]Yale University, New Haven, United States, [5]Medical Research Council Cancer Unit, University of Cambridge, United Kingdom, [6]Peter MacCallum Cancer Centre, Melbourne, Australia

Proteins are major effectors and regulators of biological processes and can elicit multiple functions depending on their interaction with other proteins. Therefore, it is of central interest in systems biology to determine the interactions and cooperation of proteins as a function of cell state. We have therefore developed an integrated experimental and computational technique for detecting in parallel hundreds of protein complexes, as well as changes in their composition and abundance in a single operation. The method consists of size exclusion chromatography (SEC) to fractionate native protein complexes, SWATH/DIA mass spectrometry to precisely quantify the proteins in each SEC fraction, and the computational framework CCprofiler to detect and quantify protein complexes by error-controlled, complex-centric analysis using prior information from generic protein interaction maps (Heusel & Bludau at al., 2019). Application of our workflow to the HEK293 cell line proteome delineates 462 complexes composed of 2,127 protein subunits, entailing 7673 unique protein-protein interactions. Our analysis further provided insights into novel sub-complexes and assembly intermediates of central regulatory complexes such as the proteasome. We have recently extended this workflow to study rearrangements of protein complex assemblies across different cell states, providing insights into assembly changes that are not captured by full proteome analyses. To increase throughput for such comparative SEC-SWATH-MS analyses, we established a fast protocol based on a 21 minute gradient on the EvoSep One HPLC system that enables the measurement of the ca. 65 SEC fractions of a biological sample per day, while minimizing loss of information. Furthermore, we extended our workflow to take advantage of available peptide-level information in the SEC-SWATH-MS data to investigate proteoform-specific complex integration. We expect our workflow to enable novel insights into the interplay between different protein variants and their impact on protein interactions and functionality on an unprecedented, system wide scale.

Moritz Heusel*, Isabell Bludau*, George Rosenberger, Robin Hafen, Max Frank, Amir Banaei-Esfahani, Ben Collins, Matthias Gstaiger, Ruedi Aebersold Complex-centric proteome profiling by SEC-SWATH-MS Molecular Systems Biology Jan 14;15(1):e8438 (2019).

MS.3

Aligning label-based discovery and global DIA validation proteomics to explore bacterial virulence phenotypes

Stuart J. Cordwell[1], Lok Man[1], Joel A. Cain[1]

[1]Charles Perkins Centre and School of Life and Environmental Sciences, The University of Sydney, Aust

The use of proteomics to inform subsequent biological validation studies requires substantial rigor in the analytical approach to ensure that the most important leads are followed. Our laboratory explores virulence determinants including an N-linked glycosylation (pgl) system and nutrient transporters in the gastrointestinal pathogen, Campylobacter jejuni. C. jejuni is a Gram negative, spiral and micro-aerophilic bacterium with a sequenced genome containing ~1620 genes. Target identification is based on the response of the proteome to environmental conditions that mimic the host, including bile salts, low iron, mucin availability and growth temperatures. The proteomics workflow includes parallel label-based liquid chromatography/tandem mass spectrometry (LC-MS/MS; minimum 3 biological growth replicates) using TMT and/or iTRAQ labelling, and system-wide validation using data independent analysis (DIA-SWATH-MS; minimum duplicate additional biological replicates). We routinely quantify ~80–90% of the predicted C. jejuni NCTC11168 genome using label-based LC-MS/MS (2 peptides; <1% FDR), and ~65–75% of these can be validated by DIA-SWATH-MS. Here, we will discuss the correlation between large-scale datasets in this biological system and how they facilitate subsequent studies, as well as highlight poorly or non-correlating data. In each case, we show how validated changes in the C. jejuni proteome reflect 'functional reality' that can be determined by molecular genetics, virulence assays and intra- and extra-cellular metabolomics.

MS.4

Advanced algorithms to assess and improve quantitative suitability in large DIA datasets

Sebastian Vaca[1], Karen Christianson[1], Nicholas Shulman[2], Karsten Krug[1], Brendan X. MacLean[2], Michael J. MacCoss[2], Steven A. Carr[1], Jacob D. Jaffe[1]

[1]Broad Institute of MIT and Harvard, Cambridge, MA, [2]University of Washington Genome Sciences, Seattle, WA

Data-Independent Acquisition (DIA) is a technique that promises to comprehensively detect and quantify all peptides above an instrument's limit of detection. Several software tools to analyze DIA data have been developed in recent years. However, several challenges still remain, like confidently identifying peptides, defining integration boundaries, dealing with interference for selected transitions, and scoring and filtering of peptide signals in order to control false discovery rates. In practice, a visual inspection of the signals is still required, which is impractical with large datasets.

