Framework for multiplicative scaling of single-cell proteomics

Many biomedical questions demand scalable, deep, and accurate proteome analysis of small samples, including single cells. A scalable framework of multiplexed data-independent acquisition for mass spectrometry enables time saving by parallel analysis of both peptide ions and protein samples, thereby realizing multiplicative gains in throughput.

This is a summary of:
Derks, J. et al. Increasing the throughput of sensitive proteomics by plexDIA. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01389-w (2021).

Published online:
Published online: 18 July 2022

The problem

Mass-spectrometry (MS) proteomics is a powerful method for accurate and specific quantification of proteins in single cells, but its throughput and depth of proteome coverage remain limited[1,2]. Although MS analysis of single human cells detects tens of thousands of peptide-like features, it can assign amino acid sequences to only a small fraction of them[1]. This limitation stems in part from the time constraint imposed by fragmenting ions (which is needed for sequence identification) from only one peptide at a time when multiplexing samples with isobaric mass tags (same total mass but different distributions of the isotopes)[1,2]. This time constraint may be relieved by analysing multiple peptides in parallel, as afforded by data-independent acquisition (DIA)[3]. If effectively combined with tagging single cells with cell-specific chemical 'barcodes', DIA might make it possible to simultaneously multiplex peptides and single cells without reducing proteome coverage or quantitative accuracy[1]. We hypothesized that such parallel multiplexing might enable multiplicative gains in throughput, as labelling single cells with n non-isobaric isotopologous mass tags (with the same chemical structure but different isotopic distributions and different total masses) may enable an n-fold increase in the number of accurate protein data points.

The solution

We tested this hypothesis by developing an optimized framework, termed plexDIA (Fig. 1a). The results from plexDIA supported the hypothesis for $n = 3$, and we expect that the plexDIA framework will scale up for $n > 3$.

For the past decade, non-isobaric mass tags (the category of barcodes used by plexDIA) have been used less often than isobaric mass tags. This preference is because in data-dependent acquisition (DDA) workflows, the use of non-isobaric mass tags results in fragmenting fewer unique peptides than when using isobaric mass tags and, therefore, reduced proteome coverage. plexDIA solved these problems by acquiring data from all detectable peptide precursor ions and fragments and using a computational framework to efficiently interpret these mass spectra despite their increased complexity. Because many peptides and their fragments are analysed in parallel (as opposed to one peptide per MS2 spectrum, as with DDA), plexDIA can afford long ion accumulation times and, therefore, high capture efficiency. This approach enabled us to count more than 1,000,000 protein molecules per single human cell.

Interpretation of plexDIA data strongly benefited from the regular structure in the mass spectra, owing to discrete mass shifts in the spectra introduced by the non-isobaric mass tags[1]. Leveraging this structure, plexDIA can estimate peptide abundance from multiple MS scans from both the peptide ions and their fragments, as shown with tubulin β-chain (TUBB) and high-mobility group protein HMG-I/HMG-Y (HMGA1) (Fig. 1b). These multiple estimates enable evaluation of the consistency of quantification and make it possible to derive reliability estimates for individual single-cell data points. Furthermore, the structure of the mass spectra enhances sequence identification of peptides with low abundance, as shown with type II cytoskeletal keratin 7 (KRT7) (Fig. 1b). This enhanced sequence identification helped to increase data completeness in both bulk and single-cell analysis with plexDIA, reaching 98% data completeness within a plexDIA set.

Future directions

plexDIA provides a proof of principle for a general approach that can be scaled by using higher plex non-isobaric mass tags that are optimized for DIA analysis[1,4]. Such tags should be much easier and cheaper to design and manufacture than isobaric tags. Thus, we expect these tags to substantially increase the throughput and accessibility of sensitive protein analysis[4]. Extrapolation of our 3-plexDIA results to a 100-plexDIA predicts the feasibility of analysing the proteomes of about 5,000 cells per day using a single MS instrument.

This extrapolation has caveats, including the potential for increased interference (due to high density of ions) and decreased ion sampling (due to limited capacities of ions traps). Nevertheless, plexDIA implemented on a decade-old Q exactive instrument demonstrated accurate quantification of about 200,000 distinct peptide precursors in a single run without signs of reaching an asymptote. Newer MS instruments should increase the ability to accurately resolve and quantify increasing numbers of peptide precursors from high-plexDIA analysis.

