

A comprehensive landscape of human organ N-glycoproteome

Corresponding Author: Professor Zhixin Tian

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)

Hu et al present an N-glycoproteomics study of 18 human organs. The glycoproteomics characterization of (healthy) human tissue material is very relevant and timely as information on this is scarce, biological relevance enormous and techniques and data analysis workflows to perform these analysis are challenging. Having this said, I do not consider the current work of high impact in the field, mainly because there are some very large over-interpretations of the data, not supported by the presented data, which I will detail below, but also because per organ only 1 sample of 1 individual is analyzed. Although this will provide some information about general trends, it is well known that interindividual glycan differences may be very large.

Major comments:

- 1) Line 87: "31,003 N-glycan structure-level" The claim that glycans are determined at the structural level is not true. Indeed some structural features are determined for part of the compositional annotations, but this should be more carefully stated.
- 2) The glycan nomenclature is not standard in the field or intuitive, and not explained.
- 3) The statement about the "new" identification of N-glycosite N1767 on protein Q07954 is not substantiated by the data. It is anyway unlikely to confirm glycan location based on HCD data, this should be done by alternative approaches such as ETD or de-glycosylation experiments.
- 4) It is unclear what the difference between BF and TF is, and how these structures can be proven by the HCD fragmentation? Is the annotation in Figure 4C regarding this correct?
- 5) Line 161: While CF and TF are mostly fucosylated HexNAc structure. – Evidence for this is missing.
- 6) Line 163: All TF glycans are identified as H2-antigen-type structures. The only N-glycan containing all the three types of fucoses is a Lewis Y structure (Fig. 4c). – Evidence for this is missing.
- 7) From the methods section it is unclear how "relative quantification" is performed. Is this per glycosylation site? And based on glycan counts or peak areas?
- 8) The biological interpretation/description of the relevance of certain findings seems a bit random to me. Eg, it is unclear why the high fucosylation in prostate tissue is "is particularly relevant given previous studies demonstrating elevated fucosylated N-glycans in neuroendocrine prostate cancer and prostate cancer during prostatectomy" (line 171). Similarly the statement in line 191 is, although true, not the only (or most specific) relevance of sialylation in biology.
- 9) Line 198: it is unclear why ST3GAL5 is mentioned here in combination with the N-glycoproteomics data as this transferase sialylates lipids, and is likely independent from the N-glycoprotein pathway.
- 10) Line 221: Please provide the evidence for the Lewis X structures.

Minor comments:

- 1) The English language can be improved, mainly in the introduction.
- 2) It is unclear what this means (and also the referred Supporting Figure does not match): The N-glycosylation motif N-x-S/T/C is highly conserved, with longer conserved sequences in muscle (Supplementary Fig. S3).
- 3) Not all figure panels match the legends and explanation of what's in the details or legends is often missing.
- 4) Line 111, It is unclear what is meant by organ-specific linkages.
- 5) The number of organs mentioned per cluster in Figure 2b is unclear to me.
- 6) Supporting figure 5 to 20 are unclear in their meaning and data presented.
- 7) Axis titles are missing in Supporting Figure 24.

Reviewer #2

(Remarks to the Author)

This work is another step towards introduction of informatics to glycobiology.

Reviewing this paper is the first time in my 30-year scientific life, when I have found no major inconsistencies in the manuscript. However my area of expertise does not intersect much with glycoproteomics. I give a highly positive mark to this timely, well structured, methodologically correct, and perfectly visualized investigation. Not less important is that authors do not attempt to cover everything in a single study, but formulate a certain definite challenge, and solve it.

Authors present a lot of statistical data, which can be utilized in further studies. They do not say what to do with feature distribution data, except general words in the introduction, however this is acceptable for papers reporting research done (in-part) by the informatics methods.

Supplementary figures might need more detailed explanations; on the other hand, they are still comprehensible for a glycoinformatician.

Reviewer #3

(Remarks to the Author)

Hu and co-authors report a comprehensive and excellent approach to construct large dataset of N-glycoproteomics providing a broad, tissue-spanning perspective of human N-glycosylation patterns and insights into the complex heterogeneity of N-glycosylation distributions from across 18 normal human tissues.

The present results may accelerate the discovery of disease-specific biomarkers when comparing with the expression levels of some target glycopeptides in some diseases. However, it should be noted that the present strategy does not pay attention to the quantitation of the glycopeptides, while relative amounts are used. In addition, it is also important to note that the results obtained here strongly depend on the glycopeptides captured by the method employed.

The followings are changes should be performed:

In line 42, HIPO may be HUPO.

In line 96, S3 may be S21.

In line 100, S4-S21 may be S3-S20.

In Fig. 1i, what does the horizontal axis represent?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The dataset presented by Hu et al is a valuable dataset containing extensive new information on the glycoprotein forms present in different human tissues. While the manuscript is much improved as compared to the previous version, I still think some over-interpretations of the data are presented, including potential miss annotations of glycopeptides, and reports on structural features that are limitedly supported by the data. What additionally is missing is an evaluation of the variation of the data over the organs of individuals and the comparison of different glycoforms of the same proteins/sites over different organs. Please find my detailed comments below:

1) In the main manuscript, information is missing on the number of biological replicates per organ. I found this information in the supporting information. However, it would be valuable to know from how many samples the numbers are derived that are presented in the first section of the results. In the discussion 219 runs are mentioned, also here the sample distribution is more relevant.

2) Related to the previous point: it is really good that multiple biological replicates were included per organ, but it is unclear if these came from different or the same individuals. And, importantly, the variation in glycopeptide abundance was not included in the data presented in the main manuscript. Differences in abundancies of glycan traits are presented between organs, but it is unclear if these were statistically significant.

3) Line 202-204: MS2 data supporting this statement was not included.

4) MS2 support for the discrimination between BF and TF is not solid. Firstly, the peaks in Supplementary Figure 31 for H1F1 fragments are very low abundant, from the current representation it is unclear if these can be reliably interpreted.

Additionally, antenna fucoses are notoriously known to migrate during HCD fragmentation (<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/rcm.2509>) making their fragment patterns unreliable in general.

5) Figure 1h and overall fragment interpretation: the glycan fragments shown to be indicative for LacNAc repeats are likely

(also) derived from the connecting mannose/GlcNAc, and thus not diagnostic for polyLacNAc. Conclusions on the presence of tri-antennarity and polyLacNAc should be more carefully made.

6) Structures shown in Figure 3c, especially in the category BF only, are only partly substantiated. For example, what is the proof of the presence of two antennas on the '6 arm, while the 3' arm is empty? And, in relation to the previous point, how to discriminate between polyLacNAc and multi-antennary glycans?

7) Figure 1: the fragment in column 12 in 1h indicates double galactosylation (alpha-gal?), this is uncommon in human and not reported in the text (same for figure 5d).

8) Line 345: bi and tri-antennary N-glycans are difficult to discriminate based on HCD fragmentation: the 3rd antenna could also be a LacNAc repeat.

9) A discussion is missing on the biological relevance of total organ data: most organs consists of many different cell types, all with their own glycosylation features. For example, blood and immune cell derived glycosylation will "contaminate" most of the organ-specific cell type glycosylation.

10) Figure 7 and related discussion: while not necessarily wrong, it is unclear what is the biological and/or scientific relevance of building a model to discriminate organs based on glycoproteomics data.

11) Figure 8b and related discussion: the glycopeptide annotations unique to a single software package should be interpreted with caution, there is a high risk of miss annotations. Importantly, for none of the tools, retention time is present in the output, while this is an important factor in validating glycoforms on the same peptide portion (they should elute close together with small increases between sialic acid clusters: <https://www.nature.com/articles/s41467-024-50338-5>). High numbers of miss annotations are a risk in this data anyway, due to the possible presence of overlapping masses due to overalkylation, oxidation, ammonia adducts and monosaccharide similarities (<https://www.sciencedirect.com/science/article/pii/S1367593123000108>). The extend of these potential artifacts in the data should be evaluated.

12) While glycoproteomics data is presented, the main focus of the data analysis is on the glycomics aspect of this data: what glycan features can be found in what organs. – In principle this information could also be obtained with, much simpler, glycomics studies. The data also contains valuable information on the proteins carrying the glycans, this part of the data could be further leveraged by reporting organ-specific glycan profiles for shared glycoproteins and protein-specific glycan profiles within one organ.

13) Line 357: what is the relevance of the SARS-CoV-2 signalling pathway here? This is just an overlapping protein annotation related to coagulation and provides no further insights into the SARS-CoV-2 disease mechanisms.

14) Line 419: "Machine learning models based on all the data also show good results." Please define what "good results" are.

15) Connection lines between datapoints in Figure 2c should be removed, data is not paired or linearly related.

16) In the methods, information is missing on the input amount for each tissue sampe (eg, sample weight). And whether this sample was taken prior or after (whole organ) homogenization.

17) Figure 5f: the "linkage" color scale is not well described.

18) Line 148: "High-mannose N-glycans and the core structure HexNAc(3)Hex(2) are predominant across organs" - HexNAc(3)Hex(2) is not the N-glycan core structure, that would be HexNAc(2)Hex(3).

19) Line 163: the use of the term "tri-antennary high-mannose" is confusing as tri-antennary is usually used for complex type N-linked glycans.

20) For example line 191: what does 01Y41Y41M(31M21M21M31G31G21G)61M(31M21M)61M refer to?

21) The manuscript should be improved language-wise.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

In my view, the manuscript in its original compact form was already suitable for publication. The authors' corrections have not diminished its quality, and most of their revisions—with a few exceptions related to nomenclature—appropriately address the requests of the other reviewers.

While the applied corrections have made the findings more substantiated, the inclusion of excessive detail has, in my opinion, reduced the article's accessibility for the broad readership of Nature Communications. I am therefore uncertain whether the extended discussion of these issues, beyond the necessary technical corrections, has ultimately improved the paper.

During the initial review, I could follow the meaning of most figures and SI figures, though sometimes a glycoinformatic context was required (knowledge, which a general reader may lack). I have not re-examined all SI figures in detail at this stage, but the authors' responses to the reviewers seem reasonable.

I am fully satisfied with the revisions made in response to the other reviewers' remarks. Although I find some additions redundant, I have not detected any errors.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The present study provides a large set of N-glycoproteome database of 18 human organs/tissues composed of intact glycopeptides on 7,782 N-glycosites of 5,078 glycoproteins. By facilitating a systemic comparison between the different analytical platforms, authors demonstrated that pPeptide has superior performance in the in-depth MS/MS data analysis. However, the results by comparative analysis also uncovered that the complementarity of current N-glycoproteomics tools is approximately 21.1% of N-glycopeptides. This human organ N-glycoproteome atlas may become a basic resource for glycobiology community. In addition, it is important to note that quantitative insight into the organ relevant glycopeptides is crucial for further investigation of disease specific N-glycosylation signatures.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #4

(Remarks to the Author)

I was brought into this review process to evaluate how Hu et al. addressed the responses from Reviewer #1 in the previous round of revision. I tracked the reviewer comments, author responses and my own evaluation of the responses. Due to this format I refer the editor and authors to the attached PDF "EvaluationOfReviewer1Comments".

