Journal of proteome research

Article

Cite This: J. Proteome Res. 2019, 18, 2493−2500

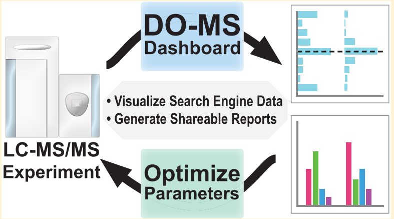
pubs.acs.org/jpr

DO-MS: Data-Driven Optimization of Mass Spectrometry Methods

R. Gray Huffman,[*,†,‡,∥] Albert Chen,[†,‡,∥] Harrison Specht,[†,‡,∥] and Nikolai Slavov[*,†,‡,§]

[†] Department of Bioengineering, Northeastern University, Boston, Massachusetts 02115, United States
[‡] Barnett Institute, Northeastern University, Boston, Massachusetts 02115, United States
[§] Department of Biology, Northeastern University, Boston, Massachusetts 02115, United States

S Supporting Information



ABSTRACT: The performance of ultrasensitive liquid chromatography and tandem mass spectrometry (LC-MS/MS) methods, such as single-cell proteomics by mass spectrometry (SCoPE-MS), depends on multiple interdependent parameters. This interdependence makes it challenging to specifically pinpoint the sources of problems in the LC-MS/MS methods and approaches for resolving them. For example, a low signal at the MS2 level can be due to poor LC separation, ionization, apex targeting, ion transfer, or ion detection. We sought to specifically diagnose such problems by interactively visualizing data from all levels of bottom-up LC-MS/MS analysis. Many software packages, such as MaxQuant, already provide such data, and we developed an open source platform for their interactive visualization and analysis: Data-driven Optimization of MS (DO-MS). We found that in many cases DO-MS not only specifically diagnosed LC-MS/MS problems but also enabled us to rationally optimize them. For example, by using DO-MS to optimize the sampling of the elution peak apexes, we increased ion accumulation times and apex sampling, which resulted in a 370% more efficient delivery of ions for MS2 analysis. DO-MS is easy to install and use, and its GUI allows for interactive data subsetting and high-quality figure generation. The modular design of DO-MS facilitates customization and expansion. DO-MS v1.0.8 is available for download from GitHub: https://github.com/SlavovLab/DO-MS. Additional documentation is available at https://do-ms.slavovlab.net.

KEYWORDS: *optimizing mass spectrometry, ultrasensitive proteomics, single-cell analysis, single-cell proteomics by mass spectrometry, quality control, visualization, MaxQuant, method development, R, Shiny*

INTRODUCTION

Analytical methods combining liquid chromatography and tandem mass spectrometry (LC-MS/MS) allow for unparalleled identification and relative quantitation of the protein components of biological systems.[1−4] Advances in LC-MS/MS have enabeled analyses of protein complexes and their functions,[5−9] regulation of protein synthesis and alternative RNA translation,[10,11] rare cells in blood,[12,13] and protein conformations.[14−16] The increasing sensitivity,[17−23] throughput, and robustness[24] of LC-MS/MS set the stage for quantifying thousands of proteins across many thousands of single cells, providing data with transformative potential for biomedical research.[25−29]

While LC-MS/MS proteomics methods are very powerful, they require extensive optimization of interdependent instrument parameters. Optimization is particularly critical for quantifying low-input samples, such as single-cell proteomes and exosomes. LC-MS/MS optimization and quality control (QC) can be performed by manually inspecting the features of peptide standards[30−32] within MS instrument software (e.g., Thermo Scientific Xcalibur) or by specialized software packages. Following the National Institute of Standards and Technology's QC data analysis pipeline,[33] numerous QC programs have been developed. These programs capitalize on advances in search engines,[34] video analysis,[35] direct analysis of raw data,[36] and database management tools[37] to track instrument performance over time and assist in duty cycle optimization. Other platforms, such as MSStatsQC 2.0, employ specialized statistical methods to differentiate normal variation in instrument performance from novel variation as a means to detect instrument problems early.[38] A comprehensive review of LC-MS/MS QC and optimization tools has been published by Bittremieux et al.[39]

We found these tools useful in developing single-cell proteomics by mass spectrometry (SCoPE-MS), which combines TMT-labeled peptides from single cells with a TMT-labeled carrier channel to enable quantifying proteins across many single cells.[25−27,40] However, none of these tools provided all of the metrics needed for optimizating our SCoPE-MS analysis.[40] This motivated us to develop DO-MS, a highly modular and interactive environment for optimizing ultrasensitive LC-MS/MS methods.

DO-MS aims to diagnose problems and suggest solutions as specifically as possible. To illustrate this point, here we describe concrete examples, including optimizing apex targeting, assessing contamination, and evaluating SCoPE-MS results. In order to enable specific diagnosis, the DO-MS dashboard juxtaposes distribution plots of data from multiple levels of LC-MS/MS analysis, including retention lengths at

