nature portfolio

Peer Review File

GlycoGenius: a streamlined high-throughput glycan composition identification tool

Corresponding Author: Mr Hector Franco Loponte

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript by Loponte et al. describes a new software platform to accommodate the processing and analysis of mass spectrometry-derived glycomics data. Developed by experts in the field, the software seems comprehensive and user-friendly. This is also something that is needed in the field, due to both the lack of general expert knowledge and data processing heterogeneity. We're generally favorable to this work in principle, yet have quite some concerns about (i) evaluation, (ii) usability, and (iii) presentation that should be addressed and that we will detail below.

1) Our main points rest on evaluation. This is especially crucial because the results portion of this work is currently a bit thin, due to it almost exclusively focusing on the description of the developed software tool. Currently, the presented software has been evaluated on only two glycomics datasets (one of which stems from the authors' own work). This does neither cover the breadth of glycomics workflows currently in use, nor the described supported features. Especially given the fact that the software is the vast majority of the value proposition of this work, we would expect a much more exhaustive validation/testing. Further, even though it is claimed that GlycoGenius can be applied to GAGs, no GAG example is present in the manuscript

2) This leads to a follow-up point that, currently, the evaluation of GlycoGenius is mostly descriptive (e.g., which compositions were detected or not and why). We would expect a much more quantitative evaluation of the performance of GlycoGenius for a full picture. This should include standard metrics (e.g., sensitivity, specificity, etc etc), but also evaluate the individual steps of the workflow (if possible)

3) Also connected to evaluation is the unfortunate fact that even the two datasets that have been evaluated have been done so post-factum. This allowed the authors to fine-tune their composition library building to an extent that would not be realistic in practice. I doubt anyone would start a standard O-glycan library with "Xyloses: 0 to 3", if that data peculiarity would not be known in advance.
If library generation is intended to be a crucial component of the workflow and constrain false-positives etc, then a fair/realistic evaluation of this is absolutely necessary. Examples could include building "standard" O-glycan / N-glycan libraries and evaluating on them. In any case, the current evaluation is overoptimistic
Further, even the retention time window selection performed for evaluation requires either other tools generating this insight or pre-existing knowledge which would not be present in a real-world scenario.
In general, we are dissatisfied by both the quality and quantity of evaluation for this software.

4) This leads us to our next point: guidance in library building. The authors use "Deoxyhexoses: 0 to 2" for plasma N-glycome analysis, but why, for example, would 3 Fuc (core Fuc + 2 antennary fucosylation events) be not permitted? This again pre-supposes extreme specialist knowledge, which--if necessary--would make this software unusable to but a few people in the world (who probably already have their respective analysis workflows set up and optimized in any case). Either this software needs to be so user-friendly that glycomics newcomers can effectively use it or there needs to be a realistic path how the authors expect the world's few experts to switch to their software.

5) This then culminates in the general point that we're uncertain about: It is currently unclear how specialized one needs to

be in order to (effectively) use this software. Since the main selling point of GlycoGenius is utility/convenience, this is important for its eventual scope.

6) The value of software stands and falls with its maintenance. What are the plans to maintain GlycoGenius at least in the medium-term?

7) We are also unsure about the stated runtime for real-world usage. Testing GlycoGenius on an arbitrary recent GlycoPOST (https://glycopost.glycosmos.org/entry/GPST000572) took 1 hour 47 minutes for a single file on the MS1 level alone (on a top-of-the-line Macbook). The MS2 level never finished (even after many hours; stuck on "Building fragments library")

8) Given the widespread usage of the Domon-Costello nomenclature for glycan fragments, we would expect GlycoGenius to support this. We realize that this decision has likely been taken because an MS2 fragment library makes the current nomenclature more convenient, but a "translation" (given a candidate structure) should be very feasible

9) It is currently very unclear to us which of the presented figures (such as in Fig. 5) are direct results from GlycoGenius and which are manually modified for the paper. For instance Fig. 5f looks like a very useful visualization but is that one generated fully automatically by GlycoGenius or not? If not, this needs to be clarified to avoid misleading readers

10) Superlative words such as "ultimate" should be avoided in titles. Very few things in science are "ultimate" (for better or worse)

11) For "custom monosaccharides", the expected format should be specified or explained via examples. In the GUI, "ion mode" is currently a bit hidden/not as prominent as one would expect, given that it affects every single calculation

12) Some of the figures (e.g., Fig 5) are very pixelated / hard to read in our copy, but we're not sure whether this is a conversion issue or the input image

13) All used software in GlycoGenius needs to be versioned in the manuscript. Currently, the used libraries are described but not versioned.

14) The claim that "full automation has been achieved for the identification and quantification of peptides and proteins" is probably up for debate, since it mostly applies to known/established peptides/proteins

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
To the Authors,

This research article presents GlycoGenius (GG), a novel open-source software tool developed for the automated analysis of glycomics data derived from liquid chromatography (LC) or capillary electrophoresis (CE) coupled with mass spectrometry (MS), including tandem MS (MS/MS). The authors assert that GG represents a significant advancement in the field by overcoming the limitations of existing tools through its fully integrated workflow and user-friendly graphical interface (GUI). The article substantiates GG's utility by comparing its performance against manual analysis and other established software using two distinct glycomics datasets.
The introduction effectively delineates the bottleneck in glycomics research arising from the inherent complexity and large volume of MS data, compounded by the scarcity of fully automated and integrated analysis tools. The authors persuasively argue for the necessity of a tool such as GlycoGenius.
The results section compellingly demonstrates GG's capacity to achieve comparable or even superior outcomes compared to manual analysis and competing software. Notably, GG identifies a greater number of glycans, including novel species, while concurrently achieving a substantial reduction in analysis time. The detailed analysis of discrepancies between GG's findings and those of the original publications underscores the software's ability to identify potential errors or low-confidence identifications. The emphasis on an intuitive GUI, coupled with the availability of a command-line interface (CLI), effectively accommodates a diverse range of users with varying levels of bioinformatics expertise.

Areas for Further Consideration:

Comment 1:
While the comparison with GlycReSoft is appreciated, a more comprehensive evaluation of GlycoGenius against other relevant software would strengthen the manuscript. Please include a feature-by-feature comparison with tools such as:
• GRITS Toolbox (complex glycan analysis at the MS/MS level; https://pubmed.ncbi.nlm.nih.gov/30913289/).
• GlycoDeNovo (elucidation of glycan topologies from MS/MS spectra).
This comparison should explicitly highlight the unique features and algorithmic approaches of GlycoGenius, including how it addresses aspects like isobaric glycans, structural isomers, and various modifications, to substantiate claims of its superiority. The current brief mention of other tools in the introduction lacks the necessary depth for a thorough evaluation.

Comment 2:
The manuscript mentions the detection of different chromatographic/electropherographic peaks for potential isomer differentiation. However, the extent to which GlycoGenius can structurally resolve isomers using MS data alone (without relying on separation or prior knowledge/MS/MS interpretation) needs clarification. Please explicitly detail the software's capabilities in this regard.
Furthermore, provide more specifics about the MS/MS annotation algorithm. Details on the fragment library generation process and the criteria used for matching fragment ions to theoretical fragments are crucial for assessing the robustness of this feature.

Comment 3:
Please clarify how GlycoGenius handles co-fragmenting spectra for glycan structure elucidation. If the software reports all possible structures with different scoring, please explain the scoring mechanism. Additionally, discuss specific parameters that users should consider when analyzing glycan data acquired from different mass analyzers, such as TOF versus Orbitrap MS, and their impact on MS/MS data interpretation within GlycoGenius.

