Title: GlyCombo enables rapid, complete glycan composition identification across diverse glycomic sample types

Authors: Maia I Kelly[1,2] and Christopher Ashwood[1*]

Affiliations: 1. Protea Glycosciences Pty Ltd, Wollongong, New South Wales 2500, Australia.

2. College of Computing. Georgia Institute of Technology, Atlanta, Georgia 30332, USA

Address reprint requests to Christopher Ashwood, Wollongong, 2500, Australia. +61466331152. chris@proteaglyco.com,

Abstract

Glycans are sugar-based polymers found to modify biomolecules including lipids and proteins, as well as occur unconjugated as free polysaccharides. Due to their ubiquitous cellular presentation, glycans mediate crucial biological processes and are frequently sought after as biomarkers for a wide range of diseases. Identification of glycans present in samples acquired with mass spectrometry (MS) is a cornerstone of glycomics research, thus, the ability to rapidly identify glycans in each acquisition is integral to glycomics analysis pipelines. Here we introduce GlyCombo (https://github.com/Protea-Glycosciences/GlyCombo), an open-source, freely available software tool designed to rapidly assign monosaccharide combinations to glycan precursor masses including those subjected to MS2 in LC-MS/MS experiments.

GlyCombo was evaluated across six diverse datasets, demonstrating MS vendor, derivatization, and glycan-type neutrality. Compositional assignments using GlyCombo are shown to be faster than the current, predominant approach, GlycoMod, a closed-source web application. Two unique features of GlyCombo, multiple adduct search and off-by-one error anticipation, reduced unassigned MS2 scans in a benchmark dataset by 40%. Finally, the comprehensiveness of glycan feature identification is exhibited in Skyline, a software that requires pre-defined transitions that are derived from GlyCombo output files.

Introduction

Glycans are comprised of individual monosaccharides that are capable of modifying diverse biomolecules such as lipids and proteins. Additionally, glycans can exist as unconjugated free polysaccharides. Glycosylation processes in nature are comprised of a wide range of monosaccharides and their derivatives, numbering in at least the hundreds[1]. There are a variety of methods that have been employed to detect and measure glycans, however mass spectrometry (MS) has become the method of choice for rapid characterisation of sample glycosylation[2]. The ability to separate glycans by mass for mass-based detection provides a relatively unbiased approach to detecting both the usual and unusual glycans in nature.

Upon completion of any glycomics MS data acquisition, the initial analysis will include the identification of the glycans detected. Unlike proteomics, which is template-derived, and therefore able to use the genome as a bioinformatic search space, the glycosylation of a given sample cannot be predicted[3–6], and requires bioinformatic confirmation.

One such approach is combinatorial monosaccharide analysis and is most popularly used in the form of GlycoMod[7,8]. GlycoMod is a Perl program designed to find all possible combinations of a glycan from an experimentally determined mass by searching a precomputed list of masses[9]. Despite being published in 2001, it remains frequently in use but is limited by its closed source nature, and online-only access. Combinatorial mass analysis is popular in the field of metabolomics, which uses high resolution accurate mass to assign chemical formulae[10–12], and rules such as the “Seven Golden Rules” restrict the number of theoretical matches based on experimental observations[13].

MS-based glycomics analyses have changed significantly as the monosaccharides and glycan structures observed experimentally challenge existing paradigms. This is compounded by improvements in data acquisition and bioinformatics that have allowed the glycoscience field to detect, quantify and share the data, describing more glycans than ever before[14–17].Considerable advancements have been made in the glycan bioinformatic search space including GlycReSoft[18], and GlycoNote[19], but throughput is lacking, and the monosaccharide search spaces are limited to routine and expected *N*-glycan structures.

To address these challenges and complement the advantages of GlycoMod, we describe an open-source, offline program called GlyCombo which combinatorially assigns monosaccharide combinations to glycan masses acquired by MS. Taking advantage of collaborative standards initiatives, we employ the mzML file format[20] to enable accessible, and complete glycan composition assignments in an automated manner while also delivering fast and efficient combinatorial analysis using dynamic programming[21].

Methods

Benchmark datasets and comparisons

Raw files were downloaded from published datasets on GlycoPost[17]. Datasets were chosen to cover negative and positive MS polarities, glycan types (NG, OG and GSL), resolutions (2D and 3D ion traps, and Orbitrap), and vendors (Bruker and Thermo Fisher Scientific). All benchmarks were performed on a consumer-grade Lenovo laptop equipped with an AMD Ryzen 6 Pro 5850U CPU, 16 GB of RAM and a 500 GB SSD. Search times for GlycoMod[7] were based on the loading timeline in Firefox network developer tools.

Raw file conversion to mzML

All downloaded files were converted to mzML file format[20] by MSConvert[22]. For low-resolution ion trap data with unassigned precursor charge states, an additional filter of "Charge State Predictor" was used without overwriting existing charges, single charge % TIC = 0.9, and only 1 charge per precursor. The mzML files were then directly read by GlyCombo.

GlyCombo design and Implementation

GlyCombo was developed in C# and is available open source under the Apache 2.0 license (https://github.com/Protea-Glycosciences/GlyCombo). The first process, mass list generation, extracts precursor *m/z* values from MS2 scans along with charge state and polarity to calculate neutral precursor mass values. This list is exported in the parameters file at the end of a given search in GlyCombo and was used for GlycoMod comparisons.

Neutral precursor mass values were adjusted for combinatorial analysis based on user-specified mass error, reducing end formats, and glycan derivatisation status. High-resolution files were searched with 50 ppm mass error, and low-resolution files were searched with 0.6 Da mass error. Based on presets or user-specified monosaccharide ranges, limits were then applied to the combinatorial analysis of monosaccharide masses to identify matches for the starting neutral precursor mass list. The same monosaccharide ranges were applied to GlycoMod for direct comparisons. Dynamic programming was used to efficiently identify all possible matching monosaccharide combinations for each precursor mass.

Skyline analysis

Output skyline transition lists were directly imported into Skyline[23–25] with minimum *m/z* 50 and maximum *m/z* 10,000. High resolution raw files were analysed with the following full-scan settings: 3 peaks for centroided mass analyser at 50 ppm mass accuracy. Low resolution raw files were analysed with the following full scan settings: 3 peaks for TOF mass analyser at 5000 resolution. Glycan precursors were filtered for quality, only including those with a minimum precursor isotopic dot product (idotp) of 0.9. The idotp value of 0.9 was empirically selected to remove poor quality MS1-matches (caused by monoisotopic peak misassignment, incorrect charge assignment, and poor signal to noise ratios) while preserving high-quality matches[26].