Received: January 17, 2019
Published: May 13, 2019

DOI: 10.1021/acs.jproteome.9b00039
J. Proteome Res. 2019, 18, 2493−2500

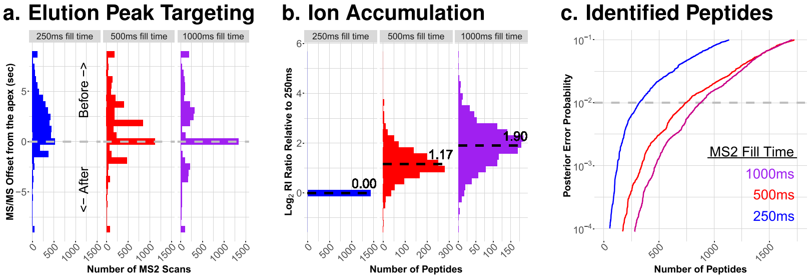


Figure 1. DO-MS-assisted optimization of instrument parameters leads to increased apex targeting, ion delivery, and peptide identification rates. (a) Distributions of apex offsets for three runs on 90 min gradients. Each injection was of 1 μL from the same vial and corresponds to 1 × M dilution of a SCoPE-MS master sample as described by Specht et al.[27] All LC parameters and instrument parameters were set to be the same except for the max fill time. (b) The relative efficiency of delivering ions for MS2 analysis was estimated by the intensities of RI. For each peptide identified across all three experiments, the RI intensity was divided by the corresponding RI intensity for 250 ms fill time, and the results for all peptides are shown as distributions on a log$_2$ scale. (c) The number of peptides identified at each PEP threshold is shown as a rank sorted list for the three fill times. This display shows the number of peptides for all levels of confidence of identification, as quantified by the PEP. The plots from panels (a−c) can be found in the "Ion Sampling" tab and in the "Peptide Identifications" tabs of DO-MS, respectively. The plot for panel (b) was normalized to the 250 ms experiment specifically for this figure to emphasize the increased ion accumulation.

the base and mid-height, intensity of all ions and of precursors selected for MS/MS, elution peak apex offset, number of MS/MS events, MS2-level coisolation (i.e., parent ion fraction), number of identified peptides at all confidence levels, and quantification benchmarks. These features are organized thematically in the dashboard for ease of reference. DO-MS has already enabled us to quickly identify problems in our methods, their exact origin in the workflow, and potential solutions. Below, we share DO-MS along with a selection of examples from our work in the hope that it will facilitate a wider adoption and advancement of single-cell proteomics. We also hope that the modular nature of our platform will enable the community to add new modules for optimizing LC-MS/MS for an expanding array of applications.

■ MATERIALS AND METHODS

Implementation

DO-MS is implemented as a Shiny app, built using R.[41] All plots are generated using the ggplot2 package.[42,43] Shiny was chosen for its interactivity, allowing data to be dynamically subset based on experiment or confidence of peptide spectral match. Additionally, this package can be run from the command line or the RStudio IDE. DO-MS v1.0.8 is available from the Slavov Lab GitHub page: https://github.com/SlavovLab/DO-MS.

DO-MS 1.0.8 requires R 3.5.2, Shiny 1.2.0, shinyWidgets 0.4.4, shinyDashboard 0.7.1, dplyr 0.7.8, tidyr 0.8.2, ggplot2 3.1.0, lattice 0.20-38, knitr 1.21, tibble 2.0.1, reshape2 1.4.3, readr 1.3.1, rmarkdown 1.11, DT 0.5, stringr 1.3.1, yaml 2.2.0, viridisLite 0.3.0, and pacman 0.5.0. The p_load() function of the pacman package will automatically check and install missing packages to ensure DO-MS has the necessary dependent packages to run. DO-MS modules are maintained for MaxQuant version 1.6.0.16. Additionally, Firefox 66.0 or Google Chrome are recommended for the best user experience. DO-MS will be maintained by the Slavov Lab to ensure continued compatibility with MaxQuant.

Data Preprocessing

The experimental data used here were generated as part of developing and optimizing minimal proteomic sample preparation (mPOP)[27] and SCoPE-MS,[40] and a full description of the experiments can be found in Specht et al.[27] Samples generated for instrument and method optimization were prepared from U937 and Jurkat cells that were lysed in HPLC-grade water according to the mPOP protocol: a 15 min freeze step at −80 °C, followed by a 10 min heating step at 90 °C. Following lysis, samples were digested at 37 °C for 3 h with 10 ng/μL of Promega Trypsin Gold in 100 mM TEAB. The bulk digested material was then serially diluted and labeled in an 11-plex design scheme with the following approximate inputs in each channel: 5000 U937 cells (126); 5000 Jurkat cells (127N); all reagents, but no cells (127C); all reagents, but no cells (128N); six alternating channels of 100 U937 cells or 100 Jurkat cells (128C-131N); no reagents or cells (131C). 1% of this bulk 100 μL sample was then injected to simulate a single SCoPE-MS experiment with two 50-cell carrier channels and six single-cell channels. Briefly, all samples were separated on a 25 cm length × 75 μm Waters nanoEase column (1.7 μm resin, Waters PN:186008795) run by a Proxeon Easy nLC1200 UHPLC (Thermo Scientific). All samples were analyzed by a Thermo Scientific Q-Exactive mass spectrometer. Prior to running DO-MS, RAW files were searched using MaxQuant 1.6.0.16.[44−46] The human SwissProt FASTA database (39,748 entries, downloaded 5/1/2018) was used for searching data from U-937 and Jurkat cells. MaxQuant searches were conducted as previously described.[27,46] Trypsin was specified as the digest enzyme, and a maximum of two missed cleavages were allowed for peptides between 5 and 26 amino acids long. Methionine oxidation (+15.99491 Da) and protein n-terminus acetylation (+42.01056 Da) were specified as variable modifications. The allPeptides.txt, evidence.txt, msmsScans.txt, and msms.txt files output by MaxQuant were imported by DO-MS for analysis and figure generation. In order to make the greatest use of the DO-MS platform, users must enable the Calculate Peak Properties option on the advanced submenu of MaxQuant's Global Parameters tab.[44,46] This option can be automatically enabled by using the mqpar.xml files from the Supporting Information of this manuscript. Searches conducted without enabling this option will not generate plots for the elution peak apex offset and peak width at full width half max panels of the DO-MS dashboard. DO-MS is currently optimized for MaxQuant search results, but users can customize it to work with alternative search engine outputs by specifying the column headers corresponding to the data selected for visualization.