(Remarks on code availability)

Version 3:

Reviewer comments:

Reviewer #4

(Remarks to the Author)

After this round of revision I recommend the manuscript for publication.

I appreciate the effort from the authors for addressing all comments. I believe that with the adaptation of the results, by toning down conclusions of fine glycan structure determination, the main concerns from Reviewer #1 have been addressed.

All my additional comments have been dutifully addressed and I appreciate the cooperative disposition from the authors.

All I have left are very minor comments (one of them caused by a mistake from my side in the initial revision):

- Text edit: "Fucosylation has three types of position isomers of core (CF) and branch (BF)."

As I understand, "three" still refers to the previous classification of CF, BF, and TF. Adapt the numbering.

- Search re-scoring strategy such as the one from PeptideShaker was mentioned, but not cited. Please add relevant citations.

- I shared the wrong DOI for what is reference 43 in the current state of the manuscript. This is the intended reference DOI: 10.1093/bioinformatics/btq054 corresponding to the initial release of Skyline. Apologies for my mistake in providing the incorrect citation for Skyline in the previous revision which referred to DIA data analysis in Skyline which is not relevant for this publication.

Looking forward to have this dataset available and future developments of gPeptide!

Cheers!

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to the reviewers' comments

Dear anonymous reviewers,

Thank you very much for time and efforts of reviewing our manuscript and providing valuable comments. We have fully addressed all of your comments and made the corresponding revision of the manuscript as well as the supplementary information as illustrated point-by-point below.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

Hu et al present an N-glycoproteomics study of 18 human organs. The glycoproteomics characterization of (healthy) human tissue material is very relevant and timely as information on this is scarce, biological relevance enormous and techniques and data analysis workflows to perform these analysis are challenging. Having this said, I do not consider the current work of high impact in the field, mainly because there are some very large over-interpretations of the data, not supported by the presented data, which I will detail below, but also because per organ only 1 sample of 1 individual is analyzed. Although this will provide some information about general trends, it is well known that interindividual glycan differences may be very large.

Major comments:

1) Line 87: "31,003 N-glycan structure-level" The claim that glycans are determined at the structural level is not true. Indeed some structural features are determined for part of the compositional annotations, but this should be more carefully stated.

Response: after incorporation of the compositional information, glycan isomers were matched by (1) a database of biologically appropriate N-glycan linkages, (2) structural diagnostic fragment ions screened to confirm monosaccharide sequence structure. Glycan structure information has been provided on <https://www.ebi.ac.uk/pride/archive/projects/PXD059886/privatereviewdataset> as shown in the "Structure-diagnostic ions" column of each excel

table. For clarity, the manuscript has been revised to a more appropriate description as “N-glycoproteomes of 18 healthy human organs revealing 58,933 plausible N-glycan structure-level and 24,135 monosaccharide composition-level intact N-glycopeptides on 7,782 N-glycosites of 5,078 N-glycoproteins”. In addition, the newly added analysis results have been included to clarify the glycosite and structural diagnosis of the IDs as described as “In addition, N-glycosite-determining fragment ions are observed for most (ranging from 66.0% to 85.5%) of the IDs, where N-glycosites are unambiguously localized; structure-diagnostic fragment ions are observed for more than 50% (ranging from 49.9% to 80.1%) of IDs, where monosaccharide sequence isomers (with the same molecular formula including the same monosaccharide compositions) are distinctly differentiated”. These ratios of the IDs across all the organs were provided in **Supplementary Figure 1f**.

2) The glycan nomenclature is not standard in the field or intuitive, and not explained.

Response: (1) in terms of the compositional level, we adopt a one-character abbreviation strategy displayed in **Supplementary Data 1** on bias of Symbol Nomenclature for Glycans (SNFG). For example, N4H4F0S1 means 4 Hexosamines (HexNAc), 4 Hexoses (Hex), zero fucose and 1 Sialic acid. (2) In terms of the linkage level, we used the one-line text formats with one-letter symbols for monosaccharides adopted from GLYCAM-Web (<https://glycam.org/docs/forcefield/glycam-naming-2/>). One-line text representation is written from left to right starting with “01” followed by one-letter symbols of monosaccharides and their linkage positions, and branches are put in parentheses in the increasing order of linkage positions (Large-scale identification and visualization of N-glycans with primary structures using GlySeeker, <https://doi.org/10.1002/rcm.8023>). It should be noted that anomeric positions of linkages are omitted from both of the representations because they are not distinguished in tandem mass spectra. (3) In terms of the nomenclature of fragment ions², theoretical fragment ions of N-glycans include canonical B/Y/C/Z ions, cross-ring A/X ions, and internal

fragment ions (pairwise coupling between {A, B, C, Y, Z} and {X, Y, Z}, except for A-X, Y-X, and Z-X ions). These internal fragment ions require at least three monosaccharides length and two fragmentation, in which cross-ring fragmentation can occur only once. In this study, the nomenclature of internal fragment ions is started with the position of location of the C-C bond, and ended with an internal fragment ion type (e.g., $^{4,5}AI3^{1+}$).

For the nomenclature of the fragment ions of the peptide backbone, b^* and y^* represent the b and y ions containing GlcNAc.

For clarity, the manuscript has been revised to add the abbreviation of monosaccharide in **Supplementary Data 1** as shown below.

Abbreviation	Monosaccharide	Icon
N	HexNAc	
H	Hex	
F	Fucose	
S	Sialic acid	
G	Glucose	
M	Mannose	
L	Galactose	
V	GalNAc	
Y	GlcNAc	

3) The statement about the “new” identification of N-glycosite N1767 on protein Q07954 is not substantiated by the data. It is anyway unlikely to confirm glycan location based on HCD data, this should be done by alternative approaches such as ETD or de-glycosylation experiments.

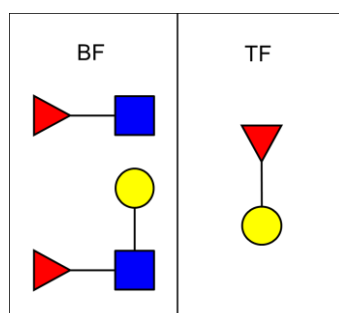
Response: the selected proteins here are those that are distributed in various organs of the human body. The site-determining ions in our study refers specifically to the GlcNAc-bearing b- or y-ions (b^* or y^* ions) generated by HCD fragmentation^{3, 4}. During HCD fragmentation of glycopeptides, the glycan (particularly the inner core GlcNAc) often remains attached to the peptide backbone fragments. This results in the production of b^* and y^* ions that still carry the GlcNAc modification. The location of the glycosite is then pinpointed by identifying pairs of these b^*

and y* fragment ions. The data on the MS2 matching situation of the new glycosites proposed in this study have been presented in the **Supplementary Data 3**.

The description has been revised to be more open expression as “N-glycosites N1767 and N3858 on the peptide backbone of LYWISSGNHTINR and NFMKTHNTCK represent novel N-glycosites unreported in the Uniprot (<https://www.uniprot.org/>)”. Nevertheless, confirmation by alternative approaches such as ETD or de-glycosylation experiments is necessary, although the authors currently do not have the experimental conditions to carry out such experiments.

4) It is unclear what the difference between BF and TF is,

Response: we distinguish BF (Branch Fucose) from TF (Terminal Fucose) mainly based on the biological synthesis rules⁵ which was stated as “the basic Gal β 1-4GlcNAc or lactosamine sequence may contain a fucose in α 1,3 linkage (31F) to the GlcNAc, as for example in human lactotransferrin” and “alternatively the galactose may be substituted with a fucose residue in α 1,2 (21F) or α 1,6 (61F) linkage”. The reference source also includes Essentials of Glycobiology (<https://www.ncbi.nlm.nih.gov/books/NBK579918/>).



and how these structures can be proven by the HCD fragmentation?

Response: for the example of the monosaccharide composition with three types of fucose (**Supplementary Figure 31**), the core fucose CF was identified by a series of Y ions with good isotopic envelope

fingerprinting, and TF was confirmed by the CI2 ion ($M = \text{Gal} + \text{Fuc}$, m/z of MH^+ is 327.13). The description has been revised as “In fat 2 (technical replicate 3), the core fucose is determined by a series of core fucose-containing Y ions including YI1, YI2, YI3 and YI4, of which the isotopic envelope fingerprint matches are displayed. The CI2 ion explained the presence of terminal fucose. And the observed mass difference (Δm) between $^{3,4}\text{Al3}$ and $^{2,5}\text{Al3}$ accounts for the presence of a branch fucose. Manual validation is performed on the matches of these two glycopeptides in representative samples from ten organs”. In terms of the strategy for distinguishing CF we are in agreement with previous research⁶ which determined CF and others using HCD.

Is the annotation in Figure 4C regarding this correct?

Response: Figure 4c has been corrected and now is **Figure 3c**.

5) Line 161: While CF and TF are mostly fucosylated HexNAc structure.
– Evidence for this is missing.

Response: the text of “Most of the CF N-glycans of the 51 glycans are bi-antennary featuring a (sialyl) LacNAc structure, which is consistent with the previous study⁷” has been added. To verify whether the CF-only glycans also exhibits this characteristic in various organs, new analyses were incorporated in **Supplementary Figure 29a**. And indeed, this characteristic holds true in a wide range of organs.

As for TF glycans, since fucose is bonded to Gal and Gal to GlcNAc, there always exists the LacNAc structure.

6) Line 163: All TF glycans are identified as H2-antigen-type structures. The only N-glycan containing all the three types of fucoses is a Lewis Y structure (Fig. 4c). – Evidence for this is missing.

