NIST National Institute of Standards and Technology U.S. Department of Commerce

Mass Spectrometry Data Center

An XIC-centric approach for improved identification, quantification, and reproducibility in proteomic data analyses

Guanghui Wang, Zheng Zhang, Yi Liu, Meghan C. Burke, Sergey L. Sheetlin, Stephen E. Stein
Mass Spectrometry Data Center, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899 USA

ASMS 2023

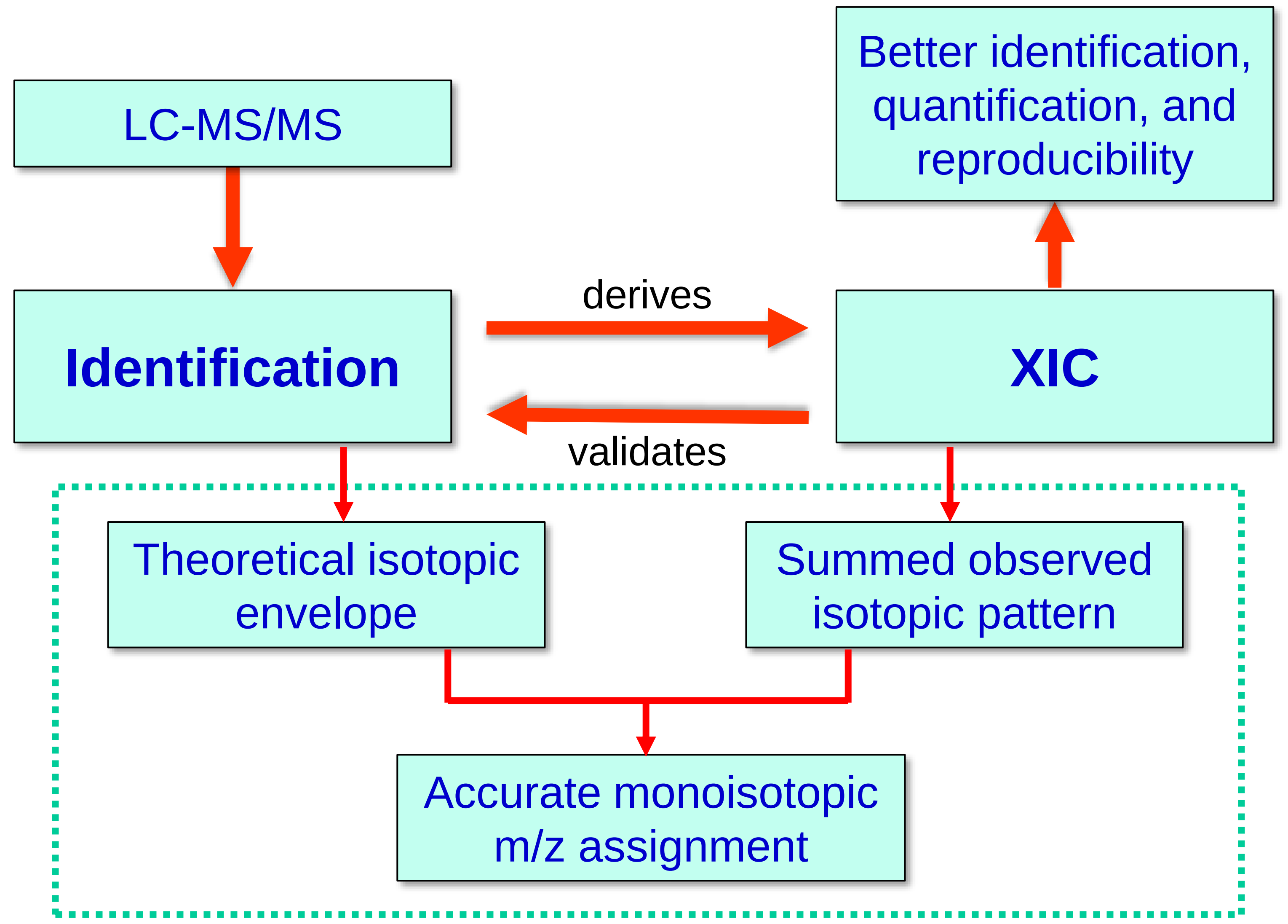
peptide.nist.gov

Introduction

Reproducibility is a “Proteomic Dream” yet to be fully realized. A significant source of variation comes from inaccuracies in data processing which include false-positive identifications. In a typical data analysis workflow for bottom-up proteomics, MS/MS spectra are first matched to peptides, followed by quantification utilizing extracted ion chromatograms (XICs) from MS scans. A drawback in this seemingly intuitive workflow is that identification-to-quantification is often considered a one-way street. Here we propose an XIC-centric approach where XICs are deemed the fundamental building blocks for both identification and quantification between which the data flow is bidirectional: identifications are used to derive XICs whose information is in turn applied to validate the identifications.