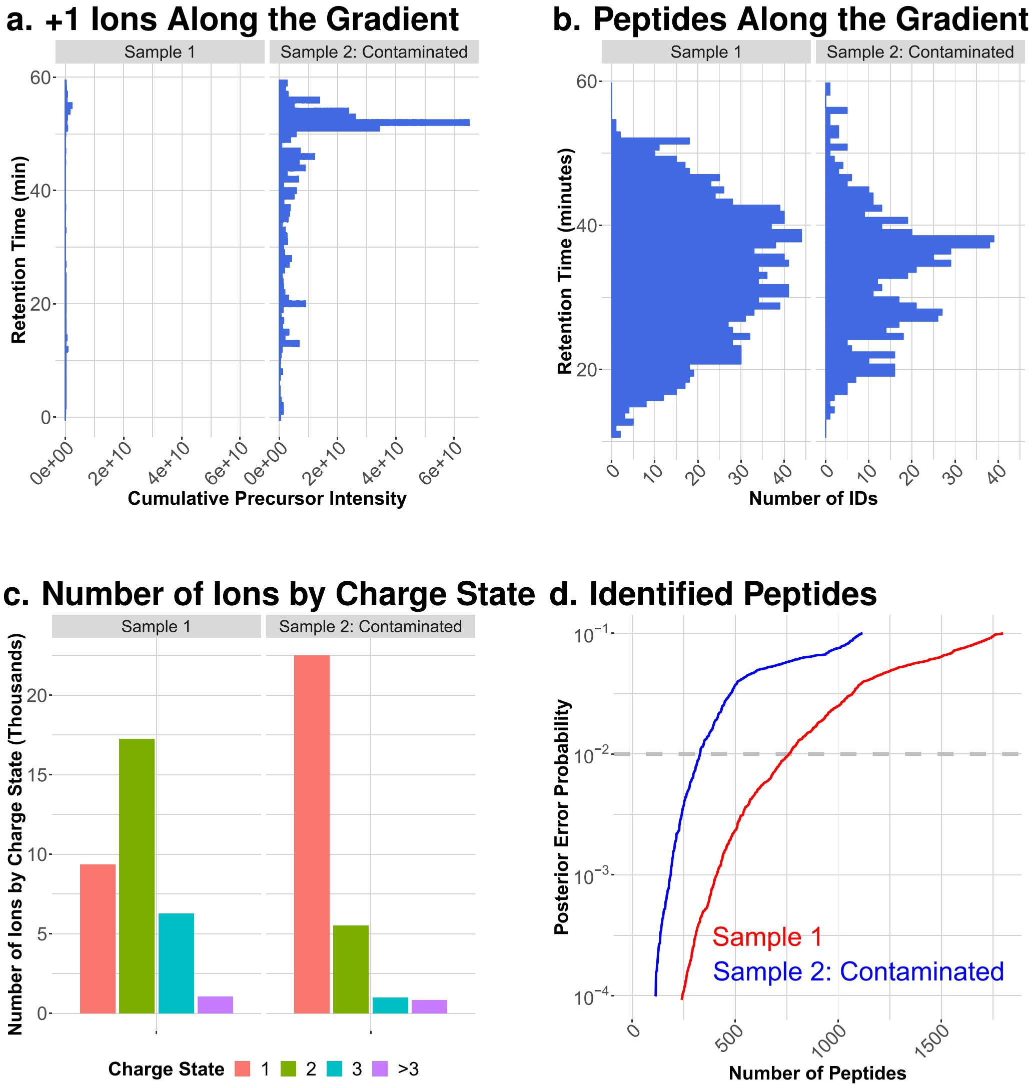


Figure 2. Diagnosing reduced peptide identification due to co-eluting contaminants. Plotting the cumulative intensities for all +1 ions detected during the survey scans (a) alongside the number of peptides identified across the gradient (b) can reveal correlations between co-eluting contaminants and reduced peptide identification. (c) Number of all detected ions by charge states. (d) Peptides were rank sorted by their PEPs to display the number of identified peptides across all levels of confidence. The plots from panels (a−c) can be found within the "Contamination" tab of DO-MS, while the plot shown in panel (d) can be found in the "Peptide Identifications" tab.

All data sets associated with this manuscript have been deposited at massIVE with the following IDs: Figure 1, apex offset data: MSV000083316; Figure 2, contamination data: MSV000083317; Figure 3, controlled comparison data: MSV000083319; and Figure 4, SCoPE-MS data: MSV000083318.

Visualization

DO-MS's diagnostic plots have been organized into the following five categories: chromatography, ion sampling, peptide identifications, contamination, SCoPE-MS diagnostics, and DART-ID.[47] Data are visualized as full distributions using vertically oriented histograms to avoid kernel-smoothing issues. This approach is advantageous, as distinct data sets may have similar summary statistics but markedly different distributions of data points.[48,49] The full distributions allow subpopulations of ions to be identified, which can be key to optimizing LC-MS/MS performance. Additionally, these distributions can be conditioned on common ions, allowing for a more principled comparison, as discussed in the Results and Discussion section.

Data imported into DO-MS can be subset based on the confidence of peptide spectral match assignment and experiment name using a slider and dynamically populated list, respectively. DO-MS relies on the posterior error probability (PEP), as estimated by MaxQuant, to indicate the confidence of a given peptide spectral match (PSM). The PEP value can be thought of as the probability that the identified peptide was not in the mass spectrometer at the time the spectra were acquired.[50] By default, DO-MS labels experiments by their corresponding raw file names. If desired, experiments can be labeled via a text-input field. Such labeling can enhance the clarity of figures and thus facilitate their analysis and broader interpretability if the figures are intended for publication.