Comment 4: Potential for False Positives/Misassignments:
The manuscript should address the potential for false positives or misassignments due to narrow mass error margins, particularly in cases where samples contain isobaric or near-isobaric glycans (e.g., distinguishing between two fucoses and a single Neu5Ac). Please discuss how GlycoGenius mitigates such issues. The reference provided (10.1093/glycob/cwae006) highlights challenges in glycopeptide assignment based on ppm error, and a similar discussion regarding glycan composition assignment within GlycoGenius would be beneficial

Minor Corrections:
• Figure 5: Please improve the resolution of Figure 5d. Additionally, the font in Figure 5g is too small and illegible; please increase the font size for better readability.
• Table 1: In the GG Monosaccharide Code, please explain the use of square brackets around "[G]" in the codes for Am[G] and E[G].
• Line 67: "The complexity associated with glycan analysis arise primarily from their unique challenges...". It should likely be "arises."
• Line 78: "...requires interfacing between various function and modules of software that do not exhibit seamless interoperability." It should be "functions"
• Line 132: The adjective should be hyphenated: "integrates feature-rich data visualization."
• Line 836: The repetition of "and" should be avoided for better flow: "...manuscript and user-guide, and manages..." or "...manuscript and user-guide; manages..."

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
GlycoGenius (GG) enables high-quality identification of glycan compositions for peaks detected within an LC-MS run, as checked against S/N ratio, consistency of isotopic envelopes, and consistent signal throughout the retention time of the eluting peaks. It has advanced glycan library building, pre-processing and EIC/EIE tracing, per spectrum EIC/EIE tracing, and post-tracing processing EIC/EIE features. It further provides quantification of AUC for the generated EIC/EIE. Notably, it uses mainly the isotopic fitting score to determine the identity quality of a glycan. Along with a shorter computational time required for data processing and analysis, GG appears to be a significant advancement from GlyReSoft. It offers solutions to streamline workflows, enhance research outcomes, and facilitates the generation of quantitative glycan tables. Its interoperability provides researchers with a powerful bioinformatic tool to expedite the analysis of larger datasets.

Short of using the software myself, running it and verifying its utility with our own datasets, I am inclined to take their words for it, and on paper, every aspect seems well considered and implemented. That’s about as many features as one would desire in so far as curating an LC-MS/MS-based glycomics dataset is concerned: accurately identifying peaks that matches glycan compositions allowed by the input of experimentally derived library or theorectically constructed one based on sets of constrains imposed at the level of the max/min number of each different monosaccharide. These features alone are welcoming and make GG a timely tool for glycomics. It would take away all the tedious aspects of manual processing and

likely constitute the most polished tools currently available.

However, as noted by the authors themselves, and considered as features to be incorporated in future, GG “currently lacks the ability to automatically determine the precise structure for a given glycan composition which is desired for generating the glycan cartoons for publication. To address this, integration of MS2 data annotation with structure elucidating powers of CandyCrunch’s into GG is being studied.” In fact, the MS2 fragment ion annotation, its similarity matching to all permissible MS2 ions derived from that precursor of assigned glycan composition, and how this is scored to provide an indication of the level of structural details that can be inferred, is not well elaborated in this manuscript.

Typically, in glycomics, due to the intrinsic nature of glycans, including their potential isomeric and isobaric variations arising from linkage, branching, and terminal substitutions at multiple allowed positions, any LC peak would still potentially comprise more than one unique structure, even with the best chromatographic compositions. This issue is compounded by the dominant fragment ions that primarily result from glycosidic cleavages, which are often insufficiently discriminative for assigning specific structures, to include or exclude the possibilities of alternative isomeric structural assignments. Linkage assignment generally requires MSn implementation or fragmentation in negative mode to favor more cross-ring cleavages. The masses of these ions can be calculated, but not without ambiguity due to identical mass numbers for different types of ions to make them sufficiently diagnostic. These challenges will not be easily resolved, and incorporating AI-driven approaches will likely be the solution - a future effort that is obviously not to be accomplished here.

However, the authors should elaborate more on what fragment ions are allowed, whether users can make judicious choices, and whether different subsets can be tailored for different modes of analysis, e.g., positive and negative mode fragmentation for non-derivatized glycans, and how permethylation or reducing end tagging with charged moiety may affect the dominant fragmentation pathways. In short, how these factors can be considered, or not considered. After matching and annotation, would scoring be implemented? Or, how would users manually choose from among the possible structures to annotate? This appears to be a feature allowed by GG, but what would be the default mode of cartoon annotation, if any?

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed all our comments to our satisfaction during their revision. We especially applaud the emphasis on a combination of user-friendly access to their software, while providing experts with the possibility for customization. We wish the authors the best with their platform

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have successfully addressed all of my concerns.

I am particularly satisfied with the inclusion of GRITS and GlycoDeNovo for glycoanalysis. The authors' detailed explanation of how these software tools handle false positives provides valuable clarification and strengthens the methodology.

With all of the modifications and revisions, I believe this article is now suitable for publication in Nature Communications.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

I appreciate the authors' response to my comments. GG still has much to further develop. I am very much looking forward to its future release versions.

(Remarks on code availability)

RESPONSE TO THE REVIEWERS'COMMENTS

Reviewer #1 (Remarks to the Author)

The manuscript by Loponte et al. describes a new software platform to accommodate the processing and analysis of mass spectrometry-derived glycomics data. Developed by experts in the field, the software seems comprehensive and user-friendly. This is also something that is needed in the field, due to both the lack of general expert knowledge and data processing heterogeneity. We're generally favorable to this work in principle, yet have quite some concerns about (i) evaluation, (ii) usability, and (iii) presentation that should be addressed and that we will detail below.

1) Our main points rest on evaluation. This is especially crucial because the results portion of this work is currently a bit thin, due to it almost exclusively focusing on the description of the developed software tool. Currently, the presented software has been evaluated on only two glycomics datasets (one of which stems from the authors' own work). This does neither cover the breadth of glycomics workflows currently in use, nor the described supported features. Especially given the fact that the software is the vast majority of the value proposition of this work, we would expect a much more exhaustive validation/testing. Further, even though it is claimed that GlycoGenius can be applied to GAGs, no GAG example is present in the manuscript

   We thank the reviewer for this valuable feedback and fully agree that broad evaluation across diverse glycomics workflows would further strengthen our manuscript. In this work, our primary aim was to benchmark GlycoGenius using datasets of the highest annotation quality, to enable a rigorous and fair comparison with manual curation, which is still the method of choice for the majority of papers presenting glycomics LC-MS/MS data.

   Identifying suitable datasets for this purpose was challenging: many published glycomics studies do not provide openly accessible, well-annotated raw data accompanied by detailed structural confirmation (e.g., MS^2 spectra or synthetic standards). To ensure that the evaluation was based on trustworthy reference annotations, we selected two datasets that:

   - Represent two different glycan classes (*N*-glycans and *O*-glycans),
   - Were generated on distinct MS platforms (Orbitrap and Q-ToF), and
   - Had undergone expert manual annotation to provide a reliable ground truth.

   We acknowledge that including glycosaminoglycan (GAG) analysis would demonstrate another important capability of GlycoGenius. Although publicly available, well-annotated GAG datasets remain scarce, we have now included a GAG dataset from Nilsson *et al*. (2023), which demonstrates that GlycoGenius can successfully detect, annotate and quantify sulfated disaccharides. Please see *Section Results – Urine GAGs Analysis by GlycoGenius* which can be found in the revised manuscript at lines 295-332.

This addition confirms that the software can be extended beyond N- and O-glycans to cover more complex glycan classes and directly addresses the reviewer's suggestion.

2) This leads to a follow-up point that, currently, the evaluation of GlycoGenius is mostly descriptive (e.g., which compositions were detected or not and why). We would expect a much more quantitative evaluation of the performance of GlycoGenius for a full picture. This should include standard metrics (e.g., sensitivity, specificity, etc etc), but also evaluate the individual steps of the workflow (if possible)

We thank the reviewer for highlighting the importance of quantitative (measurable) performance assessment of the glycan identification. We agree that standard statistical metrics provide a clearer and more objective view of the glycan identification performance, complementing the descriptive analysis presented in the original manuscript.