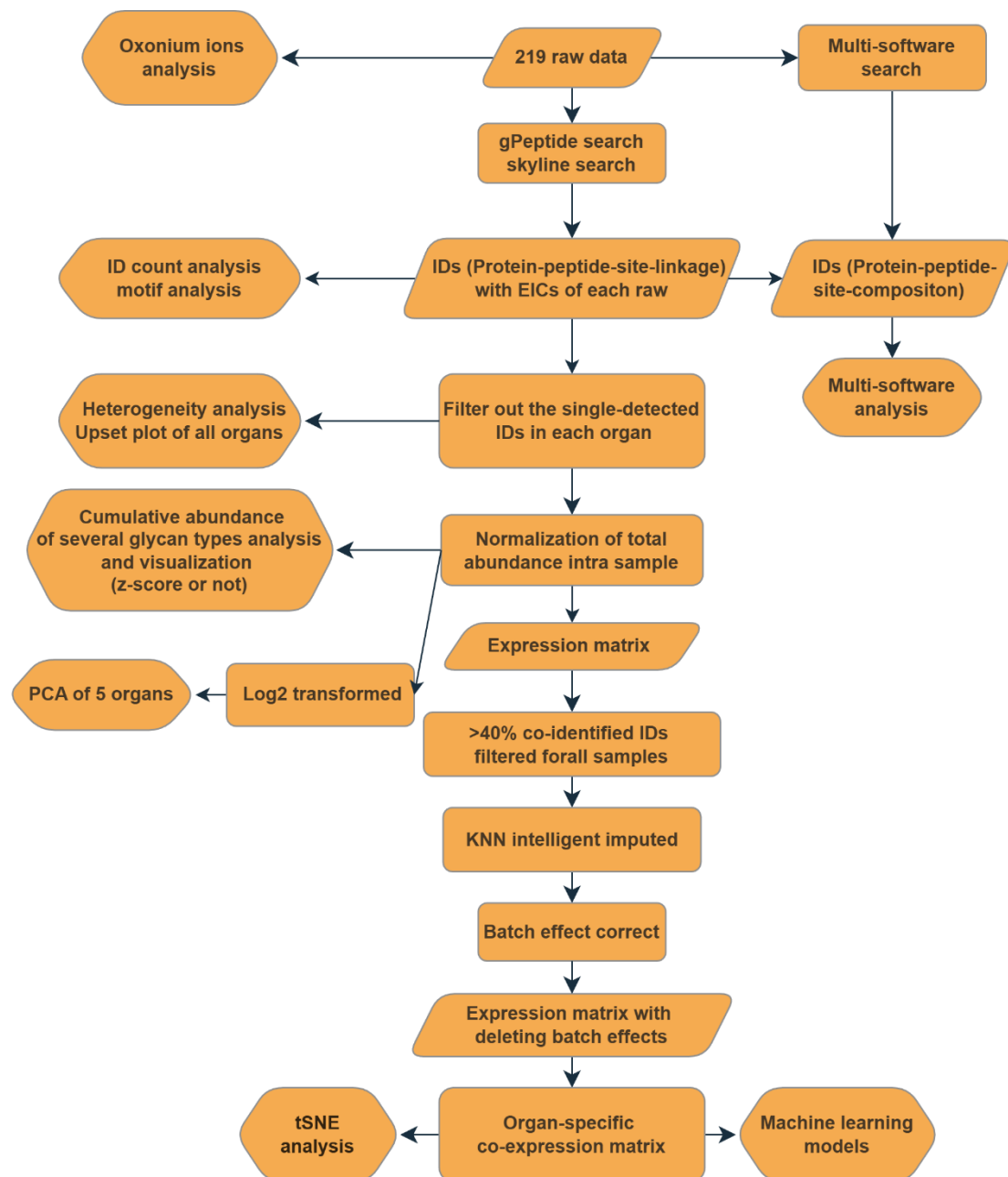
Response: we have removed the “All TF glycans are identified as H2-antigen-type structures” to avoid redundancy.

The two N-glycans containing all the three types of fucose also contain Lewis Y structure, and to avoid any ambiguity, we have directly written down the corresponding intact N-glycopeptide IDs in the manuscript and provided the corresponding MS2 matching results (**Supplementary Fig. 31, 32**). The corresponding manual verification results are listed in **Supplementary Data 4**.

7) From the methods section it is unclear how “relative quantification” is performed. Is this per glycosylation site? And based on glycan counts or peak areas?

Response: for each ID, peak area of the extracted ion chromatogram was obtained by Skyline. With removal of values observed only in one organ, the matrix was normalized by the sum of the abundance to achieve intra-sample normalization.

As shown in the following **Rev. Fig. 1**, for quantification between organs, a multi-step approach was implemented to ensure data compatibility across different samples and to achieve inter-sample normalization. To handle missing values, IDs with identifications in ≥ 85 raw files were subjected to K-nearest neighbors (KNN) imputation with the same organ and the same individual carry greater weight and missing values for which insufficient neighbors were available were filled using the minimum value. Batch effect correction (BEC) was subsequently performed on the imputed data using ComBat. The resulting batch-corrected matrix was used for tSNE analysis and model training, which were decision tree, neural network, random forest and support vector machine (SVM). This figure has been added to **Supplementary Figure 36**.



Rev. Fig.1

8) The biological interpretation/description of the relevance of certain findings seems a bit random to me. Eg, it is unclear why the high fucosylation in prostate tissue is “is particularly relevant given previous studies demonstrating elevated fucosylated N-glycans in neuroendocrine prostate cancer and prostate cancer during prostatectomy” (line 171). Similarly the statement in line 191 is,

although true, not the only (or most specific) relevance of sialylation in biology.

Response: more appropriate cross-validation with multiple studies^{1, 8, 9} is now added. The validation was conducted through the overall N-glycan composition as well as the characteristic ions in the MS2 spectra of mouse.

9) Line 198: it is unclear why ST3GAL5 is mentioned here in combination with the N-glycoproteomics data as this transferase sialylates lipids, and is likely independent from the N-glycoprotein pathway.

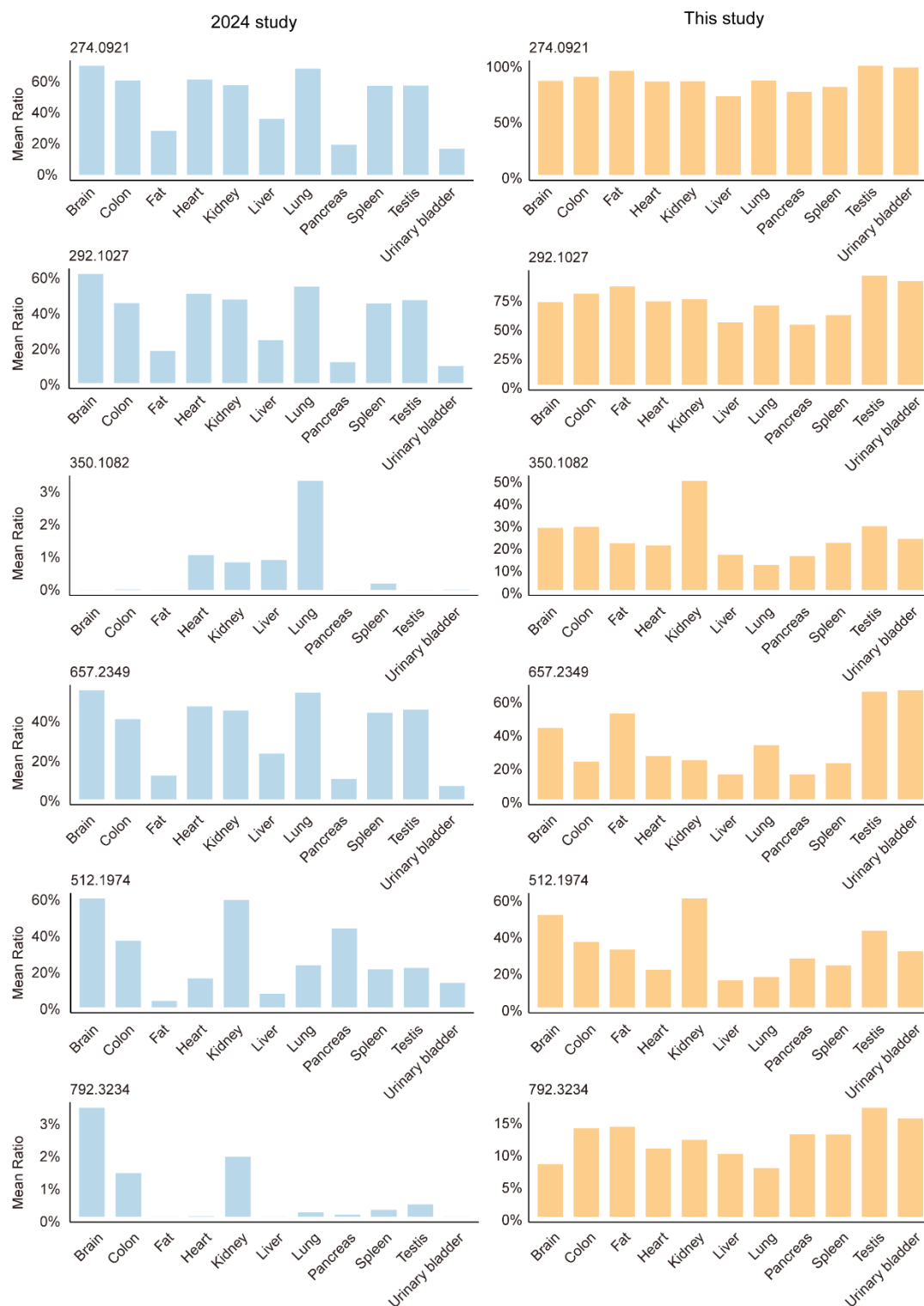
Response: removed as suggested.

10) Line 221: Please provide the evidence for the Lewis X structures.

Response: in the initial manuscript, this was intended to describe the form in which fucose and sialic acid occur together. What we advocate is the sialyl Lewis X structure. The description has been revised as “805 N-glycopeptides are found to contain the sialyl lewis X structure in the FxSy group. It is proposed from the aforementioned observation that the co-occurrence form of fucose and sialic acid in these organs is mainly the sialyl Lewis X structure.” As shown in **Supplementary Figure 31**, gPeptide distinguishes sequence isomers using structure-diagnostic fragment ions¹⁰.

However, considering the need to enhance the reliability of our data, we conducted a non-targeted search strategy to verify the relevant IDs before the count of co-occurrence forms of fucose and sialic acid. First, for the common 11 organs of our dataset (121 raw files) and the 2024 study (22 raw files), we performed a data re-analysis using a non-targeted tool reported by Johannes Stadlmann et al¹. Using oxonium ion 366 as a reference, we calculated the proportions of various N-glycan fragment ions. Our results successfully replicated the ratios reported in the 2024 study, thereby validating the accuracy of our analytical workflow (**Rev.**

Fig.2). However, the proportion of our data is overall relatively high and the difference between the organs is relatively low. Second, all the 219 raw files were searched against the non-targeted tool for validation of 350.1, 512.2 and 657.2 amu ions. Third, in all spectra containing target oxonium ions, we remove the spectra with intensities less than the lower quartile to improve data quality. At last, distribution of the spectral overlap between the 350.1, 512.2 and 657.2 amu ions and BF-only IDs from gPeptide across 18 organs showed the widespread presence of diagnostic glycan fragments (**Supplementary Figure 30**). Among them, the 657 amu containing spectra are clear evidence of the presence of sialyl Lewis X.



Rev. Fig.2

Minor comments:

- 1) The English language can be improved, mainly in the introduction.

Response: the manuscript has been all revised.

