nature portfolio

Peer Review File

A Comprehensive N-Glycoproteome Atlas Reveals Tissue-Specific Glycan Remodeling but Non-random Structural Microheterogeneities

Corresponding Author: Dr Shisheng Sun

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)
This manuscript by Wu et al. provides an extensive high-resolution N glycoproteomic comparative atlas for 24 mouse tissues. It reports the detection of distinct 3,045 N-glycan structures that are attached to 8,681 glycopeptide sites, resulting in the cumulative compilation of 74,277 glycopeptides and 5,026 glycoproteins. The overall scope and breadth of this data set is unprecedented for any mammalian species.

In Figures 1-3, and 6, the authors provide the broad comparisons of the glycan and glycopeptide species present in the 24 tissues. In Figures 4 and 5, comparative glycopeptide and glycosite analysis is provided. The demonstration of the different glycans that are detected specifically at Asn-84 in the mannose-6-phosphate receptor glycoprotein in all tissues most effectively illustrates the types of unprecedented analyses that can be done with this type of data set. Figure 7 describes co-occurrence network analysis of glycans present in each tissues. Supplemental data figures, comparative tables and glycopeptide data sets for each tissue are provided at an appropriate level.
The technical and experimental approaches used are established, and have been used extensively by the authors in prior studies. Because of the scope of the project, some experimental and study design compromises were necessary. The authors provide their rationale for these decisions, which are reasonable.

The novelty and unprecedented amounts and types of glycoproteomic data generated are impressive, and overall well summarized. There are some concerns, clarification questions and suggestions for the authors to consider, as follows.

- Including the data summary of what different search algorithms generated is important, but more details for how the different glycopeptide search algorithm comparisons were done would be helpful. Is it known why each algorithm identified unique glycopeptides relative to each other?

- There does not appear to be any sulfated or phosphorylated N-glycans glycopeptides reported by any algorithm. Is the reason for this known?

- In the Discussion, limitations, mentioning the differences in mouse relative to human (like NeuGc, alpha-Gal) could be helpful. Most mouse models do not replicate human disease in many aspects, and it has been known for a long time in the glycobiology community that the glycan differences are major factors.

- Demonstrating that the same glycoprotein has different glycans dependent on organ location is powerful, and highlights the type of comparisons that could only be done after generating the data. The authors should consider highlighting more examples of this, like glycoproteins detected in at least half of the tissues, in a supplementary table perhaps? Are there examples that are essentially invariant for site N-glycan composition across the tissues?

- In several comparisons, it would be helpful to include some additional summary tables (supplemental) of the specific identities of the shared glycopeptides. A list of the top 20 glycoproteins for each organ, for example. In particular a list of the protein identities of the 62 ubiquitous glycoproteins detected, the 24 expressed in all tissues and the 38 in all samples

except serum. And the 51 proteins detected only in neuronal tissues.

- Data for the brain and kidney were highlighted throughout as they have quite distinct N-glycomes relative to other organs, something that many other groups have noted in previous studies. While it is not possible or relevant to cite every previous study, there are a few suggestions below for the authors to consider addressing, in the context of data comparison for glycopeptides and spatial glycomics (just for mouse brain and kidney). Co-occurrence of N-glycans has been reported in multiple mouse tissues using spatial glycomic mass spectrometry imaging methods.

- Spatial Glycomic references:

Gustafsson OJ, et al. MALDI imaging mass spectrometry of N-linked glycans on formalin-fixed paraffin-embedded murine kidney. Anal Bioanal Chem. 2015;407(8):2127-2139.

Clarke HA, et al. Spatial mapping of the brain metabolome lipidome and glycome. Nat Commun. 2025 May 12;16(1):4373. doi: 10.1038/s41467-025-59487-7.

- Murine brain and kidney glycopeptide studies:

Kawade H, et al. Bisecting GlcNAc modification of angiotensin-related glycoproteins in mouse kidney. J Biochem. 2025 Jun 13:mvaf033.

Bradberry MM, et al. N-glycoproteomics of brain synapses and synaptic vesicles. Cell Rep. 2023 Apr 25;42(4):112368.
Minor comments:
- The NeuGc and NeuAc comparisonsin Figure 5 are very useful. Were any glycan species with both NeuAc and NeuGc on the same glycopeptide detected.
- Were male versus female comparisons possible for some of the most abundant tissue types?
- Can a mouse specific glycan list be provided (as compared to humans for example)?
- Ref 7 and 43 are the same paper
- Table S4, cite the source of the RNA seq data (it was cited in the main text)
- Check again for misspelled words and a few occurrences of grammatical errors

Reviewer #2

(Remarks to the Author)
Overall evaluation

The manuscript reports the identification of 8,681 glycosylation sites across 5,026 glycoproteins and 3,045 glycans, representing an impressive scale of coverage. It also attempts to provide detailed structural information beyond glycan composition—such as core structures, bisecting GlcNAc, and terminal branched motifs—using the previously developed StrucGP tool. These structural features are indeed challenging to resolve using conventional methods.

However, several major concerns remain regarding the accuracy and confidence of glycan assignments. First, multiple glycan structures reported in the dataset appear inconsistent with known biosynthetic pathways, raising serious doubts about their validity. In some cases, the MS/MS spectra are unlikely to support the assigned glycan compositions, suggesting potential misassignments during interpretation.

Second, although the manuscript emphasizes resolution of glycan isomers, it lacks methodological detail on how such isomeric differentiation was achieved and validated. Given the inherent difficulty of resolving structural isomers in glycoproteomics, this is a critical point that must be clearly addressed.

To ensure the structural reliability of glycan annotations, we strongly encourage the authors to complement the de novo interpretation tool with rigorous validation strategies. These may include manual MS/MS spectrum interpretation, cross-database comparison, assessment of reproducibility across samples or datasets, evaluation using alternative computational tools, and orthogonal structural characterization following purification and analysis by chemical or enzymatic methods where applicable. Particular attention should be paid to biosynthetic plausibility, and the manuscript should clearly explain the logic and confidence criteria used for isomer discrimination. Without such supporting evidence, novel glycan assignments risk misinterpretation or overstatement.

At present, the manuscript falls short of the standards required for publication. Substantial revisions and validation are needed before it can be considered further.

Major comments
Q1. Glycan classification is inconsistent with biosynthetic pathways.
*N2H4F1 is categorized as high-mannose, but typically, fucose does not occur in high-mannose glycans. This composition more likely represents a hybrid structure with core fucosylation.

*Similarly, structures like N2H9F1 and N2H8F1, which include fucose within high-mannose-type glycans, have also been identified; however, from a biosynthetic perspective, such structures are unlikely to be produced. Potential misassignments in MS/MS spectral interpretations were found, please see Q3 for details.

*Structures like N3H8 also frequently appear in the identification table. However, biosynthetically, glycans with three HexNAc residues should no longer be considered high-mannose-type; the pathway toward hybrid or complex types requires mannose trimming down to N2H5 before HexNAc addition, making N3H8 unlikely to form. Potential misassignments in MS/MS spectral interpretations were found, please see Q3 for details.

Q2. Potential misassignments in MS/MS spectral interpretations
*Spectra #25304 and #25306 (sample: mouse_pancreas_IGP_hilic_50cm_4h_20200122_1, peptide: FANNTSVEPQK+N2H9F1)
Although the core peptide sequence FANNTSVEPQK appears to be correctly assigned at first glance, the Y-ion signals in these spectra only extend up to the fragment corresponding to the peptide with N3H6F1 attached (Pep+N3H6F1). There are no fragment ions supporting the presence of the larger assigned glycan (N2H9F1). Moreover, this peptide contains two potential N-glycosylation sites. Additionally, the spectra show a signal at m/z 284.044, which corresponds to a HexNAc residue modified with a sulfate group. This observation suggests that, instead of the originally assigned glycan, an alternative composition—such as N3H2F1(Sul)1 + N2H3—may provide a better explanation for the data.

*Spectra #11204 and #11206 (sample: mouse_female_kidney_mix_IGP_hilic_4h_50cm_1, peptide: FM[Oxidation]KELSDDSFENSNK+N3H8)
In spectra #11204 and #11206, the peptide was assigned as FM[Oxidation]KELSDDSFENSNK with an N3H8 glycan. However, there is only a single b-ion supporting this peptide, which is insufficient evidence to confidently confirm the identification. Meanwhile, the spectra contain numerous b-ions matching a different peptide, SHTNTSHVMQYGNK. This suggests that the original peptide assignment is likely incorrect. It seems more reasonable that the spectra actually come from SHTNTSHVMQYGNK with an N4H8, which has two possible sites for glycan attachment.

Q3. In the authors' 2021 publication describing the development of the StrucGP software, they reported identifications from both standard glycoproteins and approximately 600 distinct glycans detected in mouse brain samples. In contrast, the current manuscript reports around 3,000 glycans identified across multiple mouse tissues. Given this substantial increase in the number of identified glycans and their structural complexity, I would like the authors to clarify in detail how they ensured the accuracy of glycan assignments.
Specifically, how do the authors address glycan isomers that occur on the same glycosylation site and peptide? In Figure S4B, the N5H5F1S1 isomers are shown to be present on different proteins, and the observed retention time differences are likely attributable primarily to differences in the peptide sequences. Therefore, this data alone does not sufficiently support the claim that retention time is effective for separating glycan structural isomers.

Q4. For N-glycans with more than four antennas, how do the authors determine whether the structure is a bisecting or non-bisecting isomer? In addition, how the authors determine the number of branches in the glycan backbone? For example, in Supplementary Figure S5, the authors propose a hexa-antennary structure instead of a bisecting penta-antennary.
Please provide paired MS/MS spectra of isomeric structures—one with bisecting GlcNAc and one without—to illustrate the diagnostic fragment ions and explain the logic of how you distinguished these two possibilities.

Minor comments
Q5. line 102, page 5
"This disparity is likely arising from the absence of fractionation in the gallbladder samples, highlighting the critical role of sample fractionation in enhancing the depth of analysis for complex tissues."
If the authors wish to demonstrate the effect of fractionation, comparative data from the same tissue with and without fractionation should be included.

Q6. Line 106, page 5
While the general statement that "the number of identifications can vary between tissues even under the same MS conditions" is relevant, the referenced study by Giansanti (Ref. 27) does not directly provide data on glycoproteins or glycopeptides. Therefore, the authors should consider citing a more appropriate reference.

Q7. Line 116, page 6
"likely due to more extensive mass spectrometry coverage"
The authors mainly attribute the high number of identified glycosylation sites in the kidney to differences in MS coverage. However, several publications have demonstrated that the kidney is inherently rich in glycoproteins, and this biological background should also be considered. Moreover, if differences in the number of MS runs or fractionation conditions significantly affect the number of identifications, then comparable MS conditions (e.g., injection amount, number of runs, number of fractions) should be applied when comparing different organs. This would require a reconsideration of the overall experimental design of this study.

Q8. Line 180, page 8 (Table S4)
Please provide the selection criteria for the 209 glycosyltransferases (GTs). For example, ST6GALNAC1 shows a maximum value of only 0.05 across all tissues. If, as in Ref.4, the data are expressed in TPM (log10), this gene would not display clear differential expression across tissues. If the total number of GTs included in the dataset happens to be 209, this should also

be clearly stated.

Q9. Line 190, page 8
It should be clearly stated that the annotations of subcellular localization are based on predicted cellular localization (e.g., GO terms) rather than experimental validation. Additionally, it would be advisable to note somewhere in the manuscript that predictions of localization based on GO annotations have inherent limitations.

Q10. Line 182, page 8 (Fig. 2A, F–H)
The authors suggest that high expression of Mgat3 in brain tissues explains the elevated levels of bisected N-glycans (Core III/IV). However, in Figure 2A, bisected structures also appear abundant in intestinal tissues, where Mgat3 expression is not particularly high. This raises the question of whether mRNA expression alone can sufficiently explain the observed glycan patterns across tissues. Are there additional lines of evidence—such as lectin array results, that support the proposed correlation? Moreover, is there any broader correlation with other glycosyltransferases beyond Mgat3 and Fut9?

Q11. Line 230, page 10
The order of glycan structures differs between Figure 3B and Figure S6, making it visually difficult to confirm reproducibility. Unifying the order of glycans or providing a quantitative measure of similarity between them could improve the clarity of interpretation.
Moreover, the authors emphasized that glycosylation patterns are strongly influenced by subcellular localization. Therefore, it is important to verify whether the 24 glycoproteins compared across tissues are biased toward specific cellular compartments (e.g., the endoplasmic reticulum). It is recommend that the authors indicate the subcellular localization of these proteins and discuss whether any bias might affect the observed reproducibility.

Q12. Line 242, page 10
The authors refer to "functional demands," but the manuscript does not present direct evidence or analyses clarifying what specific functional requirements drive organ-specific glycosylation. Observing correlations alone is insufficient to support causality. Therefore, if the term "functional demands" is used, it would be helpful to include supporting data or discussion—such as references to the functional roles of glycan structures in specific localizations or analyses of associations with functionally relevant genes. Otherwise, it is suggested carefully revising expressions that imply causation.

Q13. Line 248, page 10
The authors discuss the biological significance of organ-specific glycan remodeling by using the example of glycosylation at Asn-84 of M6pr. However, it is unclear whether differences in protein abundance of M6pr across tissues were considered. Since detection frequency and structural diversity of glycan modifications can depend on protein expression levels, it is recommend providing supporting data (e.g., proteomics or RNA-seq data) to examine whether expression differences might influence glycosylation patterns. This would help clarify whether the differences in glycan structures reflect true remodeling rather than biases from protein abundance.

Q14. Line 257, page 11
In Figure 4E, the majority of glycans in the seminal vesicle show zero sialylation. Therefore, the corresponding explanation in the manuscript should be reviewed for consistency.

Q15. Line 290, page 12 (paragraph 6):
Regarding the expression "prefer to perform specialized functions," although GO analysis results are presented to suggest functional roles of glycosylation, no causal relationship between glycan structures and functional outcomes is demonstrated in the manuscript. The current evidence only supports correlations. It is recommend modifying "prefer to perform specialized functions" to a more cautious expression such as "are suggested to have functional relevance."
Furthermore, to strengthen the validity of the claimed link between glycan structures and function, it would be beneficial to include references or experimental data supporting the functional importance of specific structural features, such as Lewis antigens or LacdiNAc modifications, in the relevant tissues.

Q16. Discussion
Throughout the discussion, there are several statements suggesting causal relationships between glycan structures and their functions. Since the study does not directly test functional outcomes, it would be more appropriate to phrase these descriptions in terms of associations rather than causation.

Q17. Line 464, page 18
The statement that glycosylation "adapts to extracellular functional demands" is not directly supported by the data presented in this study. Rather, the observed glycan patterns are more likely due to structural constraints or the localization of biosynthetic enzymes. Therefore, this expression should be revised.
Please note that while Ref.56 is relevant for the first part of the sentence about biosynthetic pathways, it does not provide any information to support the second part of the sentence regarding glycosylation adapting to functional demands.

Q18. Line 486, page 19
The statement on the co-occurrence of polySia and Lewisx/a structures is interesting. However, without experimental evidence of spatial proximity (e.g., whether these modifications occur on the same molecule) or functional interactions, it seems premature to conclude "functional synergy."

Q19. In methodology, what’s the reason for further High-pH HPLC fractionation after HILIC column enrichment?

Q20. Do the authors find any N-glycans with polyLacNAc extensions? If not, what could be the reason?

Q21. In Figure 2E, what does the fifth glycan from the right represent? (Does it mean non-branching?)

Q22. In Figure 3, It would be helpful if the figure legend were more detailed. It's hard to know which tissue was used for this analysis.
Q23. Figure 3D
In Figure 2A, bisecting glycans were presented as a distinguishing feature of nervous system tissues. However, in Figure 3D, bisected structures do not appear particularly enriched in the plasma membrane fraction. This discrepancy may warrant further explanation

Q24. Figure S1A
It is difficult to understand how the figure corresponds to the main text. Please consider revising the figure or adding clarifications in the caption.

Q25. Table S3
It is unclear what the numbers in this table represent. The authors appear to indicate the percentage of each subtype or structure in each organ, but please clarify this point explicitly.

Q26. Table S6
Although Lamc1 is included as an example, it is not mentioned in the main text.

Typographical or formatting issues
Q27. Line 307, page 12
"Ploy" → should be corrected to "Poly."

Q28. "Gall bladder" should be "Gallbladder" in:
Fig. 1A, H, J; Fig. 2A, C, D, E; Fig. 3D; Fig. 4A, B, D, E; Fig. 5A, B, D; Fig. 6A, E; Fig. 7A.

Q29. In Fig. 2A, there's a red line under Liver/Serum.

Q30. In Fig. 2F: "Olfactorybulb" → should be "Olfactory bulb"

Q31. In Fig. 3E and Fig. 4C: "Celluar" → should be "Cellular"

In a study where the key contribution is biological rather than technical, structural credibility becomes paramount. This is especially important given ongoing global efforts, such as the HUPO Glycoproteomics Initiative, which stress not only data quantity but interpretability. According to HUPO's cross-platform evaluation, less than 25% of glycopeptides overlap between tools, with glycan composition being the main point of disagreement (https://doi.org/10.6084/m9.figshare.27095782.v3). Hence, rigorous structural validation is essential.

Unfortunately, the current manuscript lacks sufficient evidence to support the glycan assignments. Several assigned glycans appear biosynthetically implausible, prompting us to examine representative MS/MS spectra. In multiple cases, the fragment patterns did not support the claimed structures, suggesting misassignments and undermining confidence in the overall dataset. In addition, the claimed isomer resolution and motif detection (e.g., bisecting GlcNAc) are not convincingly demonstrated. Without such evidence, the claim of "High-Resolution N-Glycoproteome" in the title is premature.

However, should the authors validate the structure carefully, a substantially revised manuscript could be reconsidered as a new manuscript. To ensure structural reliability, novel glycan assignments should not rely solely on the developed de novo interpretation tool. Robust validation—such as (1) manual MS/MS annotation, (2) cross-database comparison, (3) reproducibility accross samples or datasets, (4) cross-tool evaluation, (5) biosynthetic plausibility assessments, (6) targeted enrichment and structural characterization—is essential to avoid misassignments and overinterpretation.

