CURRENT PROTOCOLS
The Fine Art of Experimentation

Webinar Review

Dr. John Koomen on Data-Independent Analysis and Targeted MS/MS Quantification of Biomarkers

Clinical research applications of proteomics increasingly rely on MS-based approaches. Biomarker discovery, development, verification, validation, and quantification are all accessed using these approaches. Although different data collection modes exist for biomarker assessment, John Koomen, Associate Member and Scientific Director of Proteomics at the H. Lee Moffitt Cancer Center and Research Institute in Tampa, Florida, focuses on the emerging DIA mode in his webinar **Combining Data-Independent Analysis (DIA) for Broad-Scale Phenotyping and Targeted Tandem Mass Spectrometry Quantification of Specific Biomarkers.**

Data-dependent acquisition (DDA) versus Data-independent acquisition (DIA)

Koomen acknowledges the more common data-dependent acquisition (DDA) mode remains the "traditional method for discovery proteomics which relies on specific peptide identification" but asks two probing and primary questions regarding the acquisition methods: With DDA, can we sample all of the observable peptide peaks? Whereas with DIA, can we sample everything detectable? Although he is quick to point out the ongoing advances in instrumentation which will address the DDA question, the possibility to "sample all of the peptides that elute during that process that have a certain amount of intensity" using DIA is tantalizing.

In his understated manner, Koomen stresses the many players involved in generating robust research workflows that attend to the reality of limited resources in any organization or institution. The parameters he uses in his set up—precursor range of m/z 450-1400; 70,000 MS resolution and 17,500 MS/MS resolution; loop of 18 isolation windows; 90 minute gradients—are all similar to those assessed and used by other active specialists (Prakash et al., 2014). However, he questions and then carries out experiments to test these parameters: is a 90 minute gradient necessary? Is critical information lost with a shorter elution run, say 45 minutes? In his program, he's aiming for a practical limit. As an example, for different drug treatments and dose escalation studies, his colleagues want to capture all the information and data inherent in those experiments but in a practical amount of time. "And a reasonable period of time for us would be somewhere between 2 and 4 weeks to analyze a cohort of at least 100 samples, then we're generally looking at 90 minute gradients and 2 hour total run times." For the samples that Koomen and his colleagues have analyzed, DIA strategies produce label-free quantification using the peptide ion signals and do capture additional detail compared to other methods; furthermore, the DIA strategy can be supplemented by addition of targeted tandem mass spectrometry analyses of specific protein biomarkers, which can anchor research studies to a specific molecular endpoint (e.g., HER2 protein expression).

Objectives of a clinical research application

As a member of the Moffitt Cancer Center, questions from Koomen's group and those of his colleagues center around cancer biology and how proteome analysis can enhance and illuminate the current understanding of complex pathophysiology. For every question, a biological sample of finite amount is obtained—usually precious tumor samples as in one of the examples Koomen describes: formalin-fixed paraffin-embedded tissues (**Figure 1**). "DIA becomes an optimal choice for a limited amount of biological sample" where there "would be too little to do fractionation strategies to try to increase the depth of the proteome, where you wouldn't be able to enrich that much and you wouldn't see that much more by processing it in a different way."

Rather than sampling a very well-defined set of peptides with tandem mass spectrometry, DIA sampling allows a broad acquisition of peptides that can be stored and queried at later time points. The ability to re-visit the DIA samples with new or emerging questions and mine that dataset using spectral library matching is the advantage of the DIA workflow over existing techniques (Sajic et al., 2015). Koomen comments, "So as we develop, for example, a database of all the mutant peptides that might be relevant to cancer you could come back and query the existing datasets to see if you're able to observe those peptides."

Optimism about the range of problems that DIA can address

Can DIA be applied to complex protein samples?