Data availability

GlyCombo output files, converted mzML files, and Skyline assays are available on Panorama Public[27] at: https://panoramaweb.org/GlyCombo.url . Search time values and compositional identification values can be found in Supplementary File 1.

Results and Discussion

Combinatorial glycan composition determination has been demonstrated to be suitable for identifying glycans in MS acquisitions of glycan-containing samples. With a diverse range of sample preparation, acquisition methods, and vendors, profiling glycans by their precursor mass is a robust initial approach with minimal assumptions. To enhance the identification of glycan precursors in acquired MS data, we developed an open-source Windows application for the rapid extraction of precursor *m/z* values from the mzML file format, a vendor neutral file type that enables cross-platform compatibility.

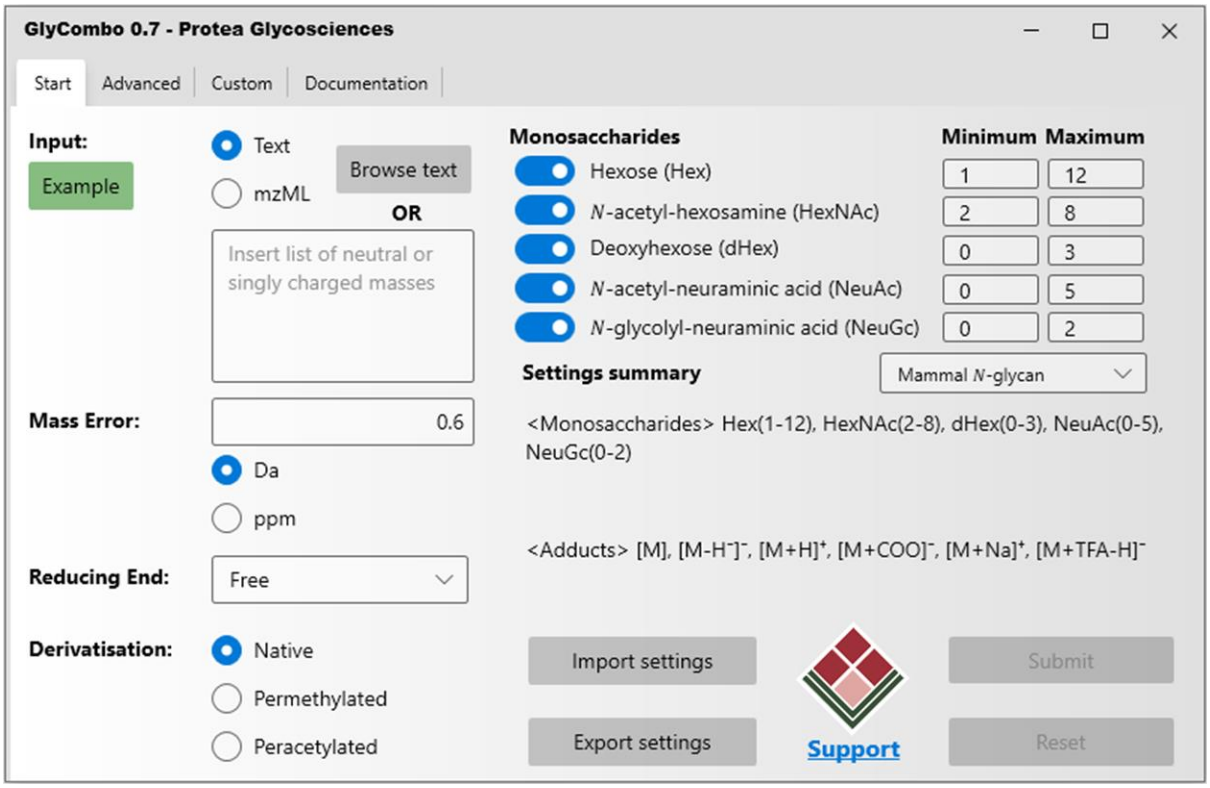

Figure 1 The Start page of the C# GlyCombo program found at https://github.com/Protea-Glycosciences/GlyCombo.

Many aspects of the glyco-analytical pipeline can affect the resulting *m/z* values observed in these mzML files, and as a result, GlyCombo is run through a GUI that requests user specifications regarding the mass error, reducing end format, glycan state, and expected monosaccharide ranges (**Fig 1**). Unlike GlycoMod, we also provide hexosamine (*i.e.* isomerically ambiguous glucosamine) as a new monosaccharide for combinatorial analysis, enabling the analysis of biomedically relevant polysaccharides such as heparan sulfate[28]. In addition to a large list of monosaccharide presets, including sialic acid derivatives, a maximum of five custom monosaccharides can be specified to ensure coverage of most applications.

Other aspects integral for correct data interpretation, such as MS polarity, charge state, and precursor *m/z* are automatically extracted from mzML files and used to build a neutral mass list that is searched, enabling direct comparisons to other platforms that require a glycan mass list as input (**Fig 2**). A comprehensive feature comparison to GlycoMod is described in **Supp Table 1**. As brute-force combinatorial analysis lacks throughput for applications such as these

(like the classic computer science coin change problem[29]), dynamic programming[21] was implemented as the solution to calculate monosaccharide combinations to match a given neutral mass. As the combinations are recursively expanded, memoization is performed to eliminate redundant calculations (simplified examples are provided in **Supp Figs 1 and 2**).