2) It is unclear what this means (and also the referred Supporting Figure does not match): The N-glycosylation motif N-x-S/T/C is highly conserved, with longer conserved sequences in muscle (Supplementary Fig. S3).

Response: our motif analysis confirmed that the canonical N-x-S/T/C motif is highly conserved across all tissues examined using R package motifStack¹¹. However, after performing probability calculations with the same parameters, it was found that the motifs of fat and urinary bladder are not limited to N-X-S/T, but rather have an extension of 17 amino acids around them. The ovary and testis have certain conserved regions around N-X-S/T near the C-terminal. We speculate that it might be that the N-glycosylation of these organs has a certain preference for the amino acid sequence around N-X-S/T. The longer conserved sequences observed in these tissues may reflect more stringent functional or structural constraints around the N-glycosylation sites. The referred Supporting Figure has been matched as **Supplementary Figure 2**. The description has been revised as “The N-glycosylation motif N-x-S/T/C is highly conserved, with longer conserved sequences in fat and urinary bladder” in this version. Thanks for your precise review.

3) Not all figure panels match the legends and explanation of what's in the details or legends is often missing.

Response: we have revised all the figures.

4) Line 111, It is unclear what is meant by organ-specific linkages.

Response: we have revised it as “organ-specific IDs”.

5) The number of organs mentioned per cluster in Figure 2b is unclear to me.

Response: “6 organs” cluster means identification in 6 organs.

6) Supporting figure 5 to 20 are unclear in their meaning and data presented.

Response: these figures have been revised based on the IDs detected in more than 60% of the raw files in each organ. These representative glycopeptides reduce inter-subject variability and serve as molecular signatures for the organ.

7) Axis titles are missing in Supporting Figure 24.

Response: after increasing the subjects, we changed the normalization method to a simpler total intensity normalization method. The graph here is no longer necessary to be shown. And the normalized data was showed in **Supplementary Figure 5** (log2 transformed).

Reviewer #2 (Remarks to the Author):

This work is another step towards introduction of informatics to glycobiology.

Reviewing this paper is the first time in my 30-year scientific life, when I have found no major inconsistencies in the manuscript. However my area of expertise does not intersect much with glycoproteomics. I give a highly positive mark to this timely, well structured, methodologically correct, and perfectly visualized investigation. Not less important is that authors do not attempt to cover everything in a single study, but formulate a certain definite challenge, and solve it.

Authors present a lot of statistical data, which can be utilized in further studies. They do not say what to do with feature distribution data, except general words in the introduction, however this is acceptable for papers reporting research done (in-part) by the informatics methods.

Supplementary figures might need more detailed explanations; on the other hand, they are still comprehensible for a glycoinformatician.

Response: supplementary figures have been all revised. We have improved our manuscript from several aspects.

Reviewer #3 (Remarks to the Author):

Hu and co-authors report a comprehensive and excellent approach to construct large dataset of N-glycoproteomics providing a broad, tissue-spanning perspective of human N-glycosylation patterns and insights into the complex heterogeneity of N-glycosylation distributions from across 18 normal human tissues.

The present results may accelerate the discovery of disease-specific biomarkers when comparing with the expression levels of some target glycopeptides in some diseases. However, it should be noted that the present strategy does not pay attention to the quantitation of the glycopeptides, while relative amounts are used.

Response: in the original manuscript, there were indeed many numerical statistics because of the limitation imposed by the number of subjects. In this version, we have revised the quantitation method (**Rev. Fig.1**). Many of the statistics have been changed to cumulative abundance. And the co-identified IDs analysis was based on the median of normalized data in each organ.

In addition, it is also important to note that the results obtained here strongly depend on the glycopeptides captured by the method employed.

Response: we added the statement of “Additionally, the current analytical workflow may not fully capture all possible N-glycans and post-glycosylation modifications (PGMs), particularly those present at

very low abundances”.

The followings are changes should be performed:

In line 42, HIPO may be HUPO.

Response: revised.

In line 96, S3 may be S21.

In line 100, S4-S21 may be S3-S20.

Response: revised.

In Fig. 1i, what does the horizontal axis represent?

Response: the horizontal axis represents the 18 organs from adrenal gland to urinary bladder. The caption has been revised in **Figure 1i**.

References

1. Helm, J. et al. Non-targeted N-glycome profiling reveals multiple layers of organ-specific diversity in mice. *Nat. Commun.* **15**, 16 (2024).
2. Chaubard, J.L. et al. Chemoenzymatic probes for detecting and imaging fucose- α (1-2)-galactose glycan biomarkers. *Journal of the American Chemical Society* **134**, 4489-4492 (2012).
3. Mersha, F.B. et al. Defining the filarial N-glycoproteome by glycosite mapping in the human parasitic nematode *Brugia malayi*. *Sci. Rep.* **13** (2023).
4. Polasky, D.A., Yu, F.C., Teo, G.C. & Nesvizhskii, A.I. Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. *Nat. Methods* **17**, 1125 (2020).
5. Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631-664 (1985).
6. Chen, Z. et al. Recognition of core-fucosylated glycopeptides based on the Y1+Fuc/Y1 ratio in low-energy HCD spectra. *Anal. Chem.* **94**, 17349-17353 (2022).
7. Li, J. et al. Site-specific analysis of core and antenna fucosylation on serum glycoproteins. *Anal. Chem.* **96**, 5741-5745 (2024).
8. Otaki, M. et al. Mouse tissue glycome atlas 2022 highlights inter-organ variation in major N-glycan profiles. *Sci. Rep.* **12** (2022).
9. Fang, P. et al. Ultradeep N-glycoproteome atlas of mouse reveals spatiotemporal signatures of brain aging and neurodegenerative diseases. *Nat. Commun.* **16**, 16 (2025).
10. Xiao, K.J. & Tian, Z.X. GPSeeker enables quantitative structural n-glycoproteomics for site- and structure-specific characterization of differentially expressed n-glycosylation in hepatocellular carcinoma. *J. Proteome Res.* **18**, 2885-2895 (2019).
11. Ou, J.H., Wolfe, S.A., Brodsky, M.H. & Zhu, L.H.J. motifStack for the analysis of transcription factor binding site evolution. *Nat. Methods* **15**, 8-9 (2018).

2/5/2026

RE: Decision on Nature Communications manuscript NCOMMS-25-05194A-Z

Dear Reviewers,

Thank you very much for your valuable comments. We have fully addressed all of them and made the corresponding revision in the manuscript as shown below.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The dataset presented by Hu et al is a valuable dataset containing extensive new information on the glycoprotein forms present in different human tissues. While the manuscript is much improved as compared to the previous version, I still think some over-interpretations of the data are presented, including potential miss annotations of glycopeptides, and reports on structural features that are limitedly supported by the data. What additionally is missing is an evaluation of the variation of the data over the organs of individuals and the comparison of different glycoforms of the same proteins/sites over different organs. Please find my detailed comments below:

1) In the main manuscript, information is missing on the number of biological replicates per organ. I found this information in the supporting information. However, it would be valuable to know from how many samples the numbers are derived that are presented in the first section of

the results. In the discussion 219 runs are mentioned, also here the sample distribution is more relevant.

Response: added as suggested as “74 subjects (2-6 per organ) from 18 healthy human organs”.

2) Related to the previous point: it is really good that multiple biological replicates were included per organ, but it is unclear if these came from different or the same individuals. And, importantly, the variation in glycopeptide abundance was not included in the data presented in the main manuscript. Differences in abundances of glycan traits are presented between organs, but it is unclear if these were statistically significant.

Response: the biological replicates were obtained from different individuals. In this study, the same amount of the enriched N-glycopeptide sample was loaded for each LC-MS/MS run, and the abundance of N-glycopeptide IDs were normalized to the total abundance in each run. The amount of the N-glycopeptides enriched from the 18 tissues are in the range of 1%-5% (w/w) of the extracted proteins with the liver and adrenal gland have the highest percentage.

3) Line 202-204: MS2 data supporting this statement was not included.

Response: N-glycosite N3858 of Q07954 was added to **Supplementary Data 3**.

4) MS2 support for the discrimination between BF and TF is not solid. Firstly, the peaks in Supplementary Figure 31 for H1F1 fragments are very low abundant, from the current representation it is unclear if these can be reliably interpreted. Additionally, antenna fucoses are notoriously known to migrate during HCD fragmentation

(<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/rcm.2509>) making their fragment patterns unreliable in general.

Response: we acknowledge the possibility of fucose migration, and the statement is added in the discussion with the reference cited. We reviewed the fragment ion intensities in the raw file, and the intensity of the C12 ion was 14377.08984, $^{3,4}\text{A}13$ equals 30914.62695, and $^{2,5}\text{A}13$ equals 185830. Both have reached the E5 level, so this matching has been retained. The low level is likely due to the presence of E7 intensity ions in this spectrogram.

5) Figure 1h and overall fragment interpretation: the glycan fragments shown to be indicative for LacNAc repeats are likely (also) derived from the connecting mannose/GlcNAc, and thus not diagnostic for polyLacNAc. Conclusions on the presence of tri-antennarity and polyLacNAc should be more carefully made.

Response: there is indeed the possibility you mentioned. We have revised **Figure 1h** and **1i** to express the fucosylated forms of HexNAc, LacNAc and sialyl LacNAc.