Within the DIA workflow, a one-size-fits-all construct in the experimental design is possible. But the more compelling aspect of the DIA approach, Koomen relates, is the ability to tailor and tune the design to provide exactly the desired measurements for the query at hand. Koomen illustrates this from his own work with breast cancer tissue sections and tissue microarray (TMA) samples, but acknowledges DIA can be applied to any complex concept—signaling pathways; phenotypic information that researchers desire; post-translational modifications; chemically modified

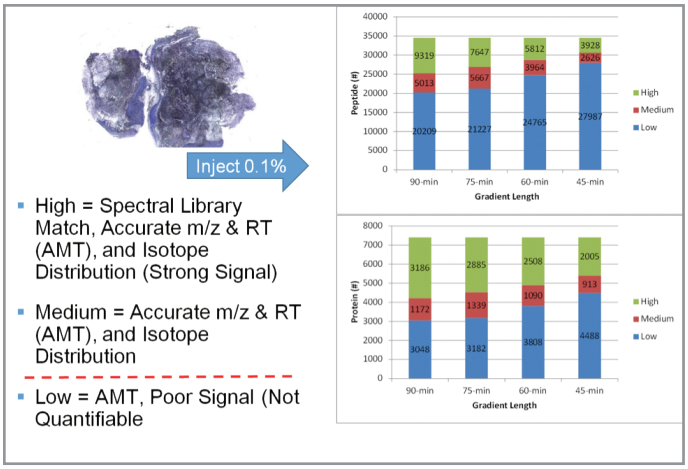

Figure 1. Comparative detection of proteins and peptides as a function of gradient length using pSMART. Each injection constitutes 0.1% (by weight) of the tissue extracted. The color coding is defined. (Koomen, 2015).

WILEY

peptide pathways. "The main focus we have is to try and simplify some of these measurements and make sure you get the quality of measurement and the depth of measurement that you want within one single LC-MS/MS analysis and that's where we think it helps the most," he states.

What is the limit for the DIA method?

One of the valid criticisms of DIA, relative to the "Gold Standard" set by DDA, is its limits in quantifying proteins. Koomen notes, "We're still trying to figure this out (i.e., the limits of the DIA method) but the most abundant proteins and the highest signal peptides from each protein are the best represented." Matching some of the DIA-analysis samples from his work with a multiple reaction monitoring (MRM)-based approach to quantify the levels of the observed proteins, is one means of improving the prospects of and confidence in DIA as a quantitative tool.

Window(s) of opportunity, peptide enrichment, and data variability

In DDA mode, usually a 2 m/z window is set around the (sampled) precursor of interest; the precursor is pre-selected. Sampling intact peptides using DIA provides an unbiased approach relative to DDA since spectral acquisition is not triggered by detection or knowledge of the precursor. Thus, different criteria must guide the windows defined with a DIA approach. For instance, the anticipated complexity of peptides in a particular region of the ion chromatogram has to be taken into consideration: lower m/z values in the ion chromatogram, copious number of peptides expected, narrow window inferred. Sampling time across the LC peak is another factor; according to Koomen, "You have to balance the number of windows you want to observe and the amount of detail you capture that way with the ability to sample your peaks in your chromatography."

Faster sampling rates make it more likely to have MS/MS data near the apex of peptide peak intensity, which provides the best potential for MS/MS data matching to the spectral library required for sequence identification. He's advocating for two samplings across the entire window to harvest ever more content with the aim to convert lower confidence measurements to the more valuable higher (confidence) quality measurements (**Figure 2**). "We're trying to balance our need to get targeted MS/MS data with the ability to get good sampling across every peak that we have."

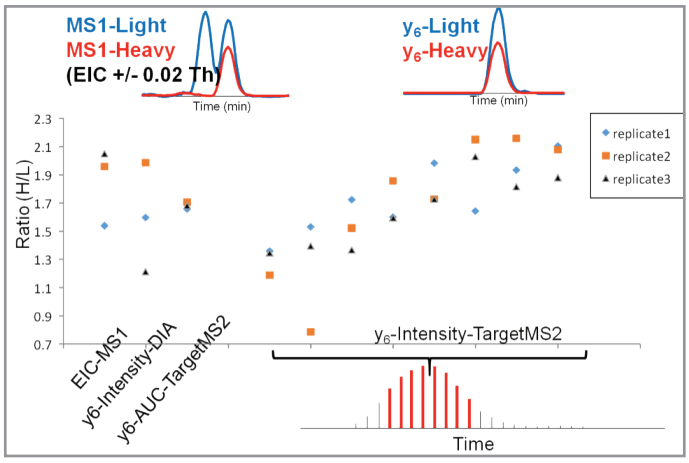

Figure 2. Comparative quantitation for the targeted peptide using MS-level analysis (shown at the left) compared to product ion quantitation using the y_6 fragment ion. (Koomen, 2015).