Methods

Materials used in this study included proteins extracted from human hairs as well as single glycoproteins obtained from commercial sources. Protein samples were denatured, reduced, and alkylated, followed by digestion with trypsin and/or other enzymes. High-resolution LC-MS/MS data for replicate injections of each digest were generated on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Database searches were carried out with MSFragger (University of Michigan) or Byonic (Protein Metrics). In-house XIC-based analysis software was developed in C++ to validate identifications and identify false positives through theoretical isotope envelope prediction, monoisotope determination, identification grouping, and ambiguity discovery. A graphical user interface, XIC Browser, was also created as a Windows program to visualize results for user inspection of analysis details.


LC-MS/MS

Identification

derives

validates

XIC

Better identification, quantification, and reproducibility

Theoretical isotopic envelope

Summed observed isotopic pattern

Accurate monoisotopic m/z assignment

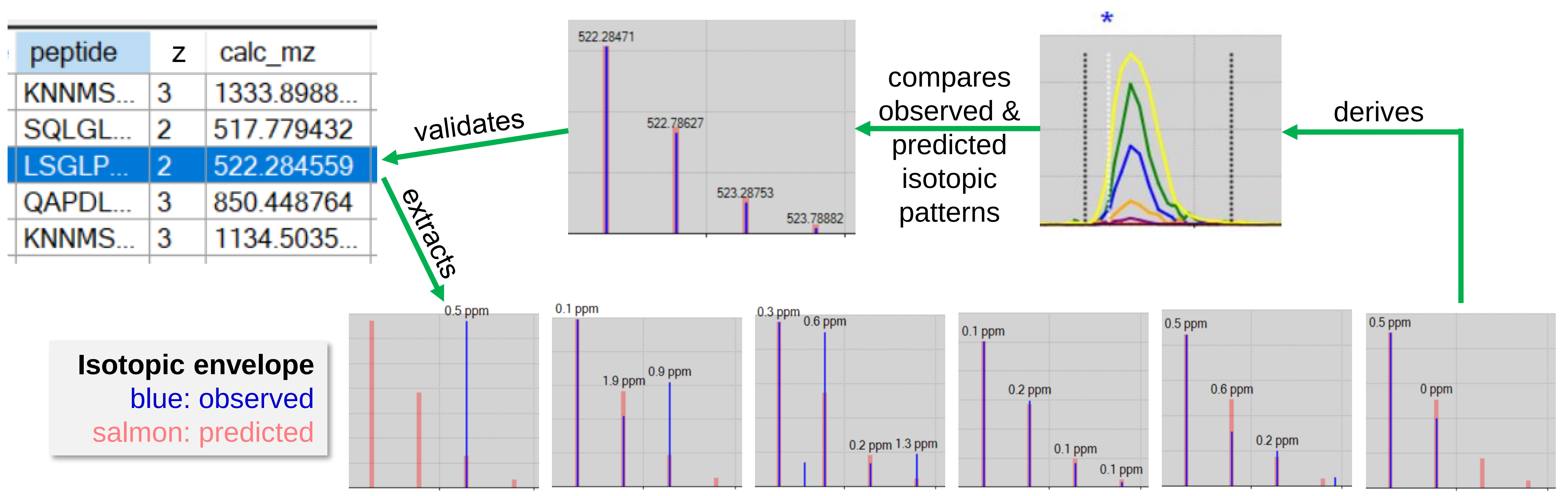
Disclaimer: Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure identified adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Results

