PIQED: Automated Identification and Quantification of Protein Modifications from DIA-MS data

Jesse G. Meyer[1], Sushanth Mukkamalla[1], Hanno Steen[2], Alexey I. Nesvizhskii[3,4], Bradford W. Gibson[1], Birgit Schilling[1]*

Affiliations

[1]Buck Institute for Research on Aging, Novato, California, USA.

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

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

[4]Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA.

*Corresponding author

Birgit Schilling, bschilling@buckinstitute.org

To the editor: Label-free quantification using mass spectrometric data-independent acquisition (DIA) is a robust method for deep and accurate proteome quantification[1]. Here we report a workflow and open-source software tool with a graphical user interface (GUI), PIQED, to enable completely automated **P**ost-translational modification (PTM) **I**dentification and **Q**uantification **E**xclusively from **D**IA **(figure 1a;** https://github.com/jgmeyerucsd/PIQEDia/releases/tag/v0.1.2**)**.

When lacking a pre-existing spectral library, as is often the case with studies of novel PTMs, samples are typically analyzed several times; one or more data-dependent acquisitions (DDA) analyses are performed to generate a spectral library, followed by DIA-MS for quantification. This combined DDA/DIA 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 PTM-modified peptides.

We developed PIQED to address this need. PTM-specific capabilities of PIQED include site-localization scoring and filtering, peptide consolidation to modification site-level, optional local or global total-ion chromatogram (TIC) 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 MSGF+, X!Tandem, and/or Comet), PeptideProphet/iProphet/PTMProphet peptide identification refinement, targeted signal re-extraction using Skyline, and fragment-level signal filtering with significance testing using

mapDIA (see **Supplementary Note and Supplementary Figures 1 and 2**). Further, PIQED enables a two-fold reduction of acquisition time because both identification and quantification are achieved with 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 Skyline[4] according to mProphet modeling **(figure 1b)**; unlike using external DDA libraries, targeted extraction in Skyline is performed using signals confidently detected by DIA-Umpire, ensuring that mProphet can easily separate the true and false peak score distributions. Rare peak integration errors are filtered post-integration by mapDIA[5], which employs 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% **(figure 1c)**. mapDIA filtering and significance testing of site-specific lysine acetylation resulted in accurate and precise quantification when comparing technical replicates of 0.5X or 1X quantity injections of enriched acetylated peptides (**figure 1d**). We identified a total of 1,385 ± 21 acetylated peptides in 3 replicates, indicating the consistency of PIQED analysis **(figure 1e)**.

We further reanalyzed a published Q-Exactive DIA dataset from human urine[6]. 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 (see 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 pSer-87 in HSA, which was significantly increased in urine from children diagnosed with urinary tract infection, but not ovarian cyst, relative to the undiagnosed pain group (**Supplementary figure 3**) using any combination of normalization options (**Supplementary figure 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 exclusively DIA-MS data.

Data Availability Statement

The current PIQED software is available from GitHub (https://github.com/jgmeyerucsd/PIQEDia/releases/tag/v0.1.2), 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 acetyl-lysine study is available at MassIVE using dataset identifier MSV000080189 (ftp://massive.ucsd.edu/MSV000080189). The raw data from the urine reanalysis is available as originally published at PeptideAtlas (dataset identifier: PASS00706, ftp://PASS00706:YP9554a@ftp.peptideatlas.org/).

Acknowledgments

This work was supported by the NIDDK (R24DK085610, B.W.G., and R01DK090242, B.W.G., both Subcontracts), NIGMS (5R01GM094231, A.I.N.), NIAID (R01AI099204, H.S.), and Thrasher Research Fund (H.S.). J.G. Meyer was supported by an NIH T32 (T32G000266, PI: J. Campisi).

We acknowledge support from the NIH shared instrumentation grant for the TripleTOF system at the Buck Institute (1S10 OD016281, B.W.G.). We thank B. MacLean for assistance with Skyline.

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, and A.I.N and B.W.G advised on software and feature optimization.