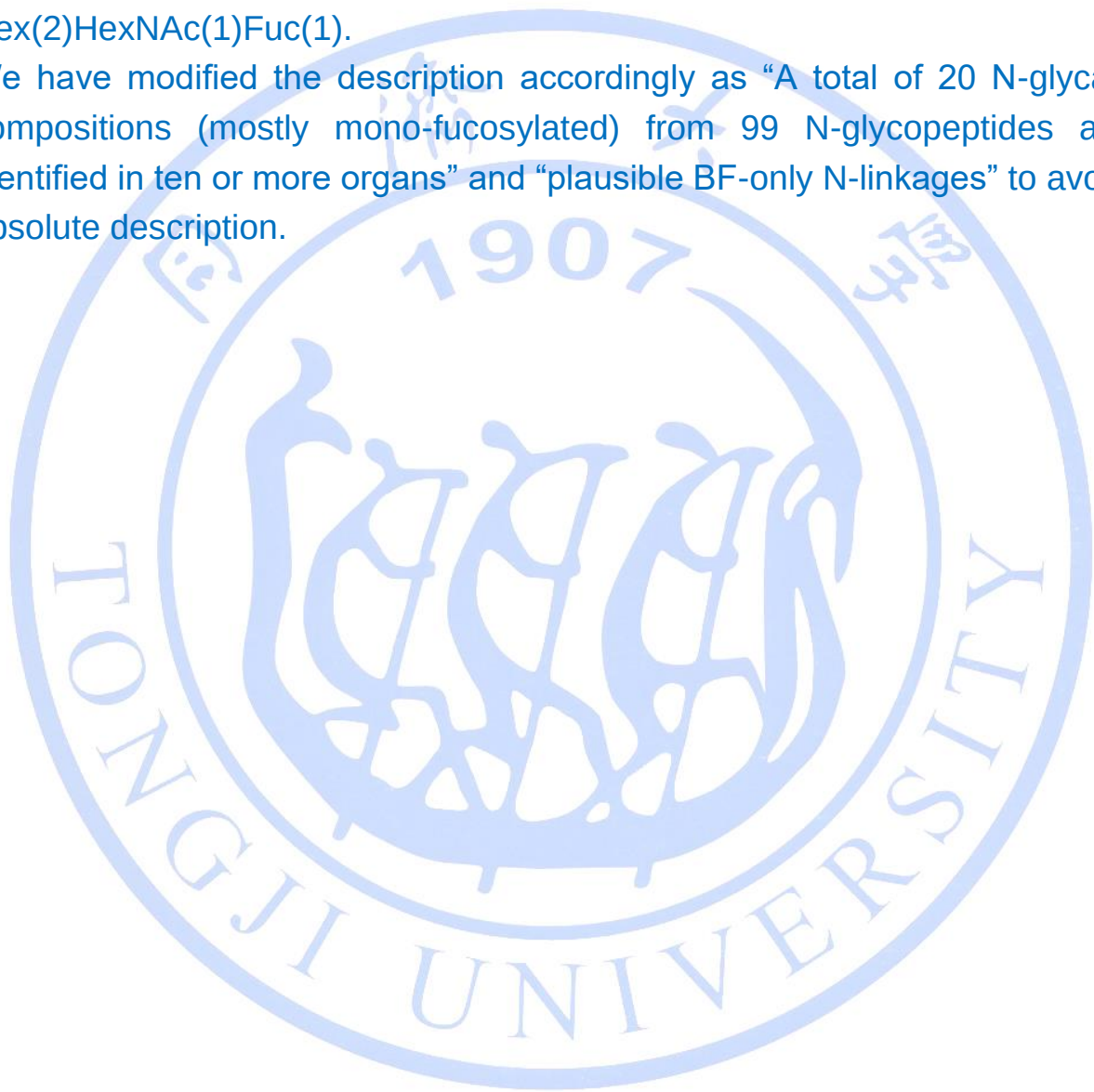
6) Structures shown in Figure 3c, especially in the category BF only, are only partly substantiated. For example, what is the proof of the presence of two antennas on the '6 arm, while the 3' arm is empty? And, in relation to the previous point, how to discriminate between polyLacNAc and multi-antennary glycans?

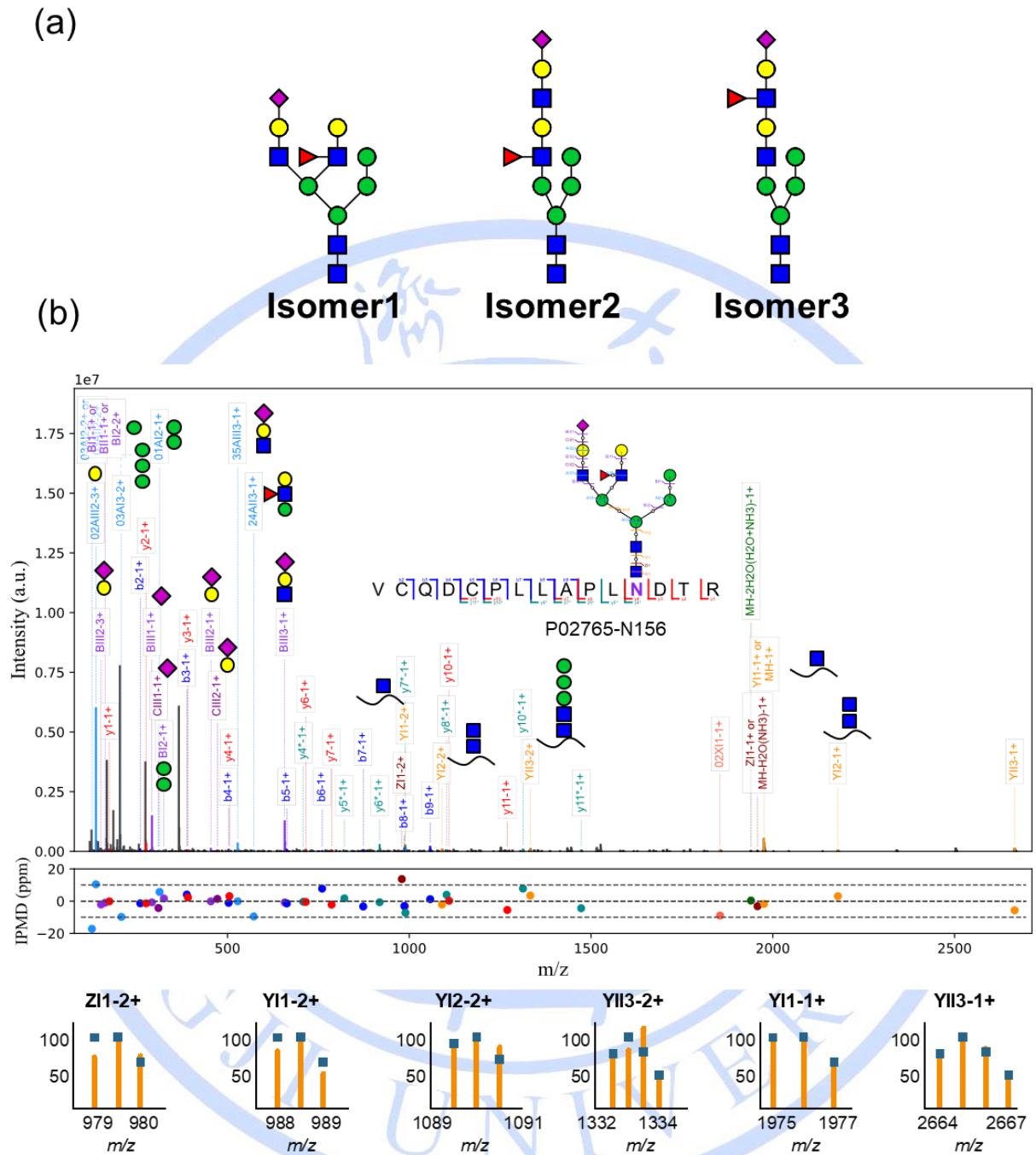
Response: thank you for your comments. there is indeed the possibility you mentioned.

gPeptide uses structure-diagnostic fragment ions to distinguish different sequence isomers of the same monosaccharide composition and gives the score of each sequence structure. Taking this figure (**Ref. Fig. 1**, adrenal gland1-2) as an example, after matching the parent ion, it can be determined that the glycan composition is HexNAc(4)Hex(6)Fuc(1)Neu5Ac(1). BIII2³⁺ and CIII2¹⁺ are the structure-diagnostic fragment ions of the sequence structure motif of Hex-Neu5Ac and likewise BIII3¹⁺ and $^{3,5}\text{A}113^{1+}$ of HexNAc-Hex-Neu5Ac. Based on our experience (as **Supplementary Figure 30**

showed), if it is a sialyl Lewis X structure, there is a considerable probability of the presence of the sialyl Lewis X ion which is 657.2 or some related ions. So, the priority of isomer 3 is relatively lower. In conclusion, the candidate with the highest software score is isomer 1. However, it is indeed impossible to rule out the possibility of additional secondary fragmentation resulting in Hex(2)HexNAc(1)Fuc(1).

We have modified the description accordingly as “A total of 20 N-glycan compositions (mostly mono-fucosylated) from 99 N-glycopeptides are identified in ten or more organs” and “plausible BF-only N-linkages” to avoid absolute description.





Ref. Fig. 1

7) Figure 1: the fragment in column 12 in 1h indicates double galactosylation (alpha-gal?), this is uncommon in human and not reported in the text (same for figure 5d).

Response: this has been previously reported in mice (<https://dx.doi.org/10.1021/acs.analchem.0c02276>); accordingly, we have made relevant changes including **Figure 1** and added **Figure 5d**.

8) Line 345: bi and tri-antennary N-glycans are difficult to discriminate based on HCD fragmentation: the 3rd antenna could also be a LacNAc repeat.

Response: revised.

9) A discussion is missing on the biological relevance of total organ data: most organs consists of many different cell types, all with their own glycosylation features. For example, blood and immune cell derived glycosylation will “contaminate” most of the organ-specific cell type glycosylation.

Response: all the tissue samples were first cut into small pieces on ice, then repeatedly washed with pre-cold 1×PBS several times until no obvious blood residue remained before being homogenized. The methods section has been revised.

10) Figure 7 and related discussion: while not necessarily wrong, it is unclear what is the biological and/or scientific relevance of building a model to discriminate organs based on glycoproteomics data.

Response: the model is built to quantify the distinctness of N-glycoproteomic signatures across different tissues. Our approach mirrors the conceptual framework used in recent landmark studies such as <https://www.nature.com/articles/s41598-022-21758-4> which highlighted inter-organ variation in major N-glycan profiles. By successfully training a model to discriminate organs based solely on N-glycoproteomic data, we

demonstrate that tissue-specific N-glycosylation patterns are conserved and distinct enough to serve as unique molecular fingerprints.

We have updated the discussion of the manuscript to explicitly state this biological relevance and minimize the focus on the model as a mere classification tool.

11) Figure 8b and related discussion: the glycopeptide annotations unique to a single software package should be interpreted with caution, there is a high risk of miss annotations. Importantly, for none of the tools, retention time is present in the output, while this is an important factor in validating glycoforms on the same peptide portion (they should elute close together with small increases between sialic acid

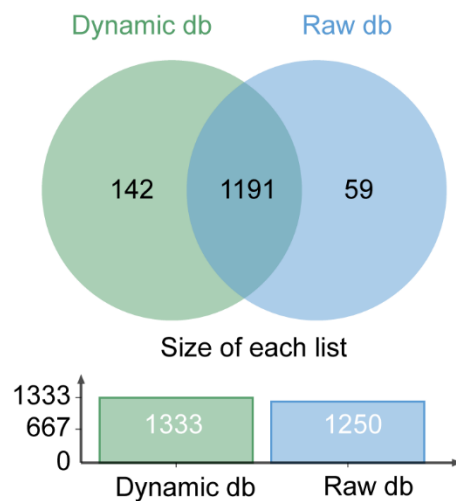
clusters: <https://www.nature.com/articles/s41467-024-50338-5>). High numbers of miss annotations are a risk in this data anyway, due to the possible presence of overlapping masses due to overalkylation, oxidation, ammonia adducts and monosaccharide similarities (<https://www.sciencedirect.com/science/article/pii/S1367593123000108>).

The extend of these potential artifacts in the data should be evaluated.