Avant-garde is a new tool to refine DIA (and PRM) by removing interfered transitions, adjusting integration boundaries and scoring peaks to control the FDR. Unlike other tools where MS experiments are scored independently from each other, *Avant-garde* uses a novel data-driven scoring strategy. DIA signals are refined by learning from the data itself, using all measurements in all samples together to achieve the best optimization. We evaluated the performance of *Avant-garde* and the results clearly showed that it is capable of improving the selectivity, accuracy, and reproducibility of the quantification results in very complex biological matrices. We have further shown that it can evaluate the suitability of a peak to be used for quantification reaching the same results obtained with manual validation. *Avant-garde* is envisioned as a tool complementary to existing DIA analysis engines that aims to establish the strongest foundation for subsequent analysis of quantitative MS data.

Our workflow was applied to a large cohort of phosphopeptide-enriched samples. The dataset presented here spanned over 6 cell lines, with 90 drug perturbations, employing drugs that span the epigenetic-, neuro- and phosphosignaling-space. The analysis of this data enabled the confident quantification of more than 5000 phosphopeptides in each of the more than 1700 DIA runs. Cumulatively we quantified more than 22000 phosphopeptides in this dataset. Our tool improved the data completeness across the sample set and was robust to retention time shifts.

MS.5

Parallel accumulation — serial fragmentation combined with data-independent acquisition (diaPASEF)

Florian Meier[1], Andreas-David Brunner[1], Max Frank[2], Annie Ha[2], Eugenia Voytik[1], Stephanie Kaspar-Schoenefeld[3], Markus Lubeck[3], Oliver Raether[3], Ruedi Aebersold[4,5], Ben C. Collins[4], Hannes L. Roest[2], Matthias Mann[1,6]

[1]Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany, [2]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada, [3]Bruker Daltonik GmbH, Bremen, Germany, [4]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland, [5]Faculty of Science, University of Zurich, Zurich, Switzerland, [6]NNF Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

Data-independent acquisition (DIA) promises reproducible and accurate protein quantification across large sample cohorts. Rather than isolating single precursors in a data-dependent manner, the mass spectrometer cycles through segments of a predefined precursor m/z range. Although these segments collectively cover the entire mass range of interest, only a few percent of all available ions are sampled due to the consecutive selection of acquisition windows. Making use of the correlation of molecular weight and ion mobility in a trapped ion mobility device (TIMS) coupled to a high-resolution quadrupole time-of-flight mass spectrometer (Bruker timsTOF Pro), we here devise a novel scan mode termed 'diaPASEF' that analyzes up to 100% of the peptide fragment ion current (Meier *et al.*, bioRxiv 2019). Precursors are accumulated in the TIMS device and released as a function of their ion mobility. DiaPASEF synchronizes the position of the quadrupole isolation window with the ion release to follow the precursor ion population in a way that ions that are not mass selected at any given time are not lost. This enables virtually complete records of the peptide signal in a four-dimensional data cuboid in retention time, mass, ion mobility and intensity dimensions. To analyze this data type, we extended the established targeted data extraction workflow for the analysis of DIA data by the additional ion mobility dimension in the open-source Mobi-DIK software package. We explored the diaPASEF principle with typical proteomics samples of different complexity using a set of acquisition schemes that balance duty cycle, precursor selectivity and sensitivity. Data acquired from simple protein mixtures verify the expected data completeness and data from single runs of a whole proteome HeLa digest demonstrate deep proteome coverage of over 7,000 proteins with a very high degree of reproducibility and quantitative accuracy. Notably, even from 10 ng sample amounts we identified over 3,000 proteins with a data completeness of 85% and a median CV below 10% in triplicate injections. We conclude that diaPASEF greatly increases the utilization of the available ion current over conventional DIA and holds particular promise for high-sensitivity and high-throughput applications.

MS.6

Mobi-DIK (Ion Mobility DIA Analysis Kit): Targeted analysis software for diaPASEF data improves proteome coverage

Annie Ha[1], Max Frank[1], Florian Meier[2], Andreas-David Brunner[2], Stephanie Kaspar-Schönefeld[3], Scarlet Koch[3], Markus Lubeck[3], Oliver Raether[3], Ben C. Collins[4], Ruedi Aebersold[4,6], Matthias Mann[2,5], Hannes Rost[1]

[1]Donnelly Centre for Cell and Molecular Research, University of Toronto, Toronto, ON, [2]Max Planck Institute of Biochemistry, Martinsried, Germany, [3]Bruker Daltonik GmbH, Bremen, Germany, [4]ETH Zurich, Zurich, Switzerland, [5]NNF Center for Protein Research University of Copenhagen, Copenhagen, Denmark, [6]Faculty of Science, University of Zurich, Switzerland

Data-independent acquisition (DIA) has gained popularity due to its reproducibility and sensitivity in high-throughput proteomics studies. Meanwhile, parallel-accumulation serial-fragmentation (PASEF) exploits trapped ion mobility spectrometry to achieve high duty cycle, efficient ion usage and improve peptide identification rates in DDA. By coupling DIA isolation windows to the precise ion mobility elution of the corresponding ions, the combination of windowed DIA with the PASEF principle allows multiplexing of DIA windows in a single 100ms ion mobility separation of precursor ions. Through multiplexing DIA windows in TIMS scans, we measured up to 32 DIA windows in 100ms, effectively increasing the duty cycle by 32x compared to traditional SWATH-MS.