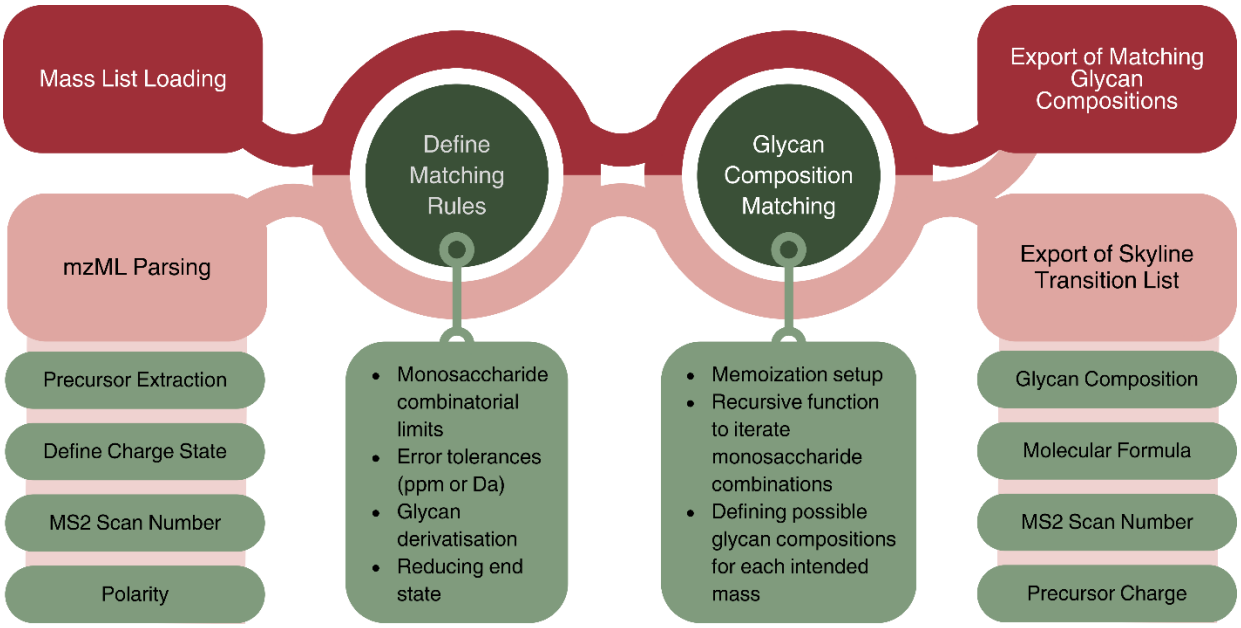

Figure 2 Process diagram describing the input, processes, and output for GlyCombo.

GlyCombo works best with mzML input as it enables the output of three files. The first is a csv file which includes glycan compositions matches, mass error of the matches, and the scan number of the respective MS2 spectrum. The second is a similar csv file which has additional columns enabling direct import into tools which generate extracted ion chromatograms (EIC) such as Skyline (an example is provided in Supplementary Table 2). The final file is a parameters file, giving the specific parameters used to generate the given output, as well as the list of precursors searched. In this work, we have used this list of precursors to benchmark our platform against one of the most popular approaches for glycan composition identification, GlycoMod. Users are also given the ability to use a text-based list of glycan masses within the GUI. In this case, the output is limited to a csv file which includes the glycan composition matches and mass error of the matches.

To benchmark our software, and ensure compatibility with existing approaches, six datasets were selected, and one raw file was downloaded from each of their respective GlycoPost accessions (**Table 1**). To demonstrate broad applicability across multi-glycomics[30], these datasets cover a range of glycan types: *N*-glycans (NG), *O*-glycans (OG) and glycans released from glycosphingolipids (GSL). For suitability of different polarities and derivatisation states, we also included permethylated and native glycans acquired in positive and negative modes. The vendor-neutrality of our approach was also assessed with dataset generated on Bruker and Thermo Fisher Scientific instruments.

Table 1 Benchmark datasets used in this study, selected to represent a wide range of glycomics approaches including glycan type, sample preparation, liquid chromatography (LC), MS polarity, MS vendor, and MS1 resolution.

Data set	LC	MS mode	MS1 resolution	Additional challenge	GlycoPost accession	Raw file name
1[31]	PGC	-	High	Native	GPST000188	20081907_Fetuin
2[32]	HILIC	+	High	Native	GPST000338	Native_CSF_03_55c_200mins_010223
3[33]	C18	+	High	Permethylated	GPST000242	Fetuin_Perm_Ng_C18_50cm
4[25]	PGC	-	Low	Thermo	GPST000029	CA_U87MG_CL_LAD_NG_090117
5[34]	PGC	-	Low	GSL, Bruker	GPST000239	54-GSL glycans - SW1116 - technical replicate 2
6[35]	PGC	-	Low	O-glycan	GPST000060	TT_190920bPRG4_2

GlyCombo was developed with three qualities in mind: broad accessibility, rapid output, and completeness of glycan detection. Benchmarking was performed by searching the same neutral mass list generated by GlyCombo between both GlycoMod and GlyCombo, using the same monosaccharide ranges, and search time was recorded in triplicate. As shown in **Fig 3A**, all datasets were compatible with our software and it generally performed faster than GlycoMod, with up to 3x faster search times. Despite this, dataset 3 was almost 2x slower with our software, causing these search speed discrepancies to be investigated.

Unlike GlyCombo, GlycoMod (and other search tools such as GlycReSoft[18] and GlycoNote[19]) utilise matching rules based on literature review and assumed biosynthetic pathways to limit the combinatorial burden. This is demonstrated by the GlycoMod *N*-glycan search function whereas the *O*-glycan search function does not have such rules beyond a mass limit of 5 kDa[8]. A comparison of the number of glycan compositions identified based on the glycan masses observed in each dataset (**Fig 3B**) confirmed that the *N*-glycan search function yielded fewer matches, and when assessed with the *O*-glycan function, similar combinations were identified at longer search times (*e.g.* dataset 3 processing time increased from 180 to 550 seconds, 50% slower than GlyCombo).