Recommended Criteria for Validation (for reference)

To enhance the reliability of glycan structural assignments, it is strongly recommended that the authors clearly distinguish between previously reported structures and novel assignments. For known glycans, validation should be performed by cross-referencing established databases such as UniCarb-DB or GlyTouCan. For novel structures, extensive evidence more than the developed de novo tool is required to avoid misassignments or overinterpretation. The following validation strategies are recommended:

1. Manual MS/MS annotation
Manual annotation of representative and structurally complex examples is required. This includes identification of diagnostic ions that clearly support the proposed isomer assignments. Resolving isomeric glycan structures at the same glycosylation site on an identical peptide backbone provides more convincing evidence than comparisons made between different peptides. Such examples should be fully presented to support the claim of "high-resolution glycoproteomics."

2. Cross-database comparison
Verify whether each assigned structure is registered in international databases such as GlyTouCan, and report matches/mismatches against UniCarb-DB.

3. Reproducibility Across Samples or Datasets
Confirm whether the identified structures are consistently observed in other tissue batches, independent datasets, or biological replicates. If structures are observed in only one sample, provide supporting data to establish their validity.

4. Cross-tool evaluation
For glycan structures uniquely identified by StrucGP, explain why they are not detected by other tools (e.g., pGlyco, Byonic). Provide MS/MS-based evidence to support that the structures are not tool-specific false positives.

5. Biosynthetic plausibility assessments
Investigate whether the assigned structures are consistent with known glycosyltransferase specificities and established biosynthetic pathways. If not, provide direct literature evidence or in vitro data to support the biological plausibility.

6. Targeted enrichment and structural characterization
For particularly novel structures, further biochemical validation is recommended, including glycopeptide isolation, specific enzymatic digestion, PMAA linkage analysis, and LC-MS^n analysis.

Reviewer's expertise
I have extensive experience in using lectin arrays and mass spectrometry to investigate inter-tissue glycan heterogeneity in mouse models (e.g., Analytical and Bioanalytical Chemistry (2025) 417:973–988). The content of the current manuscript falls well within my area of expertise.

Reviewer #3

(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.

Reviewer #4

(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.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this revised manuscript, the authors very effectively addressed questions and concerns from my previous critique. I disagree with them only in one of their rebuttal statements indicating that sulfation of N-glycans are rare, but there is no need to further address that relative to this manuscript.

Reviewer #2

(Remarks to the Author)
We thank the authors for the substantial effort in revising the manuscript and providing detailed responses to our previous comments. While some our earlier concerns have been addressed, several key points remain unresolved and prevent us from recommending acceptance of the manuscript in its current form.
Our main concern is the lack of clarity regarding structural confidence across the 3,045 reported N-glycan structures. The authors acknowledge several misassignments we identified and state that these account for less than 0.5% of all structures. However, there is still no dataset-wide, systematic quantification of structural confidence. Figures present all structures as fully determined, without visual or quantitative distinction between high- and low-confidence assignments. Additionally, a glycan-level false discovery rate supported by a realistic decoy strategy has not been demonstrated, leaving the global structure-level error rate unclear.
In our previous review, we outlined six validation criteria. The current revision addresses these only partially:
(1) Manual annotation was performed for a limited set of examples, without extending to a statistically meaningful proportion of the dataset, and without providing a clear basis for the stated misassignment rate of less than 0.5%.
(2) Cross-database matching was restricted to GlyConnect, without systematic reporting of matches/mismatches to other major databases or explicit differentiation between known and novel structures.

(3) Reproducibility was shown for selected tissues/legacy data, but not quantified across the dataset.
(4) Cross-tool analysis appears limited to selected subsets and largely at the glycan composition level; no structure-level concordance metric or spectrum-based adjudication of discrepancies is provided.
(5) Biosynthetic plausibility is discussed qualitatively, but the number/proportion of pathway-inconsistent calls is unquantified.
(6) Orthogonal validation (e.g., stepwise exoglycosidase assays, binding assays with anti-glycan probes such as antibodies and lectins) for novel/critical features is not provided.
We fully appreciate the challenges inherent to N-glycan structural analysis, particularly in backbone classification, antenna number determination, and algorithmic assignment. However, given that these dataset-level validation steps remain incomplete, further work is necessary before the study can provide the level of reliability and interpretability expected for publication in Nature Communications.
We encourage the authors to consider a systematic, dataset-wide validation—including quantitative structural confidence measures, expanded cross-database matching, structure-level cross-tool concordance, and targeted orthogonal experiments—before resubmission. Such steps would greatly strengthen the impact and reliability of this valuable dataset.

Reviewer #3

(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.

Reviewer #4

(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.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Comments to the Authors

We thank the authors for their continued efforts in revising the manuscript and for adding materials such as Supplementary Table S2 and expanded database annotations. However, our central concern remains unresolved: among the 3,045 reported N-glycan structures, which are high- versus low-confidence, and in what proportions? The manuscript still does not provide dataset-wide, quantitative measures of structural confidence, and figures continue to present all glycans as if fully determined, without distinguishing levels of certainty.
Note. Items 3 and 4 are considered addressed (within current practical limits). Items 1, 2, 5, and 6 remain outstanding.
The six validation criteria remain only partially addressed:
1. Manual annotation and misassignment rate – The reported “<0.5% error rate” is derived only from high-mannose glycans that we had highlighted as potential misassignments, and does not reflect more ambiguous features. In addition, the Highest_GlycanScore is not interpretable as a confidence metric, since no thresholds or categories are provided.
2. Cross-database matching – Expanded to GlyTouCan, but still restricted to composition-level comparisons. Structural-level confirmation remains absent.
3. Reproducibility [Addressed] – We acknowledge that the authors have provided reproducibility statistics across the dataset, including replicate-level analyses.
4. Cross-tool analysis [Addressed within current software limits] – The authors have performed cross-tool comparisons at the glycan composition level. Given the current limitations of available software, we agree that structure-level comparisons are not feasible, and we consider this point addressed as far as is reasonably possible.
5. Biosynthetic plausibility – New “red-flag” criteria were applied, but flagged glycans remain included in the overall figures and statistics. As a result, readers cannot distinguish between high- and low-confidence structures. These criteria are heuristic rather than a quantitative confidence framework.
6. Orthogonal validation – No systematic validation (e.g., exoglycosidase digests, lectin/antibody assays) is presented for novel or critical structures.

In conclusion, while transparency has improved through supplementary annotations, the fundamental issue of dataset-wide structural confidence remains to be fully addressed.

Reviewer #3

(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.

Reviewer #4

(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.

Version 3:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
We thank the authors for their continued efforts on this third revision and for substantially improving transparency by introducing a five-criterion confidence framework. However, the scientific justification of the framework remains insufficient, and the headline claim of “high-resolution characterization of 3,045 N-glycan structures” is not yet supported as written. Only 34.2% (1,024/3,045) match GlyTouCan at the structure level, and the current High/Intermediate/Low distribution (67.0%/27.5%/5.5%) requires a rigorously justified framework and systematic validation to substantiate whole-structure identification across all entries.

More specifically: (i) the five-criterion weighting lacks scientific rationale; (ii) direct MS/MS evidence (GlycanScore) contributes only 10/100 relative to largely indirect criteria; (iii) GlycanScore is under-specified (inputs, scoring function, interpretation); and (iv) the ~700 manual MS2 verifications are not propagated into confidence assignments or any auditable misassignment/FDR estimate.

In addition, analyses should be limited to structures above the defined confidence threshold; otherwise, the revision improves transparency but does not amount to validation.

Revisions requested
This revision substantially improves transparency and represents a valuable large-scale data resource. To strengthen the manuscript within the current scope, please address the items below.
1. Reframe the main focus and narrative
Shift from “identification of 3,045 structures” to a resource article centered on sub-structure–based biological insights beyond monosaccharide composition. When citing counts, report the confidence distribution (High/Intermediate/Low).
2. Provide a scientifically justified confidence framework (the most important)
Provide a scientifically justified and fully specified confidence framework.
Redefine and clearly document the framework: state the criteria, weights, thresholds, and class boundaries with rationale; mathematically specify GlycanScore (inputs, scoring function, interpretation); spell out the decision rules/data flow from evidence to confidence assignment; link the ~700 manual MS2 verifications to per-call confidence; provide dataset-level misassignment/FDR and sensitivity analyses showing that indirect criteria cannot outweigh direct MS2 evidence.
3. Apply a confidence cut-off for analyses
Restrict biological comparisons to High-confidence calls; inclusion of Intermediate may be allowed only if justified by the revised framework. Known or likely misassignments (e.g., sulfation, O-glycan co-occurrence, poly-LacNAc ambiguities) should be flagged and excluded from primary analyses where feasible.
4. State methodological limits
Explicitly connect current constraints (search space, fragmentation limits, branch number ambiguity) to what cannot be concluded at whole-structure resolution.

Status of the six validation criteria of the current manuscript
1. Manual annotation and misassignment rate [Not adequately addressed]–
As noted above, the framework’s scientific justification remains insufficient.
2. Cross-database matching [Addressed]
3. Reproducibility [Addressed]
4. Cross-tool analysis [Addressed within current software limits]
5. Biosynthetic plausibility [Addressed]
6. Orthogonal validation [Not addressed, but acceptable as a limitation]–
If the manuscript reframed as a sub-structure–centric resource, this can be stated in the limitation section.

Reviewer #3

(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.

Reviewer #4

(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.

Version 4:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Comments to the Authors

We appreciate the authors' efforts to introduce a systematic confidence framework using supervised learning, which clearly improved reliability. While the manuscript has progressed, a bit more clarification would further strengthen the scientific soundness of the framework and its alignment with the paper's overall narrative. In particular, please use confidence-aware (High/Intermediate/Low) wording for structure counts in the Abstract.

Below are two items that remain only partially addressed:

(1) Manual annotation and misassignment rate — Partially addressed
The authors describe a logistic regression model based on five features trained on 472 manually annotated examples. To further improve scientific soundness:
•Data selection criteria for the gold-standard dataset: Briefly state the criteria by which ~700 previously reported manual annotations were narrowed to 472 (e.g., spectrum quality, structure categories), so readers can assess representativeness.
•Class counts: Report the numbers of structurally reasonable vs structurally ambiguous examples used for training.
•Feature contribution: Present regression coefficients (with signs) or feature importance so readers can intuitively understand the relative influence of the five criteria on confidence estimation.
•GlycanScore interpretation: Add a short, qualitative sentence connecting the score to structural confidence and indicating how higher scores related to the tier boundaries (no formal proof needed).
•Model positioning: State explicitly that the model is not a true/false classifier, but an integrative framework that combines MS2 evidence and database consistency to assist confidence estimation.

(6) Orthogonal validation — Not addressed, acceptable as a limitation
Systematic orthogonal validation was not performed; state this explicitly as a limitation in the Discussion. Because the primary conclusions rest on sub-structure-level comparisons, make that focus explicit in the Discussion and clarify why full orthogonal validation is not required for the main claims. To avoid over-interpretation, use confidence-aware (High/Intermediate/Low) wording for structure counts in the Abstract. For example: "assigned 3,045 candidate N-glycan structures; among these, 2,687 (88.2%) met the high-confidence threshold." This wording maintains transparency without diminishing the value of the work.

Minor comment
In Fig. 5E, the confidence-level annotation appears missing—please verify and add if needed.
Visualization of confidence: In figures where color already represents tissue/system (e.g., Fig. 2A), consider using a different visual channel for confidence (e.g., color + hatching or line style).
A few typos (e.g., microheterogeneitie (line 135), Ploy-sialic/ploySia (lines 340, 348, 437,), co-existance (line 1005), thw white matter (line 558)); please proofread.

Reviewer #3

(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.

Reviewer #4

(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.

Reviewer #1 (Remarks to the Author):

This manuscript by Wu et al. provides an extensive high-resolution *N*-glycoproteomic comparative atlas for 24 mouse tissues. It reports the detection of distinct 3,045 *N*-glycan structures that are attached to 8,681 glycopeptide sites, resulting in the cumulative compilation of 74,277 glycopeptides and 5,026 glycoproteins. The overall scope and breadth of this data set is unprecedented for any mammalian species.

In Figures 1-3, and 6, the authors provide the broad comparisons of the glycan and glycopeptide species present in the 24 tissues. In Figures 4 and 5, comparative glycopeptide and glycosite analysis is provided. The demonstration of the different glycans that are detected specifically at Asn-84 in the mannose-6-phosphate receptor glycoprotein in all tissues most effectively illustrates the types of unprecedented analyses that can be done with this type of data set. Figure 7 describes co-occurrence network analysis of glycans present in each tissue. Supplemental data figures, comparative tables and glycopeptide data sets for each tissue are provided at an appropriate level.

The technical and experimental approaches used are established, and have been used extensively by the authors in prior studies. Because of the scope of the project, some experimental and study design compromises were necessary. The authors provide their rationale for these decisions, which are reasonable.

The novelty and unprecedented amounts and types of glycoproteomic data generated are impressive, and overall well summarized. There are some concerns, clarification questions and suggestions for the authors to consider, as follows.

Response: We thank the reviewer for his/her overall positive evaluations on our work, and appreciate the professional and constructive comments and suggestions. Below we have addressed each comment in a point-by-point manner and made corresponding revisions in the manuscript accordingly.

- Including the data summary of what different search algorithms generated is important, but more details for how the different glycopeptide search algorithm comparisons were done would be helpful. Is it known why each algorithm identified unique glycopeptides relative to each other?

Response: We thank the reviewer for this thoughtful comment. Based on previous studies and our own analysis, the differences in glycopeptide identifications among various search algorithms can be primarily attributed to the following four factors:

1. **Scoring and filtering differences:**

Each algorithm uses different scoring rules for glycopeptide-spectrum matches. For example, *pGlyco* requires strong glycan fragment evidence, *Byonic* allows partial glycan matches, and *MSFragger-Glyco* uses an open search approach that increases sensitivity but may reduce specificity.

2. **Glycan database differences:**

   The glycan search space varies across tools. *pGlyco* uses a full glycan structure database, while *StrucGP* uses a branch-based glycan library. *Byonic* and *MSFragger-Glyco* allow semi-combinatorial glycan generation. These differences affect what types of glycans and glycopeptides can be identified.

3. **Fragmentation model differences:**

   Tools like *StrucGP* and *pGlyco* consider more detailed glycan fragmentation patterns and linkage-specific ions, which improves accuracy but may also filter out lower-confidence matches.

4. **FDR control strategies:**

   Different tools use different methods to estimate and control FDR at both the peptide and glycan levels, leading to different thresholds for accepting results. Specifically, although both *pGlyco* and *StrucGP* follow a glycan-first identification strategy, their decoy construction methods differ. As far as we understand, *pGlyco* generates decoy spectra by adding random mass shifts to Y ions within each spectrum. This approach primarily evaluates the correctness of the peptide mass or the Y1 ion. However, such decoys often lack the conserved *N*-glycan pentasaccharide core and differ significantly from real *N*-glycopeptide spectra, which may limit their utility in accurately assessing glycan-level FDR. In contrast, our decoy strategy introduces random mass shifts at the precursor ion level for each spectrum. This generates decoy spectra that retain the conserved pentasaccharide core but differ only in glycan branching structures. As a result, our decoys more closely resemble true *N*-glycopeptide spectra and allow for more realistic and reliable estimation of glycan-level FDR, thereby improving confidence in glycan assignments.

In summary, although different algorithms report partly overlapping and partly unique results, combining them can increase glycopeptide coverage. Our dataset from 24 tissues provides a valuable resource for such integration. However, how to merge results with different confidence levels remains a challenge. Recent studies (*Nat Commun*. 2025 Jul 1;16(1):5568.) suggest that AI models trained on multi-confidence data may help address this.

In the revised manuscript, we re-searched the gallbladder dataset using other tools (*pGlyco*, *Byonic* and *MSFragger-Glyco*) and added the results to **Supplementary Figure 1**, with related updates in **Supplementary Table 2**. We added a comparative description in the *Results* section (**Page 5**), highlighting that StrucGP uniquely identifies a larger number of glycoproteins and glycosites, demonstrating its sensitivity for distinct

glycopeptide features. In the *Discussion* (**Page 18**), we acknowledged that different tools yield partially distinct identifications and emphasized that consensus results across multiple engines can improve data coverage, underscoring the dataset's value for such cross-platform analyses.

- There does not appear to be any sulfated or phosphorylated *N*-glycans glycopeptides reported by any algorithm. Is the reason for this known?

Response: We appreciate the valuable comments. In general, the low detection rate of sulfated and phosphorylated *N*-glycans can be attributed to several technical and biological factors. First, sulfation of *N*-glycans occurs at very low abundance in mammals and is often masked by highly abundant unmodified glycopeptides in MS spectra, making it more difficult to identify. Second, sulfate groups are labile and prone to dissociation during fragmentation, leading to the loss of key diagnostic ions and further complicating confident identification. Compared to sulfation, phosphorylation is relatively more stable and easier to detect. However, increasing the identification rate of both modifications, especially sulfation, typically requires specialized sample preparation workflows. For instance, phospho-glycopeptides often benefit from enrichment strategies such as IMAC to enhance specificity, whereas for sulfated glycopeptides, there is currently no well-established enrichment method to selectively improve detection.

In our study, sulfated and phosphorylated *N*-glycans were not reported in the dataset primarily because our glycopeptide search algorithms rely on a predefined glycan branch library that must explicitly include such modified structures (with defined compositions and masses) for successful spectral matching. According to current literature, sulfation of *N*-glycans in mammals occurs at extremely low levels. Therefore, to avoid inflating the search space and increasing the risk of false-positive identifications, we deliberately excluded sulfated and phosphorylated glycans from the initial search library used in this study.

Accurate identification of these rare modifications remains technically challenging and requires carefully curated search libraries along with rigorous false discovery rate (FDR) control and optimized collision energy settings. We are currently developing and validating an expanded glycan branch library that includes sulfated and phosphorylated structures, coupled with corresponding analytical workflows. In this context, we have established a lower-energy fragmentation strategy to better preserve diagnostic B ions, although the associated technology is still under development. Until this strategy is fully established and benchmarked, incorporating these modifications into routine searches could compromise data quality by introducing spurious matches.

Nevertheless, the high-coverage glycopeptide dataset derived from 24 mouse tissues in this study provides a solid foundation and can be leveraged in future targeted efforts to identify and validate such rare or regulatory

glycan modifications under optimized conditions.

In the revised manuscript, we have added a statement in the *Discussion* (**Page 18**) emphasizing that our mouse tissue dataset can support not only existing tools but also the development of new algorithms for identifying diverse glycan features, including rare modifications (e.g., sulfation, phosphorylation, *O*-acetylation), uncommon glycan structures or monosaccharides, and even *O*-glycans.

- In the Discussion, limitations, mentioning the differences in mouse relative to human (like NeuGc, alpha-Gal) could be helpful. Most mouse models do not replicate human disease in many aspects, and it has been known for a long time in the glycobiology community that the glycan differences are major factors.