Report Generation and Figure Output

To facilitate the sharing of experimental results, users can output the DO-MS dashboard plots as an HTML report which reflects all data subsetting performed in the app as well as user-

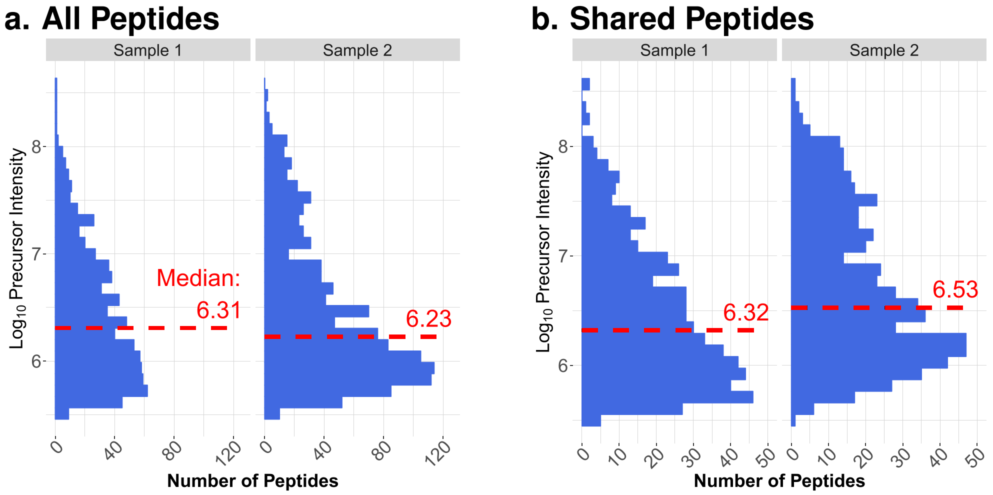


Figure 3. Controlled comparison of peptide abundances across experiments. Without controlling for the composition of the two populations being compared, trends in the data can be misread. In this case, when comparing the distribution of precursor intensities for all peptides identified in each sample (a), sample 1 appears to have more highly abundant peptides. However, when ensuring that the comparison is only based on those peptides identified in each sample (b), the opposite trend becomes apparent, namely, that the peptide species in sample 2 were more highly abundant. Both of the plots shown in Figure 3 can be found in the "Ion Sampling" tab of DO-MS.

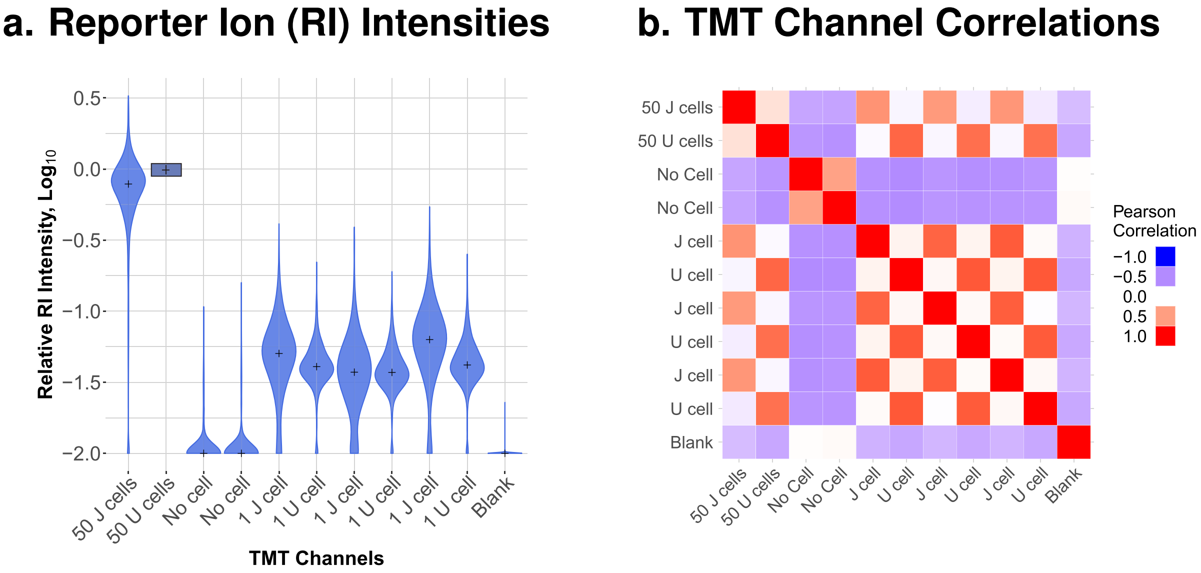


Figure 4. Evaluating low-input samples, such as SCoPE-MS sets. (a) The distributions of rRI intensities can indicate the relative amount of peptides and the efficiency of sample preparation for each channel. (b) The matrix of pairwise correlations among all all channels of a SCoPE-MS set can be used to benchmark relative quantification within that set. In (a) we expect single-cell channels to have relative rRI intensities that are 50-fold lower than the 50-cell carrier channels (about 1.7 on $\log_{10}$ scale). In (b), we expect single-cell channels to correlate positively with single-cell channels and carrier channels that contain their respective cell type, while cross-cell-type correlations for single-cell channels are expected to be negative. Both of the plots shown in Figure 4 can be found in the "SCoPE-MS Diagnostics" tab of DO-MS.

supplied experimental labels. Report generation is achieved via a button in the "Report Generation" dashboard tab. See the Supporting Information for an example HTML report. Dashboard plots can also be saved as individual .png or .pdf files for use in presentations and publications directly from the dashboard tabs on which they appear.