Scaling plexDIA will benefit from instrumentation advances (improving peptide separation, ionization, ion mobility, ion trap capacities, and MS detectors) and from new experimental designs, computational algorithms, and scalable sample preparation methods[1,4,5]. plexDIA promises to make the analysis of millions of single-cell proteomes practical while maximizing quantitative accuracy and proteome coverage[4].

Nikolai Slavov
Northeastern University, Boston, MA, USA

EXPERT OPINION

"The authors developed a multiplexed strategy to perform DIA by employing reagents that are not isobaric and, therefore, show a mass difference. The concept being tested is the ability to multiplex with these reagents, based on the fact that ion selection does not occur in DIA and, therefore, tandem MS should not be slowed in the process." **John Yates, Scripps Research Institute, La Jolla, CA, USA**

FIGURE

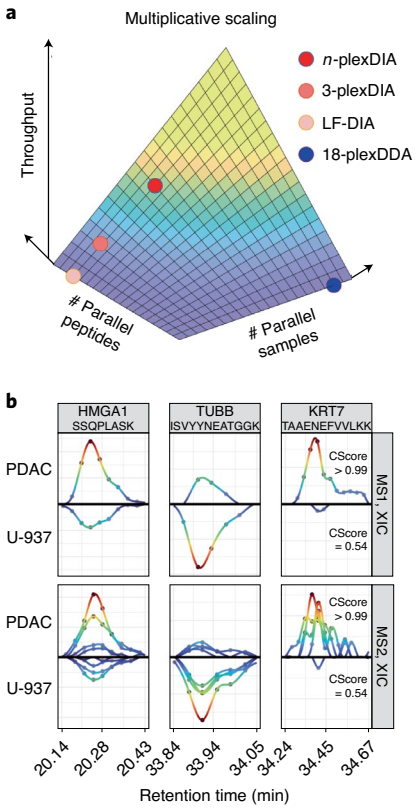

Fig. 1 | Increasing the throughput and accuracy of sensitive proteomics. a, The plexDIA framework combines parallel analysis of peptides with parallel analysis of samples to achieve a multiplicative increase in throughput (that is, the number of accurately quantified protein data points per unit time). LF, label-free. **b**, Extracted-ion chromatograms (XIC) for peptide ions (MS1 level) and their fragment ions (MS2 level). Each panel displays data from two single cells from a plexDIA set: a pancreatic ductal adenocarcinoma (PDAC) cell and a monocyte cell (U-937). CScore, a metric that quantifies the confidence of sequence identification using only data from a single mass tag channel. © 2022, Derks, J. et al.

BEHIND THE PAPER

In retrospect, the core idea behind plexDIA (Fig. 1a) seems obvious. Yet, it was less obvious prospectively. It emerged and solidified while writing a paper on optimizing the accuracy and depth of protein quantification in experiments using isobaric carriers. The ideas that we were describing had been clear to us for years, but the focused attention that was needed to make them accessible for others illuminated a key limitation of our singe-cell proteomics methods[2]: accumulating and fragmenting ions from one precursor at a time fundamentally limits the throughput and depth of sensitive MS proteomics. Appreciating this limitation motivated us to develop plexDIA. The more we developed plexDIA, the more its advantages and future potential became apparent. **N.S.**

REFERENCES

1. Slavov, N. Driving single cell proteomics forward with innovation. *J. Proteome Res.* **20**, 4915–4918 (2021).
A Perspective that explains the need and opportunities for developing plexDIA.

2. Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biol.* **22**, 50 (2021).
This article reported automated and multiplexed MS proteomics analysis of about 3,000 proteins across 1,490 single human cells.

3. Venable, J. D. et al. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).
This article introduced the concept of DIA for MS proteomics.

4. Slavov, N. Scaling up single-cell proteomics. *Mol. Cell. Proteom.* **21**, 100179 (2022).
A Perspective discussing strategies for increasing the throughput of single-cell proteomic analysis.

5. Leduc, A. et al. Exploring functional protein covariation across single cells using nPOP. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.24.441211 (2022).
A preprint introducing surface droplets for single-cell proteomics sample preparation, which prepare thousands of single cells in parallel and easily scale to different multiplexing formats.

FROM THE EDITOR

"Although multiplexing strategies have been used for a long time, their application to DIA, and in particular single-cell proteomics, is really quite a new concept." **Editorial Team, Nature Biotechnology**