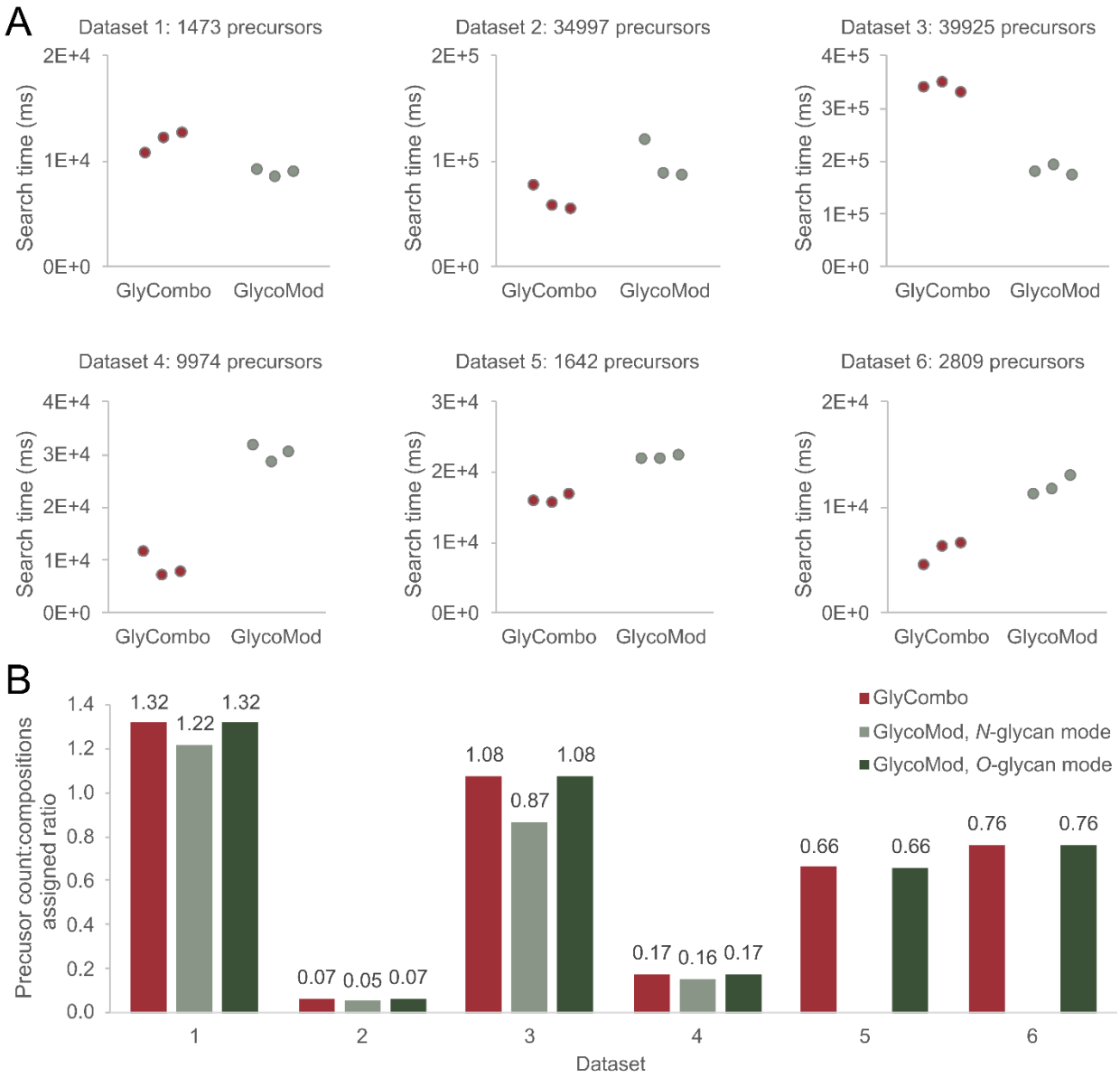

Figure 3 Benchmark comparison of compositional assignment between GlyCombo, a local application, and GlycoMod, a server-based web-app. **A** Comparison of search time for each dataset based on recommended GlycoMod settings **B** Compositional identifications made per dataset across software platforms, represented as a ratio of the number of precursors in the mzML file to approximately normalise identifications across datasets of varying scan numbers. As multiple monosaccharide combinations can be matched to a single scan, a ratio greater than 1 is possible.

The rapid analysis speed of GlyCombo could enable exciting new directions which have been realised in other MS applications such as proteomics and metabolomics. At an average search time of 7.5 milliseconds per precursor, real-time instrument acquisition methods could be devised which have been demonstrated to improve analytical depth[36] and real-time method optimisation[37]. Additionally, as downloading raw files and mzML conversion took longer than the combinatorial search itself, repository-scale reanalysis could prove a promising alternative,

enabling iterative reanalyses for new insights[38]. GlyCombo's ability to search for precursors with mass errors over 5 Da makes it ideal for data-independent glycan analysis, where precursors are isolated within windows up to 48 Da wide, potentially matching hundreds of glycan compositions to a single MS2 spectrum.

The efficiency of GlyCombo's combinatorial composition assignment algorithm enables computationally expensive features, such as multiple adduct searching and off-by-one error anticipation[39], which improve the scan annotation rate for mzML files. In **Fig 4A**, only when sodium is included as an adduct type is the correct (Hex)5 (HexNAc)2 composition assigned to the observed precursor mass. Off-by-one errors not only cause misassignments but also lead to unassigned precursors when the M+1 isotope is mistaken for the monoisotopic peak, even though both produce similar MS2 spectra due to co-isolation for fragmentation (**Fig 4B**).