In the revised manuscript, we have:

- Calculated sensitivity, specificity, accuracy, precision, and negative predictive value of glycan identification for GlycoGenius under varying isotopic fitting score thresholds.
- Generated ROC curves to visualize performance and compared these directly with the results obtained from GlycReSoft.
- Added a dedicated results subsection presenting and interpreting these standard statistical metrics, to give a more complete picture of GlycoGenius' capabilities. These results are now summarized in the newly added Figure 7, which provides a comprehensive visualization of the quantitative benchmarking and ROC-based performance evaluation.

These changes can be found at the lines 333-357 of the revised manuscript. We would like to note, however, that other metrics, such as variance between AUC values, can be influenced by factors outside the software's control (e.g., sample preparation and data acquisition variability). Furthermore, although we recognize the value of stepwise assessments, a step-by-step quantitative evaluation of individual workflow components is not feasible in practice, as no validated ground truth exists for intermediate outputs (e.g., isotope fitting before composition assignment) and individual step evaluation would not consider interactions (antagonism or synergy) between the workflow steps. For this reason, our quantitative assessment focuses on the overall, end-to-end workflow performance, which we consider to be the most relevant and informative measure for end users.

3) Also connected to evaluation is the unfortunate fact that even the two datasets that have been evaluated have been done so post-factum. This allowed the authors to fine-tune their composition library building to an extent that would not be realistic in practice. I doubt anyone would start a standard O-glycan library with "Xyloses: 0 to 3", if that data peculiarity would not be known in advance.

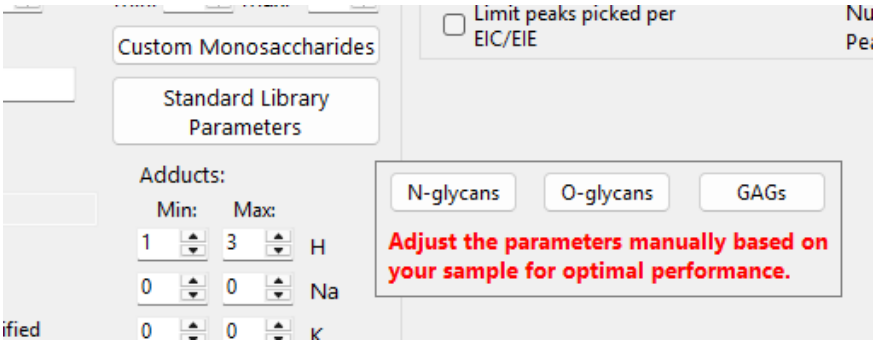
If library generation is intended to be a crucial component of the workflow and constrain false-positives etc, then a fair/realistic evaluation of this is absolutely necessary. Examples could include building "standard" O-glycan / N-glycan libraries and evaluating on them. In any case, the current evaluation is overoptimistic.
Further, even the retention time window selection performed for evaluation requires either other tools generating this insight or pre-existing knowledge which would not be present in a real-world scenario.
In general, we are dissatisfied by both the quality and quantity of evaluation for this software.

We thank the reviewer for raising these important points about ensuring a fair and realistic evaluation of GlycoGenius. We selected previously analyzed datasets precisely because they were validated and widely accepted, providing a reliable reference for benchmarking GlycoGenius against manual annotation and existing tools.

We acknowledge the reviewer's concern that the inclusion of xyloses in the *O*-glycan library involved information available at the data set's publication. Excluding xylose-containing glycans from the identifications list, GlycoGenius still detected a competitive number of glycans when compared to both the original publication and GlycReSoft (GG 34 vs. GlycReSoft 37 vs. Article 25) and assigned a good score to almost the same number of glycans as the original publication describes(GG 24 vs. 25 in the original publication). This is due to the fact that only 2 glycans detected in any of these three analyses had Xylose in its composition (H1X1 and H1X2).

We agree that incorporating standard *N*-, *O*-, and GAG-glycan libraries would improve accessibility and provide a fair baseline for evaluation. In response, we have added a "Standard Library Parameters" button to the parameters window. It provides several options (at this time, one for each glycan class: *N*-glycans, *O*-glycans and GAGs) that, when clicked, automatically adjusts the many library building parameters to default values regarding the class chosen. The user still needs to set the reducing-end tag and other relevant modifications, according to their sample preparation steps, and generate the library in the main window. These can be used as a solid starting point for the user to explore his data.

At the same time, we emphasize that GlycoGenius is not intended to be used purely as a "black box." and we also made a note on this when users want to use the standard glycan libraries, that caution is needed. Namely, prior knowledge of the sample, when available, remains highly advantageous. It allows the user to restrict the search space, streamline the workflow, and shorten runtime, while also increasing specificity. Conversely, using only predefined standard libraries may limit the discovery of unexpected or novel glycans. For this reason, we recommend that standard libraries serve as a starting point, with flexibility for user-driven refinement when the goal is to capture rare or previously unreported structures.



Regarding retention time window selection, the prior knowledge about the window where glycans are eluted is not required, since GlycoGenius can search for the glycans on the entire run time. However, such prior knowledge would shorten the time of data analysis. I addition, GlycoGenius contains a rapid built-in sub-workflow that can be used to estimate this within minutes using the Maximum Intensity Spectrum (MIS) and isotopic envelope scoring. While this allows users to define retention time windows without external tools, we appreciate the suggestion to further streamline this process. We have therefore added to our roadmap an automated one-click retention time estimation feature, which will further enhance ease of use, especially for less experienced users.

Together, we believe that the revisions already implemented, along with the planned developments, establish a more realistic and accessible evaluation framework. These changes directly address the reviewer's concerns about fairness and practicality, while ensuring that GlycoGenius continues to serve as both a powerful discovery platform and a robust solution for routine glycomics workflows.

4) This leads us to our next point: guidance in library building. The authors use "Deoxyhexoses: 0 to 2" for plasma N-glycome analysis, but why, for example, would 3 Fuc (core Fuc + 2 antennary fucosylation events) be not permitted? This again pre-supposes extreme specialist knowledge, which--if necessary--would make this software unusable to but a few people in the world (who probably already have their respective analysis workflows set up and optimized in any case). Either this software needs to be so user-friendly that glycomics newcomers can effectively use it or there needs to be a realistic path how the authors expect the world's few experts to switch to their software.

We thank the reviewer for highlighting the importance of accessibility and guidance in library building. As noted in our response to point 3, we have implemented standard *N*-, *O*-, and GAG-glycan libraries in the parameter window to provide newcomers with default, ready-to-use starting points. These defaults can be applied directly or modified as needed. This ensures that users without specialist expertise can still generate meaningful results.

It is important to emphasize that GlycoGenius does not restrict users to the example parameters shown in our manuscript. All monosaccharide counts are fully editable by the user, and there is nothing preventing the inclusion of, for example, 3 fucose residues, or any other number, if desired. Indeed, a user could set all monosaccharide counts to high upper limits (e.g., 99), and the *N*-glycan class constraint would still ensure that only structurally coherent glycans are generated, as GlycoGenius avoids certain compositions when applying class restrictions, as described at the **Extended Data Table 2**, at the Methods section.

We agree that prior knowledge of the sample, when available, is highly beneficial for tailoring library parameters. This allows the search space to be constrained, which can improve specificity, reduce false positives, and shorten runtime. However, this is not a prerequisite for using the software: newcomers can begin with standard libraries and obtain meaningful results without specialist input, while expert users can refine parameters to pursue targeted or discovery-driven analyses. In this way, GlycoGenius is designed to be accessible to beginners yet powerful enough to meet the needs of experienced glycomics researchers.

5) This then culminates in the general point that we're uncertain about: It is currently unclear how specialized one needs to be in order to (effectively) use this software. Since the main selling point of GlycoGenius is utility/convenience, this is important for its eventual scope.