1

In the XIC-centric approach, XICs are no longer just a convenient post-identification metric for quantification but play an essential role in validating identifications. This is achieved by utilizing the intensity information of each detected isotope for an identification across all MS1 scans within an XIC peak to obtain an accurate representation of the experimentally observed isotopic pattern. The resulting pattern is compared to the theoretical isotopic envelope of the identification to determine its monoisotopic m/z which is then applied to either confirm or invalidate the identification.

Note: unless stated otherwise, the y-axis for all plots in the Results section is abundance while the x-axis is either m/z (for isotope plots) or retention time (RT; for XIC plots).



peptide	z	calc_mz
KNNMS...	3	1333.8988...
SQLGL...	2	517.779432
LSGLP...	2	522.284559
QAPDL...	3	850.448764
KNNMS...	3	1134.5035...

validates

extracts

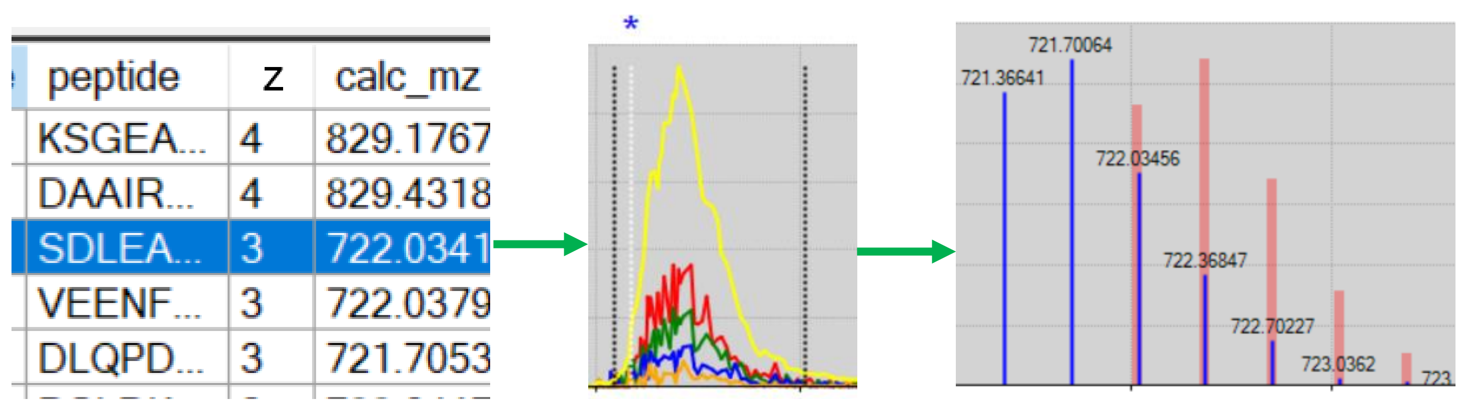
compares observed & predicted isotopic patterns