Koomen alludes to the breadth of proteomics applications used in collaborations within Moffitt. Clinical research applications—like the specific examples he expounds upon in his webinar—are a focus of his institute, but more fundamental research projects also comprise an integral part of the Moffitt mission. This was underscored in his response to the question: is peptide enrichment necessary for DIA if the modified peptides are low abundance and in a complex proteomic sample? An unequivocal yes was his answer; enriching low abundance peptides is indispensable in generating high quality outcomes for studying, say, kinase signaling (where enriching phosphopeptides is key) or ubiquitination processes for proteasome inhibitors or differential degradation studies (where enrichment of the ubiquitinated sequences would be pivotal).

His recommendation to those skilled in the art, bent on teasing out information from complex proteomic samples? "Focus the power of the mass spec on exactly what you want to study." One example using activity-based protein profiling as an enrichment tool for kinases is being pursued; DDA, DIA and LC-MRM are being compared for detection of the lower abundance kinases in the background of other ATP-utilizing enzymes.

Consistency and confidence in peptide assignment across samples in his TMA experiment are examined. Analysis of tissue microarray samples are instructive with the initial assumption that biological differences across the samples are relatively minor. For quantification, the (high resolution) MS1 measurement generates the accurate mass, a crucial parameter. Coupling that information with elution time has provided confident peptide assignment (2,000 peptides that match to 500 proteins).

Spectral library matching and confidence

An essential element in the success of a DIA approach is the availability of and access to spectral libraries that contain validated proteins of interest. Koomen mentions publically available libraries that contain both tissue-relevant and cancer-relevant proteins, and although their existence provides a crucial component, it is unlikely that centralized public libraries will keep pace with the content generated from MS-based proteomic techniques due to the expansion in the field generally and its increasing relevance in the clinical research realm (Gillet et al., 2012). The downside of a custom-built spectral library is the in-house resources needed to construct it. But for scientists querying biological systems not included in a pertinent public spectral library, the custom spectral library provides a remedy. This point is illustrated in Koomen's work—he built his own experimental spectral library from formalin-fixed paraffin-embedded tissue samples for his project characterizing protein expression in breast tumor tissue.

In response to a question from the audience, Dr. Koomen explained that spectral library matching coupled with DIA may allow interrogation of inter-protein crosslinking. Although not an area of interest in his own labs, and while acknowledging that a dominant hurdle in dealing with branched or crosslinked peptides is the difficulty in interpreting the tandem mass spectra, he suggests, "DIA as a secondary strategy would allow you to go through a larger number of samples to look for peptides that you've already identified

in the spectral library." But building such a spectral library would be no trivial task and he recognizes that "the major challenge is identifying all of those crosslinked peptides and getting them into the spectral library so you could do the matching".

Access to spectral libraries, whether available in the public sphere or of a bespoke nature, is all well and good, but what about their utilization? How much confidence in spectral library matching is there, compared with DDA? "This remains a hot button issue for a lot of people," Koomen concedes. But he affirms having a really accurate mass measurement for the intact peptide and that same accuracy for the daughter fragments, generates much higher quality and sometimes even higher content than from a MRM experiment "so having really accurate intact peptide mass and 3-4 fragments in most cases will give you almost unequivocally the peptide match that you've made from the spectral library." This data is also the underlying strategy for parallel reaction monitoring (PRM), so as that technique becomes more widely used and accepted, this question will become less important to DIA evaluation.

Sampling and Samples

How many targeted tandem MS/MS acquisitions can be interwoven with a DIA approach (**Figure 3**)? After all, the more acquisitions, the richer the resulting data set will be (Prakash et al., 2014). Currently, Koomen is comfortable with three and acknowledges that doing more would entail narrowing the DIA window significantly or using a higher frequency