We appreciate the reviewer's question regarding the level of expertise required to use GlycoGenius effectively. Our goal is to make GlycoGenius accessible to the widest possible range of users performing glycomics experiments, whether they are experienced analysts or researchers new to the field encountering glycomics data for the first time. We are aware that many users may be starting in glycomics or may receive data from third-party providers without being involved in sample preparation. In such cases, GlycoGenius can serve as an ideal first platform for learning glycomics data analysis.

To support newcomers, we provide an extensive user guide detailing all features of GlycoGenius, along with default parameter settings (including the upcoming standard *N*-, *O*-, and GAG-glycan libraries) that allow meaningful analyses without requiring advanced prior knowledge beyond the essential sample preparation information. Moreover, GlycoGenius is designed to abstract many of the technical requirements common to other tools, minimizing parameter setup burden. For example:

- No information about the specific mass spectrometer is required.

- Only two user-provided parameters are needed for analysis: PPM error and RT/MT window.
- Still, PPM error can be set conservatively (higher than actual), with results later refined using the built-in PPM error plot to identify the optimal window, and no RT/MT window may be set, which will allow GlycoGenius to examine the whole length of the samples, though it may take more time to finish the analysis.

All other information needed for proper analysis is automatically derived from the raw data file. This combination of automation, default settings, and detailed guidance ensures that GlycoGenius can be productively used by beginners, while still offering flexibility and fine-tuning options for expert users who wish to optimize their analyses. GlycoGenius offers multiple visualizations of the raw and processed data, which allow users to directly assess the impact of parameter choices, such as PPM tolerance and retention time window, and adjust them if needed.

We acknowledge, however, that certain basic experimental details must be known prior to data analysis. For example, information about reducing-end modifications and derivatization methods is necessary, as this information is required to correctly calculate the theoretical masses of each glycan composition. Such details are typically available to the user, even if they were not directly involved in data acquisition, and can be easily entered into GlycoGenius during setup. Such information is also not uncommon in other mass spectrometry related field such as proteomics when for example user should set TMT labeling as fixed modification in search engine, which information is based on knowledge from experimental details.

6) The value of software stands and falls with its maintenance. What are the plans to maintain GlycoGenius at least in the medium-term?

We thank the reviewer for emphasizing the importance of long-term maintenance and sustainability. We recognize that continuous support is essential for ensuring GlycoGenius remains functional, relevant, and compatible with evolving software environments.

Our development roadmap is publicly available via Trello (https://trello.com/b/qJU80MXM/glycogenius-dev-dashboard), where upcoming features, ongoing improvements, and planned updates are transparently documented. In the near term, our focus is on implementing the remaining planned features, after which we will prioritize bug fixes, optimization, and compatibility updates (including dependencies and Python version updates) to ensure stability.

See revised manuscript, lines 523-526: "It is important to note that the GlycoGenius is constantly being updated to address new versions of the dependencies and that the end user should not worry about them, as they are handled automatically when the program is downloaded from GitHub or installed through PyPI."

For the medium- to long-term future, we are actively seeking financial support through grants, collaborations, and potential consortium partnerships to secure resources for

continuous development. Our goal is to provide ongoing updates, incorporate community feedback, and maintain GlycoGenius as an actively supported tool for as long as possible.

7) We are also unsure about the stated runtime for real-world usage. Testing GlycoGenius on an arbitrary recent GlycoPOST (https://glycopost.glycosmos.org/entry/GPST000572) took 1 hour 47 minutes for a single file on the MS1 level alone (on a top-of-the-line Macbook). The MS2 level never finished (even after many hours; stuck on "Building fragments library")

We appreciate the reviewer raising this concern, as runtime efficiency is crucial for real-world adoption. There are many factors that can affect GlycoGenius runtime: whether the data is analyzed is profile or centroided; the retention/migration time window selected; the number of files analyzed; the size of the glycan library generated; the presence or absence of MS2; and the computer specifications, more notably the number of threads it possesses (GlycoGenius can use up to 64 threads to speed up analysis. Our User Guide (available on GitHub) provides some insights on how to better tune some of these factors.

We have tested the dataset provided by the reviewer (240604_4971_16.raw from GPST000572) using the latest version of GlycoGenius, after centroid conversion via MSConvert with peak picking. Default *O*-glycan parameters were applied with minor modifications: the minimum for each monosaccharide option was set to 0, "*O*-Glycans" was selected for the class, and reduced-end mode was used. Because the data were acquired in negative mode, the Negative Mode option was enabled. This setup generated a library of just over 500 glycan compositions. The retention time window was determined using our MIS-based workflow. *Briefly, glycans were identified in MIS, ranked by isotopic score, and then entered in "mis" on the Quick-Trace menu, which adds the m/z values of all identified glycans to the list. Tracing the first 20 compositions showed all eluting within 0–30 minutes, and this range was used for analysis. The setup procedure required ~5 minutes.* On a 12-thread desktop CPU, MS1 analysis completed in ~15 minutes, MS2 analysis in ~4 minutes, for a total runtime (including pre-processing) of ~20 minutes. This analysis identified 47 *O*-glycan compositions (24 with high isotopic and curve scores), compared to 41 reported in the original publication, with retention times in close agreement. Analysis of the same file from profile spectra took ~30 minutes. While examining glycans reported in the original publication, we noted that some *N*-glycans and Kdn-containing glycans were included there but were not considered in our GlycoGenius run.

For larger datasets (e.g., 50+ files), we recommend using a dedicated high-performance computer, as runtimes can extend to tens of hours, comparable to or shorter than state-of-the-art proteomics workflows for datasets of similar size. It is important to note that these computationally heavy analyses are best suited for high-thread-count analysis dedicated systems, and they should provide substantially faster runtimes.

As emphasized in earlier responses, defining a retention time window is key to reducing runtime. To further support users, we have added to our development roadmap a feature

that will automatically estimate the retention time window within minutes, enabling faster setup and processing.

8) Given the widespread usage of the Domon-Costello nomenclature for glycan fragments, we would expect GlycoGenius to support this. We realize that this decision has likely been taken because an MS2 fragment library makes the current nomenclature more convenient, but a "translation" (given a candidate structure) should be very feasible.

We thank the reviewer for this valuable suggestion. We agree that it would be valuable for GlycoGenius to support the Domon-Costello nomenclature. While partial support could be implemented within our current framework, GlycoGenius does not yet perform automatic structural analysis of glycans and their fragments. Without this capability, the benefit of full Domon–Costello integration is limited. At present, GlycoGenius annotates oxonium ions (which are always B-fragments), but for more complex fragments, multiple possible break points must be considered, something the current version does not yet handle. We have clarified this aspect in the revised manuscript. Specifically, we have added the following sentence: *"At present, GG's MS2 annotation is limited to oxonium ions"* (lines 280-281). This ensures that the current scope of MS2 annotation is transparent to the reader.

As we recognize the importance of this nomenclature, since submission of the manuscript, we have already made substantial improvements to MS2 annotation, including the inclusion of sulphated and phosphorylated fragments (essential for GAGs analysis) and major algorithm optimizations, and further enhancements are planned. When structural analysis of glycans and fragments is implemented in a future update, we also added the incorporation of Domon–Costello nomenclature support to our roadmap. We view this as a natural next step in ensuring that GlycoGenius annotations align with established community standards.

9) It is currently very unclear to us which of the presented figures (such as in Fig. 5) are direct results from GlycoGenius and which are manually modified for the paper. For instance Fig. 5f looks like a very useful visualization but is that one generated fully automatically by GlycoGenius or not? If not, this needs to be clarified to avoid misleading readers.