Response: we appreciate the reviewer's emphasis on the importance of Retention Time (RT) for validation. We would like to clarify the data processing presented in **Figure 8b**. This data represents unique N-glycopeptide identifications grouped by protein, N-glycosite, and monosaccharide composition. To facilitate a clear comparison of software overlap and manage the millions of GPSM entries generated by the four tools, we aggregated the data by retaining only the unique N-glycopeptide IDs with the highest confidence score for each group. We note that the overlap between gPeptide and other software exceeds 50%, and we have updated **Supplementary Data 7** to clearly list the unique IDs for each software and added the retention time of each ID from corresponding spectrum.

We agree that unique IDs found by a single tool require careful interpretation. In order to evaluate the unique IDs, we conducted verification in two aspects.

First, after performing the oxidation modification with Met and Trp, we conducted a new search for adrenal gland2-1. While adding these variable modifications increased the search space, 95.3% IDs we reported were successfully replicated (**Ref. Fig. 2**). The remaining 59 IDs that were not replicated might be related to the update of our software version. It is also possible that the existence of dynamic modifications has led to higher matching scores for GPSM that contain dynamic modifications. Besides, we have performed oxidation modification with Met and Trp, with overalkylation with Met in Fragpipe. There were 131 carrying such dynamic modifications or their combinations in 4,479 GPSMs, only counting 2.9%. Although it is impossible to rule out the possibility of these modifications, it can be estimated that the proportion of such IDs is relatively low.

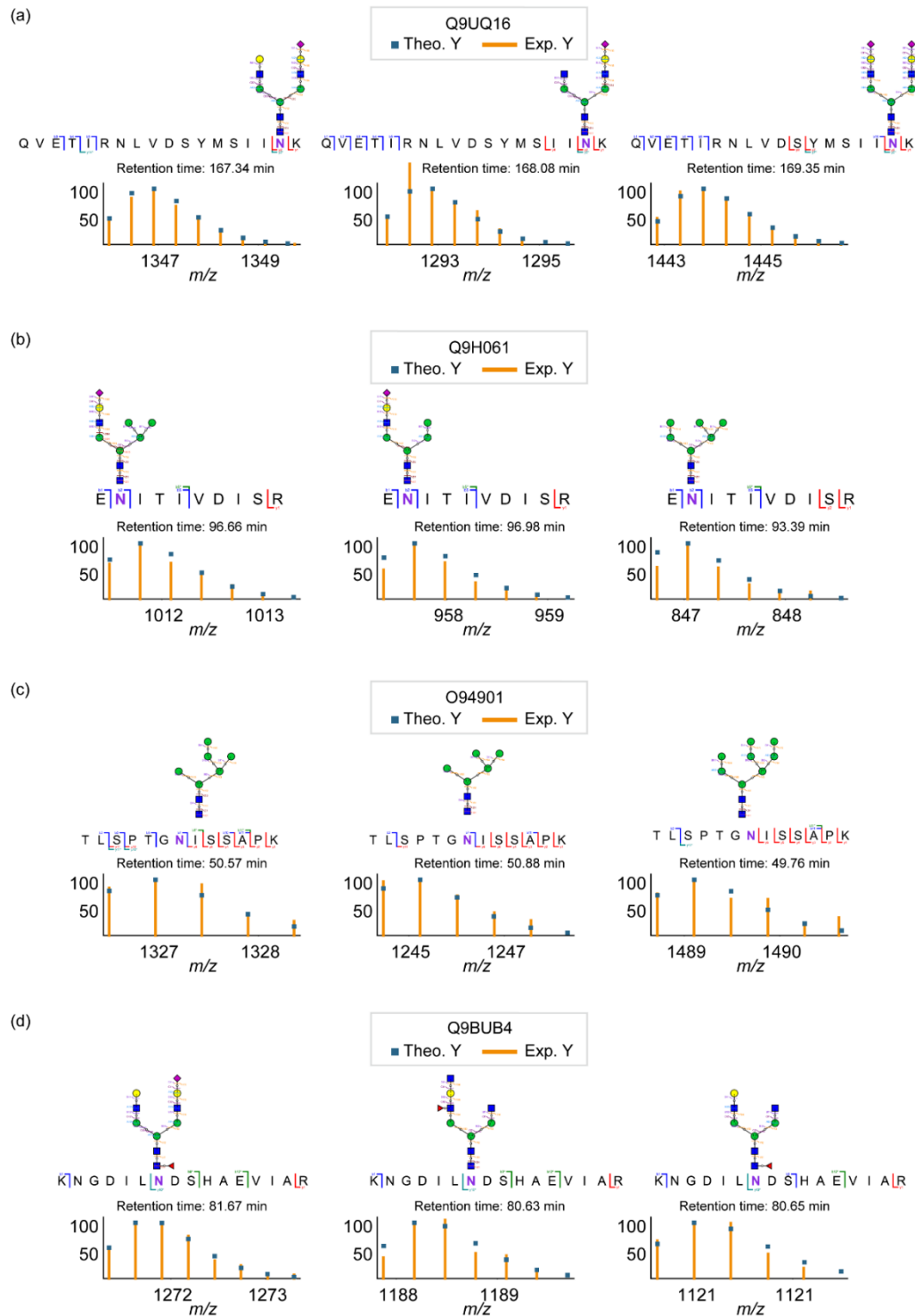


Ref. Fig. 2

Second, taking adrenal gland2-1 as an example, the matches of the top several N-glycoproteins were manually verified in the MS1 and MS2 spectra. Three random N-glycopeptides separately from Q9UQ16, Q9H061, O94901

and Q9BUB4 were showed in **Ref. Fig. 3**. Among these 12 glycopeptides, 11 of them have been identified to contain site-diagnostic ions from MS2. Besides, the retention time of each N-glycopeptide was reasonable. Therefore, we believe that these IDs have some value.

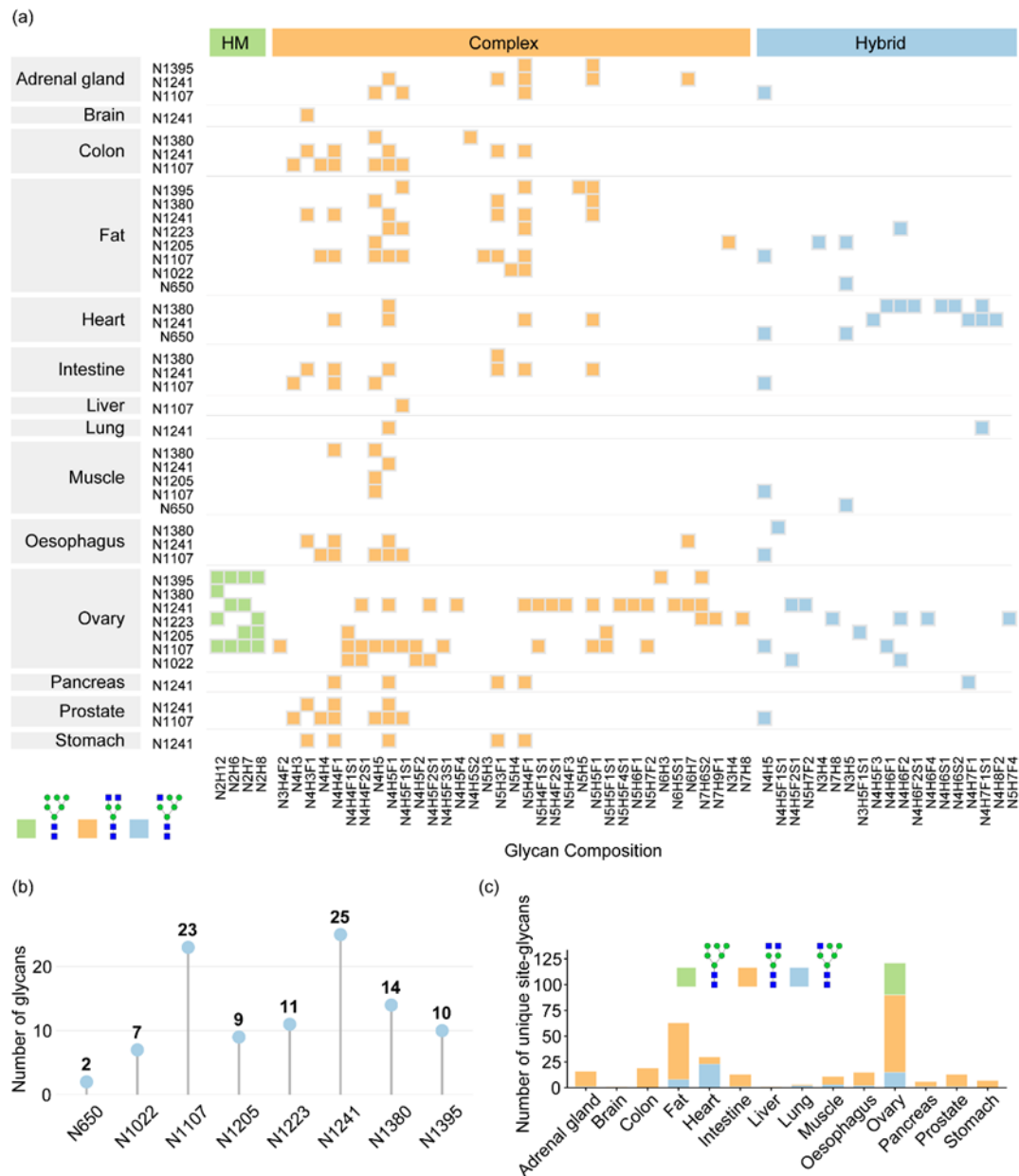
In summary, while we acknowledge that no software can completely eliminate all chemical noise, the combination of strict FDR control and the manual confirmation of parent ions and fragment ions provides credibility for the IDs. We believe that the credibility of the majority of these IDs is relatively high. It is also important to note that each software can produce multiple candidates for a given spectrogram; due to differing scoring mechanisms, each also generates a unique ID. Besides, future version of gPeptide will incorporate more extensive dynamic modification modeling to further automate this validation. We thank the reviewer for these valuable suggestions.



Ref. Fig. 3 Manual verification of unique IDs from adrenal gland2-1.

12) While glycoproteomics data is presented, the main focus of the data analysis is on the glycomics aspect of this data: what glycan features can be found in what organs. – In principle this information could also be obtained with, much simpler, glycomics studies. The data also contains valuable information on the proteins carrying the glycans, this part of the data could be further leveraged by reporting organ-specific glycan profiles for shared glycoproteins and protein-specific glycan profiles within one organ.

Response: in addition to the existing protein-related analysis including the upset analysis, heat map of co-identification analysis, GO and KEGG and the machine learning models, we have added some new analysis of the shared N-glycoproteins. The TOP3 glycoproteins that were identified with the most of IDs were CD91, lumican and laminin subunit gamma-1. Since the identification results of the former two have been discussed separately in the results section, we have included the site-specific analysis of laminin subunit gamma-1 in **Supplementary Figure 36**.



