Published in final edited form as:

Nat Methods. 2017 June 29; 14(7): 646–647. doi:10.1038/nmeth.4334.

PIQED: automated identification and quantification of protein modifications from DIA-MS data

Jesse G Meyer¹, Sushanth Mukkamalla¹, Hanno Steen², Alexey I Nesvizhskii^{3,4}, Bradford W Gibson¹, and Birgit Schilling¹

¹Buck Institute for Research on Aging, Novato, California, USA

²Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA

³Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA

⁴Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA

To the Editor

Label-free quantification using mass spectrometric data-independent acquisition (DIA) is a robust method for deep and accurate proteome quantification¹. Here we report PIQED, a work-flow and open-source software tool with a graphical user interface (GUI), that enables completely automated post-translational modification (PTM) identification and quantification exclusively from DIA (Fig. 1a) (<https://github.com/jgmeyerucsd/PIQEDia/releases/tag/v0.1.2>).

When a pre-existing spectral library is not available, as is often the case with novel PTMs, samples are typically analyzed several times. One or more data-dependent acquisition (DDA) analyses are done to generate a spectral library, then DIA-MS is performed for quantification. This combined strategy is prohibitive when many biological replicates are required or sample quantity is limited. A recently developed software tool, DIA-Umpire, allows combined peptide identification and quantification from a single DIA analysis without a pre-existing spectral library^{2,3}. However, DIA-Umpire was originally designed for protein-level quantification and lacks the capability to quantify post-translationally modified peptides.

We developed PIQED to address this need. The PTM-specific capabilities of PIQED include site localization scoring and filtering, peptide consolidation to modification site level,

Correspondence should be addressed to B.S. (bschilling@buckinstitute.org).

AUTHOR CONTRIBUTIONS

J.G.M. and B.S. designed the study and experiments; J.G.M. and S.M. developed the software code; H.S. contributed additional data sets; A.I.N. and B.W.G. advised on software and feature optimization.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

optional local or global total-ion chromatogram normalization, and optional PTM-level correction by protein-level changes. PIQED also provides automatic, parallelized execution of all workflow steps (not included in the original DIA-Umpire program), including file conversion to mzXML, integrated identification (using MS-GF+, X! Tandem, and/or Comet), PeptideProphet, iProphet, or PTMProphet peptide identification refinement, targeted signal re-extraction using Skyline, and fragment-level signal filtering with significance testing using mapDIA (Supplementary Note and Supplementary Figs. 1 and 2). Further, PIQED enables a twofold reduction of acquisition time, because identification and quantification are achieved in a single DIA analysis.

We benchmarked the PIQED workflow using acetylated peptides enriched by immunoprecipitation (Supplementary Methods). Peptides identified from pseudo-MS/MS spectra are selected for integration by Skyline⁴ according to mProphet modeling (Fig. 1b). Unlike methods using external DDA libraries, targeted extraction in Skyline is performed using signals confidently detected by DIA-Umpire, ensuring that mProphet can easily separate true and false peak score distributions. Rare peak integration errors are filtered after integration by mapDIA⁵, which uses several fragment-level interference filters. Coefficients of variation (CVs) for these peptides among three technical replicates were excellent; nearly 70% of CVs were below 10%, and 90% of CVs were below 20% (Fig. 1c). mapDIA filtering and significance testing of site-specific lysine acetylation resulted in accurate and precise quantification when we compared technical replicates using 0.5-fold or 1-fold volume injections of enriched acetylated peptides (Fig. 1d). We identified a total of $1,385 \pm 21$ acetylated peptides in 3 replicates, indicating the consistency of PIQED analysis (Fig. 1e).

We also reanalyzed a published Q-Exactive DIA data set from human urine⁶. Although no phosphopeptide enrichment was performed, PIQED confidently identified and quantified 448 unique phosphopeptides, localizing to 159 phosphorylation sites in 81 proteins. When the optional protein quantity correction was used (Supplementary Note), we retained 134 phosphorylation sites from 67 proteins. Although human serum albumin (HSA) was not changed at the protein level, PIQED analysis automatically identified and quantified phosphorylation of Ser82 (p-Ser82) in HSA, which was significantly increased in urine from children diagnosed with urinary tract infection (39-fold increase with $FDR = 1.1 \times 10^{-6}$), but not ovarian cyst, relative to the control group (Supplementary Fig. 3) using any combination of normalization options (Supplementary Fig. 4). Additional examples of annotated phosphopeptide peptide-spectrum matches are presented in Supplementary Figure 5.

In conclusion, PIQED provides a complete, automated workflow for PTM identification, quantification, and statistical testing from DIA-MS data.

Data availability statement

The current PIQED software is available from GitHub (<https://github.com/jgmeyerucsd/PIQEDia/releases/tag/v0.1.2>) under GNU General Public License v3.0. Updated and future software versions will be available from <https://github.com/jgmeyerucsd/PIQEDia>. The raw data for the acetyllysine study are available at MassIVE using data set identifier MSV000080189. The raw data from the urine reanalysis are available as originally

published at PeptideAtlas (data set identifier PASS00706). Source data for Figure 1 are available online.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank B. MacLean for assistance with Skyline. This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (R24DK085610 and R01DK090242, B.W.G.), the National Institute of General Medical Sciences (5R01GM094231 to A.I.N.), the National Institute of Allergy and Infectious Disease (R01AI099204 to H.S.), and the Thrasher Research Fund (to H.S.). J.G. Meyer was supported by a US National Institutes of Health (NIH) grant (T32G000266; J. Campisi, principal investigator). We acknowledge support from the NIH shared instrumentation grant for the TripleTOF system at the Buck Institute (1S10 OD016281, B.W.G.).

References

1. Gillet LC, et al. Mol Cell Proteomics. 2012; 11:016717. [PubMed: 22261725]
2. Tsou C-C, et al. Nat Methods. 2015; 12:258–264. [PubMed: 25599550]
3. Li Y, et al. Nat Methods. 2015; 12:1105–1106. [PubMed: 26436481]
4. Egertson JD, MacLean B, Johnson R, Xuan Y, MacCoss MJ. Nat Protoc. 2015; 10:887–903. [PubMed: 25996789]
5. Teo G, et al. J Proteomics. 2015; 129:108–120. [PubMed: 26381204]
6. Muntel J, et al. J Proteome Res. 2015; 14:4752–4762. [PubMed: 26423119]

Figure 1.

Automated identification and quantification of PTMs using PIQED. **(a)** PIQED workflow for identification and quantification of PTMs. Modified peptides from affinity enrichment, or peptides from proteome digestion without enrichment, are analyzed by DIA-MS. Raw instrument files are then used to automatically identify, quantify, filter, normalize, and test for statistically significant regulation of arbitrary PTMs. UTI, urinary tract infection; OC, ovarian cyst. **(b)** mProphet composite score distributions of target and second-best peaks picked by Skyline showing high-confidence peak picking of peptides identified using pseudo-MS/MS spectra from DIA-MS. **(c)** Distribution of coefficient of variations for peak areas observed from three technical replicates for 1,760 acetylated peptide precursors using $1\times$ injection volume. **(d)** Observed distributions of $\log_2$ (fold change) computed by mapDIA using 3 technical replicates of $1\times$ injection volume compared to 3 technical replicates of

0.5× injection volume; expected $\log_2$ (fold change) = 1. FC, fold change. (e) Number of identified peptides from each injection replicate (1–3) and from combinations of replicates 1 and 2 or 1–3.