Response: We appreciate this insightful comment. We agree that species-specific glycosylation patterns, such as the expression of Neu5Gc and α-Gal in mice but not in humans, are important considerations when interpreting the translational relevance of our findings.

In the revised manuscript, we have discussed the limitations of directly extrapolating mouse glycosylation patterns to humans due to species-specific differences in the *Discussion* section (**Page 21**).

- Demonstrating that the same glycoprotein has different glycans dependent on organ location is powerful, and highlights the type of comparisons that could only be done after generating the data. The authors should consider highlighting more examples of this, like glycoproteins detected in at least half of the tissues, in a supplementary table perhaps? Are there examples that are essentially invariant for site *N*-glycan composition across the tissues?

Response: We appreciate the reviewer's insightful suggestion. To better illustrate tissue-dependent heterogeneity and conservation of *N*-glycan structures at specific glycosites, we performed a new analysis by calculating a similarity score matrix for 49 glycosites derived from 38 commonly expressed glycoproteins across 23 tissues (excluding serum). In this matrix, a higher score indicates greater similarity of glycan structures at the same site across tissues, while a lower score reflects more site-specific heterogeneity. The scoring criteria and statistical methods are detailed in the revised *Methods* section (**Page 24-25**).

In the revised manuscript, we have added incorporated the new analysis into **Supplementary Table 7** (updated from the original Supplementary Table 6), which provides the full list of site-specific glycan similarity scores across tissues, ranked from high to low. Additionally, we have added **Supplementary Figure 8**, which illustrates representative examples of (1) glycosites exhibiting high glycan heterogeneity across tissues (**Lamc1, Asn-1221**) and (2) glycosites with conserved glycan compositions (**Cemip2, Asn-914**). The

corresponding descriptions, which highlight the tissue-dependent heterogeneity observed at Lamc1 Asn-1221 and the conserved glycosylation pattern at Cemip2 Asn-914, have also been added in the *Results* section of the revised manuscript (**Page 11**).

- In several comparisons, it would be helpful to include some additional summary tables (supplemental) of the specific identities of the shared glycopeptides. A list of the top 20 glycoproteins for each organ, for example. In particular a list of the protein identities of the 62 ubiquitous glycoproteins detected, the 24 expressed in all tissues and the 38 in all samples except serum. And the 51 proteins detected only in neuronal tissues.

Response: We thank the reviewer for this constructive suggestion. In the revised supplementary materials, we have now included several new summary tables to facilitate deeper comparison across tissues. Specifically, we provide: (1) A list of the top 20 most abundant glycoproteins identified in each tissue (**Supplementary Table S8**); (2) A detailed list of the 24 proteins detected in all tissues, the 38 proteins identified in all tissues except serum, and the 51 glycoproteins uniquely detected in neuronal tissues (**Supplementary Table S6**).

- Data for the brain and kidney were highlighted throughout as they have quite distinct *N*-glycomes relative to other organs, something that many other groups have noted in previous studies. While it is not possible or relevant to cite every previous study, there are a few suggestions below for the authors to consider addressing, in the context of data comparison for glycopeptides and spatial glycomics (just for mouse brain and kidney). Co-occurrence of *N*-glycans has been reported in multiple mouse tissues using spatial glycomic mass spectrometry imaging methods.

- Spatial Glycomic references:

Gustafsson OJ, et al. MALDI imaging mass spectrometry of N-linked glycans on formalin-fixed paraffin-embedded murine kidney. Anal Bioanal Chem. 2015;407(8):2127-2139.

Clarke HA, et al. Spatial mapping of the brain metabolome lipidome and glycome. Nat Commun. 2025 May 12;16(1):4373. doi: 10.1038/s41467-025-59487-7.

- Murine brain and kidney glycopeptide studies:

Kawade H, et al. Bisecting GlcNAc modification of angiotensin-related glycoproteins in mouse kidney. J Biochem. 2025 Jun 13:mvaf033.

Bradberry MM, et al. *N*-glycoproteomics of brain synapses and synaptic vesicles. Cell Rep. 2023 Apr 25;42(4):112368.

Response: We thank the reviewer for these valuable references and suggestions. We agree that the brain and

kidney exhibit highly distinct *N*-glycosylation features, which are consistent with previous spatial glycomics and glycopeptide profiling studies. Notably, the cited studies further validate the accuracy of our overall glycan structure assignments by independently supporting key tissue-specific patterns observed in our dataset.

In the revised manuscript, we have now cited the suggested publications in the *Result* section (**Page 8**) to better contextualize our findings and highlight their consistency with established tissue-specific glycosylation patterns reported in the literature.

Minor comments:

- The NeuGc and NeuAc comparisons in Figure 5 are very useful. Were any glycan species with both NeuAc and NeuGc on the same glycopeptide detected.

Response: We thank the reviewer for this insightful question. Upon re-examining our dataset, we found that glycan containing both NeuAc and NeuGc on the same structure were indeed present and accounted for approximately 2%–25% of all unique glycans identified in each tissue. The five mouse tissues with the highest proportions of such dual-sialylated glycans were heart (25%), liver (18.4%), muscle (17.1%), testis (17.0%), and spleen (15.5%). These results suggest that co-occurrence of NeuAc and NeuGc, while not dominant, is not uncommon in certain tissues.

In the revised manuscript, we have added this result on **Page 13**, and included the corresponding spectra and statistics in a new supplementary figure (**Supplementary Figure 9A and 9B**).

- Were male versus female comparisons possible for some of the most abundant tissue types?

Response: We thank the reviewer for this valuable suggestion. Given that the kidney yielded the highest number of glycoprotein and glycopeptide identifications in our dataset, we had indeed performed a comparative analysis of the renal *N*-glycoproteome between male and female mice. While the overall profiles—such as the numbers and identities of glycopeptides, glycoproteins, and glycosites—were highly similar between sexes, with only subtle structural differences observed.

Specifically, we observed subtle but significant differences between male and female kidneys: female kidneys showed enriched LacdiNAc structures and predominantly Neu5Ac-containing glycosylated glycans, whereas male kidneys showed a higher prevalence of Neu5Gc-modified *N*-glycans. In addition, female-enriched glycans were functionally associated with reproductive and immune processes, whereas male-enriched glycans were associated with protein localization and metabolic functions. These findings provide new insights into gender-related glycosylation profiles and underscore the importance of considering gender

as a biological variable in glycoproteomics studies.

In the revised manuscript, we have added a summary of the sex-specific differences in the *Results* section (**Page 7**), and included detailed data and visualization in **Supplementary Figure 6**.

- Can a mouse specific glycan list be provided (as compared to humans for example)?

Response: We appreciate the reviewer's suggestion regarding a mouse-specific glycan list. Currently, we do not possess a human glycan structural database generated under identical experimental conditions as our mouse dataset. Moreover, existing public glycan databases may not fully correspond to the glycan structures identified in our mouse samples.

In the revised manuscript, we have expanded the *Discussion* section to include a brief comparison highlighting key differences in *N*-glycosylation between humans and mice (**Page 20**), which we hope will help clarify species-specific glycosylation features and assist readers in interpreting our findings. Importantly, this dataset also provides a foundation for future efforts aimed at systematically defining human–mouse glycan differences, which will be critical for translational research and comparative glycobiology.

- Ref 7 and 43 are the same paper
- Table S4, cite the source of the RNA seq data (it was cited in the main text)
- Check again for misspelled words and a few occurrences of grammatical errors

Response: We thank the reviewer for the careful reading of our manuscript. We have corrected the duplicated citation by removing the redundant entry and updating the reference list accordingly. In **Table S4**, we have now clearly cited the source of the RNA-seq dataset, consistent with the citation already provided in the manuscript. Additionally, we have carefully rechecked the manuscript for typographical and grammatical errors and made necessary corrections throughout the manuscript.

We sincerely thank the reviewers again for their constructive and insightful comments. These suggestions have greatly helped us improve the quality and clarity of the manuscript.

Reviewer #2 (Remarks to the Author):

Overall evaluation

The manuscript reports the identification of 8,681 glycosylation sites across 5,026 glycoproteins and 3,045 glycans, representing an impressive scale of coverage. It also attempts to provide detailed structural information beyond glycan composition—such as core structures, bisecting GlcNAc, and terminal branched motifs—using the previously developed StrucGP tool. These structural features are indeed challenging to resolve using conventional methods.

However, several major concerns remain regarding the accuracy and confidence of glycan assignments. First, multiple glycan structures reported in the dataset appear inconsistent with known biosynthetic pathways, raising serious doubts about their validity. In some cases, the MS/MS spectra are unlikely to support the assigned glycan compositions, suggesting potential misassignments during interpretation.

Second, although the manuscript emphasizes resolution of glycan isomers, it lacks methodological detail on how such isomeric differentiation was achieved and validated. Given the inherent difficulty of resolving structural isomers in glycoproteomics, this is a critical point that must be clearly addressed.

To ensure the structural reliability of glycan annotations, we strongly encourage the authors to complement the de novo interpretation tool with rigorous validation strategies. These may include manual MS/MS spectrum interpretation, cross-database comparison, assessment of reproducibility across samples or datasets, evaluation using alternative computational tools, and orthogonal structural characterization following purification and analysis by chemical or enzymatic methods where applicable. Particular attention should be paid to biosynthetic plausibility, and the manuscript should clearly explain the logic and confidence criteria used for isomer discrimination. Without such supporting evidence, novel glycan assignments risk misinterpretation or overstatement.

At present, the manuscript falls short of the standards required for publication. Substantial revisions and validation are needed before it can be considered further.

Response: We sincerely thank the reviewer for the thorough and professional evaluation of our manuscript. As correctly pointed out, there are important concerns regarding the validation of glycan structural annotations, which we agree are critical for the reliability of our study. In response to these constructive suggestions, we have carefully revised the manuscript and implemented substantial improvements. Detailed point-by-point responses and corresponding corrections are provided below.

Major comments

Q1. Glycan classification is inconsistent with biosynthetic pathways.

*N2H4F1 is categorized as high-mannose, but typically, fucose does not occur in high-mannose glycans. This composition more likely represents a hybrid structure with core fucosylation.

*Similarly, structures like N2H9F1 and N2H8F1, which include fucose within high-mannose-type glycans, have also been identified; however, from a biosynthetic perspective, such structures are unlikely to be produced. Potential misassignments in MS/MS spectral interpretations were found, please see Q3 for details.

*Structures like N3H8 also frequently appear in the identification table. However, biosynthetically, glycans with three HexNAc residues should no longer be considered high-mannose-type; the pathway toward hybrid or complex types requires mannose trimming down to N2H5 before HexNAc addition, making N3H8 unlikely to form. Potential misassignments in MS/MS spectral interpretations were found, please see Q3 for details.

Response: Thank you for this insightful comment. We agree that certain glycan compositions such as N2H4F1, N2H8F1, and N2H9F1, though classified by software as high-mannose-type structures, are not consistent with known biosynthetic pathways due to the presence of core fucose, which is typically added during later stages of glycan maturation. Similarly, the structure N3H8, which includes three HexNAc residues, is unlikely to be a true high-mannose glycan, as its biosynthesis would require prior mannose trimming.

StrucGP primarily adopts a modular analysis strategy that focuses on key structural features—such as core structures, branch structures, and glycan types—and combines these modules based on accurate molecular weight. This design enhances structural interpretability but in rare cases may allow matches to glycan compositions that deviate from canonical biosynthetic pathways. Such mismatches can arise from special glycan modifications, fucose migration within a given glycopeptide, or co-occurrence of *N*- and *O*-glycans on the same peptide.

To assess the validity of these assignments, we manually examined representative MS/MS spectra, including that of N2H8F1, and found that the diagnostic B- and Y-ions supporting these structures were present, indicating that the software annotations are technically plausible **(Fig. R1)**. However, we suspect that some of these cases may arise from co-elution artifacts. For N3H8 structure, it may result from the combination of *N*-glycans and *O*-glycans on the same peptide, such as N2H6 + NH or N2H7 + N **(Fig. R2)**. We have included representative spectra and annotated key fragment ions below to support this interpretation. Notably, these potentially misassigned structures account for only a small fraction of the identified high-mannose glycopeptides: N2H4F1 (0.49%), N2H8F1 (0.14%), N2H9F1 (0.03%), and N3H8 (0.09%). Given their very limited abundance, these exceptions are unlikely to substantially affect the overall conclusions of our study.

Given that updating the software requires substantial engineering effort and that manual validation of all questionable cases is not feasible, we have instead focused on discussing the current limitations of the existing software version. In the revised manuscript, we have added a note to the *Discussion* section to acknowledge this limitation of software-based glycan assignment and to emphasize the importance of careful interpretation, especially for low-abundance or biosynthetically ambiguous structures (**Page 20**).

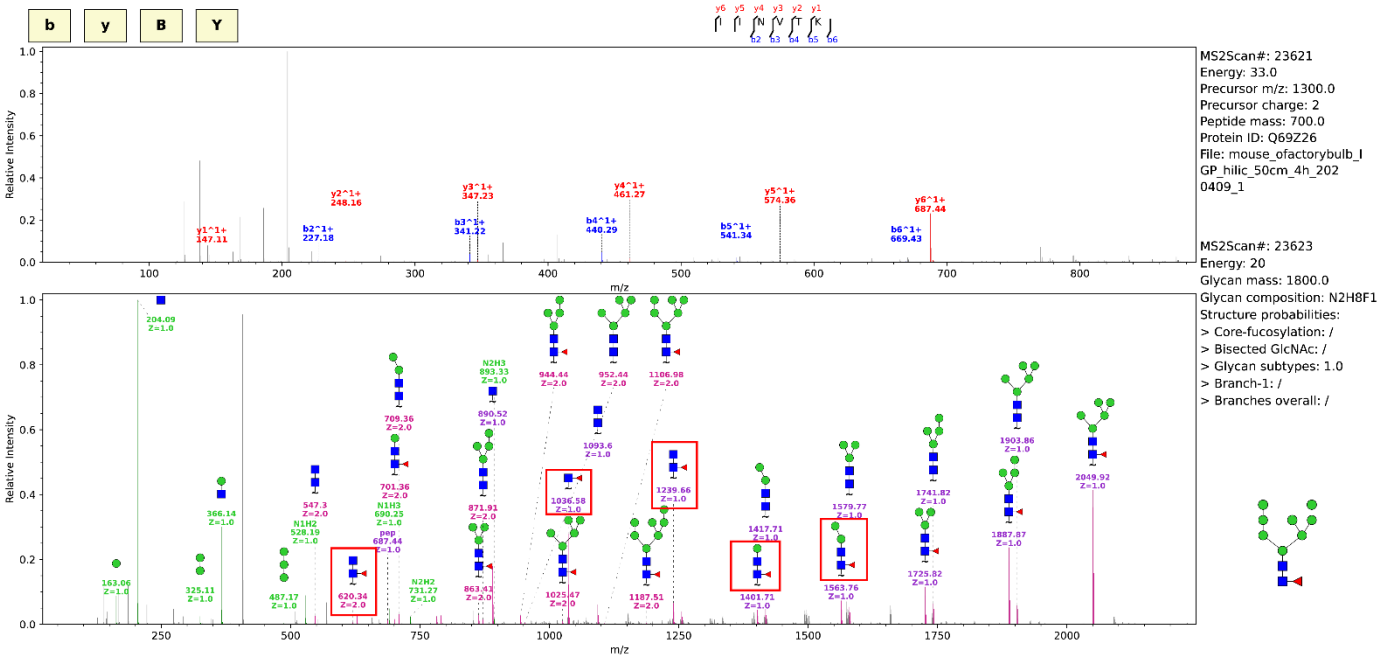

Fig. R1 Manual interpretation of N2H8F1 spectrum.

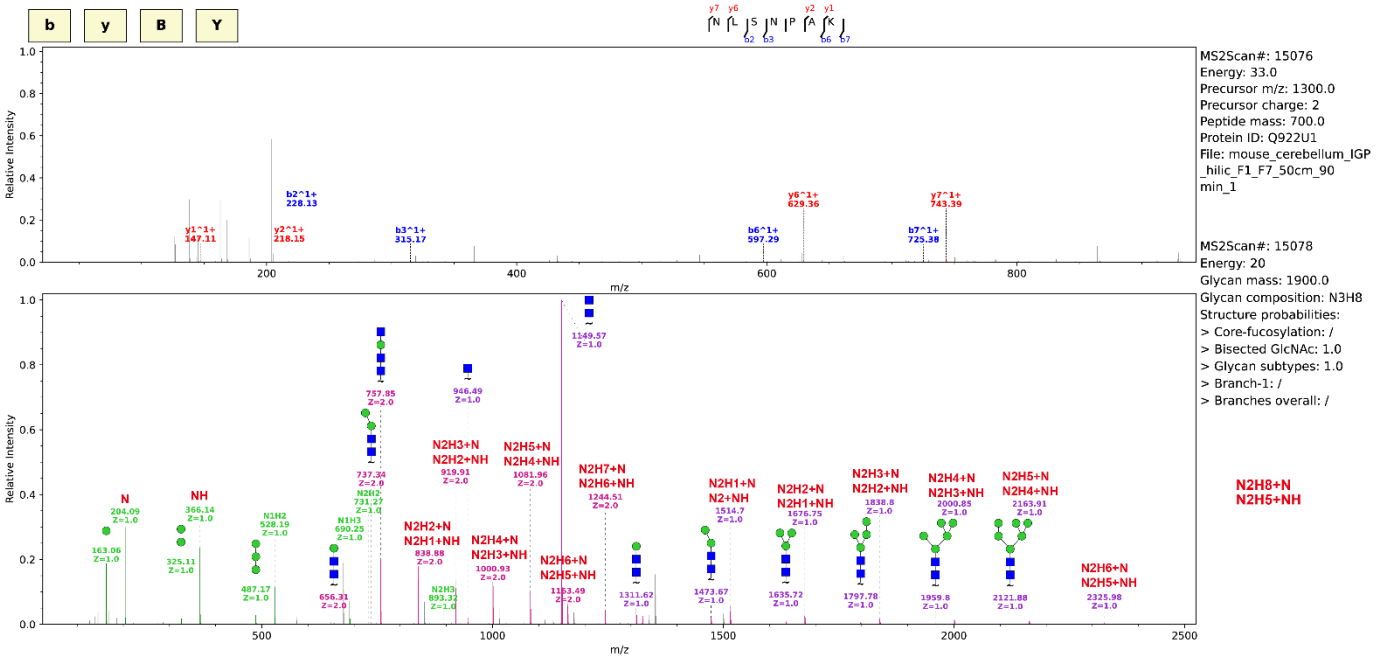

Fig. R2 Manual interpretation of N3H8 spectrum.

