

## PROTEOMICS

# Increasing proteomics throughput

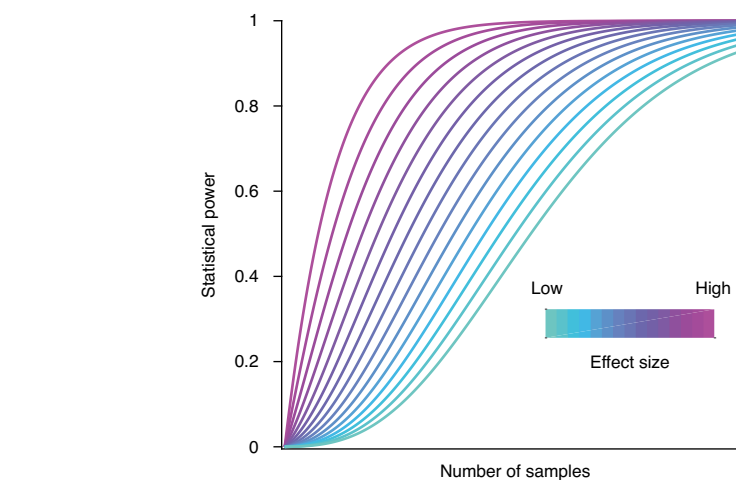
A new dimension for analyzing mass spectrometry data allows rapid quantification of up to 70% more peptides.

Nikolai Slavov

Which proteins are disease biomarkers? What is the impact of diseases on the proteome? How does genetic variation affect protein abundances? Answering these and many other questions requires methods to quantify proteins in large numbers of samples. To address this demand, Messner et al.<sup>1</sup> report in this issue a mass spectrometry (MS) method that quantifies thousands of proteins in mere minutes. Such increases in the throughput of proteomics can empower applications that are impractical to carry out with slower methods (Fig. 1).

Two decades ago, a landmark method, multidimensional protein identification technology (MudPIT), enabled quantification of about 1,500 proteins<sup>2</sup>. MudPIT set this record by adding a dimension to peptide separation: it used a two-dimensional liquid chromatography approach to separate peptides before tandem MS analysis<sup>2</sup>. This added dimension was essential to the success of MudPIT, but it also extended the time needed to analyze each sample. In the ensuing two decades, advances in peptide separation methods, MS instruments and methods have reduced the analysis time while simultaneously increasing the number of quantified proteins<sup>3</sup>.

Despite this progress, the analysis time per sample continues to limit the number of analyzed samples and thus the scope and statistical power of investigations<sup>4</sup>. To relax this limitation, Messner et al. also added a dimension to protein MS analysis: they used a scanning-quadrupole-based data-independent acquisition (DIA) method previously introduced by Moseley et al.<sup>5</sup>. The scanning quadrupole method allows continuous movement of the precursor isolation window. Messner et al. exploited this movement to assign precursor masses to the MS2 fragment traces, effectively adding another dimension to peptide sequence determination. This dimension allows improved discrimination between correct and incorrect peptide identifications based on the degree of correspondence between the precursors and their matching MS2 fragments. Notably, this dimension does not



**Fig. 1 | Higher throughput of proteomic analysis increases statistical power.** As the number of samples analyzed per unit time increases, the cost per sample decreases. The increased throughput and reduced cost allow the analysis of more samples, which in turn increases the statistical power of many applications, such as discovering biomarkers and genetic associations. The higher throughput is particularly important for smaller effect sizes that may demand many thousands of samples to be adequately powered.

increase the time needed to analyze a sample and therefore enables the identification of more peptides (and thus proteins) per unit time. Using their method, termed scanning SWATH (sequential window acquisition of all theoretical fragment ion spectra), the authors analyzed a human cell line and report identifying about 2,000 proteins in 30 s of active gradient and about 5,000 proteins in 5 min of active gradient. The method also enhanced quantification precision compared with traditional SWATH, as shown by reduced variability between replicate measurements.