We thank the reviewer for pointing out the need to clarify which figures are generated directly by GlycoGenius and which include manual modifications. In the revised manuscript, we now state explicitly that all panels in that, for figures 5d-h, all panels were generated automatically by GlycoGenius, with one exception: in Figure 5f, the glycan cartoon in the top-right corner with highlighted colors was manually added to illustrate how GlycoGenius identifications correspond to the original structure. This clarification has been included in both the figure legend and main text to ensure it is unambiguous to readers.

See revised manuscript, lines: 277-282: "It is important to note that, while the starred annotations were generated automatically by GG, the color-coding and the glycan cartoon

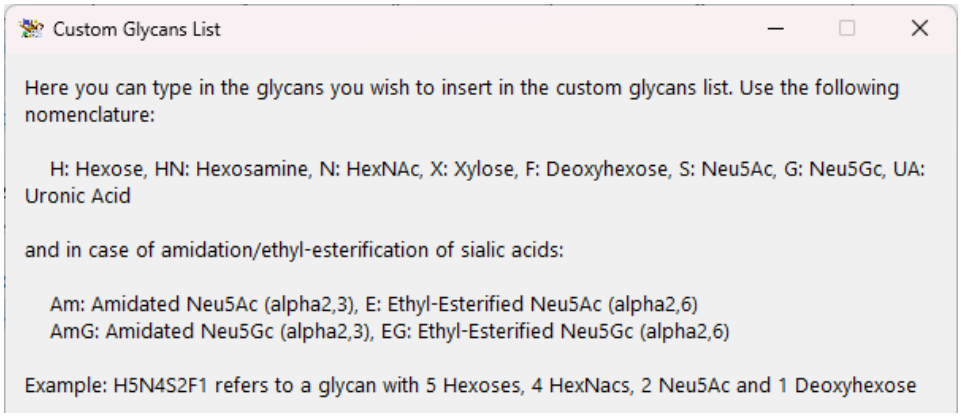
with highlighted fragments were added manually for this publication. At present, GG's MS2 annotation is limited to oxonium ions."

10) Superlative words such as "ultimate" should be avoided in titles. Very few things in science are "ultimate" (for better or worse)

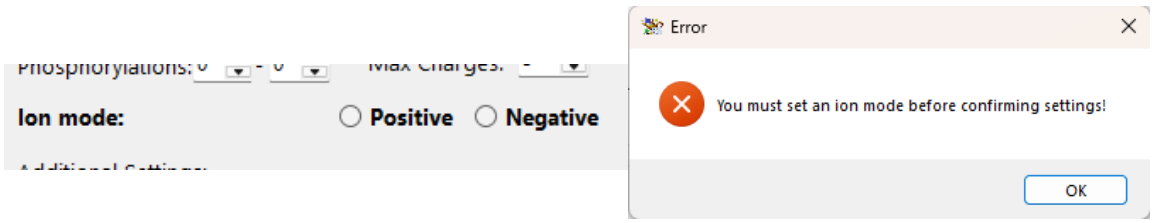
We agree with this point and have revised the title to "GlycoGenius: **a streamlined** high-throughput glycan composition identification tool".

11) For "custom monosaccharides", the expected format should be specified or explained via examples. In the GUI, "ion mode" is currently a bit hidden/not as prominent as one would expect, given that it affects every single calculation

The expected format for custom monosaccharides is described in detail with examples at the top of the "Custom Monosaccharides" window (see figure below). This information is also included in the User Guide and in the GlycoGenius monosaccharide code provided in the manuscript.



Regarding the "Ion Mode" option, we have acted on the suggestion to make it more prominent in the performance window, ensuring it is clearly visible given its impact on all calculations. Besides being bold and split into two, clear separate options, it now starts unselected and, if the user attempts to confirm the parameters without selecting it, an error message window will indicate that it needs to be set.



12) Some of the figures (e.g., Fig 5) are very pixelated / hard to read in our copy, but we're not sure whether this is a conversion issue or the input image

We thank the reviewer for pointing this out. The pixelation observed is due to PDF conversion during the manuscript upload process. The original figures are of substantially

higher quality and will be provided in their native high-quality format to ensure clarity in the final publication.

13) All used software in GlycoGenius needs to be versioned in the manuscript. Currently, the used libraries are described but not versioned.

The versions of all dependencies used by GlycoGenius have now been added to the manuscript. We note that users of the provided executable or the PyPI version of GlycoGenius do not need to manage dependency versions themselves, as these are pre-configured. Furthermore, all files generated by GlycoGenius (results and library) include metadata, accessible via the GUI, that records the version, origin, analysis and library parameters, and contents of the file, enabling full traceability.

14) The claim that "full automation has been achieved for the identification and quantification of peptides and proteins" is probably up for debate, since it mostly applies to known/established peptides/proteins(Remarks on code availability).

While many articles in the literature describe full automation for peptide and protein identification and quantification, we agree that this primarily applies to known protein sequences with expected chemical modifications, and that analyzing protein variants or modification outside of the known ones remains a significant challenge. The text has been revised accordingly to reflect this nuance.

Revised manuscript, lines: 64-67: *"While automated workflows requiring minimal user intervention for peptide and protein identification and quantification have existed for over 15 years, an equivalent level of automation has yet to be achieved for glycans."*

Reviewer #3 (Remarks to the Author)

To the Authors,

This research article presents GlycoGenius (GG), a novel open-source software tool developed for the automated analysis of glycomics data derived from liquid chromatography (LC) or capillary electrophoresis (CE) coupled with mass spectrometry (MS), including tandem MS (MS/MS). The authors assert that GG represents a significant advancement in the field by overcoming the limitations of existing tools through its fully integrated workflow and user-friendly graphical interface (GUI). The article substantiates GG's utility by comparing its performance against manual analysis and other established software using two distinct glycomics datasets.

The introduction effectively delineates the bottleneck in glycomics research arising from the inherent complexity and large volume of MS data, compounded by the scarcity of fully automated and integrated analysis tools. The authors persuasively argue for the necessity of a tool such as GlycoGenius.

The results section compellingly demonstrates GG's capacity to achieve comparable or even superior outcomes compared to manual analysis and competing software. Notably, GG identifies a greater number of glycans, including novel species, while concurrently achieving a substantial reduction in analysis time. The detailed analysis of discrepancies between GG's findings and those

of the original publications underscores the software's ability to identify potential errors or low-confidence identifications. The emphasis on an intuitive GUI, coupled with the availability of a command-line interface (CLI), effectively accommodates a diverse range of users with varying levels of bioinformatics expertise.

We thank the reviewer for her/his thoughtful and constructive feedback, as well as for her/his kind and encouraging words about our work. We greatly appreciate the time taken to provide such detailed comments, which have helped us clarify our manuscript and strengthen the presentation of GlycoGenius.

Areas for Further Consideration:

1) While the comparison with GlycReSoft is appreciated, a more comprehensive evaluation of GlycoGenius against other relevant software would strengthen the manuscript. Please include a feature-by-feature comparison with tools such as:

    • GRITS Toolbox (complex glycan analysis at the MS/MS level; https://pubmed.ncbi.nlm.nih.gov/30913289/).

    • GlycoDeNovo (elucidation of glycan topologies from MS/MS spectra).

    This comparison should explicitly highlight the unique features and algorithmic approaches of GlycoGenius, including how it addresses aspects like isobaric glycans, structural isomers, and various modifications, to substantiate claims of its superiority. The current brief mention of other tools in the introduction lacks the necessary depth for a thorough evaluation.