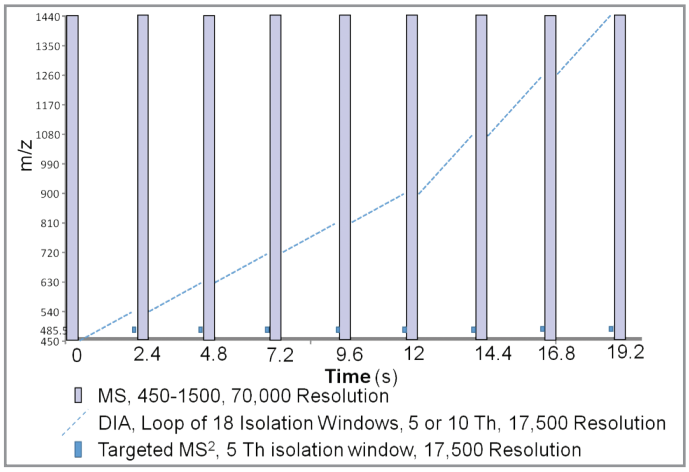

Figure 3. Pictorial representation of extending the capabilities of pSMART for mixed mode quantitation. The original pSMART method is represented by MS acquisition with one targeted PRM event with the idea to stratify which peptides can be quantified by MS and which ones may need the added sensitivity of PRM. (Koomen, 2015).

scanning instrument. "We would really like to have two inclusion lists, one for the DIA and one for targeted MS/MS where you could just say, 'I know this particular peptide elutes at 30 minutes so I need a 2-minute window to monitor this with targeted MS/MS.' Then you could stack a very large number of those measurements in parallel."

Tissue microarrays (TMAs) increase throughput and enable comparisons of sample groups within the same experiment by providing a large cohort of tissue samples with linked clinical information. Although the number of detectable peptides (and by inference, the proteins they represent) is increasing due to new techniques and improvement in the capabilities of mass spectrometers, quantifying the protein content of TMA samples remains a very big challenge. TMA samples do not provide enough sample to perform a parallel protein assay, and indirect procedures for quantitation—such as applying algorithms to evaluate tissue cellularity and stromal content then linking that information to observed intracellular proteins or extracellular proteins—are used. In response to this deficit, Thermo Fisher Scientific has a method in development for quantifying the amount of digested peptide in very small samples.

Acceleration with an iterative collaboration

Ever present in the research workflow descriptions is the interplay of in-house, industry-based scientists with researchers like Koomen. Advances in instrumentation, data collection and analysis; spectral library formation, access, and annotation; and biological sample preparation are all components of a complex proteomic-based approach (Bruce et al., 2013; Zhang et al., 2014). The desire and need to achieve higher sensitivity, higher throughput, and more accurate quantification is propelled by the need to provide deeper, crucial insights to the questions in cancer biology he and his colleagues continue to probe. A classic partnership with academic scientists and industry scientists exists; input from all players is essential and reveals an iterative, curiosity-driven paradigm which is revealed in software, instrumentation and related advances in the proteomics realm.

References

Bruce, C., Stone, K., Gulcicek, E., and Williams, K. **2013**. Proteomics and the analysis of proteomic data: 2013 overview of current protein-profiling technologies. *Current Protocols in Bioinformatics 41,* 13.21.1-13.21.17. DOI: 10.1002/0471250953.bi1321s41

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. **2012**. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Molecular & Cellular Proteomics 11*. DOI: 10.1074/mcp.O111.016717

Koomen, John. Combining Data-Independent Analysis for Broad-Scale Phenotyping and Targeted Tandem Mass Spectrometry Quantification of Specific Biomarkers. Webinar, April 29, 2015, **http://www.currentprotocols.com/WileyCDA/Section/id-810325.html**.

Prakash, A., Peterman, S., Ahmad, S., Sarracino, D., Frewen, B., Vogelsang, M., Byram, G., Krastins, B., Vadali, G., and Lopez, M. **2014**. Hybrid data acquisition and processing strategies with increased throughput and selectivity: pSMART analysis for global qualitative and quantitative analysis. *Journal of Proteome Research 13,* 5415-5430. DOI: 10.1021/pr5003017

Sajic, T., Liu, Y. and Aebersold, R. **2015**. Using data-independent, high-resolution mass spectrometry in protein biomarker research: Perspectives and clinical applications. *Proteomics - Clinical Applications 9,* 307-321. DOI: 10.1002/prca.201400117

Zhang, G., Annan, R.S., Carr, S.A., and Neubert, T.A. **2014**. Overview of peptide and protein analysis by mass spectrometry. *Current Protocols in Molecular Biology 108,* 10.21.1-10.21.30. DOI: 10.1002/0471142727.mb1021s108

Webinar was presented online on April 29, 2015 with sponsorship by Current Protocols and Thermo Fisher Scientific Inc.