Mobi-DIK (Ion Mobility DIA Analysis Kit) is a novel software capable of analyzing highly multiplexed diaPASEF data, where ions are separated by ion mobility in each MS and MS/MS scan. The software generates ion mobility-enabled spectral libraries directly from from MaxQuant output of highly fractionated DDA PASEF runs and stores them in the standardized TraML format. For analysis, Mobi-DIK automatically calibrates mass (non-linear), retention time (non-linear) and drift time (linear) between our assay library and experimental diaPASEF runs, achieving less than 1% deviation in drift time values. Targeted extraction using only the relevant ion mobility space increases signal to noise and rates of peptide identifications, while highly accurate reference drift times improve scoring of peaks and filter isobaric interferences. We show separation of isobaric compounds, increase in signal to noise and increased sensitivity directly through our analysis. Here, we are able, for the first time, to directly distinguish and quantify the analytical improvements achieved by the focusing effect of ion mobility, the ion mobility selection and the DIA multiplexing, which respectively increase peptide identifications by 45%, 55%, and 45% respectively, yielding an impressive 3-fold overall improvement of peptide identifications relative to DIA without ion mobility separation under otherwise comparable conditions. Our Mobi-DIK software combining DIA with IM in a targeted platform is capable of quantifying 7000 proteins at 1% FDR from 200 ng of total peptide mass on column.

MS.7

Applications of DIA for PTM research and specific DIA workflows

Birgit Schilling[1]

[1]Buck Institute for Research on Aging, Novato, CA 94945

Proteomic analysis of posttranslational modifications is highly relevant as PTM signaling plays a large role in diseases and metabolism. However, PTM-related projects pose unique challenges to quantitative workflow, such as enrichment strategies, PTM site localization etc. Recent work is presented using data-independent acquisitions to identify and quantify PTM containing peptides in a high-throughput format. Several projects will be shown assessing protein acylation, such as acetylation and succinylation in liver and brown adipose fat tissues, and other PTMs. Simultaneous enrichment of multiple PTMs using antibody enrichment DIA-MS workflows (one-pot enrichment) are performed to assess PTM crosstalk and to obtain comprehensive coverage. PTM site localization is determined using various software tools, such as Skyline and Spectronaut. Acquiring PTM samples in DIA mode provides capabilities to identify multiple different site localization isomers from peptides with the same peptide sequence and precursor ion m/z. Thus DIA acquisitions for PTM profiling and quantification provides unique opportunities to overcome challenges that are associated with protein PTM workflows.

MS.8

In-silico generated proteome-wide spectral libraries of (un)modified (non)tryptic peptides at your fingertips: Prosit

Siegfried Gessulat[1], Tobias Schmidt[1], Daniel Zolg[1], Karsten Schnatbaum[2], Johannes Zerweck[2], Tobias Knaute[2], Bernard Delanghe[3], Andreas Huhmer[4], Ulf Reimer[2], Bernhard Kuster[1], Mathias Wilhelm[1]

[1]Proteomics and Bioanalytics, Technical University of Munich, Freising, Germany, [2]JPT Peptide Technologies GmbH, Berlin, Germany, [3]Thermo Fisher Scientific, Bremen, Germany, [4]Thermo Fisher Scientific, San Jose, CA, USA

The fragmentation pattern and retention time of a peptide are (currently) two of the most important features for any mass spectrometry-based proteomics method. However, the generation of comprehensive proteome-wide spectral libraries is time consuming and arguable not possible. Because of this, spectral libraries generated from prior DDA experiments are an essential step before conducting DIA or PRM experiments. Here, we show how Prosit, our deep-learning framework for predicting fragment intensities and retention times of peptides, can be used to generate proteome-wide in-silico spectral libraries with near reference data quality for virtually any (un)modified (non-)tryptic peptide irrespective of its origin. Because Prosit was trained on systematically acquired data from the ProteomeTools project, it is able to predict spectra at any commonly used collision energy. This allows users to calibrate predictions to their mass spectrometer avoiding time-consuming re-training. We show that our predictions allow the confident identification of peptides when investigating excessively large search spaces by DDA, perform on par with custom spectral libraries when analyzing DIA data and can be used to speed up assay development using MRM or PRM acquisition schemes.