User Customization

We built DO-MS as a modular application to make dashboard customization and expansion as easy as possible. Each plot is generated from a separate R file, and a template file has been provided as a guide for users interested in including additional plots in their dashboard. Adding plots to the DO-MS dashboard can be accomplished by adding a customized template.r file to a new folder in the "modules" directory. After reloading the app, the new plot will automatically appear in the DO-MS dashboard in the user-specified tab. More details for this process can be found on the Github project page: https://github.com/SlavovLab/DO-MS.

■ RESULTS AND DISCUSSION

Sampling the Elution Peak Apex

LC-MS/MS methods aim to sample the elution profile of each peptide at its apex because such sampling maximizes the number and purity of sampled ions.[51] However, no existing method can target the apex of every ion. Rather, instrument parameters can be optimized to maximize the fraction of ions sent for MS/MS at or close to their apexes. DO-MS facilitates this optimization by visualizing the distribution of apex offsets (i.e., the time offset between the apex of an ion and the time when it is fragmented) as estimated by MaxQuant.

The visualization of apex offsets, as shown in Figure 1a, enabled us to rationally optimize the instrument methods for analyzing SCoPE-MS sets. To resolve the reporter ions (RI) of a TMT 11-plex, we perform the MS2 scans at 70,000 resolving power, which on our Q-exactive takes 256 ms.[52] Since Q-exactive instruments can perform the ion accumulation and MS scan in parallel, we started by setting our max fill (ion accumulation) times for MS2 scans to be 250 ms.[52] This setting resulted in sampling most ions for MS2 scans too early, significantly before the apexes of their elution peaks (Figure 1a). We reasoned that such premature sampling could be alleviated by elongating the duty cycle either by increasing the number of MS/MSed ions per cycle or by increasing the max fill times. Indeed, increasing the fill times to 500 ms and 1000 ms, while keeping all other parameters constant, increased the fraction of ions whose elution peaks are sampled at or near the apex (Figure 1a).

The increased fill times improved apex targeting and increased ion delivery for MS2 analysis (Figure 1b), leading to an increased number of identified peptides at all levels of confidence, as shown in Figure 1c. Rather than displaying the number of identified peptides at an arbitrary confidence cutoff, DO-MS plots all peptides rank-sorted by the posterior error probability (PEP) of their identification (Figure 1c). These curves show the number of identifications for all levels of confidence. The dashed gray line on the plot denotes peptides with a 99% probability of having been correctly identified. Such curves offer insight into low-confidence peptide identifications which might be boosted by incorporating additional features, such as retention time.[46,47]

This example is consistent with previous observations that longer accumulation times can increase the number of confident peptide identifications for lowly abundant samples.[17,40] Furthermore, this example underscores that fill times can strongly influence apex targeting. Such optimization of duty cycle and apex targeting is sample and system dependent, and thus it requires methods that allow for rational optimization of all levels of MS analysis, such as DO-MS.

Characterizing Contamination

Contaminants of nonprotein origin are very common in proteomics experiments.[34,53] At best, they are lowly abundant and elute separately from peptides; at worst, they are highly abundant and elute with peptides, undermining ionization, charge determination, and identification. Low-input samples are especially sensitive to contaminants, as the ratio between target and contaminant ions is more likely to be lower. DO-MS displays contaminants across the LC gradient by plotting the intensity of ions with a +1 charge state ($z = 1$) for each minute of the gradient. This type of data presentation allows users to distinguish between hydrophilic and hydrophobic contaminants.

Due to the mosaic structure of the DO-MS dashboard, potential relationships between factors such as contaminant ion intensity and peptide identifications can be easily seen, as exemplified in Figure 2. The juxtaposition of summed precursor intensities for contaminant ions (Figure 2a) and for peptides (Figure 2b) shows a clear correlation: Sample 2 has more hydrophobic contaminants, and their elution coincides with reduced peptide identifications. This correlation immediately suggested that the lower identification rates for sample 2 were due to its contamination, which we subsequently identified to be polyethylene glycol (PEG).

As an additional diagnostic plot for contamination, DO-MS displays the number of ions detected by the instrument by charge state, as shown in Figure 2c. This compact display is particularly useful for comparing ions likely to be contaminants (with charge +1) and those likely to be peptides (with charge ≥ 2) across many runs. The negative impact of these numerous contaminant ions on peptide identifications can be seen in Figure 2d. Once the presence of contaminants has been diagnosed with DO-MS, several existing software tools can be applied to more fully characterize the type of contamination present in each sample.[34,53]

Controlled Sample Comparisons

While distributions are much more informative than their summary statistics, comparing distributions for different populations can still be misleading. Thus, when comparing the distributions plotted by DO-MS, it is important to control for (condition on) the composition of the distributions, for example, the peptides comprising each distribution. The importance of such controlled comparison is exemplified in Figure 3 with an experiment testing the effect of calcium addition on trypsin digestion. For this experiment, a sample was split into two equal parts, samples 1 and 2. They were processed in identical ways, except that 50 mM $CaCl_2$ was added to sample 2 during its digestion. Both samples were digested with 20 ng/μL Promega Trypsin/LysC mix. The distributions of all peptide abundances and the corresponding median abundances (Figure 3a) may be interpreted to suggest that sample 1 resulted in more efficient delivery of peptides to MS analysis, and the addition of calcium was detrimental. This comparison, however, is complicated by the differential number of PSMs in each sample (sample 1: 817; sample 2: 1258). By conditioning the comparison on the common peptides, the opposite conclusion is indicated: The addition of calcium chloride to sample 2 during its digestion resulted in more efficient delivery of peptides to the instrument.