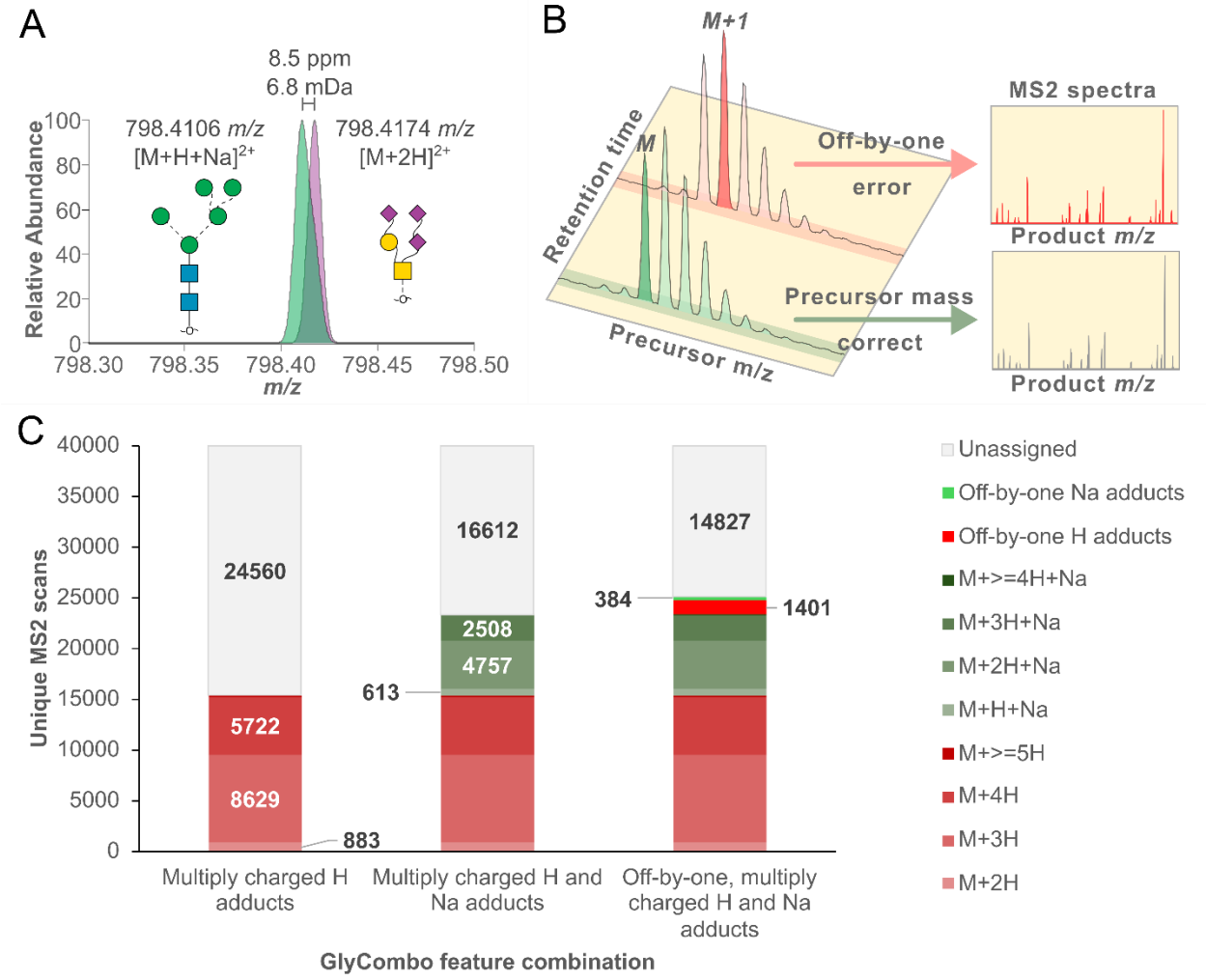

Figure 4 Combinatorial glycan composition assignment is aided by including all observed charge states, adducts, and MS misassignments using Dataset 3 as an example. **A** Utilisation of all expected adducts improves accuracy of glycan composition assignment. **B** The off-by-one error caused by MS misassignment of the monoisotopic peak can impact accurate-mass precursor searching. **C** MS2 scan assignment rate across GlyCombo features including multiply

charged adducts, inclusion of more than one adduct type, and allowing for off-by-one monoisotopic mass assignment errors.

Re-searching Dataset 3 with both off-by-one error anticipation, and sodium adduction enabled in GlyCombo reduced the number of unassigned MS2 scans by 40% compared to a protonated adduct search (**Fig 4C**). Although this complicated search comes at a cost of a 4x greater search time, the total search time was a fraction of the 200-minute-long LC-MS run. As the off-by-one search only yielded an 11% decrease in unassigned MS2 scans, this option could be skipped for data generated by instruments with accurate monoisotopic peak detection[40]. A limitation of these more complex searches is the increase in false positives due to mistaking protonated adducts for sodiated adducts and correct precursor masses for off-by-one errors. As GlyCombo is the initial step in filtering candidate precursor ions for quantitation and characterization, users can opt for a higher false positive rate to ensure a high true positive rate. Therefore, it can be crucial to detect a comprehensive range of glycans to leverage subsequent quality assessments, including retention time filtering (Skyline), MS2 scoring (GlycoWorkBench), and precursor isotopic distribution evaluation (Skyline).

Characterisation and quantitation of glycan signal are common steps in the glyco-analytical data analysis workflow, yielding information about the abundance and identity of the glycan signals detected. GlycoWorkbench is a characterisation-oriented software tool, useful for matching MS1 and MS2 signals to drawn structures, or those found in databases. Like GlycoMod, it remains widely used demonstrating its enduring value for composition and structure characterisation[41]. Orthogonal to the characterisation approach of GlycoWorkbench, Skyline is a quantitation tool that uses predefined molecule transitions to generate EICs, measuring glycan signal in LC-MS data and providing scores based on signal relationships. A useful score used in this work is the isotopic dot product (idotp) which describes how closely the isotopic distribution of an observed precursor isotopic envelope matches the theoretical abundances of a given chemical formula[23,24,26].

Here, we used Skyline[23–25] to remove poor quality combinatorial matches (idotp > 0.9 for the first three isotopes) and visualise the glycan profile of three datasets which feature *N*-glycans and glycans released from glycosphingolipids (**Fig 5**). In all three datasets, qualitatively identical plots of detected glycans were observed when these EICs were compared to base peak chromatograms (BPC) of the original raw files, demonstrating that GlyCombo is effective at glycan detection across these diverse datasets. As idotp is based on the isotopic distribution of a chemical formula, it informatively scores composition assignments but is unsuitable at discriminating glycan structures due to the isotopic distribution of isomers being identical.

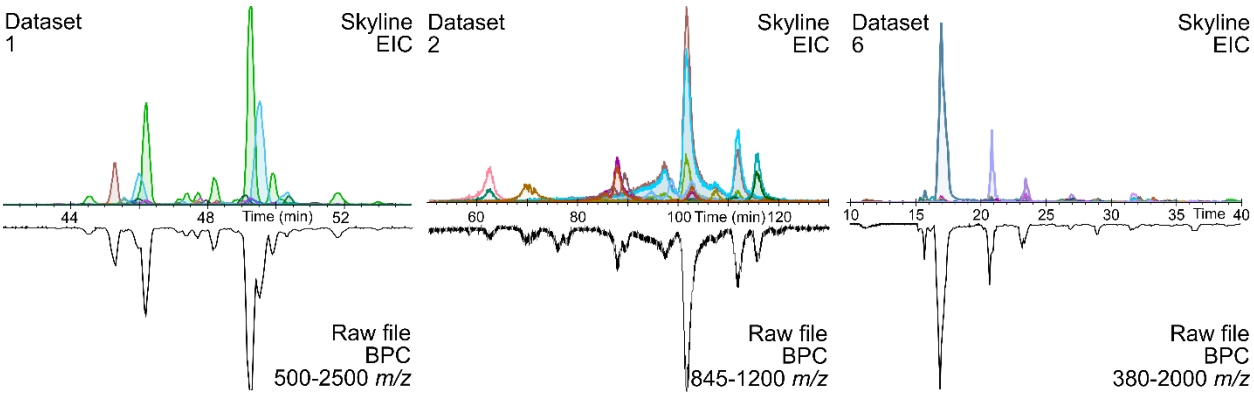

Figure 5 Comparison of glycan composition extracted ion chromatograms (EIC) generated by GlyCombo for Skyline to raw file base peak chromatograms (BPC) qualitatively recapitulate the feature profiles

Although our software extracts and rapidly searches mzML files, instrument acquisition schemes can incorrectly identify the monoisotopic peak (known as off-by-one errors[42]), leading to incorrectly calculated neutral glycan masses. This is also exacerbated by our requirement of an MS2 scan for a given glycan composition to be detected. The use of glycan precursor isotopic distribution filtering ensures that these effects are mitigated, due to scoring the abundance of the first three isotopes of each glycan and subsequent removal of glycan compositions returning poor scores.