derives

Isotopic envelope
blue: observed
salmon: predicted

2

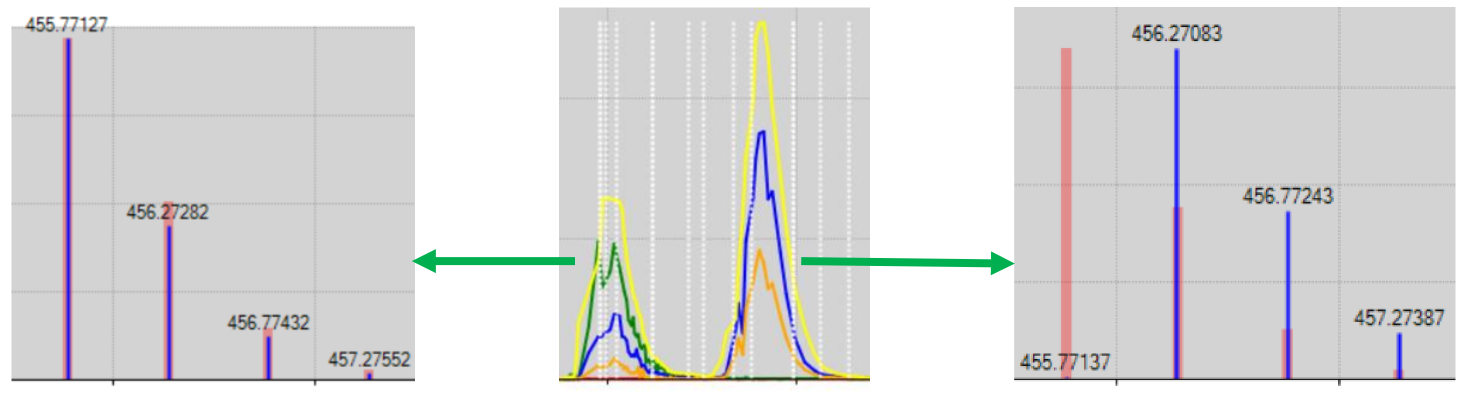
Identifies potential false positives. When an MS/MS is acquired at the beginning or ending of an XIC peak or at a place where signals are low, its precursor’s monoisotopic m/z determined from a neighboring MS1 scan is error-prone and will likely result in an incorrect identification. As shown in the example below, a peptide identification (left panel) was made for an MS/MS spectrum taken at the early rising phase (starred white line; middle panel) of its associated XIC peak, but appeared to be a false positive as its theoretical isotopic envelope (salmon; right panel) shifted 2 Da from the actual isotopic pattern (blue) established for the precursor ion from the XIC peak. Identifications whose theoretical m/z values differ from experimentally observed ones are likely false positives.



peptide	z	calc_mz
KSGEA...	4	829.1767
DAAIR...	4	829.4318
SDLEA...	3	722.0341
VEENF...	3	722.0379
DLQPD...	3	721.7053