Q2. Potential misassignments in MS/MS spectral interpretations

*Spectra #25304 and #25306 (sample: mouse_pancreas_IGP_hilic_50cm_4h_20200122_1, peptide: FANNTSVEPQK+N2H9F1)

Although the core peptide sequence FANNTSVEPQK appears to be correctly assigned at first glance, the Y-ion signals in these spectra only extend up to the fragment corresponding to the peptide with N3H6F1 attached (Pep+N3H6F1). There are no fragment ions supporting the presence of the larger assigned glycan (N2H9F1). Moreover, this peptide contains two potential *N*-glycosylation sites. Additionally, the spectra show a signal at m/z 284.044, which corresponds to a HexNAc residue modified with a sulfate group. This observation suggests that, instead of the originally assigned glycan, an alternative composition—such as N3H2F1(Sul)1 + N2H3—may provide a better explanation for the data.

*Spectra #11204 and #11206 (sample: mouse_female_kidney_mix_IGP_hilic_4h_50cm_1, peptide: FM[Oxidation]KELSDDSFENSNK+N3H8)

In spectra #11204 and #11206, the peptide was assigned as FM[Oxidation]KELSDDSFENSNK with an N3H8 glycan. However, there is only a single b-ion supporting this peptide, which is insufficient evidence to confidently confirm the identification. Meanwhile, the spectra contain numerous b-ions matching a different peptide, SHTNTSHVMQYGNK. This suggests that the original peptide assignment is likely incorrect. It seems more reasonable that the spectra actually come from SHTNTSHVMQYGNK with an N4H8, which has two possible sites for glycan attachment.

Response: We thank the reviewer for the detailed inspection of specific spectra and the insightful suggestions. We have carefully re-examined the cited MS/MS spectra and agree that they raise valid concerns about potential misassignments in certain cases.

For spectra #25304 and #25306 (peptide: FANNTSVEPQK + N2H9F1), our manual re-annotation suggests that the most plausible glycan composition is N4H6F1(Sul)1, rather than the originally assigned N2H9F1 **(Fig. R3)**. The annotated MS/MS spectra and corresponding diagnostic fragment ions supporting this assignment have been provided below. Due to the extremely low reported frequency of sulfation on *N*-glycans in mammals, we intentionally did not include chemically modified glycans (e.g., sulfation or phosphorylation) in our glycan branch library. This decision was made to reduce the search space and minimize the false discovery rate (FDR) caused by introducing a large number of low-probability modifications. As a result, rare structures containing such modifications could have been misassigned to unmodified but mass-similar glycan compositions.

For spectra #11204 and #11206 (peptide: FM[Oxidation]KELSDDSFENSNK + N3H8), we confirmed

that the assigned peptide, FM[Oxidation]KELSDDSFENSNK, with an N3H8 glycan, is supported by only a limited number of b/y ions—specifically, 4 b-ions and 4 y-ions (8 in total). This level of fragmentation is indeed insufficient to confidently validate the peptide backbone. To systematically evaluate the quality of peptide backbone fragmentation in our glycopeptide identifications, we analyzed the distribution of matched b/y ions across all glycopeptides identified by StrucGP in the 24 mouse tissues. As shown in **Fig. R4**, most glycopeptides were supported by more than 10 b/y ions, while only 6.2% had less than 7 b/y ions, reflecting the overall high confidence of peptide assignment. The distribution dropped sharply beyond 20 matched ions but exhibited a long tail extending to higher values, reflecting the presence of glycopeptides with rich fragmentation information across the dataset. This overall pattern demonstrates the robustness of StrucGP in assigning glycopeptides with sufficient fragment evidence.

Regarding the glycan structure of N3H8, as discussed in our response to Comment Q1, we also considered the possibility of glycan misassignment due to the co-occurrence of *N*-glycan and *O*-glycan, such as N2H6 + NH or N2H7 + N. Manual interpretation of the spectra supports this scenario **(Fig. R5)**. These examples reflect important cases where the current glycopeptide identification strategy—while broadly effective—can be limited by the exclusion of chemical modifications and the presence of *O*-glycans.

In the revised manuscript, we have added relevant content to the *Discussion* section (**Page 20**) to acknowledge these limitations and to illustrate how certain modifications (e.g., sulfation) or the co-occurrence of *O*-glycans may lead to incorrect glycopeptide assignments when such features are not included in the search space. We also note that the glycan database and spectral resource provided in this study may serve as a valuable foundation for future algorithm development aimed at identifying chemically modified glycans and distinguishing between *N*- and *O*-glycan combinations.

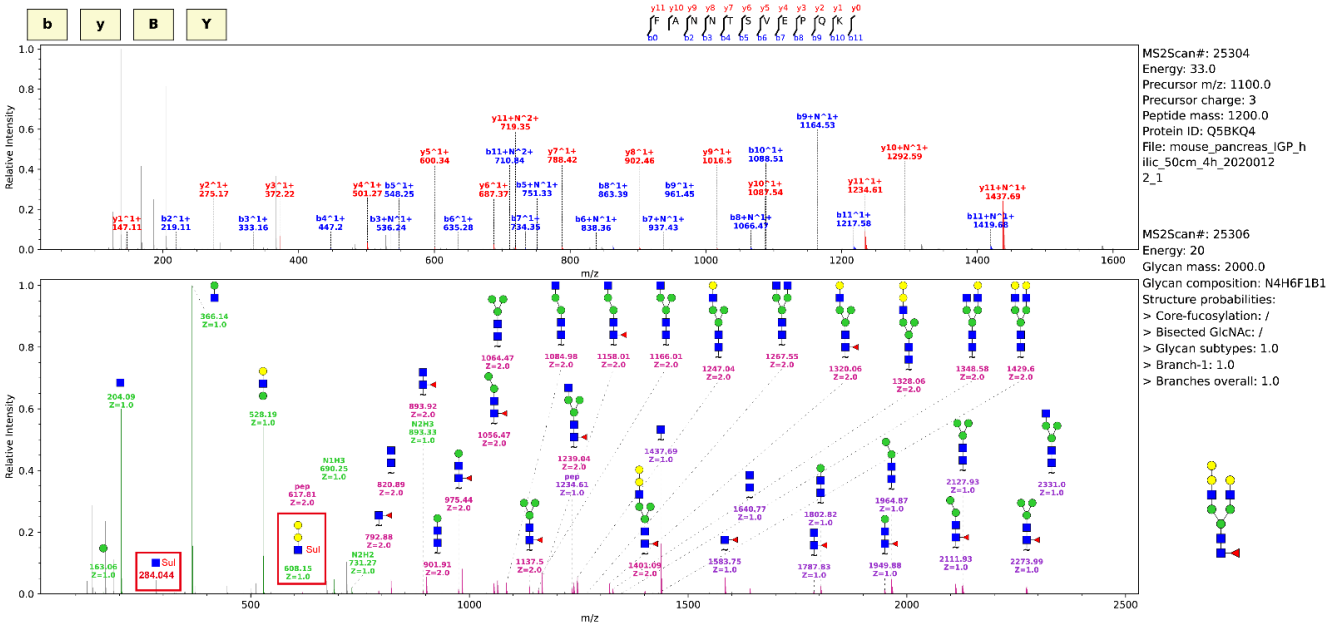

Fig. R3 Manual interpretation of N4H6F1(Sul)1 spectrum.

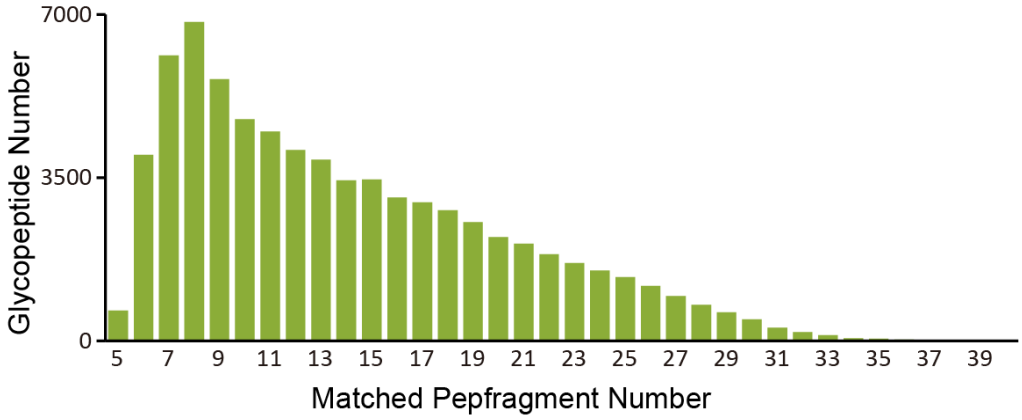

Fig. R4 Distribution of matched b/y ions per glycopeptide.

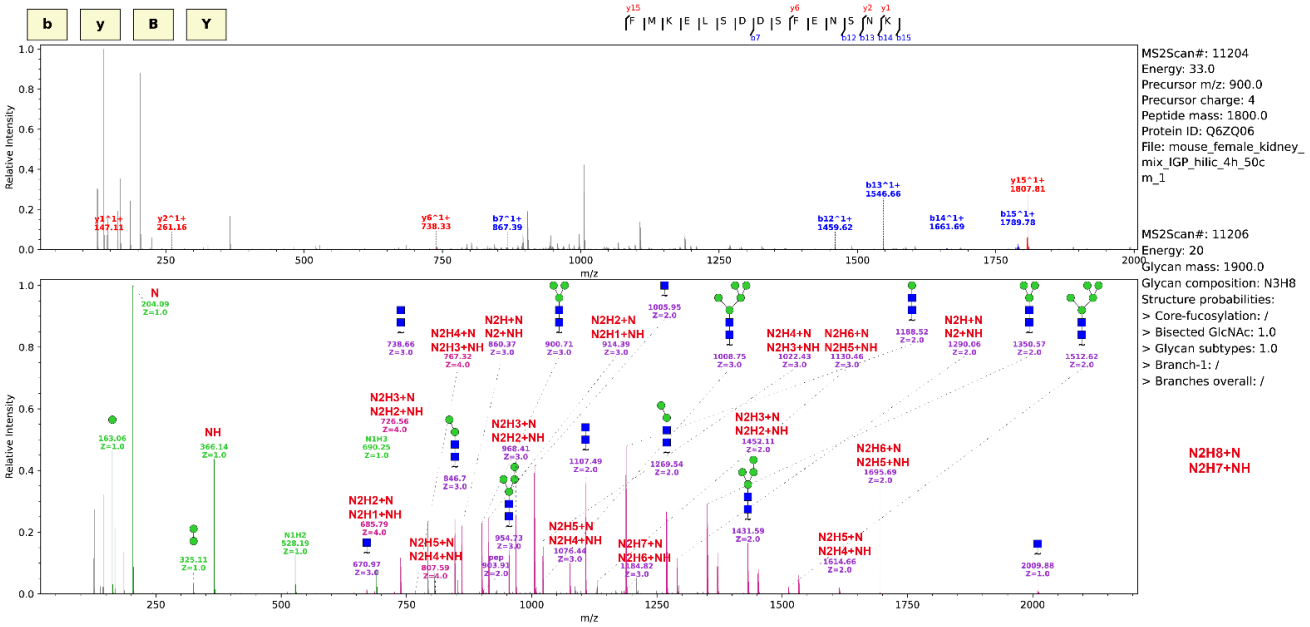

Fig. R5 Manual interpretation of N3H8 spectrum.

Q3. In the authors' 2021 publication describing the development of the StrucGP software, they reported identifications from both standard glycoproteins and approximately 600 distinct glycans detected in mouse brain samples. In contrast, the current manuscript reports around 3,000 glycans identified across multiple mouse tissues. Given this substantial increase in the number of identified glycans and their structural complexity, I would like the authors to clarify in detail how they ensured the accuracy of glycan assignments. Specifically, how do the authors address glycan isomers that occur on the same glycosylation site and peptide? In Figure S4B, the N5H5F1S1 isomers are shown to be present on different proteins, and the observed

retention time differences are likely attributable primarily to differences in the peptide sequences. Therefore, this data alone does not sufficiently support the claim that retention time is effective for separating glycan structural isomers.

Response: We thank the reviewer for this important question. The increase in the number of glycan structures identified in the current study—from ~600 in our previous 2021 publication to ~3,000—is primarily attributable to improvements in sample preparation and data acquisition, rather than changes in the core StrucGP algorithm. Specifically, (1) the current dataset encompasses 24 different mouse tissues, whereas the earlier study focused solely on the brain. As different tissues exhibit distinct glycosylation profiles, expanding tissue coverage naturally increases glycan diversity. (2) Most of the tissue samples in the current study underwent peptide fractionation prior to enrichment and LC-MS/MS analysis, whereas the original brain dataset did not. To evaluate the impact of this step, we directly compared identification results from the same tissue with and without fractionation, and observed a substantial increase in glycopeptide and glycan identifications when fractionation was applied (**Supplementary Figure 1F**). (3) Many of the tissues analyzed in this study contain glycopeptides with Neu5Gc, which were not present in the original brain dataset. These combined factors contributed significantly to the expanded glycan identification in this study.

Regarding the identification of glycan isomers, we acknowledge the reviewer's concern that the examples shown in Figure S4B involve isomers assigned to different peptides, where observed retention time (RT) differences are largely attributable to differences in peptide sequence. Indeed, most glycan isomers were identified either on different peptide backbones or in different tissues. For those isomers assigned to the same peptide, their identification was primarily enabled by chromatographic separation, as distinct retention times were observed. These cases—such as sialylated isomers like N4H5G2—are supported by both MS/MS fragmentation patterns and RT differences, and are now added in the revised **Supplementary Figure 4**. Additionally, the underlying algorithm used for isomer identification follows the same general framework described in our original 2021 publication. Although minor adjustments have been made in scoring and filtering, the core structure of the algorithm remains consistent. We agree that current approaches still face challenges in accurately distinguishing glycan isomers, and we appreciate the reviewer's comment, which provides valuable direction for further optimizing the software.

In the revised manuscript, we have added new example supporting the claim that retention time is effective for separating glycan structural isomers derived from the same peptide sequence, and incorporated the relevant description into the *Result* section (**Page 6**).

Q4. For *N*-glycans with more than four antennas, how do the authors determine whether the structure is a bisecting or non-bisecting isomer? In addition, how the authors determine the number of branches in the glycan backbone? For example, in Supplementary Figure S5, the authors propose a hexa-antennary structure instead of a bisecting penta-antennary.

Please provide paired MS/MS spectra of isomeric structures—one with bisecting GlcNAc and one without—to illustrate the diagnostic fragment ions and explain the logic of how you distinguished these two possibilities.

Response: We thank the reviewer for raising important points regarding the identification of bisecting versus non-bisecting *N*-glycans and the determination of glycan antenna number. The assignment of bisecting GlcNAc structures in our study was based on previously published algorithms (*Anal Chem*. 2019 May 7;91(9):5478-5482.), which primarily rely on one or two characteristic Y ions - peptide + $HexNAc_3Hex_1$ and, if core fucose is present, peptide + $HexNAc_3Hex_1Fuc_1$ - to distinguish bisected from non-bisected structures. As shown in the paired MS/MS spectra we now provide (**Fig. R6** and **Fig. R7**), when two isomeric glycan structures with identical precursor masses are co-detected, only those spectra containing the specific diagnostic ions (m/z 744.3 and m/z 920.4) are annotated as bisected glycans. This approach enables broad and efficient profiling of bisecting structures across tissues. However, we acknowledge that this method may result in a certain level of false positives, particularly in cases of highly branched glycans where diagnostic ion patterns may overlap. To address this, we plan to improve accuracy by incorporating additional diagnostic ions, leveraging machine learning-based spectral interpretation, and validating assignments using bisecting-deficient or -overexpressing cell models in the future. Importantly, the number of bisected structures identified varies across tissues, with notably higher expression in brain and kidney, which is consistent with previous studies and further supports the reliability of most glycan structures identifications.

Regarding antenna number determination, we first enumerate all plausible branch structures based on the fragment ions, and then infer the most likely antenna number by combinatorically assembling these branches. However, for glycans annotated as penta- or hexa-antennary, there remains a possibility of misidentification due to poly-LacNAc extensions. In many of these cases, diagnostic B-ions of poly-LacNAc were not observed, likely due to over-fragmentation under the current collision energy (HCD 20%), which we have previously shown to compromise the detection of poly-LacNAc-specific ions. This limitation is also discussed in our response to Q20.

In the revised manuscript, we have cited the related literature in the *Result* section (**Page 8**) to better contextualize our findings and highlight their consistency with established tissue-specific glycosylation patterns, and we have also added a discussion on poly-LacNAc-related misidentification in the *Discussion*

section (**Page 21**) to address potential overestimation of antenna numbers.

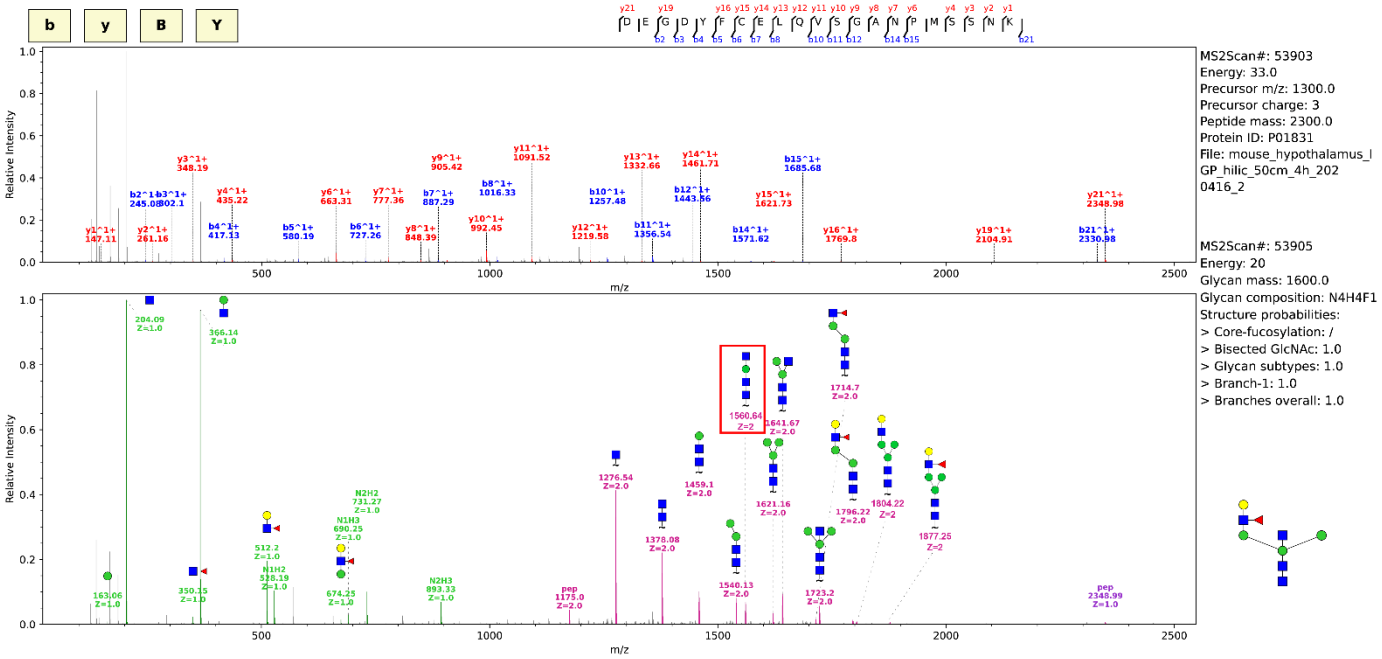

Fig. R6 Representative MS/MS spectrum of a bisected *N*-glycan.