Another notable limitation of this precursor mass-based combinatorial approach is the inability to assess glycan topology or glycosidic linkages. As glycans are frequently present in multiple isomeric forms or structures, and each isomer has differing biomarker potential, additional forms of separation beyond precursor mass are needed including liquid chromatography stationary phases (such as porous graphitised carbon and hydrophilic interaction columns benchmarked here), and diagnostic product ions from MSn spectra[43]. GlyCombo output includes scan numbers to aid in subsequent downstream MS2 annotation and structure elucidation by software such as GlycoWorkBench[41] and spectral library searching[19].

Conclusions

The identification of glycans in raw files serves as the first foundational step of glycomics data analysis. We describe a new open-source computational tool, GlyCombo, that is capable of rapidly elucidating possible glycan compositions from MS analyses. In this paper we utilize GlyCombo and benchmark its performance against the current state-of-the-art combinatorial solution for glycomics, GlycoMod. The experimental results exemplify the speed and robustness of GlyCombo for use with *N*-glycans, *O*-glycans, glycans released from glycosphingolipids, in combination with multiple polarities and derivatisation states. In addition to faster search times, broader accessibility, and completeness of annotation, processing is automated, and output are specifically formatted to connect with downstream processes including GlycoWorkBench for structural annotation, and Skyline for quality control and quantitation.

Conflict of Interest statement

C Ashwood is the director of Protea Glycosciences, a company which provides fee-for-service glycomics assays, analytical standards, and spectral libraries. The remaining authors declare no competing financial interest.

Supporting information

1. GlyCombo_Supplementary_Table.xlsx – Excel spreadsheet containing glycan preset numbers, search times for GlyCombo, composition identification rates across datasets, and MS2 scan annotations corresponding to GlyCombo features
2. GlyCombo_SupplementaryFigures.pdf – Supplementary tables and figures associated with the manuscript: Feature comparison between GlycoMod and Glycombo, example Skyline output file from GlyCombo, simplified GlyCombo process by recursive loop memoisation, and the GlyCombo recursive tree framework structure to prevent recursive composition calculations.

References

(1) Cummings, R. D. A Periodic Table of Monosaccharides. *Glycobiology* **2024**, *34* (1). https://doi.org/10.1093/GLYCOB/CWAD088.

(2) Harvey, D. J. Analysis of Protein Glycosylation by Mass Spectrometry. *Analysis of Post-Translational Modifications Using Mass Spectrometry* **2016**, 89–159. https://doi.org/10.1002/9781119250906.CH3.

(3) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom* **1994**, *5* (11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2.

(4) Binz, P. A.; Shofstahl, J.; Vizcaíno, J. A.; Barsnes, H.; Chalkley, R. J.; Menschaert, G.; Alpi, E.; Clauser, K.; Eng, J. K.; Lane, L.; Seymour, S. L.; Sánchez, L. F. H.; Mayer, G.; Eisenacher, M.; Perez-Riverol, Y.; Kapp, E. A.; Mendoza, L.; Baker, P. R.; Collins, A.; Van Den Bossche, T.; Deutsch, E. W. Proteomics Standards Initiative Extended FASTA Format. *J Proteome Res* **2019**, *18* (6), 2686–2692. https://doi.org/10.1021/acs.jproteome.9b00064.

(5) Craig, R.; Cortens, J. P.; Beavis, R. C. The Use of Proteotypic Peptide Libraries for Protein Identification. *Rapid Communications in Mass Spectrometry* **2005**, *19* (13), 1844–1850. https://doi.org/10.1002/rcm.1992.

(6) Verheggen, K.; Ræder, H.; Berven, F. S.; Martens, L.; Barsnes, H.; Vaudel, M. Anatomy and Evolution of Database Search Engines—a Central Component of Mass Spectrometry Based Proteomic Workflows. *Mass Spectrometry Reviews*. John Wiley and Sons Inc. May 1, 2020, pp 292–306. https://doi.org/10.1002/mas.21543.

(7) Cooper, C. A.; Gasteiger, E.; Packer, N. H. GlycoMod-A Software Tool for Determining Glycosylation Compositions from Mass Spectro-Metric Data. *Proteomics* **2001**, *1*, 340–349. https://doi.org/10.1002/1615-9861.

(8) Cooper, C. A.; Gasteiger, E.; Packer, N. H. Predicting Glycan Composition from Experimental Mass Using GlycoMod. *Handbook of Proteomic Methods* **2003**, 225–231. https://doi.org/10.1007/978-1-59259-414-6_14.

(9) Mariethoz, J.; Alocci, D.; Gastaldello, A.; Horlacher, O.; Gasteiger, E.; Rojas-Macias, M.; Karlsson, N. G.; Packer, N. H.; Lisacek, F. Glycomics@ExPASy: Bridging the Gap. *Molecular and Cellular Proteomics* **2018**, *17* (11), 2164–2176. https://doi.org/10.1074/mcp.RA118.000799.

(10) Pluskal, T.; Uehara, T.; Yanagida, M. Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Anal Chem* **2012**, *84* (10), 4396–4403. https://doi.org/10.1021/ac3000418.

(11) Kind, T.; Fiehn, O. Metabolomic Database Annotations via Query of Elemental Compositions: Mass Accuracy Is Insufficient Even at Less than 1 Ppm. *BMC Bioinformatics* **2006**, *7*. https://doi.org/10.1186/1471-2105-7-234.

(12) Brown, M.; Wedge, D. C.; Goodacre, R.; Kell, D. B.; Baker, P. N.; Kenny, L. C.; Mamas, M. A.; Neyses, L.; Dunn, W. B. Automated Workflows for Accurate Mass-Based Putative Metabolite Identification in LC/MS-Derived Metabolomic Datasets. *Bioinformatics* **2011**, *27* (8), 1108–1112. https://doi.org/10.1093/bioinformatics/btr079.

(13) Kind, T.; Fiehn, O. Seven Golden Rules for Heuristic Filtering of Molecular Formulas Obtained by Accurate Mass Spectrometry. *BMC Bioinformatics* **2007**, *8*. https://doi.org/10.1186/1471-2105-8-105.