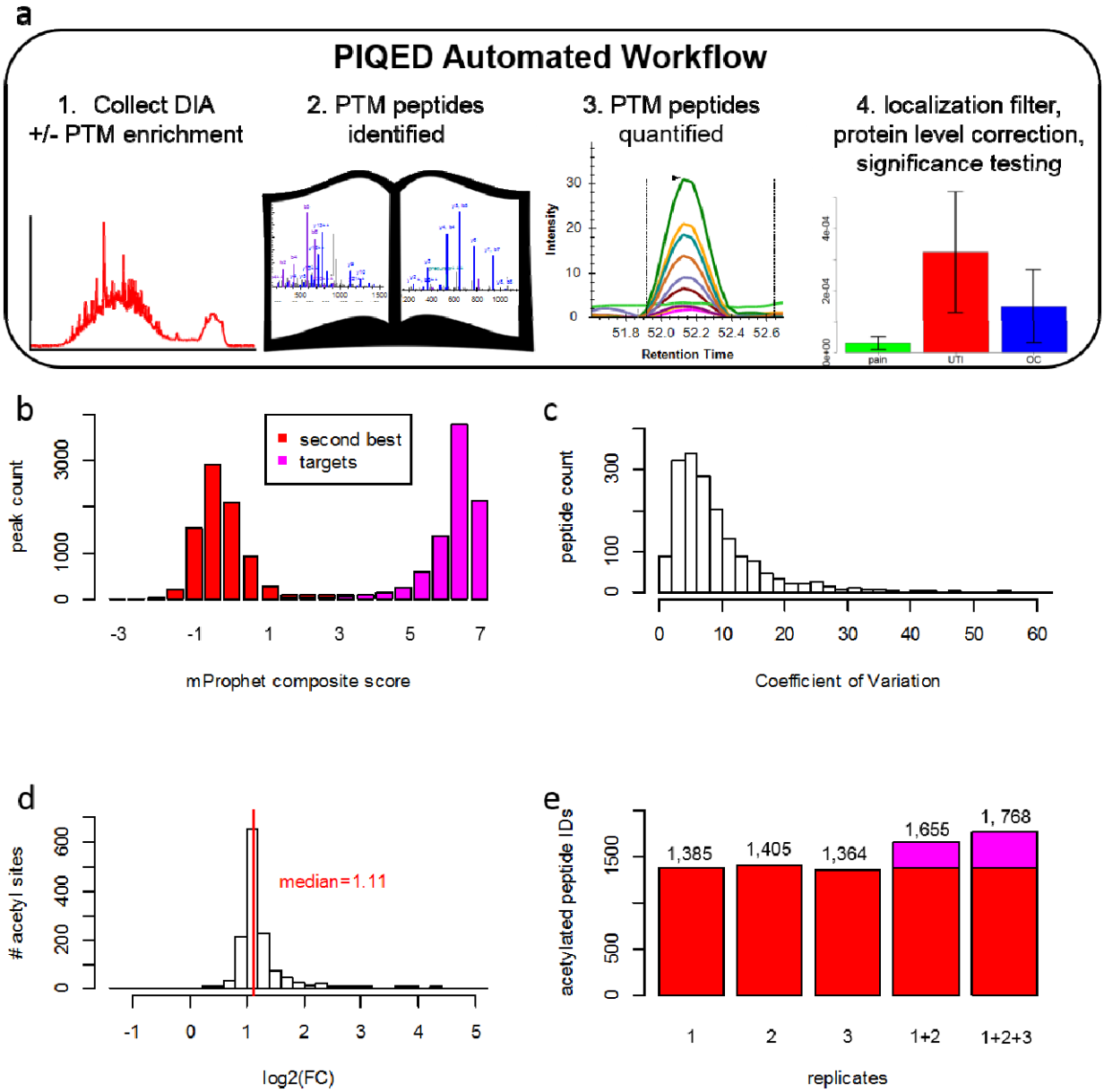
COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References

1. Gillet, L. C. *et al.* Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteomics* **11,** (2012).
2. Tsou, C.-C. *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nat Meth* **12,** 258–264 (2015).
3. Li, Y. *et al.* Group-DIA: analyzing multiple data-independent acquisition mass spectrometry data files. *Nat Meth* **12,** 1105–1106 (2015).
4. Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y. & MacCoss, M. J. Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nat Protoc.* **10,** 887–903 (2015).

5. Teo, G. *et al.* mapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *J. Proteomics* **129,** 108–120 (2015).
6. Muntel, J. *et al.* Advancing Urinary Protein Biomarker Discovery by Data-Independent Acquisition on a Quadrupole-Orbitrap Mass Spectrometer. *J. Proteome Res.* **14,** 4752–4762 (2015).


a
PIQED Automated Workflow
1. Collect DIA
+/- PTM enrichment
2. PTM peptides
identified
3. PTM peptides
quantified
4. localization filter,
protein level correction,
significance testing
Intensity
Retention Time
pain
UTI
OC
b
peak count
second best
targets
mProphet composite score
c
peptide count
Coefficient of Variation
d
# acetyl sites
median=1.11
log2(FC)
e
acetylated peptide IDs
1,385
1,405
1,364
1,655
1, 768
1
2
3
1+2
1+2+3
replicates

Main figure Legend

Figure 1: Automated Identification and Quantification of Post-Translational Modifications 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. (**B**) mProphet composite score distributions of target and second-best peaks picked by Skyline showing high-confidence peak picking by Skyline 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 the 1X injection volume . CVs from the 0.5x injections were similar (data not shown). (**D**) Observed distributions of $\log_2$(fold change) computed by mapDIA using three technical replicates of 1X injection volume compared to three technical replicates of 0.5X injection volume; expected $\log_2$(fold change) value =1. (**E**) Number of identified peptides from each single replicate injection and from the combination of two or three replicates.