Supplementary Figure 36.

13) Line 357: what is the relevance of the SARS-CoV-2 signalling pathway here? This is just an overlapping protein annotation related to coagulation and provides no further insights into the SARS-CoV-2 disease mechanisms.

Response: what this sentence intends to convey is that NGP cluster 1 is a cluster enriched in fucosylation and sialylation, so the pathways may be related to these specific glycoproteins. It has been deleted.

14) Line 419: “Machine learning models based on all the data also show good results.” Please define what “good results” are.

Response: revised. We have included a detailed comparison.

15) Connection lines between datapoints in Figure 2c should be removed, data is not paired or linearly related.

Response: revised.

16) In the methods, information is missing on the input amount for each tissue sample (eg, sample weight). And whether this sample was taken prior or after (whole organ) homogenization.

Response: for each organ, 50-100 mg tissue was used for homogenization, and 10 µg enriched N-glycopeptides were used for each LC-MS/MS run.

17) Figure 5f: the “linkage” color scale is not well described.

Response: added in the figure.

18) Line 148: “High-mannose N-glycans and the core structure HexNAc(3)Hex(2) are predominant across organs” - HexNAc(3)Hex(2) is not the N-glycan core structure, that would be HexNAc(2)Hex(3).

Response: revised.

19) Line 163: the use of the term “tri-antennary high-mannose” is confusing as tri-antennary is usually used for complex type N-linked glycans.

Response: revised.

20) For example line 191: what does
01Y41Y41M(31M21M21M31G31G21G)61M(31M21M)61M refer to?

Response: revised to HexNAc(2)Hex(11).

21) The manuscript should be improved language-wise.

Response: The overall manuscript has been revised.

Reviewer #2 (Remarks to the Author):

In my view, the manuscript in its original compact form was already suitable for publication. The authors' corrections have not diminished its quality, and most of their revisions—with a few exceptions related to nomenclature—appropriately address the requests of the other reviewers.

While the applied corrections have made the findings more substantiated, the inclusion of excessive detail has, in my opinion, reduced the article's accessibility for the broad readership of Nature Communications. I am therefore uncertain whether the extended discussion of these issues, beyond the necessary technical corrections, has ultimately improved the paper.

During the initial review, I could follow the meaning of most figures and SI figures, though sometimes a glycoinformatic context was required (knowledge, which a general reader may lack). I have not re-examined all SI figures in detail at this stage, but the authors' responses to the reviewers seem reasonable.

I am fully satisfied with the revisions made in response to the other

reviewers' remarks. Although I find some additions redundant, I have not detected any errors.

Response: thank you for your comments.

Reviewer #3 (Remarks to the Author):

The present study provides a large set of N-glycoproteome database of 18 human organs/tissues composed of intact glycopeptides on 7,782 N-glycosites of 5,078 glycoproteins. By facilitating a systemic comparison between the different analytical platforms, authors demonstrated that pPeptide has superior performance in the in-depth MS/MS data analysis. However, the results by comparative analysis also uncovered that the complementarity of current N-glycoproteomics tools is approximately 21.1% of N-glycopeptides. This human organ N-glycoproteome atlas may become a basic resource for glycobiology community. In addition, it is important to note that quantitative insight into the organ relevant glycopeptides is crucial for further investigation of disease specific N-glycosylation signatures.

Response: thank you for your comments.

Detailed response to reviewers:

I was brought into this review process to evaluate how Hu et al. addressed the responses from Reviewer #1 in the previous round of revision.

The data presented by Hu et al. is of great interest to the glycoproteomics community, particularly on how it can contrast the diversity of glycosylation patterns that can be observed in different human tissues. I believe the data is convincing for the classification of tissues based on glycosylation profiles and this work could catalyze further interest in the study the importance of finite glycosylation profiles in tissue differentiation. The work is original and the monumental effort implemented in its analysis is recognized.

However, before I can recommend publication, several major issues must be seriously addressed regarding the validity of the mass spectrometry data interpretation. This is particularly important given the broad readership that this publication may reach, including readers who may not be fully aware of the limitations of fragment ion spectra in accurately assigning glycan compositions, structures, and glycosylation sites.

I share the same concerns of Reviewer #1 with respect to the overinterpretation of the glycan structures for some compositions. Additionally, I have concerns with the information provided for novel glycosylation sites and the lacking standardization in which peptide identification results are presented.

This work aims to serve as a glycoprotein atlas encompassing multiple human tissues. It represents an important source of mass spectrometry data for future research, and the reported glycosite and glycan composition annotations appear reasonably reliable. However, I am not convinced by the glycan structure assignments claimed by the authors based on the data presented, nor by their response to the concerns raised by Reviewer #1. Their reply suggests, at best, a choice to ignore the limitations of HCD spectra for resolving glycan fine structures, which undermines the confidence in the structural interpretations provided. Although highly desirable, defining glycan structure is not trivial and the certainty used to report these structures should be toned down to avoid wrong interpretation in future studies from this human tissue glycoprotein atlas.

Below are specific evaluations to the comments suggested by the editors for further evaluation:

Comment #3:

Reviewer 1:

Line 202-204: MS2 data supporting this statement was not included.

Author's response:

N-glycosite N3858 of Q07954 was added to Supplementary Data 3." Evaluation:

The authors have included a table with annotated intensities of arbitrarily selected ions. However, the modified peptide sequence, file, and scan number are still missing. Without this information Supplementary Data 3 holds no value since it cannot be independently confirmed. Additionally, a legend of the columns should be provided.

Response:

Thanks for your comments. The raw file, modified peptide sequence, and scan number are added to Supplementary Data 3 as a USI format. Besides, the column legend has been added to it, too.

Comment #4:

Reviewer #1:

MS2 support for the discrimination between BF and TF is not solid. Firstly, the peaks in Supplementary Figure 31 for H1F1 fragments are very low abundant, from the current representation it is unclear if these can be reliably interpreted. Additionally, antenna fucoses are notoriously known to migrate during HCD fragmentation

(<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/rcm.2509>) making their fragment patterns unreliable in general.

Author's Response:

we acknowledge the possibility of fucose migration, and the statement is added in the discussion with the reference cited. We reviewed the fragment ion intensities in the raw file, and the intensity of the C12 ion was 14377.08984, 3,4A13 equals 30914.62695, and 2,5A13 equals 185830. Both have reached the E5 level, so this matching has been retained. The low level is likely due to the presence of E7 intensity ions in this spectrum.

Evaluation:

The authors have added a statement in the discussion that points toward caution in the distinction between branch fucosylation (BF) and terminal fucosylation (TF). However, this does not address the concerns of Reviewer #1 as the glycan classification remains the same during the manuscript. Here I agree with Reviewer #1 where distinction between BF and TF cannot be easily claimed. A safer approach would be to merge BF and TF glycan classifications and adapt all subsequent analysis. If this is not required by the editors, then at best a strong compelling argument should be introduced prior to the results presentation so readers are aware of the ambiguity.

With regards to Supplementary Figure 31, reporting based on absolute intensities is not reliable. Orbitrap data is heavily denoised during data acquisition which prevents independent evaluation of S/N and the intensity thresholds vary between instruments. It is more useful to evaluate the detection of lower abundant isotopes. In this regard, the display of the Y-ions is a useful confirmation

of the isotopic envelope. However, this is missing for the B-ion H1F1 that was questioned by Reviewer #1. Why are the B-ions not zoomed in like the Y-ions? Additionally, the authors should provide file and scan number of the annotated spectra for external revision. This is recommended for all displayed spectra.

Response:

Thank you for your comments. We fully agree with your viewpoint. In Figure 3, the data and linkages of BF and TF have been combined and presented together. The description of the corresponding result section has also been revised.

Regarding the cross-validation in the Supplemental Figures, the description has been changed to "Fucose-HexNAc-containing glycans in BF-only glycans". This is used to express the verification of the part of the glycopeptides that contains HexNAc-fucose rather than Galactose-Fucose.

Besides, the isotope peaks of BII2¹⁺, BII3¹⁺, and BII1¹⁺ were not shown because theoretically they do not have isotope peaks; there is only a single peak corresponding to m/z values of 163.0601, 366.139472, and 528.192296 respectively. We examined the matching accuracy and indeed all the matches fell within 5 ppm. The same applies to many small glycan fragment ions.

Comment #5:

Reviewer #1:

Figure 1h and overall fragment interpretation: the glycan fragments shown to be indicative for LacNAc repeats are likely (also) derived from the connecting mannose/GlcNAc, and thus not diagnostic for polyLacNAc. Conclusions on the presence of tri-antennarity and polyLacNAc should be more carefully made.

Author's response:

there is indeed the possibility you mentioned. We have revised Figure 1h and 1i to express the fucosylated forms of HexNAc, LacNAc and sialyl LacNAc.

Evaluation:

The response from the authors is not related to the concern expressed by Reviewer #1, which addressed the ambiguity between annotating an additional antenna or a LacNAc repeat. No comments on the addition of fucosylated forms. Concerns of Reviewer #1 were not addressed and goes in line with the ambiguous interpretation of glycan structure classification pointed out earlier.

Response:

Thanks for your detailed comments. We agree that fragment ions from the mannose core and the branching GlcNAc can overlap with those from polyLacNAc extensions, making it difficult to definitively distinguish tri-antennary structures from LacNAc repeats. In the previous version of the manuscript, we have revised the glycan icons in Figure 1h and 1i to a format that explicitly shows the observed

compositions (e.g., “HexNAc(1)Fuc(1)” or “HexNAc(1)Fuc(1)+Hex(1)HexNAc(1)”) rather than implying a specific connectivity. This representation conveys the possibility of both tri-antennarity and polyLacNAc, and we have accordingly softened the corresponding descriptions in the text to adopt a more cautious interpretation. And the Ref. Fig. 1 before is provided to explain the identification and scoring of tri-antennarity and polyLacNAc structure for similar candidates by gPeptide in certain cases.