(14) Egorova, K. S.; Toukach, P. V. Glycoinformatics: Bridging Isolated Islands in the Sea of Data. *Angewandte Chemie - International Edition* **2018**, *57* (46), 14986–14990. https://doi.org/10.1002/anie.201803576.

(15) Fujita, A.; Aoki, N. P.; Shinmachi, D.; Matsubara, M.; Tsuchiya, S.; Shiota, M.; Ono, T.; Yamada, I.; Aoki-Kinoshita, K. F. The International Glycan Repository GlyTouCan Version 3.0. *Nucleic Acids Res* **2021**, *49* (D1), D1529–D1533. https://doi.org/10.1093/NAR/GKAA947.

(16) Rojas-Macias, M. A.; Mariethoz, J.; Andersson, P.; Jin, C.; Venkatakrishnan, V.; Aoki, N. P.; Shinmachi, D.; Ashwood, C.; Madunic, K.; Zhang, T.; Miller, R. L.; Horlacher, O.; Struwe, W. B.; Watanabe, Y.; Okuda, S.; Levander, F.; Kolarich, D.; Rudd, P. M.; Wuhrer, M.; Kettner, C.; Packer, N. H.; Aoki-Kinoshita, K. F.; Lisacek, F.; Karlsson, N. G. Towards a Standardized Bioinformatics Infrastructure for N- and O-Glycomics. *Nat Commun* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-11131-x.

(17) Watanabe, Y.; Aoki-Kinoshita, K. F.; Ishihama, Y.; Okuda, S. GlycoPOST Realizes FAIR Principles for Glycomics Mass Spectrometry Data. *Nucleic Acids Res* **2021**, *49* (D1), D1523–D1528. https://doi.org/10.1093/NAR/GKAA1012.

(18) Maxwell, E.; Tan, Y.; Tan, Y.; Hu, H.; Benson, G.; Aizikov, K.; Conley, S.; Staples, G. O.; Slysz, G. W.; Smith, R. D.; Zaia, J. GlycReSoft: A Software Package for Automated Recognition of Glycans from LC/MS Data. *PLoS One* **2012**, *7* (9). https://doi.org/10.1371/JOURNAL.PONE.0045474.

(19) Liu, M. Q.; Treves, G.; Amicucci, M.; Guerrero, A.; Xu, G.; Gong, T. Q.; Davis, J.; Park, D.; Galermo, A.; Wu, L.; Cao, W.; Lebrilla, C. B. GlycoNote with Iterative Decoy Searching and Open-Search Component Analysis for High-Throughput and Reliable Glycan Spectral Interpretation. *Anal Chem* **2023**, *95* (21), 8223–8231. https://doi.org/10.1021/ACS.ANALCHEM.3C00083.

(20) Deutsch, E. MzML: A Single, Unifying Data Format for Mass Spectrometer Output. *Proteomics* **2008**, *8* (14), 2776–2777. https://doi.org/10.1002/PMIC.200890049.

(21) Morin, T. L. COMPUTATIONAL ADVANCES IN DYNAMIC PROGRAMMING. *Dynamic Programming and its Applications* **1978**, 53–90. https://doi.org/10.1016/B978-0-12-568150-6.50009-X.

(22) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: Open Source Software for Rapid Proteomics Tools Development. *Bioinformatics* **2008**, *24* (21), 2534–2536. https://doi.org/10.1093/BIOINFORMATICS/BTN323.

(23) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**, *26* (7), 966–968. https://doi.org/10.1093/BIOINFORMATICS/BTQ054.

(24) Adams, K. J.; Pratt, B.; Bose, N.; Dubois, L. G.; St. John-Williams, L.; Perrott, K. M.; Ky, K.; Kapahi, P.; Sharma, V.; Maccoss, M. J.; Moseley, M. A.; Colton, C. A.; Maclean, B. X.; Schilling, B.; Thompson, J. W. Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *J Proteome Res* **2020**, *19* (4), 1447–1458. https://doi.org/10.1021/ACS.JPROTEOME.9B00640/ASSET/IMAGES/LARGE/PR9B00640_0001.JPEG.

(25) Ashwood, C.; Pratt, B.; Maclean, B. X.; Gundry, R. L.; Packer, N. H. Standardization of PGC-LC-MS-Based Glycomics for Sample Specific Glycotyping. *Analyst* **2019**, *144* (11), 3601. https://doi.org/10.1039/C9AN00486F.

(26) Tsantilas, K. A.; Merrihew, G. E.; Robbins, J. E.; Johnson, R. S.; Park, J.; Plubell, D. L.; Huang, E.; Riffle, M.; Sharma, V.; Maclean, B. X.; Eckels, J.; Wu, C. C.; Bereman, M. S.; Spencer, S. E.; Hoofnagle, A. N.; Maccoss, M. J. A Framework for Quality Control in Quantitative Proteomics. *bioRxiv preprint* **2024**. https://doi.org/10.1101/2024.04.12.589318.

(27) Sharma, V.; Eckels, J.; Schilling, B.; Ludwig, C.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B. Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline. *Molecular and Cellular Proteomics* **2018**, *17* (6), 1239–1244. https://doi.org/10.1074/mcp.RA117.000543.

(28) Karlsson, R.; Chopra, P.; Joshi, A.; Yang, Z.; Vakhrushev, S. Y.; Clausen, T. M.; Painter, C. D.; Szekeres, G. P.; Chen, Y. H.; Sandoval, D. R.; Hansen, L.; Esko, J. D.; Pagel, K.; Dyer, D. P.; Turnbull, J. E.; Clausen, H.; Boons, G. J.; Miller, R. L. Dissecting Structure-Function of 3-O-Sulfated Heparin and Engineered Heparan Sulfates. *Sci Adv* **2021**, *7* (52), 6026. https://doi.org/10.1126/SCIADV.ABL6026/SUPPL_FILE/SCIADV.ABL6026_SM.PDF.

(29) Van Cott, C. A.; Zhang, Q. The Change-Making Problem for Six Coin Values and Beyond. *ArXiv* **2023**.

(30) Moh, E. S. X.; Dalal, S.; DeBono, N. J.; Kautto, L.; Wongtrakul-Kish, K.; Packer, N. H. SSSMuG: Same Sample Sequential Multi-Glycomics. *Anal Chem* **2023**. https://doi.org/10.1021/ACS.ANALCHEM.3C04928/SUPPL_FILE/AC3C04928_SI_002.XLSX.