    We added a feature-by-feature comparison with the GRITS Toolbox to **Figure 1**, including discussion of the GELATO and GlycoDeNovo algorithms, their workflows, and their respective advantageous features and limitations. *See revised manuscript, lines 120-127: "The GRITS Toolbox is a GUI-based tool with rich metadata storage capabilities that offers two algorithms for analysis: GELATO, which leverages GlycoWorkbench to generate fragments and align their m/z values with those in MS2 spectra; and GlycoDeNovo, which generates fragments from a single MS2 spectrum and outputs a GlycoWorkbench-compatible file for visualization. The main limitation of the GRITS Toolbox, is the lack of chromatogram-based visualization and MS1-level analysis, meaning it cannot perform quantification beyond fragmentation. Although GRITS Toolbox produces a cohesive output report, it requires reporting each MS2 data of interest individually, which can be cumbersome."*

    Additional details on the unique features of GlycoGenius algorithms have been incorporated into the introduction to clarify that the unique features we claim are indeed implemented. *See revised manuscript, lines 145-148: "It supports the addition of any reducing-end tag, monosaccharide modifications, as well as phosphorylation and sulfation,*

and can detect multiple peaks within the same chromatogram. Each peak is quantified separately, enabling accurate quantification of isobaric compounds that do not co-elute."

As before, all GlycoGenius algorithms are described comprehensively in full in the methodology section of the manuscript.

[Figure Redacted]

2) The manuscript mentions the detection of different chromatographic/electropherographic peaks for potential isomer differentiation. However, the extent to which GlycoGenius can structurally resolve isomers using MS data alone (without relying on separation or prior knowledge/MS/MS interpretation) needs clarification. Please explicitly detail the software's capabilities in this regard.

   Furthermore, provide more specifics about the MS/MS annotation algorithm. Details on the fragment library generation process and the criteria used for matching fragment ions to theoretical fragments are crucial for assessing the robustness of this feature.

   The peak-picking implementation details and MS/MS annotation, together with all the other algorithms employed by GlycoGenius, are thoroughly detailed in the methodology section of the paper (lines 678-711 for the peak-picking algorithm and lines 712-741 for the MS/MS interpretation). However, we have added further details in response to the reviewer's request (lines 682-684, more precise description adjacent peaks combination and lines 730-741 describing in detail the MS2 spectra scoring and exploration features). Still, here is a short description of the two algorithms of interest for the reviewer:

- **Peak-picking:** GlycoGenius verifies that peaks are monoisotopic and with the correct charge state in MS1 spectra, close to the expected molecular mass, and trace corresponding EIC/EIE traces. Peaks are detected by identifying local maxima and minima; adjacent peaks with minimal difference between local minima and maxima (local minima within 10% of the intensity of a neighboring local maxima) are merged to avoid artificial splitting of irregular peak shapes.
- **MS/MS annotation:** For each glycan in the library, a fragment library is generated by producing all possible fragment compositions within the constraints of the glycan composition. If an MS/MS spectrum contains a precursor matching an isotopic envelope peak for a given glycan within a mass tolerance, GlycoGenius annotates the fragment peaks by matching them with the generated fragment library.

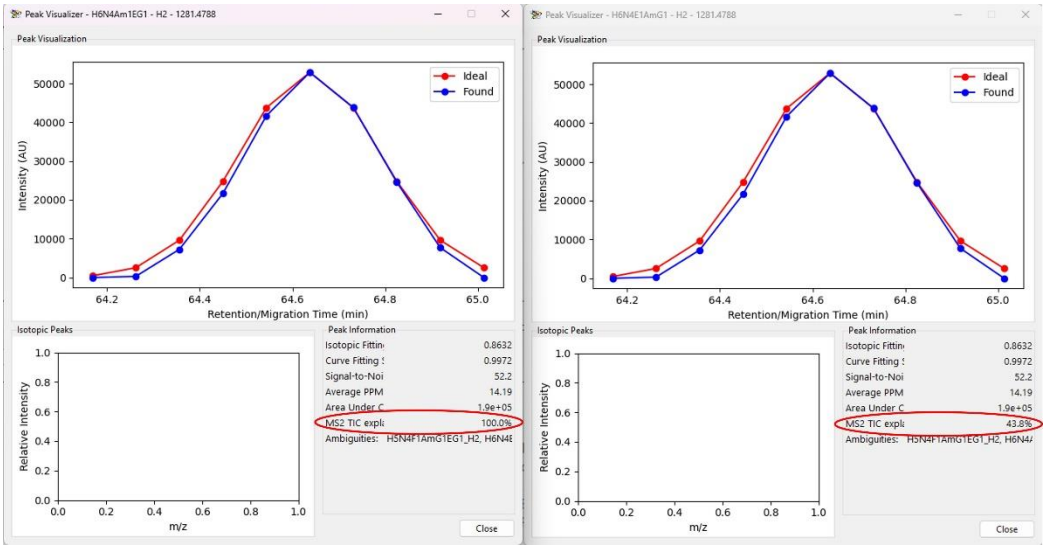
We note that the current MS/MS annotation:

- Does not yet assign **structures** to fragments;
- Generates only oxonium fragments;
- Does not reconstruct the glycan structure from fragments.

These features are planned for future releases, as outlined in our development roadmap: https://trello.com/b/qJU80MXM/glycogenius-dev-dashboard.

3) Please clarify how GlycoGenius handles co-fragmenting spectra for glycan structure elucidation. If the software reports all possible structures with different scoring, please explain the scoring mechanism. Additionally, discuss specific parameters that users should consider when analyzing glycan data acquired from different mass analyzers, such as TOF versus Orbitrap MS, and their impact on MS/MS data interpretation within GlycoGenius.

GlycoGenius annotates each MS/MS spectrum independently against all candidate glycan compositions. For example, when analyzing spectrum x for glycan y, the software may annotate z% of the peaks. If the same spectrum is then evaluated for another glycan w, the percentage of annotated peaks may differ. This side-by-side reporting of annotation

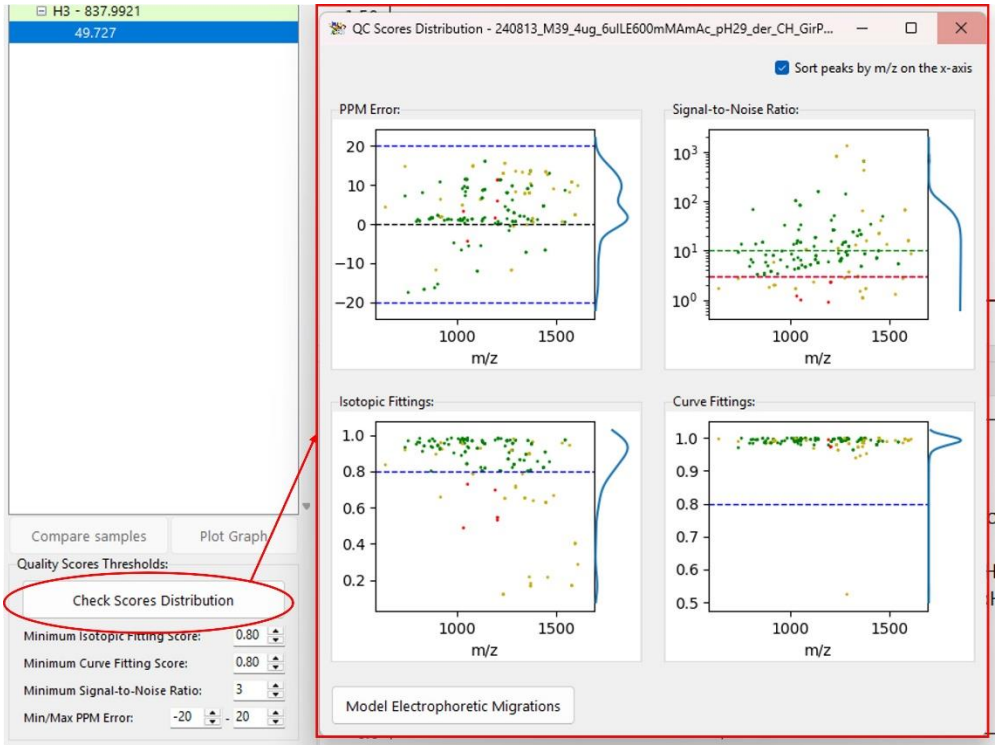


coverage allows the user to directly compare competing assignments in the GUI and determine which composition provides the most consistent explanation of the spectrum. Rather than discarding alternatives, GlycoGenius reports all plausible candidates with their respective TIC coverage (which is the score we provide), ensuring that potential co-fragmentation events are transparent to the user.