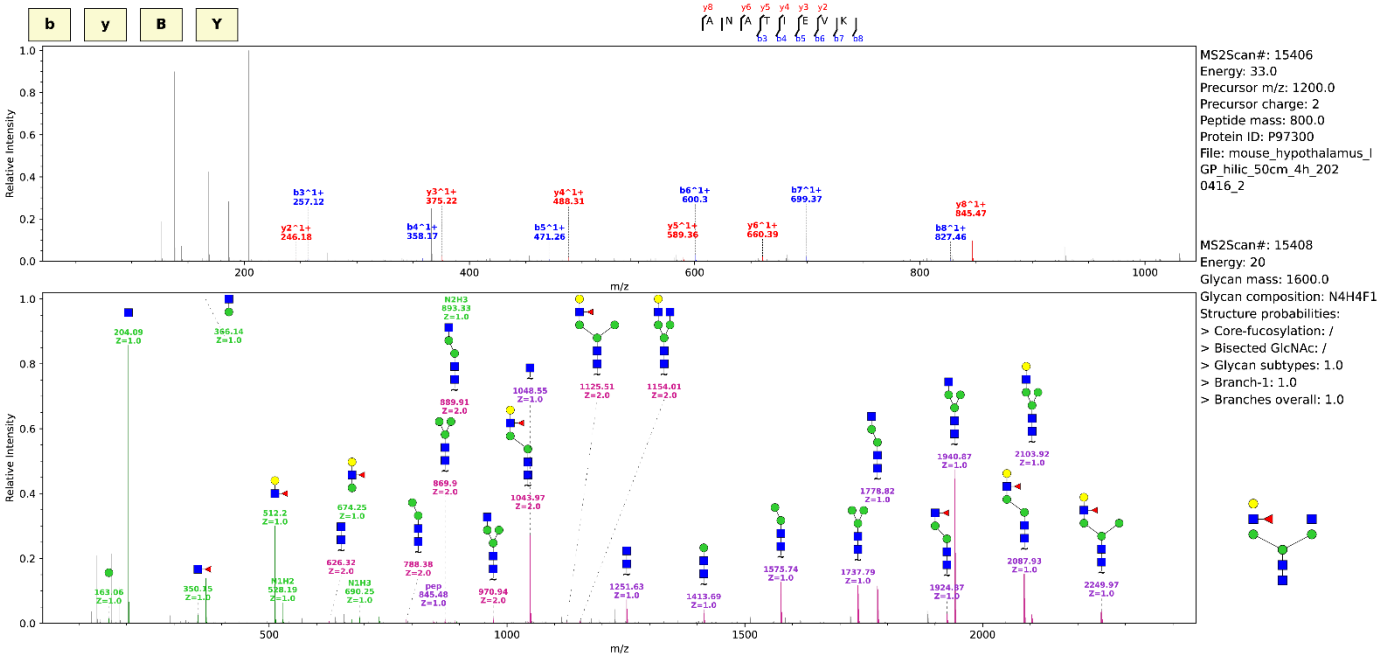

Fig. R7 MS/MS spectrum of the corresponding isomeric structure lacking the bisecting GlcNAc.

Minor comments

Q5. line 102, page 5

"This disparity is likely arising from the absence of fractionation in the gallbladder samples, highlighting the critical role of sample fractionation in enhancing the depth of analysis for complex tissues."

If the authors wish to demonstrate the effect of fractionation, comparative data from the same tissue with and

without fractionation should be included.

Response: We appreciate this insightful comment. To better illustrate the effect of fractionation, we have now reanalyzed the dataset by comparing, for each tissue (excluding ovary and gallbladder), the number of glycopeptides, glycoproteins, glycan structures and glycosites identified in unfractionated samples versus those after high-pH reversed-phase fractionation into six fractions following HILIC enrichment.

The results clearly show that sample fractionation substantially enhances analytical depth for complex tissues. On average, glycopeptide identifications increased by 1.2 - 45.6 fold, glycoprotein identifications by 1.3 - 13 fold, glycan structure identifications by 1.2 - 18.2 fold, and glycosite identifications by 1.2 - 16.8 fold in fractionated samples compared to their unfractionated counterparts. Notably, the degree of improvement varied across tissues, which can be attributed to both the intrinsic complexity of tissue glycoproteomes and differences in sample input. Overall, these findings underscore the critical role of fractionation in achieving comprehensive glycoproteomic coverage.

In the revised manuscript, we have added a new supplementary figure (**Supplementary Figure 1E**) to present these results with additional description in the revised *Results* section (**Page 5**).

Q6. Line 106, page 5

While the general statement that "the number of identifications can vary between tissues even under the same MS conditions" is relevant, the referenced study by Giansanti (Ref. 27) does not directly provide data on glycoproteins or glycopeptides. Therefore, the authors should consider citing a more appropriate reference.

Response: We thank the reviewer for pointing this out. We have now replaced Ref. 27 with two more appropriate citations that directly support the variability of glycoprotein/glycopeptide identifications across tissues under comparable mass spectrometry (MS) settings.

First, **Okumura et al.** (*Anal Bioanal Chem.* (2025) Feb;417(5):973-988.) applied a site-specific glycoform profiling approach (Glyco-RIDGE) to six mouse tissues, identifying 11,325 site-specific glycoforms from 800 glycoproteins. Their results revealed extensive glycan heterogeneity across tissues, even when processed with identical MS conditions and analysis pipelines. Second, **Liu et al.** (*Nat Commun*. (2017) Sep 5;8(1):438.), in their development of the pGlyco 2.0 algorithm, reported comparative glycopeptide identification across five mouse tissues. They found that only ~1% of intact glycopeptides were shared among tissues, despite using consistent sample preparation and MS parameters. This underscores the strong tissue-specificity of the glycoproteome.

In the revised manuscript, we have replaced the previous reference with these two citations at **Page 5**.

Q7. Line 116, page 6

"likely due to more extensive mass spectrometry coverage"

The authors mainly attribute the high number of identified glycosylation sites in the kidney to differences in MS coverage. However, several publications have demonstrated that the kidney is inherently rich in glycoproteins, and this biological background should also be considered. Moreover, if differences in the number of MS runs or fractionation conditions significantly affect the number of identifications, then comparable MS conditions (e.g., injection amount, number of runs, number of fractions) should be applied when comparing different organs. This would require a reconsideration of the overall experimental design of this study.

Response: We appreciate the reviewer's thoughtful comment. We agree that the high number of glycosites in the kidney likely reflects not only the relatively extensive mass spectrometry analysis but also the intrinsic glycoprotein richness of this organ.

In the revised manuscript, we now cite relevant literature supporting this view (*Front Mol Biosci.* (2022) Aug 17:9:976298.) and have added corresponding clarification in the *Results* section (**Page 6**). Additionally, we have acknowledged in the *Discussion* section that variability in experimental design across tissues—including differences in fractionation depth and LC-MS/MS coverage—is a limitation of the current study (**Page 20**).

Q8. Line 180, page 8 (Table S4)

Please provide the selection criteria for the 209 glycosyltransferases (GTs). For example, ST6GALNAC1 shows a maximum value of only 0.05 across all tissues. If, as in Ref.4, the data are expressed in TPM (log10), this gene would not display clear differential expression across tissues. If the total number of GTs included in the dataset happens to be 209, this should also be clearly stated.

Response: We thank the reviewer for the constructive comment. In the study, we initially retrieved 728 mouse glycosyltransferase (GT) entries from the UniProt database. After removing entries with redundant gene names, we obtained a non-redundant list of 384 GT genes. Among these, 209 GTs were successfully matched with the transcriptomic dataset reported in Ref. 4. The gene expression levels shown in **Figure 2F** represent these 209 GTs. As RNA and protein expression levels may sometimes show weak or even negative correlation, our use of RNA-seq data here is not intended to infer functional glycosylation activity directly, but rather to demonstrate that GTs as a group also display tissue-specific expression patterns at the transcript level, as described in the *Results* section of the original manuscript.

In the revised manuscript, we have now added a detailed description of the GT selection and filtering process in the *Methods* section (**Page 26**), and included corresponding content in the *Results* section to clarify this analysis (**Page 8**).

Q9. Line 190, page 8

It should be clearly stated that the annotations of subcellular localization are based on predicted cellular localization (e.g., GO terms) rather than experimental validation. Additionally, it would be advisable to note somewhere in the manuscript that predictions of localization based on GO annotations have inherent limitations.

Response: We thank the reviewer for the helpful comment. The subcellular localization analysis in this study was performed using bioinformatics-based annotation, rather than experimental validation. Specifically, we assigned subcellular localization based on established protein databases and computational prediction tools, which provide expected protein locations rather than experimentally confirmed compartmentalization.

In the revised manuscript, we have specified that the subcellular localization annotation was based on the UniProtKB database (**Page 9**).

Q10. Line 182, page 8 (Fig. 2A, F–H)

The authors suggest that high expression of Mgat3 in brain tissues explains the elevated levels of bisected *N*-glycans (Core III/IV). However, in Figure 2A, bisected structures also appear abundant in intestinal tissues, where Mgat3 expression is not particularly high. This raises the question of whether mRNA expression alone can sufficiently explain the observed glycan patterns across tissues. Are there additional lines of evidence—such as lectin array results, that support the proposed correlation? Moreover, is there any broader correlation with other glycosyltransferases beyond Mgat3 and Fut9?

Response: We thank the reviewer for the insightful comments. We would like to clarify that our use of RNA-seq data is not intended to directly infer glycosylation activity, but rather to demonstrate that glycosyltransferases (GTs) as a group also exhibit tissue-specific expression patterns at the transcript level. Given that RNA and protein expression levels can sometimes show weak or even negative correlation, we acknowledge the limitations of transcriptomic data in predicting enzymatic activity.

To support the presence of bisected *N*-glycans in intestinal tissues, we have reviewed the literature and found reports confirming that bisected structures are indeed relatively abundant in the intestine (*Nat Commun*. 2024 Nov 9;15(1):9725.). The finding is consistent with our MS-based results.

In the revised manuscript, we have added this reference to the *Results* section (**Page 8**) and included a corresponding explanation to highlight the consistency between previously reported intestinal bisected *N*-glycans and our mass spectrometry findings.

Q11. Line 230, page 10

The order of glycan structures differs between Figure 3B and Figure S6, making it visually difficult to confirm reproducibility. Unifying the order of glycans or providing a quantitative measure of similarity between them could improve the clarity of interpretation.

Moreover, the authors emphasized that glycosylation patterns are strongly influenced by subcellular localization. Therefore, it is important to verify whether the 24 glycoproteins compared across tissues are biased toward specific cellular compartments (e.g., the endoplasmic reticulum). It is recommended that the authors indicate the subcellular localization of these proteins and discuss whether any bias might affect the observed reproducibility.

Response: We appreciate the reviewer's insightful comment regarding the glycan order differences between Figure 4B and Figure S7 (originally labeled as Figure S6 in the previous version). Indeed, both heatmaps display z-score normalized abundances to facilitate cross-tissue comparison. However, the glycan substructure sets visualized in these two panels are not identical due to differences in the underlying glycoprotein groups: Figure 4B focuses on glycopeptides from glycoproteins commonly identified in at least 12 tissues, while Figure S7 extends this to a more stringently defined set of 24 universally expressed glycoproteins. As a result, the specific glycan substructures and their frequencies naturally differ between the two analyses, leading to distinct hierarchical clustering and ordering in the heatmaps. Nevertheless, the key conclusion remains consistent across both figures: despite the conservation of protein backbones, the attached *N*-glycan structures exhibit marked tissue-specific variation. This finding highlights extensive remodeling of glycosylation in a tissue-dependent manner, and is reproducibly observed regardless of the selection criteria for conserved glycoproteins.

In the revised manuscript, we have added corresponding note to explain the observed ordering differences (**Page 10**).

Q12. Line 242, page 10

The authors refer to "functional demands," but the manuscript does not present direct evidence or analyses clarifying what specific functional requirements drive organ-specific glycosylation. Observing correlations

alone is insufficient to support causality. Therefore, if the term "functional demands" is used, it would be helpful to include supporting data or discussion—such as references to the functional roles of glycan structures in specific localizations or analyses of associations with functionally relevant genes. Otherwise, it is suggested carefully revising expressions that imply causation.

Response: We thank the reviewer for this insightful comment. We agree that our previous use of the term "functional demands" may have implied causality without sufficient direct evidence.

In the revised manuscript, we have revised the relevant sentences to avoid causal language and to present the associations more cautiously (**Page 11**).

Q13. Line 248, page 10

The authors discuss the biological significance of organ-specific glycan remodeling by using the example of glycosylation at Asn-84 of M6pr. However, it is unclear whether differences in protein abundance of M6pr across tissues were considered. Since detection frequency and structural diversity of glycan modifications can depend on protein expression levels, it is recommend providing supporting data (e.g., proteomics or RNA-seq data) to examine whether expression differences might influence glycosylation patterns. This would help clarify whether the differences in glycan structures reflect true remodeling rather than biases from protein abundance.

Response: We thank the reviewer for this insightful comment. We fully agree that protein expression levels can influence both the detection frequency and the apparent structural diversity of glycan modifications. Unfortunately, matched transcriptomic or proteomic data for our glycoproteomic samples are not available, and we acknowledge this as a limitation of the current study. Importantly, this is not an isolated case: we observed similar tissue-specific glycan remodeling patterns across 24 glycoproteins (49 glycosylation sites) that were commonly expressed in all 24 mouse tissues analyzed. For each site, glycan structural compositions differed markedly across tissues, suggesting that local glycan remodeling, rather than detection bias due to protein abundance, underlies the observed heterogeneity. To support this conclusion, we have added a glycan similarity score table for these shared proteins as **Supplementary Table 7**. These findings further strengthen the notion that tissue-specific regulation of glycosylation occurs even when protein expression is broadly consistent.

In the revised manuscript, we have also included a statement in the *Discussion* (**Page 19**) acknowledging that differential protein expression may still contribute to glycosylation heterogeneity.

Q14. Line 257, page 11

In Figure 4E, the majority of glycans in the seminal vesicle show zero sialylation. Therefore, the corresponding explanation in the manuscript should be reviewed for consistency.

Response: We thank the reviewer for pointing out this inconsistency in our manuscript. As shown in Figure 4E, the liver, as a metabolic organ, features abundant sialylation, which may help protect modified proteins from rapid degradation and thereby extend their half-life. In contrast, sialic acid is largely absent in the seminal vesicle.

In the revised manuscript, we have corrected the corresponding description to ensure consistency between the text and the figure (**Page 12**).

Q15. Line 290, page 12 (paragraph 6):

Regarding the expression "prefer to perform specialized functions," although GO analysis results are presented to suggest functional roles of glycosylation, no causal relationship between glycan structures and functional outcomes is demonstrated in the manuscript. The current evidence only supports correlations. It is recommend modifying "prefer to perform specialized functions" to a more cautious expression such as "are suggested to have functional relevance."

Furthermore, to strengthen the validity of the claimed link between glycan structures and function, it would be beneficial to include references or experimental data supporting the functional importance of specific structural features, such as Lewis antigens or LacdiNAc modifications, in the relevant tissues.

Response: We thank the reviewer for the insightful comment. We agree that the original phrase may overstate the evidence and have revised “prefer to perform specialized functions” to “may perform specialized functions” to better reflect the correlative nature of our findings.

We also agree that citing relevant literature is important for highlighting the potential functional relevance of glycan structures in specific tissues. Accordingly, we have already included such references (**Ref. 48-52**) in the *Discussion* section of the original manuscript (**Page 19**).

Q16. Discussion

Throughout the discussion, there are several statements suggesting causal relationships between glycan structures and their functions. Since the study does not directly test functional outcomes, it would be more appropriate to phrase these descriptions in terms of associations rather than causation.

Response: We thank the reviewer for this important observation. We agree that our study does not directly

establish causality between glycan structures and functional outcomes.

In the revised manuscript, we have carefully revised the *Discussion* to rephrase relevant statements in terms of associations rather than causal relationships, ensuring a more accurate interpretation of our data (**Page 18-20**).

Q17. Line 464, page 18

The statement that glycosylation "adapts to extracellular functional demands" is not directly supported by the data presented in this study. Rather, the observed glycan patterns are more likely due to structural constraints or the localization of biosynthetic enzymes. Therefore, this expression should be revised.

Please note that while Ref.56 is relevant for the first part of the sentence about biosynthetic pathways, it does not provide any information to support the second part of the sentence regarding glycosylation adapting to functional demands.

Response: We thank the reviewer for the helpful comment. In the revised manuscript, we have replaced the original phrasing with a more cautious description to avoid implying a direct causal relationship (**Page 19**). Additionally, Ref. 56 (now updated to Ref. 57 due to the addition of new references) has been moved to a more appropriate position where it specifically supports the discussion of biosynthetic pathways.

Q18. Line 486, page 19

The statement on the co-occurrence of polySia and Lewisx/a structures is interesting. However, without experimental evidence of spatial proximity (e.g., whether these modifications occur on the same molecule) or functional interactions, it seems premature to conclude "functional synergy."

Response: We thank the reviewer for the insightful comment. In the revised manuscript, we have modified the wording to avoid implying functional synergy without direct supporting evidence (**Page 20**).

Q19. In methodology, what's the reason for further High-pH HPLC fractionation after HILIC column enrichment?