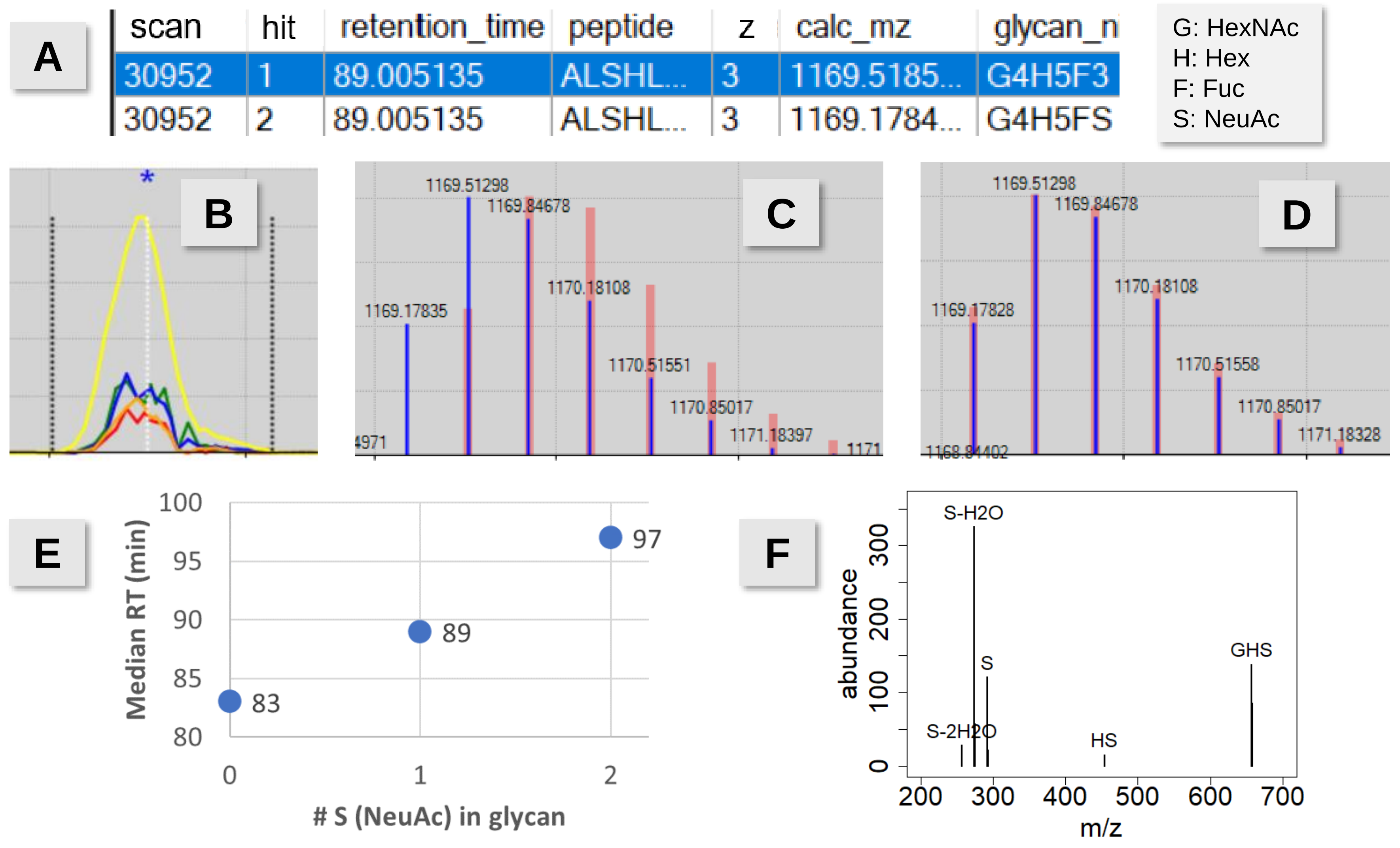
3

Improves quantification. When multiple instances (hits) of a peptide ion are identified across several distinct XIC peaks, either a sum of the peak areas or the largest one could be used to represent its abundance. But both methods would give a wrong answer in cases similar to the one illustrated below. Hits (white lines; middle panel) associated with the second XIC peak were false identifications as their theoretical mass (salmon; right panel) was 1 Da lower than the experimentally determined monoisotope (blue), while hits for the first XIC peak were validated by the observed isotopic pattern (left panel). Picking the correct XIC peak is vital in generating reproducible quantification results.



4

Rescues glycopeptide identifications. Sugars differing by 1 or 2 Da (e.g., F2 & S, G2H2 & FS2) in masses add another layer of complexity to the already challenging glycopeptide identification process. The XIC-centric approach comes to the rescue by providing accurate assignment of monoisotopes, which is exemplified by a case where the top hit for an MS/MS scan (starred white line; B) was a peptide with the glycan G4H5F3 (A). The hit was incorrect, evidenced by a mismatch between predicted and measured isotopic envelopes (C). An alternative assignment of G4H5FS (2nd entry; A) appeared correct (D), supported by RT (E) and the presence of S-containing oxonium ions in the MS/MS spectrum (F).



A

scan	hit	retention_time	peptide	z	calc_mz	glycan_n
30952	1	89.005135	ALSHL...	3	1169.5185...	G4H5F3
30952	2	89.005135	ALSHL...	3	1169.1784...	G4H5FS

G: HexNAc
H: Hex
F: Fuc
S: NeuAc

Summary & Conclusions

- ❑ We propose an XIC-centric approach where the rich information contained in XICs is exploited to increase identification confidence and detect potential false positives.
- ❑ Taking advantage of isotopic information across the whole range of an XIC peak uncovers misidentifications originating from mass assignment errors, which particularly benefits the analysis of glycopeptides owing to their relatively low abundances and glycan diversity.
- ❑ This will ultimately lead to improved reproducibility in quantitative proteomic data analyses.