Low-Input Sample Diagnostics

In the process of developing mPOP[27] and SCoPE-MS,[40] we discovered a number of helpful metrics for diagnosing SCoPE-MS sample preparations and optimizing instrument parameters for low-input samples. To optimize SCoPE-MS, we used standards from which a single 1 μL injection (1% of the bulk sample volume) corresponds to the peptide input present in a single SCoPE-MS 11-plex set. Such samples allowed us to assess and optimize instrument performance independent of sample variability (since we could inject multiple aliquots of the same sample) and to assess quantification, as we had a strong expectation that the pseudo-single-cell channels should correlate positively to their corresponding carrier channel. Below we demonstrate diagnosis of such a sample by DO-MS; the sample preparation is described in detail by Specht et al.[27] Briefly, the sample was composed of 10 TMT channels containing serial dilutions of digested cell lysate from two cell types: U-937 (monocytes, denoted by U) and Jurkat (T-cells, denoted by J). This reference standard is diluted so that 1 μL corresponds to a SCoPE-MS set and contains peptide input equivalent to about 106 single cells (20−50 ng of total protein). The reference samples have two carrier channels, each of which contains peptide input comparable to 50 cells of each type, and 6 channels contain peptide inputs comparable to individual single cells, 3 of each type.[27]

The distribution of RI intensities from a SCoPE-MS set can be an informative indicator. Low RI intensities may be due to failed cell isolation, digestion, or labeling. Higher than expected RI intensities may be due to background contamination or cross-labeling. Such deviations may be diagnosed simply from the distributions of RI intensities. However, these distributions are quite broad since the RI intensities for quantified peptides often span several orders of magnitude. To decrease this dynamic range, DO-MS plots the distribution of relative RI (rRI) intensities: The RI intensities of each peptide are normalized (divided) by the RI intensity in the most abundant channel (the one with the highest median RI intensity, Figure 4a). This visualization clearly indicates whether the single-cell channels have about a 50-fold lower median intensity than a 50-cell carrier channel as well as the background signal (including isotopic contamination) in the channels without a cell. Such a diagnostic is helpful for determining label-quenching efficiency as well as the efficiency of cell sorting by FACS. By examining the rRI intensities present in channels without a cell, one can also assess the amount of background signal present and the degree of isotopic carryover.

The quantitative accuracy of SCoPE-MS sets can be benchmarked by comparing the relative quantification from the carrier channels and single-cell channels as quantified by the correlations among them. DO-MS computes all possible Pearson pairwise correlations, that is, the correlation matrix for the column and row-normalized RI intensities. This matrix can serve as a complementary diagnostic for relative quantification in SCoPE-MS sets based on the expectation that cells of the same type should correlate positively with each other but not with different cell types. Furthermore, blank channels should not correlate positively with either the single-cell or carrier channels. This expectation is consistent with the correlations shown in Figure 4b. This correlation matrix can also be useful for identifying cross-contamination, cross-labeling, and on-column carryover. In the context of our control sets, the correlations between channels corresponding to the cells from the same cell type can be interpreted as reliability estimates, that is, estimating the fraction of variance due to signal.[54] This interpretation provides a concrete and objective benchmark for the reliability of the LC-MS/MS measurements.

These SCoPE-MS plots should be analyzed in the context of distribution plots reporting on all levels of the LC-MS/MS analysis so that the origin of problems can be identified and interdependent parameters optimized. We hope that these metrics will assist the wide adoption of low-input sample preparation and analysis methods.

DART-ID Diagnostics

We recently developed a Bayesian framework for global retention time (RT) alignment and for incorporating RT estimates toward improved confidence estimates of peptide-spectrum matches (PSMs): Data-driven Alignment of Retention Times for Identification (DART-ID)[47] To visualize the performance of DART-ID, we added a dedicated tab to DO-MS. The DART-ID tab allows users to assess the impact of DART-ID on their data sets through such metrics as the number of upgraded PSMs per raw file and the residual error from global RT alignment, that is, the difference between the measured RT and reference RT inferred by DART-ID. This dashboard tab also illustrates the ease with which new modules can be added to DO-MS as users develop new methods and want to visualize new informative features of their data and analysis.

■ CONCLUSION

Optimal instrument parameters are context dependent and thus should be determined systematically by data-driven approaches for each set of samples and LC-MS/MS configurations. DO-MS enables such optimization. Sometimes, this data-driven approach results in counterintuitive results, as demonstrated in Figure 1 with the increased number of identified peptides at longer max fill times. By increasing our MS2 injection time, and consequently our duty cycle length, we managed to increase the number of confidently identified peptides in our sample. This result contrasts with the strategy commonly employed by bulk proteomics methods, namely seeking to increase peptide identification by increasing MS2 sampling frequency and thus decreasing the fill time for each MS/MS.

The DO-MS dashboard is an open-source, GUI-based tool for quickly assessing LC-MS/MS parameter optimization strategies and sample quality in single-cell proteomics experiments. This diagnostic platform can assist other laboratories in adopting ultrasensitive, low-input LC-MS/MS methods, such as SCoPE-MS, and can serve as a highly adaptable data visualization tool for proteomics researchers.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.9b00039.

DO-MS Getting Started Guide pdf: This document provides users with a step-by-step guide to using the DO-MS dashboard, and example HTML report (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: huffman.r@husky.neu.edu.
*E-mail: nslavov@northeastern.edu.

ORCID

R. Gray Huffman: 0000-0002-1579-4216
Albert Chen: 0000-0002-5387-0208
Harrison Specht: 0000-0003-3151-6803
Nikolai Slavov: 0000-0003-2035-1820

Author Contributions

‖These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The catalyst for creating this package was a breakout discussion group led by Jürgen Cox at the 2018 Single-Cell Proteomics conference. Special thanks goes to Jürgen Cox for structuring this discussion, as well as to Ezra Levy, Guillaume Harmange, and Hung-Yi Wu for their part in outlining the characteristics that an instrument-parameter optimization and sample-quality analysis software package should possess. We would also like to thank Ed Emmott, David Perlman, and Toni

Koller for their valuable feedback and troubleshooting efforts during the package development process. This project was supported by a New Innovator Award from the NIGMS from the National Institutes of Health to N.S. under award number DP2GM123497.