Response: We thank the reviewer for the question. After HILIC enrichment, further high-pH reversed-phase HPLC fractionation was performed to reduce sample complexity and improve MS/MS identification depth. While HILIC enriches glycopeptides based on hydrophilicity, it does not provide sufficient separation for complex mixtures. The additional high-pH fractionation, which separates peptides based on hydrophobicity under orthogonal conditions, enhances the resolution of glycopeptide species and allows for more

comprehensive and confident identification in downstream LC-MS/MS analysis. Compared with the glycopeptides identified by Max column from unfractionated samples, those enriched by HILIC exhibited higher complexity, thus we prioritized fractionation of HILIC-enriched samples to maximize identification coverage. High-pH HPLC is a commonly used sample preparation method; however, we acknowledge that more advanced approaches are emerging. For example, a recent study (*Nat Commun*. 2025 Jul 1;16(1):5568) employed a combination of Sepharose CL-4B and ZIC-HILIC enrichment for mouse brain glycoproteomics and achieved excellent identification depth. We plan to further optimize our workflow in future studies.

In the revised manuscript, we have added clarifications in the *Methods* section to highlight the role of high-pH fractionation in improving identification depth (**Page 23**).

Q20. Do the authors find any *N*-glycans with polyLacNAc extensions? If not, what could be the reason?

Response: We thank the reviewer for this important question. In our dataset, we rarely observed characteristic fragment ions indicative of polyLacNAc extensions in the MS/MS spectra acquired at HCD energy of 20%. Even when such ions were present, their intensities were generally very low, which likely contributed to the lack of confident identifications by the software.

Based on our experience, polyLacNAc-containing *N*-glycans require lower collision energies for optimal fragmentation. HCD 20% is relatively high for these structures and often leads to excessive fragmentation, resulting in the loss of key diagnostic ions. As mentioned in our previous responses, we suspect that some of the penta-antennary or hexa-antennary structure may contain polyLacNAc extensions, but further experimental validation and improvements in identification algorithms are needed to confirm this. The identification of polyLacNAc structures remains an active area of research in our laboratory.

In the revised manuscript, we have added a discussion on poly-LacNAc-related misidentification in the *Discussion* section (**Page 21**).

Q21. In Figure 2E, what does the fifth glycan from the right represent? (Does it mean non-branching?)

Response: We thank the reviewer for the question. The fifteenth branch structure "Ø" in Figure 2E represents a lack of one branch structure. This explanation was included in the original manuscript (**Page 8**) and has now been further clarified in the Figure 2E legend for better clarity (**Page 39**).

Q22. In Figure 3, It would be helpful if the figure legend were more detailed. It's hard to know which tissue was used for this analysis.

Response: We thank the reviewer for the helpful suggestion. Figure 3 is based on all glycoproteins identified across the 24 mouse tissues analyzed in this study. We have revised the figure legend to clarify the tissue source used for the analysis (**Page 40**).

Q23. Figure 3D

In Figure 2A, bisecting glycans were presented as a distinguishing feature of nervous system tissues. However, in Figure 3D, bisected structures do not appear particularly enriched in the plasma membrane fraction. This discrepancy may warrant further explanation

Response: We thank the reviewer for this insightful observation. We would like to clarify that the glycan statistics shown in Figure 2A are based on the top ten glycans with the highest PSM counts identified specifically in nervous system tissues, thereby representing the dominant glycan features within those tissues. In contrast, Figure 3D focuses solely on glycan substructures localized to the plasma membrane and endoplasmic reticulum fractions, which represent only specific subcellular compartments rather than the overall glycan profile of a tissue. Therefore, the differences between these figures arise from the distinct data subsets and contexts analyzed.

Q24. Figure S1A

It is difficult to understand how the figure corresponds to the main text. Please consider revising the figure or adding clarifications in the caption.

Response: We appreciate the reviewer's feedback. In response, we have revised **Figure S1A** to better reflect and correspond with the descriptions provided in the main text.

Q25. Table S3

It is unclear what the numbers in this table represent. The authors appear to indicate the percentage of each subtype or structure in each organ, but please clarify this point explicitly.

Response: We thank the reviewer for this important point. The percentages in the table were calculated based on the number of unique glycopeptides containing each type of glycan structural feature relative to the total number of glycopeptides identified. We have now explicitly clarified this calculation in the **Supplementary Table 3**.

Q26. Table S6

Although Lamc1 is included as an example, it is not mentioned in the main text.

Response: We thank the reviewer for the helpful suggestion. We have reviewed **Supplementary Table 6** (currently **Supplementary Table 7**) and added additional more representative examples. Additionally, we added a detailed list of glycan similarity scores for 38 commonly expressed glycoproteins across 23 tissues (excluding serum), along with two new examples—one showing glycan heterogeneity and another showing glycan conservation of the same glycoprotein across tissues. These two new examples are presented in the newly added **Supplementary Figure 8**.

Typographical or formatting issues

Q27. Line 307, page 12

"Ploy" → should be corrected to "Poly."

Q28. "Gall bladder" should be "Gallbladder" in:

Fig. 1A, H, J; Fig. 2A, C, D, E; Fig. 3D; Fig. 4A, B, D, E; Fig. 5A, B, D; Fig. 6A, E; Fig. 7A.

Q29. In Fig. 2A, there's a red line under Liver/Serum.

Q30. In Fig. 2F: "Olfactorybulb" → should be "Olfactory bulb"

Q31. In Fig. 3E and Fig. 4C: "Celluar" → should be "Cellular"

Response: Thank you for pointing out these issues. We have corrected the relevant spelling and labeling errors accordingly.

In a study where the key contribution is biological rather than technical, structural credibility becomes paramount. This is especially important given ongoing global efforts, such as the HUPO Glycoproteomics Initiative, which stress not only data quantity but interpretability. According to HUPO's cross-platform evaluation, less than 25% of glycopeptides overlap between tools, with glycan composition being the main point of disagreement (https://doi.org/10.6084/m9.figshare.27095782.v3). Hence, rigorous structural validation is essential.

Unfortunately, the current manuscript lacks sufficient evidence to support the glycan assignments. Several assigned glycans appear biosynthetically implausible, prompting us to examine representative MS/MS spectra. In multiple cases, the fragment patterns did not support the claimed structures, suggesting misassignments and undermining confidence in the overall dataset. In addition, the claimed isomer resolution and motif detection (e.g., bisecting GlcNAc) are not convincingly demonstrated. Without such evidence, the claim of "High-Resolution N-Glycoproteome" in the title is premature

However, should the authors validate the structure carefully, a substantially revised manuscript could be reconsidered as a new manuscript. To ensure structural reliability, novel glycan assignments should not rely solely on the developed de novo interpretation tool. Robust validation—such as (1) manual MS/MS annotation, (2) cross-database comparison, (3) reproducibility accross samples or datasets, (4) cross-tool evaluation, (5) biosynthetic plausibility assessments, (6) targeted enrichment and structural characterization—is essential to avoid misassignments and overinterpretation.

Recommended Criteria for Validation (for reference)

To enhance the reliability of glycan structural assignments, it is strongly recommended that the authors clearly distinguish between previously reported structures and novel assignments. For known glycans, validation should be performed by cross-referencing established databases such as UniCarb-DB or GlyTouCan. For novel structures, extensive evidence more than the developed de novo tool is required to avoid misassignments or overinterpretation. The following validation strategies are recommended:

1. Manual MS/MS annotation

Manual annotation of representative and structurally complex examples is required. This includes identification of diagnostic ions that clearly support the proposed isomer assignments. Resolving isomeric glycan structures at the same glycosylation site on an identical peptide backbone provides more convincing evidence than comparisons made between different peptides. Such examples should be fully presented to support the claim of "high-resolution glycoproteomics."

2. Cross-database comparison

Verify whether each assigned structure is registered in international databases such as GlyTouCan, and report matches/mismatches against UniCarb-DB.

3. Reproducibility Across Samples or Datasets

Confirm whether the identified structures are consistently observed in other tissue batches, independent datasets, or biological replicates. If structures are observed in only one sample, provide supporting data to establish their validity

4. Cross-tool evaluation

For glycan structures uniquely identified by StrucGP, explain why they are not detected by other tools (e.g., pGlyco, Byonic). Provide MS/MS-based evidence to support that the structures are not tool-specific false positives.

5. Biosynthetic plausibility assessments

Investigate whether the assigned structures are consistent with known glycosyltransferase specificities and

established biosynthetic pathways. If not, provide direct literature evidence or in vitro data to support the biological plausibility.

6. Targeted enrichment and structural characterization

For particularly novel structures, further biochemical validation is recommended, including glycopeptide isolation, specific enzymatic digestion, PMAA linkage analysis, and LC-MS^n analysis.

Response: We sincerely thank the reviewer for providing all these constructive guidance. We fully agree that structural credibility is paramount for the biological interpretation of glycoproteomics data, especially in the context of ongoing global efforts such as the HUPO Glycoproteomics Initiative. The concerns raised regarding biosynthetic plausibility, MS/MS spectral support, isomer resolution, and potential misassignments have prompted us to critically re-evaluate our dataset and analysis strategy.

We manually inspected representative spectra of glycopeptides carrying biosynthetically questionable structures (e.g., N2H9F1, N3H8), and re-annotated several of them based on fragment ion evidence. In cases where sulfation or *O*-glycan co-occurrence provided a better explanation of the data (e.g., N4H6F1(Sul)1 vs. N2H9F1). While chemical modifications (e.g., sulfation) and *O*-glycan co-occurrence can offer alternative explanations for certain glycan assignments, the lack of comprehensive identification strategies for such structures means that a small number of misassignments may be unavoidable. Fortunately, these cases constitute a very small proportion of the dataset and do not compromise the main conclusions of the study. To raise awareness of this issue, we have explicitly discussed this limitation and provided concrete examples in the revised Discussion section. Additionally, further details have also been provided in our responses to Q1 and Q2.

We clarified our method for distinguishing glycan isomers by diagnostic fragment ions. To illustrate this approach, we now provide paired MS/MS spectra of representative isomers—with and without bisecting GlcNAc—in Figs. 6 and 7, along with a detailed explanation of the ion-based criteria (e.g., peptide + $HexNAc_3Hex_1$, $+Fuc_1$) based on previously published methods. For glycans initially annotated as penta- or hexa-antennary, we carefully considered alternative interpretations involving poly-LacNAc extensions. We also acknowledged the inherent limitations in resolving these possibilities, as current HCD fragmentation settings do not readily generate poly-LacNAc-specific ions. Further details can be found in our responses to Q3 and Q4.

Additionally, we sincerely appreciate the reviewer's six thoughtful and constructive guidance. Below, we provide a point-by-point response, detailing the methods we adopted during revision, as well as the rationale behind decisions where certain suggestions could not be fully implemented:

1. Manual MS/MS annotation

As described above, we primarily adopted manual MS/MS annotation to carefully examine the spectra flagged by the reviewer and other structurally complex or representative examples. We focused on identifying diagnostic fragment ions that support or refute the proposed glycan isomer assignments. In cases where misassignments were detected, we investigated the underlying causes and further quantified the proportion of such cases across the dataset. Fortunately, these low-abundance mismatches represent only a small fraction of total identifications and do not affect the overall conclusions of the study. Given that updating the software requires substantial engineering effort and that manual validation of all questionable cases is not feasible, we have instead focused on discussing the current limitations of the existing software version. We sincerely thank the reviewer for highlighting very insightful and representative examples.

2. Cross-database comparison

In the original manuscript, we included comparisons with GlyConnect. The results showed that only 47.2% of identified glycan structures matched entries in the database. This highlights the current limitation that no existing glycan database is sufficiently comprehensive for validating large-scale glycoproteomic datasets at the structural level. As a result, cross-database matching does not offer a definitive means to confirm or reject glycan assignments, unlike manual annotation. Nevertheless, our lab has been actively working to address this issue by continuously collecting and curating glycan structures from both in-house and external datasets to construct a more complete and structurally resolved glycan database.

3. Reproducibility across samples or datasets

Although some glycan structures appear only in specific samples, we emphasize that the tissue-specific glycan features reported in this study are not based on a single replicate. For example, brain tissue was repeatedly used during method development due to its high glycan diversity but relatively simple glycan branching patterns. As described in the manuscript, the structural features we highlight were consistently observed across biological replicates and are supported by our earlier dataset published in 2021 using StrucGP. Additionally, we conducted literature-based validation for key tissue-enriched glycans, citing references (e.g., refs. 31–33) that independently reported similar structures, despite using different analytical platforms.

4. Cross-tool evaluation

We acknowledge that cross-tool comparisons are important but challenging due to several technical reasons. Many commonly used software tools (e.g., Byonic, MSFragger-Glyco) identify glycan

compositions rather than intact structures, limiting direct comparison. Even when tools such as pGlyco3 also provide glycan structures, the overlap of identified glycopeptides between tools remains low, largely due to the following four factors:

1) **Scoring and filtering criteria**: Each tool uses different matching and filtering strategies. For instance, pGlyco requires strong glycan fragment evidence, while Byonic may allow partial matches.
2) **Glycan database design**: StrucGP uses a modular, branch-based glycan library, whereas other tools use full or semi-combinatorial databases. These differences significantly influence identification results.
3) **Fragmentation model**: StrucGP incorporates linkage-specific and branching-specific ions, which enhances accuracy but may also reduce the number of accepted matches compared to tools with less stringent models.
4) **FDR control strategy**: Although both pGlyco and StrucGP apply glycan-first identification, they differ in decoy generation. pGlyco adds random mass shifts to Y ions, which may fail to reflect the conserved N-glycan core and result in unrealistic decoys. In contrast, StrucGP introduces precursor-level mass shifts while retaining the pentasaccharide core, thereby generating more biologically plausible decoys and allowing more realistic FDR estimation at the glycan level.

Due to these fundamental differences, integrating and comparing results across tools remains technically difficult, particularly at the structural level.

5. Biosynthetic plausibility assessments

We recognize that a small subset of glycan assignments may deviate from canonical biosynthetic pathways. As discussed in the revised manuscript, these mismatches often result from StrucGP's modular analysis strategy, which assembles glycan structures from individual core and branch modules based on accurate precursor mass. While this enhances interpretability and scalability, it may occasionally allow combinations that are not biosynthetically feasible. Additionally, certain mismatches could be caused by rare events such as fucose migration, unaccounted glycan modifications, or the co-occurrence of *N*- and *O*-glycosylation on the same peptide. We have carefully analyzed and discussed these potential sources of structural mismatch in the revised Discussion, and identified them as important targets for future improvements in software development.

6. Targeted enrichment and structural characterization

We greatly appreciate the reviewer's constructive recommendation. We fully agree that for particularly novel or unexpected glycan structures, further chemical and biochemical validation is essential before

functional studies can proceed. Techniques such as glycopeptide isolation, specific enzymatic digestion, PMAA linkage analysis, and LC-MS^n analysis are indispensable for confirming glycan connectivity and linkages. In the revised discussion, we explicitly emphasize that while our large-scale dataset provides a valuable reference framework, targeted experimental validation is a critical next step for studying specific glycan functions.

We hope that these revisions and clarifications adequately address the reviewer's concerns. We are grateful for the opportunity to improve the rigor and transparency of our study through these thoughtful suggestions.

Reviewer's expertise

I have extensive experience in using lectin arrays and mass spectrometry to investigate inter-tissue glycan heterogeneity in mouse models (e.g., Analytical and Bioanalytical Chemistry (2025) 417:973–988). The content of the current manuscript falls well within my area of expertise.

Response: We sincerely thank the reviewer for sharing their relevant expertise and realize that the reviewer is a pioneer in glycoproteome field. We have carefully read the publication (*Analytical and Bioanalytical Chemistry* (2025) 417:973–988) and found it highly relevant to the comments presented in our manuscript. Accordingly, we have now cited this work in the revised version to support the related discussion, as detailed in our response to Comment Q6.

Thanks again for all your professional and constructive suggestions and comments.

Reviewer #3 (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.

Response: We sincerely thank the Reviewer #3 for his/ her valuable evaluation on our manuscript, and appreciate his/ her professional comments and constructive suggestions, which has helped us improve the quality and clarity of our study.

Reviewer #4 (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.

Response: We sincerely thank the Reviewer #4 for his/ her valuable evaluation on our manuscript, and appreciate his/ her professional comments and constructive suggestions, which has helped us greatly improve the quality and clarity of our study.

Reviewer #1 (Remarks to the Author):

In this revised manuscript, the authors very effectively addressed questions and concerns from my previous critique. I disagree with them only in one of their rebuttal statements indicating that sulfation of *N*-glycans are rare, but there is no need to further address that relative to this manuscript.

Response: Thank you for your careful review of our revised manuscript. We are pleased to hear that you think we have effectively addressed the previous concerns and improved the manuscript. Regarding your comments on the sulfation of *N*-glycans, we appreciate your valuable feedback. In fact, we are developing a new method to enhance the detection of these type of chemical modifications and some special glycan structures such as di- or poly-LacNAc by using much lower HCD energies, as we found that the characteristic B ions for sulfated glycans or polyLacNAc were easily lost under HCD energy of 20%. We will investigate the sulfated glycopeptides and some other chemical modifications from these datasets once the method is developed.

Once again, we sincerely thank you for your constructive and insightful suggestions, which have contributed significantly to enhancing the quality of our manuscript.

Reviewer #2 (Remarks to the Author):

We thank the authors for the substantial effort in revising the manuscript and providing detailed responses to our previous comments. While some our earlier concerns have been addressed, several key points remain unresolved and prevent us from recommending acceptance of the manuscript in its current form.

Our main concern is the lack of clarity regarding structural confidence across the 3,045 reported *N*-glycan structures. The authors acknowledge several misassignments we identified and state that these account for less than 0.5% of all structures. However, there is still no dataset-wide, systematic quantification of structural confidence. Figures present all structures as fully determined, without visual or quantitative distinction between high- and low-confidence assignments. Additionally, a glycan-level false discovery rate supported by a realistic decoy strategy has not been demonstrated, leaving the global structure-level error rate unclear.

In our previous review, we outlined six validation criteria. The current revision addresses these only partially:

(1) Manual annotation was performed for a limited set of examples, without extending to a statistically meaningful proportion of the dataset, and without providing a clear basis for the stated misassignment rate of less than 0.5%.

(2) Cross-database matching was restricted to GlyConnect, without systematic reporting of matches/mismatches to other major databases or explicit differentiation between known and novel structures.

(3) Reproducibility was shown for selected tissues/legacy data, but not quantified across the dataset.

(4) Cross-tool analysis appears limited to selected subsets and largely at the glycan composition level; no structure-level concordance metric or spectrum-based adjudication of discrepancies is provided.