Regarding instrument types, the primary difference from GlycoGenius' perspective, both for MS1 and MS2, is the ppm error range used in matching. Orbitrap instruments typically allow for narrower ppm error windows due to higher resolution and accuracy, which increases specificity in both precursor (MS1) and fragment (MS2) matching. TOF instruments generally require broader ppm tolerances, which can slightly increase the number of potential matches for a given fragment peak, making discrimination between isobaric fragments more challenging.

The underlying matching and annotation algorithms in GlycoGenius are identical for all instrument types; what changes is the confidence level and potential ambiguity depending on the ppm setting. Users can optimize this by initially setting a broad ppm range, analyzing the data, plotting the ppm error distribution (a one-click operation in the GUI, example below), and then refining the error bounds (upper and lower) to balance sensitivity and specificity. By enabling user-defined scoring thresholds and transparent visualization of annotated coverage, GlycoGenius accommodates both analyzer types and facilitates robust interpretation across different experimental platforms.

These steps are detailed in the article (revised manuscript lines 712-741).

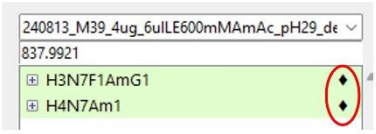


4) Potential for False Positives/Misassignments: The manuscript should address the potential for false positives or misassignments due to narrow mass error margins, particularly in cases where samples contain isobaric or near-isobaric glycans (e.g., distinguishing between two fucoses and a single Neu5Ac). Please discuss how GlycoGenius mitigates such issues. The reference provided (10.1093/glycob/cwae006) highlights challenges in glycopeptide assignment based on ppm error, and a similar discussion regarding glycan composition assignment within GlycoGenius would be beneficial

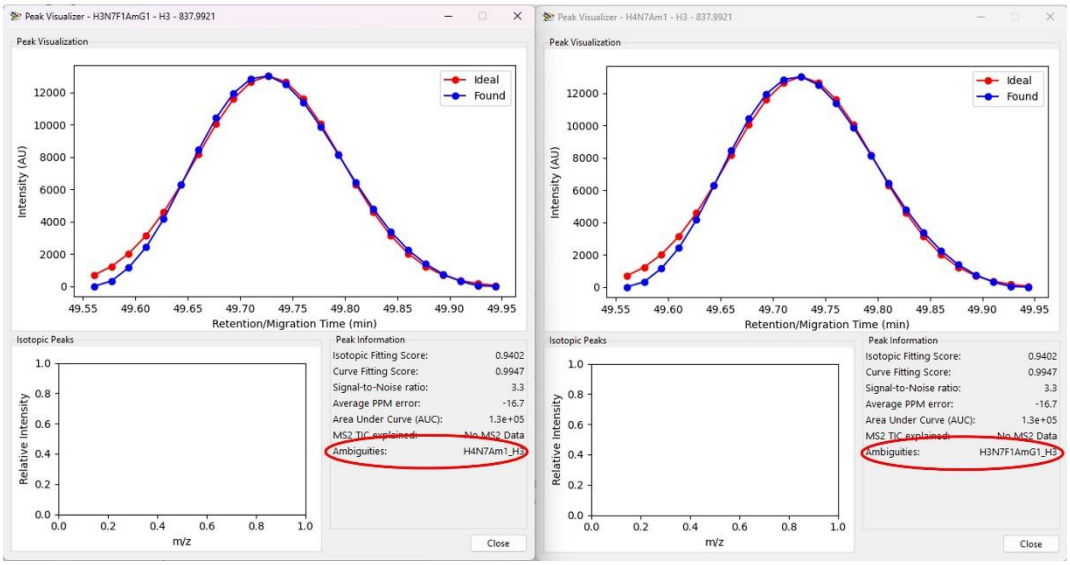
We thank the reviewer for raising this important point. We have added a new section ("Performance Meta-Analysis of GlycoGenius") to the revised manuscript presenting a meta-analysis of GlycoGenius performance, including sensitivity, accuracy, and specificity. It can be found in the revised manuscript, at lines 333-357. This analysis demonstrates that our scoring framework effectively balances false-positive reduction with high-confidence detection.

To specifically address ambiguities from isobaric or near-isobaric glycans, GlycoGenius incorporates several safeguards. First, the software automatically flags different glycans with identical masses (screenshot below, showing the GUI and the excel report). Second, composition-level scoring is not based solely on ppm error but integrates multiple quality parameters, including isotopic envelope fitting, curve fitting, and signal-to-noise ratio. This multi-parameter scoring reduces the likelihood that a low-mass-error match alone is accepted as correct. Third, MS/MS annotation provides an additional layer of evidence, allowing users to evaluate whether diagnostic fragments support one composition over another.

Looking forward, the ambiguity reporting feature will be expanded to incorporate mass error margins directly, improving differentiation of compositions with close but distinct



%_TIC_explained	Ambiguity
0	H10N3E1F1AmG1_H3, H9N3F2AmG1EG1_H3
0	H10N3Am1F1EG1_H3, H9N3F2AmG1EG1_H3
0	H9N4Am1F2AmG1_H3
0	No



masses. This enhancement has been added to our development roadmap. We thank the reviewer for this valuable suggestion which will help us make GlycoGenius even more robust against potential false positives.

Minor Corrections:

- Figure 5: Please improve the resolution of Figure 5d. Additionally, the font in Figure 5g is too small and illegible; please increase the font size for better readability.

  The resolution was a conversion limitation during the submission. Regarding 5g, as that figure constitutes a screenshot of one of the program windows, the font size itself can't be increased. Instead, we increased the size of Figure 5g as much as possible to increase clarity. The goal of that figure is to show one of the windows accessible by the user, and the information presented at that screenshot is not essential for the comprehension of the article itself.
- Table 1: In the GG Monosaccharide Code, please explain the use of square brackets around "[G]" in the codes for Am[G] and E[G].

  We thank the reviewer for this observation. In the GG monosaccharide code, the square brackets denote "optional" parameters (e.g. Am= amidated acetyl sialic acids; AmG = amidated glycolyl sialic acids). A clarifying note has been added to the table.
- Line 67: "The complexity associated with glycan analysis arise primarily from their unique challenges...". It should likely be "arises."

  Corrected.
- Line 78: "...requires interfacing between various function and modules of software that do not exhibit seamless interoperability." It should be "functions"

  Corrected.
- Line 132: The adjective should be hyphenated: "integrates feature-rich data visualization."

  Corrected.
- Line 836: The repetition of "and" should be avoided for better flow: "...manuscript and user-guide, and manages..." or "...manuscript and user-guide; manages..."

  Corrected.

Reviewer #4 (Remarks to the Author)

GlycoGenius (GG) enables high-quality identification of glycan compositions for peaks detected within an LC-MS run, as checked against S/N ratio, consistency of isotopic envelopes, and consistent signal throughout the retention time of the eluting peaks. It has advanced glycan library building, pre-processing and EIC/EIE tracing, per spectrum EIC/EIE tracing, and post-tracing processing EIC/EIE features. It further provides quantification of AUC for the generated EIC/EIE. Notably, it uses mainly the isotopic fitting score to determine the identity quality of a glycan. Along with a shorter computational time required for data processing and analysis, GG appears to be a significant advancement from GlyReSoft. It offers solutions to streamline workflows, enhance research outcomes, and facilitates the generation of quantitative glycan tables. Its interoperability provides researchers with a powerful bioinformatic tool to expedite the analysis of larger datasets.