REFERENCES

(1) Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1998**, *33*, 1−19.

(2) Yates, J. R.; Ruse, C. I.; Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49−79.

(3) Nesvizhskii, A. I. Proteogenomics: concepts, applications and computational strategies. *Nat. Methods* **2014**, *11*, 1114.

(4) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347−355.

(5) Cravatt, B. F.; Simon, G. M.; Yates, J. R., Iii The biological impact of mass-spectrometry-based proteomics. *Nature* **2007**, *450*, 991.

(6) Leitner, A.; Faini, M.; Stengel, F.; Aebersold, R. Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. *Trends Biochem. Sci.* **2016**, *41*, 20−32.

(7) Slavov, N.; Semrau, S.; Airoldi, E.; Budnik, B.; van Oudenaarden, A. Differential stoichiometry among core ribosomal proteins. *Cell Rep.* **2015**, *13*, 865−873.

(8) Malioutov, D.; Chen, T.; Airoldi, E.; Jaffe, J.; Budnik, B.; Slavov, N. Quantifying homologous proteins and proteoforms. *Mol. Cell. Proteomics* **2019**, *18*, 162.

(9) Emmott, E. P.; Jovanovic, M.; Slavov, N. Ribosome stoichiometry: from form to function. *Trends Biochem. Sci.* **2019**, *44*, 95.

(10) Trendel, J.; Schwarzl, T.; Horos, R.; Prakash, A.; Bateman, A.; Hentze, M. W.; Krijgsveld, J. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell* **2019**, *176*, 391−403.

(11) Slavoff, S. A.; Mitchell, A. J.; Schwaid, A. G.; Cabili, M. N.; Ma, J.; Levin, J. Z.; Karger, A. D.; Budnik, B. A.; Rinn, J. L.; Saghatelian, A. Peptidomic discovery of short open reading frame−encoded peptides in human cells. *Nat. Chem. Biol.* **2013**, *9*, 59.

(12) Li, S.; Plouffe, B. D.; Belov, A. M.; Ray, S.; Wang, X.; Murthy, S. K.; Karger, B. L.; Ivanov, A. R. An Integrated Platform for Isolation, Processing, and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood. *Mol. Cell. Proteomics* **2015**, *14*, 1672−1683.

(13) Shraibman, B.; et al. Identification of Tumor Antigens Among the HLA Peptidomes of Glioblastoma Tumors and Plasma. *Mol. Cell. Proteomics* **2018**, *17*, 2132−2145.

(14) Savitski, M. M.; Reinhard, F. B. M.; Franken, H.; Werner, T.; Savitski, M. F.; Eberhard, D.; Molina, D. M.; Jafari, R.; Dovega, R. B.; Klaeger, S.; Kuster, B.; Nordlund, P.; Bantscheff, M.; Drewes, G. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **2014**, *346*, 1255784.

(15) Yates, J. R. Innovation: Structural Proteomics Goes Global. *J. Proteome Res.* **2018**, *17*, 3613−3613.

(16) Kaur, U.; Meng, H.; Lui, F.; Ma, R.; Ogburn, R. N.; Johnson, J. H. R.; Fitzgerald, M. C.; Jones, L. M. Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *J. Proteome Res.* **2018**, *17*, 3614−3627.

(17) Kelstrup, C. D.; Young, C.; Lavallee, R.; Nielsen, M. L.; Olsen, J. V. Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* **2012**, *11*, 3487−3497.

(18) Virant-Klun, I.; Leicht, S.; Hughes, C.; Krijgsveld, J. Identification of maturation-specific proteins by single-cell proteomics of human oocytes. *Mol. Cell. Proteomics* **2016**, *15*, 2616−2627.

(19) Lombard-Banek, C.; Moody, S. A.; Nemes, P. Single-Cell Mass Spectrometry for Discovery Proteomics: Quantifying Translational Cell Heterogeneity in the 16-Cell Frog (Xenopus) Embryo. *Angew. Chem., Int. Ed.* **2016**, *55*, 2454−2458.

(20) Cifani, P.; Kentsis, A. High Sensitivity Quantitative Proteomics Using Automated Multidimensional Nano-flow Chromatography and Accumulated Ion Monitoring on Quadrupole-Orbitrap-Linear Ion Trap Mass Spectrometer. *Mol. Cell. Proteomics* **2017**, *16*, 2006−2016.

(21) Shao, X.; Wang, X.; Guan, S.; Lin, H.; Yan, G.; Gao, M.; Deng, C.; Zhang, X. Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. *Anal. Chem.* **2018**, *90*, 14003−14010.

(22) Shishkova, E.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. Ultra-High Pressure (>30,000 psi) Packing of Capillary Columns Enhancing Depth of Shotgun Proteomic Analyses. *Anal. Chem.* **2018**, *90*, 11503−11508.

(23) Macron, C.; Lane, L.; Núñez Galindo, A.; Dayon, L. Deep Dive on the Proteome of Human Cerebrospinal Fluid: A Valuable Data Resource for Biomarker Discovery and Missing Protein Identification. *J. Proteome Res.* **2018**, *17*, 4113−4126.

(24) Bache, N.; Geyer, P. E.; Bekker-Jensen, D. B.; Hoerning, O.; Falkenby, L.; Treit, P. V.; Doll, S.; Paron, I.; Müller, J. B.; Meier, F.; Olsen, J. V.; Vorm, O.; Mann, M. A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics. *Mol. Cell. Proteomics* **2018**, *17*, 2284−2296.