(5) Biosynthetic plausibility is discussed qualitatively, but the number/proportion of pathway-inconsistent calls is unquantified.

(6) Orthogonal validation (e.g., stepwise exoglycosidase assays, binding assays with

anti-glycan probes such as antibodies and lectins) for novel/critical features is not provided.

We fully appreciate the challenges inherent to *N*-glycan structural analysis, particularly in backbone classification, antenna number determination, and algorithmic assignment. However, given that these dataset-level validation steps remain incomplete, further work is necessary before the study can provide the level of reliability and interpretability expected for publication in Nature Communications.

We encourage the authors to consider a systematic, dataset-wide validation—including quantitative structural confidence measures, expanded cross-database matching, structure-level cross-tool concordance, and targeted orthogonal experiments—before resubmission. Such steps would greatly strengthen the impact and reliability of this valuable dataset.

Response: We sincerely thank the reviewer for their thorough and thoughtful evaluation of our revised manuscript, as well as for acknowledging the improvements we made in response to your previous comments. We greatly appreciate your concerns regarding the accuracy of the glycan structures, and we agree with your emphasis on structural confidence. In addressing this concern, as well as the six validation criteria you outlined, we have added a new **Supplementary Table** (**Table S2;** with subsequent tables renumbered), including the detailed annotation of the 3,045 *N*-glycans. These revisions are as follows:

1. **Validation of Glycan Structures**: In response to **Comments 1** and **5**, we have provided more detailed annotations for glycans that do not align with known biosynthetic pathways in order to highlight the potential misassignments. To systematically evaluate such cases, we applied predefined filtering criteria for glycans unlikely to arise from canonical biosynthetic pathways. These criteria were derived from established knowledge of *N*-glycan biosynthesis (Essentials of Glycobiology, 4th edition; *Nat Rev Mol Cell Biol*. 2020 Dec;21(12):729-749; *Annu Rev Biochem*. 1985:54:631-64; *Glycobiology*. 2003 Jul;13(7):41R-53R; *Zoology* (Jena). 2004;107(1):49-64), and included the following thresholds: N < 2 or N > 8; H < 3 or

H > 9; S > 4; F > 3 (especially F > 4, which rarely conforms to canonical pathways); structures with disproportionately high sialylation relative to N/H (e.g., N2H3S5); and G > 2. It should be emphasized that these criteria are not "absolute exclusion rules" but rather serve as "red-flag indicators." Certain atypical structures (e.g., highly sialylated or heavily fucosylated glycans) may occur under extreme conditions, such as in tumors or specific tissues (*Mol Cell Proteomics*, 2022, 21(4):100214; *J Proteome Res*. 2022, 21(7):1664-1674), and many such cases have also been reported in public databases such as GlyTouCan. For transparency, we have annotated all such glycans in Supplementary Table 2, together with explanations for why they were flagged. Importantly, their overall frequency in the dataset remains relatively low, supporting the robustness of our identifications.

Additionally, we apologize for not providing more detailed justification for the misassignment rate of less than 0.5% in our initial response. Specifically, based on the identification statistics of unique glycopeptides across 24 tissues (**Supplementary Table 1**), we identified a total of 71,190 high-mannose glycopeptides. Among these, the glycan N2H4F1 appeared 352 times (0.49%), N2H9F1 appeared 24 times (0.03%), N2H8F1 appeared 102 times (0.14%), and N3H8 appeared 61 times (0.08%).

2. **Cross-database Matching**: In response to **Comment 2**, we expanded our database search beyond GlyConnect and included other widely-used glycan databases. We found that most of these databases (including GlyConnect, Glygen, and Glycosmos) are linked via the GlyToucan ID, which serves as a universal token across these platforms. We utilized the GlyToucan API to annotate all glycan compositions and clearly indicated whether each glycan has been reported in these public databases. For the UniCarb-DB database, we found that it had not been updated since 2016 and contained primarily common structures, constrained by the technologies available at that time. Since UniCarb-DB is also linked to GlyToucan, we relied mainly on GlyToucan for our annotations. Accordingly, we added results in the revised manuscript (**Page 5**), indicating that 71% of the identified glycan compositions have been reported in GlyToucan (since a large number of glycan structures were incomplete in the

GlyToucan database, it is still difficult to compare our results with databases at the glycan structural level).

3. **Glycan identification and reproducibility across the dataset**: In response to **Comment 3**, we conducted a comprehensive statistical analysis of glycan identifications across the entire dataset, including information on the corresponding tissues, glycoproteins, glycosites, and PSMs for each glycan. This ensures that the reproducibility of glycan identifications is evaluated comprehensively across all tissues, rather than being limited to a few selected examples.

4. **Structural confidence**: Since all glycan structures reported in this study are the combinations of 4 known core and 17 known branch structures with three major glycan types, there should not be any fundamental distinction between known and novel structures in our dataset. In addition, as our false discovery rate (FDR) is controlled at the PSM level, transforming it to the unique glycopeptide level might lead to a relatively higher FDR. However, based on our observations, even a spectrum has been identified as a decoy peptide, the matched glycan structure can still be a common one in lots of cases. And therefore the estimation of FDR at the glycan structure level is quiet difficult. Instead, we have provided the highest score for each glycan structure for facilitating other researchers' judgement when using our datasets. All four points above constitute the main annotations have been added in Supplementary Table 2.

In response to **Comment 4**, we would like to clarify that almost all other search algorithms provide only composition-level data rather than structural information, which restricts cross-tool comparisons mainly to glycan compositions. For instance, *MSFragger-Glyco* and *Glyco Decipher* infer only possible glycan masses to assign compositions, while *pGlyco* introduces artificial masses on Y1–Y5 ions to judge Y1 correctness. However, this strategy does not represent a genuine glycan-level FDR, and our benchmarking during *StrucGP* development showed that such approaches fail to remove false positives. Indeed, relaxing glycan-level FDR control can nearly double the number of identified PSMs, but at the expense of reliability. These methodological limitations explain why cross-tool evaluation is inherently challenging, especially at the

structural level. Even when putative structures are reported (e.g., *pGlyco3* assigning the most common structure for each composition), substantial discrepancies in glycopeptide identifications remain across tools. This issue has also been highlighted in recent studies (*Nat Commun*. 2025 Jul 1;16(1):5568.), which discusses the difficulty of merging results obtained with different FDR thresholds across tools. As we mentioned in our initial response, the discrepancies in the identification of unique glycopeptides across different software tools may be attributed to four major factors: (1) Scoring and Filtering Differences; (2) Glycan Database Differences; (3) Fragmentation Model Differences; (4) FDR Control Strategies. Together, these factors explain the lack of consistency in glycan identifications across tools and underscore the challenges of cross-tool evaluation, especially at the structural level.

In response to **Comment 6**, we fully agree that targeted orthogonal experiments are necessary before performing more in-depth functional analysis. It should be noted that the main conclusions in our manuscript—such as tissue-specific glycan patterns and glycan heterogeneity on shared proteins—are based on glycans that were repeatedly identified across multiple tissues. Glycans with few identifications, which may include potential misassignments, do not affect these conclusions. Furthermore, some observations, such as the enrichment of bisected glycans in brain and kidney tissues and the prevalence of diantennary glycans with terminal Neu5Gc modifications in liver and serum, have been confirmed in previous glycomics studies, supporting the reliability of the majority of glycan structures reported here. In addition, some rare structures, such as LacdiNAc, have been validated in our other projects (*Redox Biol*. 2025 Apr:81:103584; *Adv Sci* (Weinh). 2025 Jul 24: e02013.) using lectin-based assays (e.g., WFA-Cy5), indicating that the identification of these less common glycans is also relatively trustworthy. For the numerous other low-abundance glycans mentioned in the manuscript, their verification would be best performed by biological experts prior to functional studies. It should also be noticed that all glycan structures reported in this study are actually the various combinations of 4 known core and 17 known branch structures with three known glycan types, and therefore it may be quite difficult to

validate these special glycan structures especially at the glycosite-specific level using lectins or antibodies (this information has been included in the manuscript, Page 5). Accordingly, we have carefully revised the manuscript to emphasize that researchers should perform appropriate experimental validation before conducting in-depth functional and mechanical studies based on our dataset. These points have been added to the *Discussion* section (**Page 20**).

Once again, we sincerely thank the reviewer for their constructive suggestions. We hope the additional supplementary tables and revisions we have made will address your concerns regarding the structural accuracy of the glycans.

Reviewer #3 (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.

Response: We sincerely thank the Reviewer #3 for his/ her valuable evaluation on our manuscript, and appreciate his/ her professional comments and constructive suggestions, which has helped us improve the quality and clarity of our study.

Reviewer #4 (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.

Response: We sincerely thank the Reviewer #3 for his/ her valuable evaluation on our manuscript, and appreciate his/ her professional comments and constructive suggestions, which has helped us improve the quality and clarity of our study.

Reviewer #2 (Remarks to the Author):

We thank the authors for their continued efforts in revising the manuscript and for adding materials such as Supplementary Table S2 and expanded database annotations. However, our central concern remains unresolved: among the 3,045 reported *N*-glycan structures, which are high- versus low-confidence, and in what proportions? The manuscript still does not provide dataset-wide, quantitative measures of structural confidence, and figures continue to present all glycans as if fully determined, without distinguishing levels of certainty.

Note. Items 3 and 4 are considered addressed (within current practical limits). Items 1, 2, 5, and 6 remain outstanding.

The six validation criteria remain only partially addressed:

1. Manual annotation and misassignment rate – The reported "<0.5% error rate" is derived only from high-mannose glycans that we had highlighted as potential misassignments, and does not reflect more ambiguous features. In addition, the Highest_GlycanScore is not interpretable as a confidence metric, since no thresholds or categories are provided.

2. Cross-database matching – Expanded to GlyTouCan, but still restricted to composition-level comparisons. Structural-level confirmation remains absent.

3. Reproducibility [Addressed] – We acknowledge that the authors have provided reproducibility statistics across the dataset, including replicate-level analyses.

4. Cross-tool analysis [Addressed within current software limits] – The authors have performed cross-tool comparisons at the glycan composition level. Given the current limitations of available software, we agree that structure-level comparisons are not feasible, and we consider this point addressed as far as is reasonably possible.

5. Biosynthetic plausibility – New "red-flag" criteria were applied, but flagged glycans remain included in the overall figures and statistics. As a result, readers cannot distinguish between high- and low-confidence structures. These criteria are heuristic

rather than a quantitative confidence framework.

6. Orthogonal validation – No systematic validation (e.g., exoglycosidase digests, lectin/antibody assays) is presented for novel or critical structures.

In conclusion, while transparency has improved through supplementary annotations, the fundamental issue of dataset-wide structural confidence remains to be fully addressed.

Response: Thank you for your careful review of our revised manuscript. We are pleased to hear that we have effectively addressed some of your concerns and improved the manuscript. In response to additional comments on glycan confidence classification, we further compared our data with the GlyTouCan database at the glycan level and added key information to **Supplementary Table 2**, including the classification into high-, intermediate-, and low-confidence categories as well as the detailed scoring criteria. These revisions are as follows:

Response to Comment 1, 2, and 5 — Confidence-level assignment of glycans

we have established a systematic confidence framework for all 3,045 reported *N*-glycan structures based on five criteria, integrating both biological and statistical considerations:

1) **Cross-database matching at the structural level (GlyTouCan, 30 points; highest weight)**: Any glycan matched to GlyTouCan was automatically assigned the full 30 points.

2) **Biosynthetic plausibility (30 points; highest weight)**: Glycans consistent with known pathways were given 30 points, and inconsistent glycans were given 0 points.

3) **Reproducibility across biological replicates (20 points; moderate weight)**: Glycans detected in ≥12 tissues out of 24 received 20 points; glycans detected in 2–11 tissues received 10 points; glycans found in only one tissue or a single site received 0 points.

4) **Highest_GlycanScore (10 points; auxiliary weight)**: Log-transformed GlycanScores (base 10) were classified using the three-sigma rule ($\mu \pm 3\sigma$): glycans with log (GlycanScore) > 3.70 received 10 points; glycans within 1.04–3.70 received 5 points; glycans <1.04 received 0 points.

5) **Probability of GlycanType and All Branches (10 points; auxiliary weight)**: Glycans with both probabilities >0.99 received 10 points; those with either probability >0.99 received 5 points; glycans with neither >0.99 received 0 points. It should be noted that for high-mannose glycans, since they lack branching structures, the branch probability is inherently zero. In such cases, the confidence evaluation relied solely on the probability of GlycanType assignment.

Based on these five criteria (total 100 points), glycans were classified into three confidence categories: high confidence (≥75 points), intermediate confidence (40–74 points), and low confidence (< 40 points). To ensure robustness, glycans structurally matched to GlyTouCan were automatically assigned at least intermediate confidence. Applying this classification, 67% of glycans (2,041/ 3,045) were high-confidence, 27.5% (837/ 3,045) were intermediate-confidence, and 5.5% (167/ 3,045) were low-confidence. To ensure reliability, we manually interpreted 1–3 spectra for each low-confidence glycan as well as for intermediate-confidence glycans that did not conform to canonical biosynthetic pathways, totaling 352 glycans and more than 700 spectra. Manual interpretation showed that 4.3% of glycans (131/ 3,045) remained structurally ambiguous. The results of these manual annotations, labeled as either "structurally reasonable" or "structurally ambiguous," have been added to **Supplementary Table 2**. In addition, we manually interpreted all glycans (106 spectra) mentioned in the manuscript. The results confirm that all glycans presented in this paper are sufficiently supported by characteristic B and Y ions. We have compiled all corresponding spectra into **Supplementary Figure 2**, with characteristic ions highlighted in red boxes.

In the revised manuscript, the confidence classification results for all 3,045 glycans are now described in the *Results* section (**Page 5-6**), while the detailed criteria are

presented in the *Methods* section (**Page 25-26**). The corresponding statistical summary has been also added as **Supplementary Figure 1F**. To provide readers with an intuitive overview, all figures presenting glycan structures (**Figure 1, 2, 6, 7, and S12**) have been updated with color gradients to denote confidence levels, with corresponding explanations in the legends. In addition, we added new **Supplementary Figure 2** (with subsequent tables renumbered), which presents representative spectra for all glycans displayed in the main and supplementary figures. It should be noted that although some glycans did not conform to canonical biosynthetic pathways, the results of manual interpretation were reasonable (e.g., the spectra of the glycans in Figure 6D, as shown in Supplementary Figure 2).

Response to Comment 2 — Cross-database matching (structural level)

We performed composition-level and structure-level matching against GlyTouCan (covering major resources including GlyConnect, GlyGen, and GlyCosmos). Overall, 76.3% (2,324/ 3,045) at the composition levels and 34.2% (1,024/ 3,045) at the structural levels have exact matches registered in GlyTouCan. The relatively lower proportion of structural-level matches may be explained by two factors. First, StrucGP does not rely on pre-existing glycan databases, and is therefore more capable of identifying novel glycans. Second, all glycan structures reported in this study are in fact combinatorial assemblies of four known core structures and seventeen known branches structures within three *N*-glycan types. Based on these considerations, and in line with the reviewer's suggestion, we have now classified all identified glycans into high, medium, and low confidence categories according to a set of stringent scoring criteria. To further validate these assignments, we manually annotated over 800 spectra, explicitly marking each structure as either "reasonable" or "ambiguous." Importantly, we extended this manual verification to every glycan structure discussed in the manuscript, thereby reinforcing the reliability of conclusions derived from structural-level analyses. These overlap rates have been added to the *Results* section (**Page 5**), and per-glycan GlyTouCan IDs and match status are listed in **Supplementary Table S2**.

Response to Comment 6 — Orthogonal validation

We agree that targeted orthogonal experiments are necessary before performing more in-depth functional analyses. However, verifying such a large number of structures individually is challenging, and all glycan structures reported in this study are combinatorial assemblies of four known cores and seventeen known branches within three *N*-glycan types; therefore, glycosite-specific validation using lectins or antibodies is inherently difficult (this information is included in the manuscript, **Page 5**). Fortunately, the main conclusions in our manuscript—such as tissue-specific glycan patterns and glycan heterogeneity on shared proteins—are based on glycans that were repeatedly identified across multiple tissues. The newly added confidence classification together with manual spectral interpretation further supports that these glycans fall into the intermediate-to-high confidence categories, with sufficient diagnostic B- and Y-ions corroborating their structural assignments (**Supplementary Figure 2**). Furthermore, most of the conclusions in the manuscript result from proportion-based differential analyses of substructural features, such as different core and branch structures, rather than whole glycan structures. This means that rare erroneous identifications exert minimal influence on the trends underpinning biological interpretation. In addition, observations such as the enrichment of bisected glycans in brain and kidney tissues and the prevalence of diantennary glycans with terminal Neu5Gc modifications in liver and serum have been confirmed in previous glycomics studies, supporting the reliability of the majority of glycan structures reported here. Some rare structures, such as LacdiNAc, have been validated in our other projects (Redox Biol. 2025 Apr:81:103584; Adv Sci (Weinh). 2025 Jul 24: e02013.) using lectin-based assays (e.g., WFA-Cy5), indicating that the identification of these less common glycans is also relatively trustworthy. For the numerous other low-abundance glycans mentioned in the manuscript, their verification would be best performed by biological experts prior to functional studies. Accordingly, these limitations have been added to the *Discussion* section (**Page 20-21**).

Once again, we sincerely thank you for your constructive and insightful suggestions, which have contributed significantly to enhancing the quality of our

manuscript. Based on these comments, we will also be actively improving our StrucGP software.

Reviewer #2 (Remarks to the Author):

We thank the authors for their continued efforts on this third revision and for substantially improving transparency by introducing a five-criterion confidence framework. However, the scientific justification of the framework remains insufficient, and the headline claim of "high-resolution characterization of 3,045 *N*-glycan structures" is not yet supported as written. Only 34.2% (1,024/3,045) match GlyTouCan at the structure level, and the current High/Intermediate/Low distribution (67.0%/27.5%/5.5%) requires a rigorously justified framework and systematic validation to substantiate whole-structure identification across all entries.

More specifically: (i) the five-criterion weighting lacks scientific rationale; (ii) direct MS/MS evidence (GlycanScore) contributes only 10/100 relative to largely indirect criteria; (iii) GlycanScore is under-specified (inputs, scoring function, interpretation); and (iv) the ~700 manual MS2 verifications are not propagated into confidence assignments or any auditable misassignment/FDR estimate.

In addition, analyses should be limited to structures above the defined confidence threshold; otherwise, the revision improves transparency but does not amount to validation.