Short of using the software myself, running it and verifying its utility with our own datasets, I am inclined to take their words for it, and on paper, every aspect seems well considered and implemented. That's about as many features as one would desire in so far as curating an LC-MS/MS-based glycomics dataset is concerned: accurately identifying peaks that matches glycan compositions allowed by the input of experimentally derived library or theorectically constructed one based on sets of constrains imposed at the level of the max/min number of each different monosaccharide. These features alone are welcoming and make GG a timely tool for glycomics. It would take away all the tedious aspects of manual processing and likely constitute the most polished tools currently available.

However, as noted by the authors themselves, and considered as features to be incorporated in future, GG "*currently lacks the ability to automatically determine the precise structure for a given glycan composition which is desired for generating the glycan cartoons for publication. To address this, integration of MS2 data annotation with structure elucidating powers of CandyCrunch's into GG is being studied.*" In fact, the MS2 fragment ion annotation, its similarity matching to all permissible MS2 ions derived from that precursor of assigned glycan composition, and how this is scored to provide an indication of the level of structural details that can be inferred, is not well elaborated in this manuscript.

Typically, in glycomics, due to the intrinsic nature of glycans, including their potential isomeric and isobaric variations arising from linkage, branching, and terminal substitutions at multiple allowed positions, any LC peak would still potentially comprise more than one unique structure, even with the best chromatographic compositions. This issue is compounded by the dominant fragment ions that primarily result from glycosidic cleavages, which are often insufficiently discriminative for assigning specific structures, to include or exclude the possibilities of alternative isomeric structural assignments. Linkage assignment generally requires MSn implementation or fragmentation in negative mode to favor more cross-ring cleavages. The masses of these ions can be calculated, but not without ambiguity due to identical mass numbers for different types of ions to make them sufficiently diagnostic. These challenges will not be easily resolved, and incorporating AI-driven approaches will likely be the solution - a future effort that is obviously not to be accomplished here.

However, the authors should elaborate more on what fragment ions are allowed, whether users can make judicious choices, and whether different subsets can be tailored for different modes of analysis, e.g., positive and negative mode fragmentation for non-derivatized glycans, and how permethylation or reducing end tagging with charged moiety may affect the dominant fragmentation pathways. In short, how these factors can be considered, or not considered. After matching and annotation, would scoring be implemented? Or, how would users manually choose from among the possible structures to annotate? This appears to be a feature allowed by GG, but what would be the default mode of cartoon annotation, if any?(Remarks on code availability)

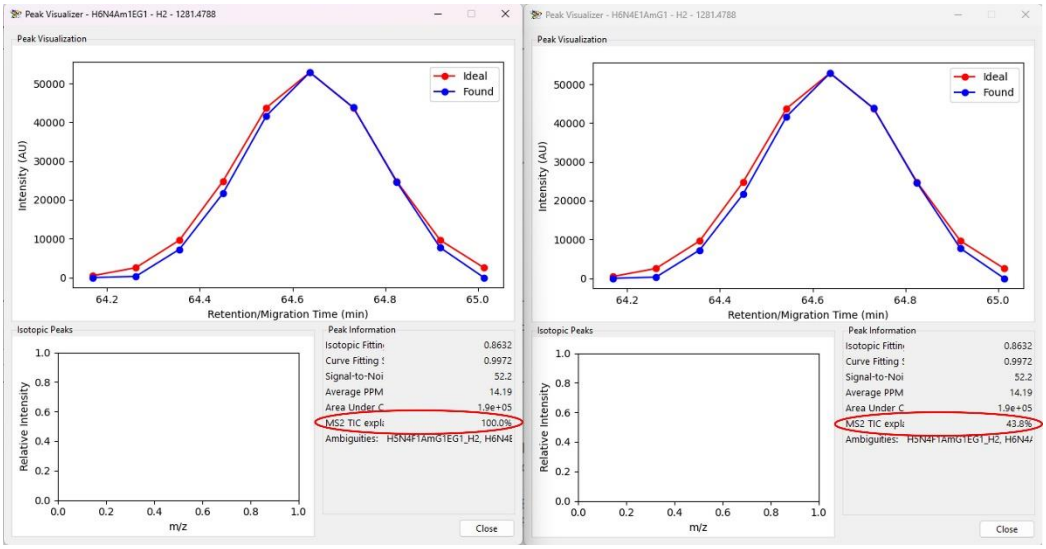
We sincerely thank the reviewer for their thoughtful and encouraging feedback on GlycoGenius, as well as for their insightful comments regarding MS2 fragment ion annotation and structural interpretation. These points are highly valuable, as they address one of the most challenging aspects of glycomics data analysis.

Below, we address each of these issues in turn:

Fragment ions and mode-specific tailoring: At present, GlycoGenius generates and annotates oxonium ions for each glycan in the library. Other fragment ion types (e.g., cross-ring cleavages, Y/Z ions) are not yet implemented. The ions generated are automatically adjusted for the adducts found on the precursor and the ion mode (positive or negative) used in the analysis. This means that mode-specific fragmentation differences are already taken into account during annotation. Development is already underway to expand support for additional fragment ion types, as well as user-defined fragment sets, is planned for a future release.

Scoring and resolving ambiguous MS2 spectra: Since GUI version 1.0.12 of GlycoGenius (a version that came after the initial submission), we have added MS2 spectrum scoring for non-annotated MS2 spectra that works as an inference of the probability that the spectra is originated from glycan fragmentation, enabling users to identify glycans that are not present in the library. This functionality also supports instantaneous, automatic annotation of any MS2 spectrum selected by the user in the GUI, using the fragments library that's built during analysis.
See revised manuscript, lines 733-744: "Additionally, GG provides the percentage of the TIC of each annotated MS2 spectrum separately for each composition it annotates (i.e. if two compositions share the same mass, the same MS2 spectra will be annotated targeting both compositions, and this will generate different TIC annotation percentage, tied to each composition), and ranks the most annotated fragments among all the annotated MS2 spectra. Using the top 10 most annotated fragments from this ranking, GG analyzes the MS2 spectra that had not been selected for annotation in the first run (due to not having a matching precursor m/z value, for example) and scores them from 0 to 10, depending on the number of fragments annotated. This provides a score that estimates the likelihood that a MS2 spectrum originates from a glycan, and is automatically adaptable across any datasets analyzed. In the GUI, users can make GG annotate one of these spectra using the fragments library built during analysis with a simple click of a button, even allowing to restrict the annotation to specific compositions. This feature facilitates further exploration of MS2 spectra in an efficient and user-directed manner."

Additionally, GlycoGenius annotates each MS/MS spectrum independently against all candidate glycan compositions. For example, when analyzing spectrum x for glycan y, the software may annotate z% of the peaks. If the same spectrum is then evaluated for another glycan w, the percentage of annotated peaks may differ. This side-by-side reporting of annotation coverage allows the user to directly compare competing assignments in the GUI and determine which composition provides the most consistent explanation of the spectrum. Rather than discarding alternatives, GlycoGenius reports all plausible candidates with their respective TIC coverage, ensuring that potential co-fragmentation events are transparent to the user.



Our long-term plan is to integrate structural scoring by leveraging CandyCrunch's deep-learning algorithms, and we are also evaluating GlycoDeNovo and GELATO. Should these approaches not yield satisfactory performance, we will develop our own algorithmic solution.

Cartoon annotation: Currently, GlycoGenius generates default glycan cartoons for each composition, with several style options from which the user can select (See **Extended Data Figure 1d**) .While user-uploaded cartoons are not yet supported, exported figures can easily be customized in external programs (e.g., to PowerPoint) can be adapted for customized presentation. Once automated structure determination is implemented as described above, GlycoGenius will also be able to automatically select an appropriate default cartoon for a given glycan composition.

Future development: Features under active consideration for future versions include:

- Support for additional fragment ion types and user-defined fragment libraries.
- Integration of structural inference and scoring for MS2 spectra.
- Automatic default cartoon selection based on structure prediction.