The authors demonstrated the method's capabilities with two use cases. First, they analyzed the proteomes of budding yeast treated with cytostatic and antifungal drugs. With 7.3 min total analysis time per sample, they analyzed quadruplicates of 16 different drug treatments in mere hours. The data revealed class-dependent changes in protein abundance, suggesting which enzymes are affected by each class of drug. The power of scanning SWATH was also demonstrated

by the analysis of plasma samples from 30 patients with COVID-19. The approach quantified 43 biomarkers approved by the US Food and Drug Administration, and a principal component analysis of the data allowed the stratification of patients based on disease severity.

Although these capabilities are exciting and will increase the statistical power of many applications (Fig. 1), the number of identified proteins is about half of the quantifiable proteome of human cell lines<sup>3</sup>. Furthermore, the maximum gains were achieved with relatively large samples—10  $\mu$ g of human cell line tryptic digest—and thus may not be applicable to all samples, such as biopsy samples obtained via fine needle aspiration. Thus, highly sensitive and comprehensive proteome quantification on the time scale of a minute remains an open challenge.

To have impact, good ideas need good and accessible implementations. Messner et al. implemented scanning SWATH with accessible hardware and software: the

complexity of data acquisition is similar to that of traditional SWATH, and data analysis is performed by open-source algorithms that are made broadly available with the new version of the DIA-NN software suite<sup>6</sup>. Thus, scanning SWATH is ready for wide deployment. It makes large-scale analysis of many samples cheaper and more practical, which is needed for many applications, such as identifying genetic polymorphisms causing variation in protein abundances<sup>7</sup>.

The higher throughput afforded by scanning SWATH is an important addition to rapidly advancing MS methods and applications. It can be extended to applications beyond quantifying protein abundances, such as MS methods for quantifying protein structures, interactions and activities. Higher throughput is particularly needed for single-cell proteomics<sup>8</sup>. So far, the throughput of single-cell protein analysis has benefited from multiplexing methods allowing simultaneous analysis of many samples in parallel<sup>9</sup>. Combining multiplexing with

shorter analysis time can further enhance throughput and increase robustness of relative quantification.

As the throughput and power of proteomics methods increase, so does the need for rigorous experimental designs. The throughput afforded by scanning SWATH and other recent methods<sup>10</sup> can enable the analysis of thousands of samples within days. This quantitative change in throughput can lead to a qualitative change in the biomedical applications employing proteomic technologies (Fig. 1). Associations between genetic variants and protein abundance have already provided insights undetectable by RNA associations<sup>7</sup>. Increased throughput will afford the analysis of more samples, which in turn will increase statistical power<sup>4</sup>. Investigations that traditionally have been confined to the transcriptome can expand to incorporate proteomics data. Taking full advantage of these exciting prospects requires careful experimental designs that mitigate batch effects and other systematic artifacts<sup>7</sup>. □

Nikolai Slavov <sup>1,2</sup> ✉

<sup>1</sup>Department of Bioengineering, Northeastern University, Boston, MA, USA. <sup>2</sup>Barnett Institute, Northeastern University, Boston, MA, USA.

✉e-mail: nslavov@northeastern.edu

Published online: 25 March 2021

<https://doi.org/10.1038/s41587-021-00881-z>

#### References

1. Messner, C. et al. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-00860-4> (2021).
2. Washburn, M. P., Wolters, D. & Yates, J. R. III *Nat. Biotechnol.* **19**, 242–247 (2001).
3. Muntel, J., Gandhi, T., Verbeke, L. & Bernhardt, O. M. *Mol. Omics* **15**, 348–360 (2019).
4. Sham, P. C. & Purcell, S. M. *Nat. Rev. Genet.* **15**, 335–346 (2014).
5. Moseley, M. A. et al. *J. Proteome Res.* **17**, 770–779 (2018).
6. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. *Nat. Methods* **17**, 41–44 (2020).
7. Battle, A. et al. *Science* **347**, 664–667 (2015).
8. Slavov, N. *Science* **367**, 512–513 (2020).
9. Specht, H. et al. *Genome Biol.* **22**, 50 (2021).
10. Meyer, J. G., Niemi, N. M., Pagliarini, D. J. & Coon, J. J. *Nat. Methods* **17**, 1222–1228 (2020).

#### Competing interests

The author declares no competing interests.