(25) Levy, E.; Slavov, N. Single cell protein analysis for systems biology. *Essays Biochem.* **2018**, *62*, 595−605.

(26) Specht, H.; Slavov, N. Transformative opportunities for single-cell proteomics. *J Proteome Res* **2018**, *17*, 2565−2571.

(27) Specht, H.; Harmange, G.; Perlman, D. H.; Emmott, E.; Niziolek, Z.; Budnik, B.; Slavov, N. Automated sample preparation for high-throughput single-cell proteomics. 2018 https://www.biorxiv.org/content/10.1101/399774v1 (accessed May 28, 2019).

(28) Yates, J. R. Innovations in Proteomics: The Drive to Single Cells. *J. Proteome Res.* **2018**, *17*, 2563−2564.

(29) Packer, J.; Trapnell, C. Single-Cell Multi-omics: An Engine for New Quantitative Models of Gene Regulation. *Trends Genet.* **2018**, *34*, 653−665.

(30) Burkhart, J. M.; Premsler, T.; Sickmann, A. Quality control of nano-LC-MS systems using stable isotope-coded peptides. *Proteomics* **2011**, *11*, 1049−1057.

(31) Abbatiello, S. E.; et al. Design, Implementation and Multisite Evaluation of a System Suitability Protocol for the Quantitative Assessment of Instrument Performance in Liquid Chromatography-Multiple Reaction Monitoring-MS (LC-MRM-MS). *Mol. Cell. Proteomics* **2013**, *12*, 2623−2639.

(32) Gallien, S.; Bourmaud, A.; Domon, B. A Simple Protocol To Routinely Assess the Uniformity of Proteomics Analyses. *J. Proteome Res.* **2014**, *13*, 2688−2695.

(33) Rudnick, P. A.; et al. Performance Metrics for Liquid Chromatography-Tandem Mass Spectrometry Systems in Proteomics Analyses. *Mol. Cell. Proteomics* **2010**, *9*, 225−241.

(34) Bielow, C.; Mastrobuoni, G.; Kempa, S. Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J. Proteome Res.* **2016**, *15*, 777−787.

(35) Scheltema, R. A.; Mann, M. SprayQc: A Real-Time LCMS/MS Quality Monitoring System To Maximize Uptime Using Off the Shelf Components. *J. Proteome Res.* **2012**, *11*, 3458−3466.

(36) Trachsel, C.; Panse, C.; Kockmann, T.; Wolski, W. E.; Grossmann, J.; Schlapbach, R. rawDiag: An R Package Supporting Rational LCMS Method Optimization for Bottom-up Proteomics. *J. Proteome Res.* **2018**, *17*, 2908−2914.

(37) Bittremieux, W.; Willems, H.; Kelchtermans, P.; Martens, L.; Laukens, K.; Valkenborg, D. iMonDB: Mass Spectrometry Quality Control through Instrument Monitoring. *J. Proteome Res.* **2015**, *14*, 2360−2366.

(38) Dogu, E.; Taheri, S. M.; Olivella, R.; Marty, F.; Lienert, I.; Reiter, L.; Sabido, E.; Vitek, O. MSstatsQC 2.0: R/Bioconductor Package for Statistical Quality Control of Mass Spectrometry-Based Proteomics Experiments. *J. Proteome Res.* **2019**, *18*, 678−686.

DOI: 10.1021/acs.jproteome.9b00039

(39) Bittremieux, W.; Valkenborg, D.; Martens, L.; Laukens, K. Computational quality control tools for mass spectrometry proteomics. *Proteomics* **2017**, *17*, 1600159.

(40) Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **2018**, *19*, 161.

(41) Team, R. C. R. *A language and environment for statistical computing*; The R Foundation for Statistical Computing: Vienna, Austria, 2013; http://www.R-project.org/.

(42) Wickham, H. *ggplot2: Elegant Graphics for Data Analysis (Use R!)*, 2nd ed.; Springer International Publishing: New York, 2016.

(43) Wilkinson, L. *The Grammar of Graphics (Statistics and Computing)*, 2nd ed.; Springer-Verlag: New York, 2005.

(44) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.

(45) 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*, 1794−1805.

(46) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301−2319.

(47) Chen, A.; Franks, A.; Slavov, N. DART-ID increases single-cell proteome coverage, *PLoS Comput. Biol.* 2019, in press DOI: 10.1371/journal.pcbi.1007082.

(48) Anscombe, F. J. Graphs in Statistical Analysis. *Am. Stat.* **1973**, *27*, 17−21.

(49) Gatto, L.; Breckels, L. M.; Naake, T.; Gibb, S. Visualization of proteomics data using R and bioconductor. *Proteomics* **2015**, *15*, 1375−1389.

(50) Käll, L.; Storey, J. D.; MacCoss, M. J.; Noble, W. S. Posterior Error Probabilities and False Discovery Rates: Two Sides of the Same Coin. *J. Proteome Res.* **2008**, *7*, 40−44.

(51) Savitski, M. M.; Sweetman, G.; Askenazi, M.; Marto, J. A.; Lang, M.; Zinn, N.; Bantscheff, M. Delayed Fragmentation and Optimized Isolation Width Settings for Improvement of Protein Identification and Accuracy of Isobaric Mass Tag Quantification on Orbitrap-Type Mass Spectrometers. *Anal. Chem.* **2011**, *83*, 8959−8967.

(52) Michalski, A.; Damoc, E.; Hauschild, J.-P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **2011**, *10*, M111.011015.

(53) Rardin, M. J. Rapid Assessment of Contaminants and Interferences in Mass Spectrometry Data Using Skyline. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1327−1330.

(54) Franks, A.; Airoldi, E.; Slavov, N. Post-transcriptional regulation across human tissues. *PLoS Comput. Biol.* **2017**, *13*, No. e1005535.