Revisions requested

This revision substantially improves transparency and represents a valuable large-scale data resource. To strengthen the manuscript within the current scope, please address the items below.

1. Reframe the main focus and narrative

Shift from "identification of 3,045 structures" to a resource article centered on sub-structure–based biological insights beyond monosaccharide composition. When citing counts, report the confidence distribution (High/Intermediate/Low).

2. Provide a scientifically justified confidence framework (the most important)

Provide a scientifically justified and fully specified confidence framework.

Redefine and clearly document the framework: state the criteria, weights, thresholds, and class boundaries with rationale; mathematically specify GlycanScore (inputs, scoring function, interpretation); spell out the decision rules/data flow from evidence to confidence assignment; link the ~700 manual MS2 verifications to per-call confidence; provide dataset-level misassignment/FDR and sensitivity analyses showing that indirect criteria cannot outweigh direct MS2 evidence.

3. Apply a confidence cut-off for analyses

Restrict biological comparisons to High-confidence calls; inclusion of Intermediate may be allowed only if justified by the revised framework. Known or likely misassignments (e.g., sulfation, *O*-glycan co-occurrence, poly-LacNAc ambiguities) should be flagged and excluded from primary analyses where feasible.

4. State methodological limits

Explicitly connect current constraints (search space, fragmentation limits, branch number ambiguity) to what cannot be concluded at whole-structure resolution.

Status of the six validation criteria of the current manuscript

1. Manual annotation and misassignment rate [Not adequately addressed]–

As noted above, the framework's scientific justification remains insufficient.

2. Cross-database matching [Addressed]

3. Reproducibility [Addressed]

4. Cross-tool analysis [Addressed within current software limits]

5. Biosynthetic plausibility [Addressed]

6. Orthogonal validation [Not addressed, but acceptable as a limitation]–

If the manuscript reframed as a sub-structure–centric resource, this can be stated in the limitation section.

Response: We thank Reviewer #2 for the thorough and constructive evaluation of our third revision. We greatly appreciate the reviewer's recognition of the improved transparency brought by the introduction of the five-criterion confidence framework and for acknowledging the value of this large-scale *N*-glycan dataset. In response to the reviewer's suggestions, we have made the following adjustments:

Response to Request 2.1-Provide a scientifically justified confidence framework (most important)

In this revision, we have completely redefined the confidence assignment system based on a data-driven and quantitatively validated model that integrates both direct MS/MS evidence and indirect structural criteria (as the reviewer suggested, we used manually verified spectra as a gold-standard reference to calibrate the discrimination power of each criterion, thereby establishing a quantifiable and interpretable confidence system).

1. Gold-standard dataset construction

To calibrate the framework, we used manually annotated spectra as a gold-standard subset containing both positive and negative examples. In total, 472 *N*-glycan structures were manually interpreted and labeled as either structurally reasonable (1) or structurally ambiguous (0). The subset included 352 spectra from the previous revision (mainly medium/low-confidence spectra) and an additional ~200 randomly selected spectra (to avoid sampling bias in the gold subset and ensure a balanced distribution of

positive and negative examples).

2. Feature selection and preprocessing

We used six quantitative indicators derived from the five-criterion system:

(1) Highest_GlycanScore (continuous),

(2) GlyTouCan match (binary),

(3) GlycanType_probability (continuous),

(4) BranchesOverall_probability (continuous),

(5) Pathway_status (binary), and

(6) Reproducibility (excluded in this revision).

We explicitly excluded Reproducibility across biological replicates as an input variable, because manual inspection revealed that repeated identifications do not necessarily indicate correctness—for example, the misassigned structure N2H9F1 appeared 24 times but remained incorrect, whereas many structures identified only once were still accurate (77% (115/149) were "structurally reasonable" through manual inspection), which demonstrates that repeated identification frequency cannot effectively discriminate between correct and incorrect structures, and thus should not be included as a confidence criterion. Moreover, several glycans that were identified only once exhibited clear and well-matched B- and Y-ion fragmentation patterns, as shown in Figure R1.

All continuous features (Highest_GlycanScore, GlycanType_probability, BranchesOverall_probability) were z-score–standardized before model fitting to eliminate scale effects.

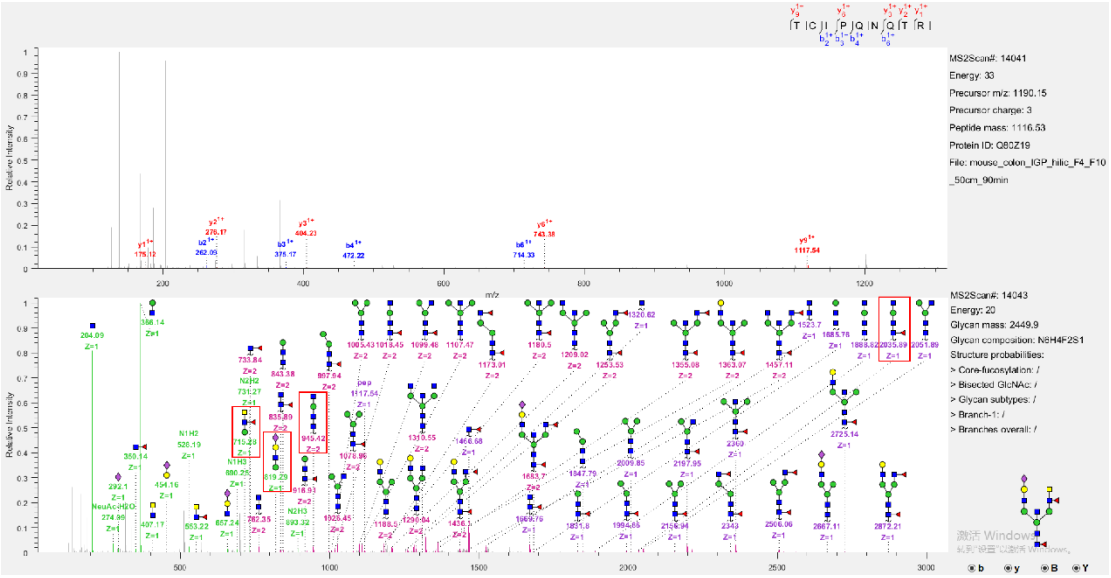

Figure R1. Representative MS/MS spectrum of an *N*-glycan identified only once, showing clearly matched B- and Y-ion fragments that support the structural assignment.

3. **Model training and performance evaluation**

We trained a logistic regression model to predict structural validity, using five-fold cross-validation (5-fold CV) combined with grid search (GridSearchCV) for hyperparameter optimization. The final configuration was:

LogisticRegression(max_iter=100, C=1, class_weight='balanced', penalty='l1', solver='saga').

The model achieved stable and satisfactory performance across folds: **AUC = 0.79; Accuracy = 0.69; Precision = 0.77; Recall = 0.69; F1 score = 0.70.** These results indicate that the model provides good discrimination and robustness in distinguishing reasonable from ambiguous structures (demonstrating solid and reproducible discriminative performance).

To ensure probabilistic interpretability, we further calibrated the output probabilities using isotonic regression (CalibratedClassifierCV, method='isotonic') with five-fold cross-validation. (Note: logistic regression is already well-calibrated by design; therefore, calibration caused negligible change in model performance. For brevity, we do not report post-calibration metrics separately but confirm that the probabilities can be directly interpreted as true confidence levels.)

4. **Threshold determination and class boundaries**

To define interpretable confidence categories, we systematically scanned the calibrated probability outputs and calculated Positive Predictive Value (PPV) and Negative Predictive Value (NPV) at each threshold. High confidence was defined as PPV ≥ 0.95. Low confidence was defined as NPV ≥ 0.95. The automatically determined thresholds were: t_high = 0.82; t_low = 0.46. Since no threshold achieved NPV ≥ 0.95, we adopted the cutoff that maximized NPV (t_low = 0.46) (ensuring that the classification boundaries are statistically optimal and interpretable).

We then applied the calibrated model to the 2,573 *N*-glycan structures that had not undergone manual MS2 verification, and automatically assigned confidence levels according to the optimized probability thresholds. The resulting confidence distribution was: High-confidence (probability ≥ 0.82): 2,355 glycans; Intermediate-confidence (0.46 ≤ probability < 0.82): 218 glycans; Low-confidence (probability < 0.46): none (Consistent with the previous scoring system, these 2,573 glycans were previously classified as medium- to high-confidence structures.). The minimum predicted probability among all glycans was 0.33, indicating that all entries exhibit reasonably high modeled confidence. This is consistent with manual MS2 verification, confirming that the model-based confidence stratification aligns with biological plausibility and expert evaluation.

Response to Request 2.2-Mathematically specify GlycanScore

The specific calculation method for GlycanScore has been detailed in a previous article (*Nature Methods*, 2021) and is briefly supplemented in the revised *Methods* section. Briefly, it is calculated as a function of matched ion ratios and precursor accuracy: $GlycanScore=(Ratioi)^{0.6}\times(RatioY)^{1.6}$, where RatioY is the ratio of identified Y5–Yn ions to theoretical Y5–Yn ions, Ratioi is the intensity of Y ions relative to the total ion intensity in the corresponding m/z region, and Errp denotes the precursor mass error (included in the search tolerance Tolp). The formula and exponents were optimized using a parameter-learning strategy previously applied in StrucGP glycopeptide identification, based on 24 MS datasets generated from mouse brain and validated via target–decoy estimation at 1% FDR. We confirmed that small perturbations (±0.2 in

exponents) change total identifications by < 2%, indicating robustness.

In the revised manuscript, the newly confidence classification results for all 3,045 glycans are now described in the *Results* section (**Page 5**), while the detailed criteria are presented in the *Methods* section (**Page 25-26**). The machine learning–based confidence framework has been updated in **Supplementary Table 2**.

Response to Request 1, 3, and 4-Reframing, Confidence Cut-off, and Methodological Limits

Since these three comments are all derived from the implementation of the revised, data-driven confidence framework described in Request 2, we address them together below.

In this revision, we have thoroughly updated both the manuscript and figures to reflect the implementation of the new, data-driven confidence framework. The title has been adjusted from the original "High-Resolution" to "Comprehensive *N*--glycoproteome atlas" to ensure rigorous scientific narration. The distribution of high-, medium-, and low-confidence glycans was clearly reported in the results and legend upon the first occurrence of 3,045 glycans. The model construction process and evaluation metrics are described in detail in the revised Methods.

All downstream biological analyses based on complete glycan structures are now derived exclusively from high-confidence glycans. Medium- and low-confidence glycans-including previously mentioned cases such as sulfated forms, *O*-glycan co-occurrence, or poly-LacNAc ambiguities-were excluded from these analyses. Conversely, analyses based on substructures were retained, as all substructures were identified through clearly matched B- and Y-ion fragment features. In addition, statistical results and figures have been updated accordingly, and figure legends have been revised to reflect the new data scope and confidence classifications. Importantly, despite the stringent filtering criteria applied, 88.2% (2,687/3,045) of glycans were still classified as high confidence, indicating that the capability of StrucGP to resolve complete glycan structures should not be ignored. Therefore, this study should not be narrowly characterized as limited to substructure analysis. Recognizing the importance

of structural accuracy, we have carefully removed all biological conclusions derived from medium- and low-confidence glycans in the previous version and retained only analyses supported by high-confidence structures.

Finally, the *Discussion* section has been expanded to describe the methodological limitations of the current workflow, including constraints related to the glycan search space, fragmentation coverage, and branch-number ambiguity. Collectively, these revisions ensure that all conclusions are supported by reproducibly validated data and that the scope of the study accurately reflects the achievable structural resolution of the current MS/MS-based approach.

In the revised manuscript, the results based solely on high-confidence glycans have been updated in the *Results* section (**Page 6, 12, 13, 14**). The corresponding figures and legends have also been revised accordingly, including **Figures 1, 2, 4, 5, 6, 7, S1, S4, S6, S9, S11, S12 and S13**, and the associated supplementary tables (**Table S1, S2, S3, S8, S10, and S11**) have been updated to reflect the new high-confidence statistics. The technical limitations of the current approach have been described in detail in the *Discussion* section (**Page 21**).

Once again, we sincerely thank you for your constructive and insightful suggestions, which have contributed significantly to enhancing the quality of our manuscript.

Reviewer #2 (Remarks to the Author):

We appreciate the authors' efforts to introduce a systematic confidence framework using supervised learning, which clearly improved reliability. While the manuscript has progressed, a bit more clarification would further strengthen the scientific soundness of the framework and its alignment with the paper's overall narrative. In particular, please use confidence-aware (High/Intermediate/Low) wording for structure counts in the Abstract.

Below are two items that remain only partially addressed:

(1) Manual annotation and misassignment rate — Partially addressed

The authors describe a logistic regression model based on five features trained on 472 manually annotated examples. To further improve scientific soundness:

•Data selection criteria for the gold-standard dataset: Briefly state the criteria by which ~700 previously reported manual annotations were narrowed to 472 (e.g., spectrum quality, structure categories), so readers can assess representativeness.

•Class counts: Report the numbers of structurally reasonable vs structurally ambiguous examples used for training.

• Feature contribution: Present regression coefficients (with signs) or feature importance so readers can intuitively understand the relative influence of the five criteria on confidence estimation.

•GlycanScore interpretation: Add a short, qualitative sentence connecting the score to structural confidence and indicating how higher scores related to the tier boundaries (no formal proof needed).

•Model positioning: State explicitly that the model is not a true/false classifier, but an integrative framework that combines MS2 evidence and database consistency to assist confidence estimation.

(6) Orthogonal validation — Not addressed, acceptable as a limitation

Systematic orthogonal validation was not performed; state this explicitly as a limitation in the Discussion. Because the primary conclusions rest on sub-structure-level comparisons, make that focus explicit in the Discussion and clarify why full orthogonal validation is not required for the main claims. To avoid over-interpretation, use confidence-aware (High/Intermediate/Low) wording for structure counts in the Abstract. For example: "assigned 3,045 candidate *N*-glycan structures; among these, 2,687 (88.2%) met the high-confidence threshold." This wording maintains transparency without diminishing the value of the work.

Minor comment

In Fig. 5E, the confidence-level annotation appears missing—please verify and add if needed.

Visualization of confidence: In figures where color already represents tissue/system (e.g., Fig. 2A), consider using a different visual channel for confidence (e.g., color + hatching or line style).

A few typos (e.g., microheterogeneitie (line 135), Ploy-sialic/ploySia (lines 340, 348, 437), co-existance (line 1005), thw white matter (line 558)); please proofread.

Response: We thank Reviewer #2 for the thorough and constructive evaluation of our revised manuscript. We have now revised the *Abstract* to adopt confidence-aware wording for reported structure counts and added additional methodological details to improve transparency:

Response to Comment 1- Manual annotation and misassignment rate

1. Data selection criteria for the gold-standard dataset

We thank the reviewer for pointing this out and apologize for the confusion caused by our previous wording. To clarify, the gold-standard dataset was not narrowed down from ~700 glycans to 472. Instead, the number of glycan structures represented in the manually annotated set actually expanded from 352 to 472. The previously mentioned "~700 spectra" referred to the fact that, for the original set of 352 glycans, we manually interpreted 1–3 spectra per glycan to ensure structural reliability. Because many of these glycans were originally assigned low confidence in the scoring framework, multiple independent manual interpretations were necessary to support the correctness of the structures. As stated in our prior response:

"*To ensure reliability, we manually interpreted 1–3 spectra for each low-confidence glycan as well as for intermediate-confidence glycans that did not conform to canonical biosynthetic pathways, totaling 352 glycans and more than 700 spectra.*"

In addition, to avoid sampling bias and to ensure a balanced representation of positive and negative examples for the logistic regression model, we subsequently added ~200 randomly selected spectra from the medium-to-high confidence portion of the original scoring results. This led to a final training set comprising 472 glycans and their associated spectra, as previously summarized:

"*The subset included 352 spectra from the previous revision (mainly medium/low-confidence spectra) and an additional ~200 randomly selected spectra (to avoid sampling bias in the gold subset and ensure a balanced distribution of positive and negative examples)*."

To address the reviewer's request and further improve transparency, we have now added a more explicit description of the data selection criteria in the *Methods* section

(**Page 26**) of the revised manuscript.

2. Class counts

Specifically, among the 472 manually annotated *N*-glycan structures used for model training, 332 were classified as structurally reasonable and 140 as structurally ambiguous. This information has been added to the *Methods* section (**Page 26**) to improve transparency.

3. Feature contribution

In the revised manuscript, we now report the contribution of each feature to the logistic regression classifier. Specifically, the model coefficients show the following order of contribution: Pathway (1.25) > All Branches probability (0.49) > GlyTouCan match (0.45) > Highest_GlycanScore (0.41) > GlycanType probability (0.03). These values have been added to the *Methods* section (**Page 26**) to provide a clearer understanding of how each feature contributes to the overall confidence prediction.

4. GlycanScore interpretation

We thank the reviewer for this helpful suggestion. In the revised manuscript, we have added a brief explanation to clarify the interpretation of GlycanScore in *Methods* section (**Page 26**). Although GlycanScore is only one of several features contributing to the integrative confidence model, it reflects the overall quality of MS/MS spectral evidence supporting a glycan structure. Higher GlycanScores correspond to spectra with more complete and higher-quality Y-ion series, and therefore are generally associated with higher structural confidence.

5. Model positioning

In the revised manuscript, we have now explicitly clarified that the model is not intended to function as a binary true/false classifier, but rather as an integrative confidence-estimation framework. This clarification has been added to the *Results* section (**Page 5**) of the revised manuscript.

Response to Comment 2- Orthogonal validation

We thank the reviewer for this thoughtful comment. In the revised manuscript, we

have explicitly acknowledged the lack of systematic orthogonal validation as a limitation in the *Discussion* (**Page 21**). We now clarify that all statistical analyses involving glycan structures are restricted to high-confidence assignments, whereas the other biological conclusions of the study rely on sub-structure–level comparisons. As suggested, we have also revised the *Abstract* (**Page 2**) by incorporating confidence-aware wording and explicitly reporting the number of high-confidence glycans.

Minor Comments

In the revised manuscript, we have added the missing confidence-level annotation to Fig. 5E and improved the visualization of confidence levels in figures where color is already used to denote tissue or system (e.g., Fig. 2A) by replacing color coding with alternative visual channels, specifically double solid lines, single solid lines, and dashed lines. In addition, all noted typographical errors have been corrected, and the entire manuscript has been thoroughly re-proofread for accuracy and consistency.

Once again, we sincerely thank you for your constructive and insightful suggestions, which have contributed significantly to enhancing the quality of our manuscript.