(31) Kondo, K.; Harada, Y.; Nakano, M.; Suzuki, T.; Fukushige, T.; Hanzawa, K.; Yagi, H.; Takagi, K.; Mizuno, K.; Miyamoto, Y.; Taniguchi, N.; Kato, K.; Kanekura, T.; Dohmae, N.; Machida, K.; Maruyama, I.; Inoue, H. Identification of Distinct N-Glycosylation Patterns on Extracellular Vesicles from Small-Cell and Non-Small-Cell Lung Cancer Cells. *J Biol Chem* **2022**, *298* (6). https://doi.org/10.1016/J.JBC.2022.101950.

(32) Daramola, O.; Gutierrez-Reyes, C. D.; Wang, J.; Nwaiwu, J.; Onigbinde, S.; Fowowe, M.; Dominguez, M.; Mechref, Y. Isomeric Separation of Native N-Glycans Using Nano Zwitterionic- Hydrophilic Interaction Liquid Chromatography Column. *J Chromatogr A* **2023**, *1705*, 464198. https://doi.org/10.1016/J.CHROMA.2023.464198.

(33) Wang, J.; Dong, X.; Yu, A.; Huang, Y.; Peng, W.; Mechref, Y. Isomeric Separation of Permethylated Glycans by Extra-Long Reversed-Phase Liquid Chromatography (RPLC)-MS/MS. *Analyst* **2022**, *147* (10), 2048–2059. https://doi.org/10.1039/D2AN00010E.

(34) Wang, D.; Madunić, K.; Zhang, T.; Mayboroda, O. A.; Lageveen-Kammeijer, G. S. M.; Wuhrer, M. High Diversity of Glycosphingolipid Glycans of Colorectal Cancer Cell Lines Reflects the Cellular Differentiation Phenotype. *Molecular and Cellular Proteomics* **2022**, *21* (6), 100239. https://doi.org/10.1016/j.mcpro.2022.100239.

(35) Flowers, S. A.; Thomsson, K. A.; Ali, L.; Huang, S.; Mthembu, Y.; Regmi, S. C.; Holgersson, J.; Schmidt, T. A.; Rolfson, O.; Björkman, L. I.; Sundqvist, M.; Karlsson-Bengtsson, A.; Jay, G. D.; Eisler, T.; Krawetz, R.; Karlsson, N. G. Decrease of Core 2 O- Glycans on Synovial Lubricin in Osteoarthritis Reduces Galectin-3 Mediated Crosslinking. *J Biol Chem* **2020**, *295* (47), 16023–16036. https://doi.org/10.1074/JBC.RA120.012882.

(36) McGann, C. D.; Barshop, W. D.; Canterbury, J. D.; Lin, C.; Gabriel, W.; Huang, J.; Bergen, D.; Zabrouskov, V.; Melani, R. D.; Wilhelm, M.; McAlister, G. C.; Schweppe, D. K. Real-Time Spectral Library Matching for Sample Multiplexed Quantitative Proteomics. *J Proteome Res* **2023**, *22* (9), 2836–2846. https://doi.org/10.1021/ACS.JPROTEOME.3C00085.

(37) Shuken, S. R.; Yu, Q.; Gygi, S. P. Inserting Pre-Analytical Chromatographic Priming Runs Significantly Improves Targeted Pathway Proteomics with Sample Multiplexing. *J Proteome Res* **2024**. https://doi.org/10.1021/ACS.JPROTEOME.4C00096/SUPPL_FILE/PR4C00096_SI_001.PDF.

(38) Choi, M.; Carver, J.; Chiva, C.; Tzouros, M.; Huang, T.; Tsai, T. H.; Pullman, B.; Bernhardt, O. M.; Hüttenhain, R.; Teo, G. C.; Perez-Riverol, Y.; Muntel, J.; Müller, M.; Goetze, S.; Pavlou, M.; Verschueren, E.; Wollscheid, B.; Nesvizhskii, A. I.; Reiter, L.; Dunkley, T.; Sabidó, E.; Bandeira, N.; Vitek, O. MassIVE.Quant: A Community Resource of Quantitative Mass Spectrometry–Based Proteomics Datasets. *Nature Methods 2020 17:10* **2020**, *17* (10), 981–984. https://doi.org/10.1038/s41592-020-0955-0.

(39) Lermyte, F.; Dittwald, P.; Claesen, J.; Baggerman, G.; Sobott, F.; O'Connor, P. B.; Laukens, K.; Hooyberghs, J.; Gambin, A.; Valkenborg, D. MIND: A Double-Linear Model to Accurately Determine Monoisotopic Precursor Mass in High-Resolution Top-Down Proteomics. *Anal Chem* **2019**, *91* (15), 10310–10319. https://doi.org/10.1021/acs.analchem.9b02682.

(40) Hebert, A. S.; Thöing, C.; Riley, N. M.; Kwiecien, N. W.; Shiskova, E.; Huguet, R.; Cardasis, H. L.; Kuehn, A.; Eliuk, S.; Zabrouskov, V.; Westphall, M. S.; McAlister, G. C.; Coon, J. J. Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Anal Chem* **2018**, *90* (3), 2333–2340. https://doi.org/10.1021/acs.analchem.7b04808.

(41) Ceroni, A.; Maass, K.; Geyer, H.; Geyer, R.; Dell, A.; Haslam, S. M. GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *J Proteome Res* **2008**, *7* (4), 1650–1659. https://doi.org/10.1021/PR7008252/SUPPL_FILE/PR7008252-FILE003.PDF.

(42) Radziński, P.; Valkenborg, D.; Startek, M. P.; Gambin, A. Envemind: Accurate Monoisotopic Mass Determination Based on Isotopic Envelope. *J Am Soc Mass Spectrom* **2022**, *33* (11), 2063–2069. https://doi.org/10.1021/JASMS.2C00176/ASSET/IMAGES/LARGE/JS2C00176_0006.JPEG.

(43) Ashwood, C.; Lin, C. H.; Thaysen-Andersen, M.; Packer, N. H. Discrimination of Isomers of Released N- and O-Glycans Using Diagnostic Product Ions in Negative Ion PGC-LC-ESI-MS/MS. *J Am Soc Mass Spectrom* **2018**, *29* (6), 1194–1209. https://doi.org/10.1007/S13361-018-1932-Z/SUPPL_FILE/JS8B05838_SI_006.CSV.

For Table of Contents Only