Comment #6

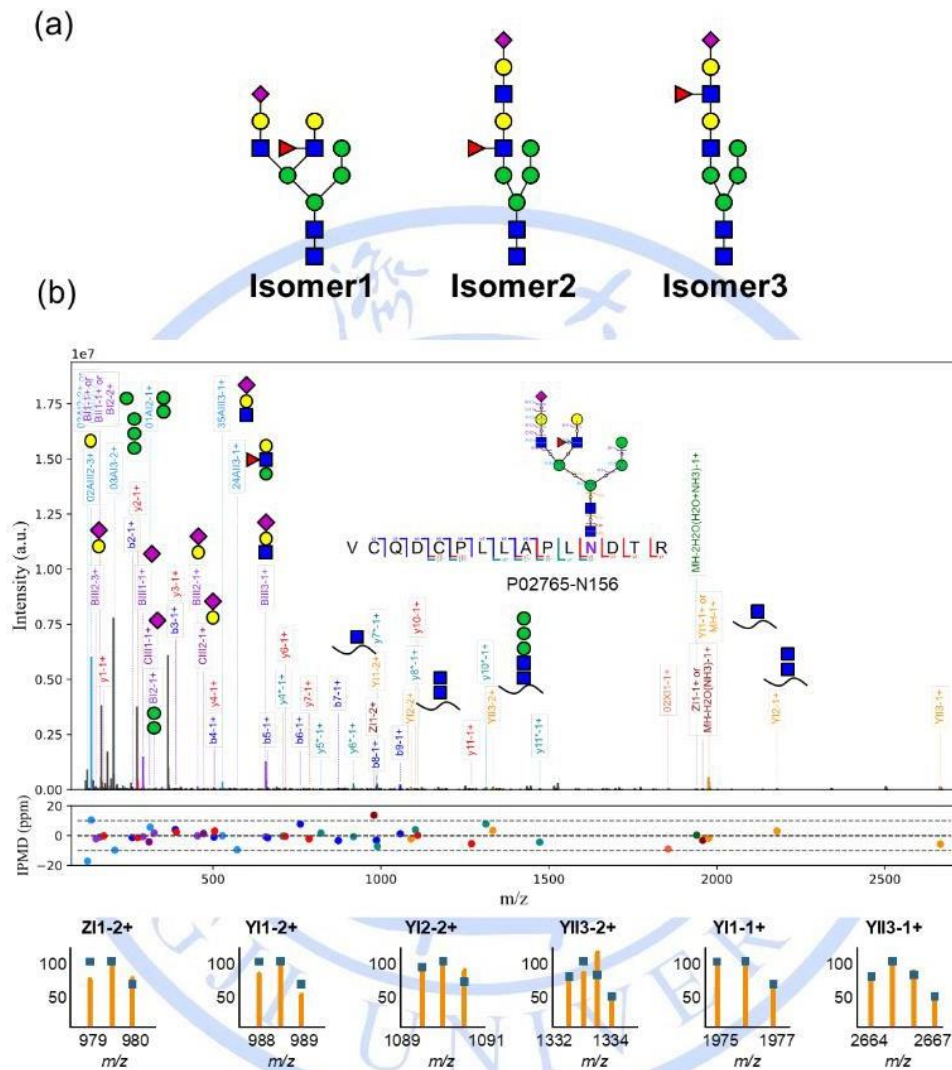
Reviewer #1:

Structures shown in Figure 3c, especially in the category BF only, are only partly substantiated. For example, what is the proof of the presence of two antennas on the ‘6 arm, while the 3’ arm is empty? And, in relation to the previous point, how to discriminate between polyLacNAc and multi-antennary glycans?

Author’s response:

thank you for your comments. there is indeed the possibility you mentioned. gPeptide uses structure-diagnostic fragment ions to distinguish different sequence isomers of the same monosaccharide composition and gives the score of each sequence structure. Taking this figure (Ref. Fig. 1, adrenal gland1-2) as an example, after matching the parent ion, it can be determined that the glycan composition is HexNAc(4)Hex(6)Fuc(1)Neu5Ac(1). BIII23+ and CIII21+ are the structure-diagnostic fragment ions of the sequence structure motif of Hex- Neu5Ac and likewise BIII31+ and 3,5AIII31+ of HexNAc-Hex-Neu5Ac. Based on our experience (as Supplementary Figure 30 shown), if it is a sialyl Lewis X structure, there is a considerable probability of the presence of the sialyl Lewis X ion which is 657.2 or some related ions. So, the priority of isomer 3 is relatively lower. In conclusion, the candidate with the highest software score is isomer 1. However, it is indeed impossible to rule out the possibility of additional secondary fragmentation resulting in Hex(2)HexNAc(1)Fuc(1).

We have modified the description accordingly as “A total of 20 N-glycan compositions (mostly mono-fucosylated) from 99 N-glycopeptides are identified in ten or more organs” and “plausible BF-only N-linkages” to avoid absolute description.



Ref. Fig. 1

Evaluation:

Response from authors don't seem to address the questions about antenna localization in '6 or 3' arm or how to discriminate between polyLacNAc and multi-antennary features in glycans.

It would be good to include in Supplementary Data 7, for gPeptide and Glyco-Decipher, a column that indicates that structure is glycan structure is ambiguous and the competing structures (at least the number) so readers are aware of the ambiguity of the structure assignment.

Response:

We appreciate the reviewer's questions, which allow us to clarify the limitations and strengths of the structural assignments in our workflow.

1) Regarding antenna localization on the 3' vs. 6' arm, we completely agree with the reviewer that unambiguously distinguishing the localization of antennas between the 3' and 6' branches is fundamentally challenging using standard bottom-up LC-MS/MS without specific isotopic labeling or sequential exoglycosidase digestion. We have ensured our manuscript reflects this limitation

in discussion part. It has been revised as “the output of 3’ and 6’ glycans is based on the matching conditions in the database, which remain challenging to differentiate using the current LC-MS/MS-based pipeline”.

2) Regarding the discrimination between polyLacNAc and multi-antennary glycans, the figure was to explain that single cleavage from Isomer1 could produce $^{2,5}AII3^{1+}$ (Hex(2)HexNAc(1)Fuc(1)). Because secondary fragmentation events are statistically less probable the scoring algorithm will assign a higher probability score to Isomer 1. However, it is indeed impossible to completely rule out Isomer2.

3) At last, in the 18 identification lists of each organ, we have added new columns (isomer count and other isomers) for the gPeptide results. This addition includes the number of competing structures considered by the software, ensuring readers are fully aware of the structural ambiguity and the inherent confidence level of the top-ranked assignment. For the IDs in the "Structure-diagnostic ions" column that are not NA, it indicates that structural-diagnostic ions have been found in the MS2. We consider such IDs to have a higher credibility in the "g-linkage." column, and they are less likely to represent other isomers. It has been uploaded to GitHub (<https://github.com/HuXiaoyu-TJU/A-comprehensive-landscape-of-human-organ-N-glycoproteome.git>).

Besides, the same information was also added to Supplementary Data 7. However, this does not mean that there are many uncertain linkages in the output results of gPeptide in Supplementary Data 7. It is simply because Supplementary Data 7 only retains one ID at the Glycoprotein-glycosite-composition level, so the isomers listed in the "other isomers" column only represent the isomer corresponding to this one identification of one row, rather than all the isomers that were not identified throughout the entire study.

We sincerely appreciate your valuable comments, which helps improve the quality of our manuscript.

Comment #7:

Reviewer #1:

Figure 1: the fragment in column 12 in 1h indicates double galactosylation (alpha-gal?), this is uncommon in human and not reported in the text (same for figure 5d).

Author’s response:

this has been previously reported in mice (<https://dx.doi.org/10.1021/acs.analchem.0c02276>); accordingly, we have made relevant changes including Figure 1 and added Figure 5d.

Evaluation:

The response of the authors is insufficient to address the concerns of Reviewer #1. Proof of detection in other mammals does not support its detection in humans. The citation provided specifically alludes that the alpha-gal N-terminal moiety is not presented in humans and is recognized as an antigen that induces immunogenic response. If the alpha-galactosyl moiety is indeed detect across several tissues, it should be addressed more in detail, present convincing annotated spectra, and at least confirm if the strategy (ratio $I_{m/z528}/I_{m/z366} > 0.3$) used in <https://dx.doi.org/10.1021/acs.analchem.0c02276> for structural classification of this epitope.

However, the figures presented (1h and 5d) at this stage of the revision don't display the alpha-gal moiety. Was this removed from the study? If so, this should at least be addressed in detail to the reviewers.

Response:

Yes, we have removed the di-gal terminal IDs in the results, updated the figures and Supplementary Table 2 before. Since the proportion is very low, there have been no significant changes in the figures. Thank you for your kind suggestion.

Comment #8:

Reviewer #1

Line 345: bi and tri-antennary N-glycans are difficult to discriminate based on HCD fragmentation: the 3rd antenna could also be a LacNAc repeat.

Author's response:

revised.

Evaluation:

Authors' response to Reviewer #1 are insufficient to assess what was "revised". Based on previous responses, the author's are ignoring the limitations of CID fragmentation patterns in determining fine glycan structure. This also applies to glycosylation site localization in glycopeptides carrying two tentative glycosylation sites. See major comment at the end.

Response:

Thanks for your comments. In the previous version, we removed the description regarding antennas in NGP cluster 1, and softened many descriptions of glycan antennas.

After re-analyzing the data in Figure 6, we found that NGP clusters 2 and 3 correspond to glycopeptides mainly carrying N2H6 to N2H10, which are mainly tri-antennary glycans. However, there remains some uncertainty regarding NGP clusters 1 and 4. This annotation has been removed, and glycan type has been added instead.

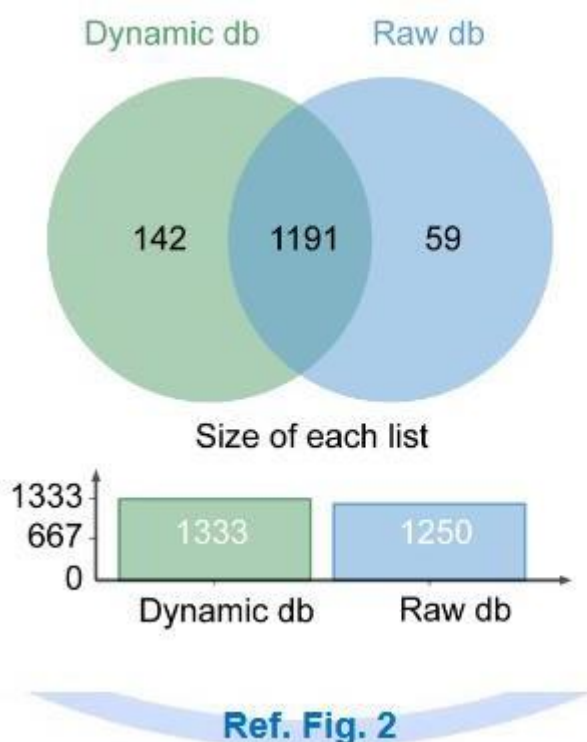
Comment #11:

Figure 8b and related discussion: the glycopeptide annotations unique to a single software package should be interpreted with caution, there is a high risk of miss annotations. Importantly, for none of the tools, retention time is present in the output, while this is an important factor in validating glycoforms on the same peptide portion (they should elute close together with small increases between sialic acid clusters: <https://www.nature.com/articles/s41467-024-50338-5>). High numbers of miss annotations are a risk in this data anyway, due to the possible presence of overlapping masses due to overalkylation, oxidation, ammonia adducts and monosaccharide similarities (<https://www.sciencedirect.com/science/article/pii/S1367593123000108>). The extend of these potential artifacts in the data should be evaluated.

Author's Response:

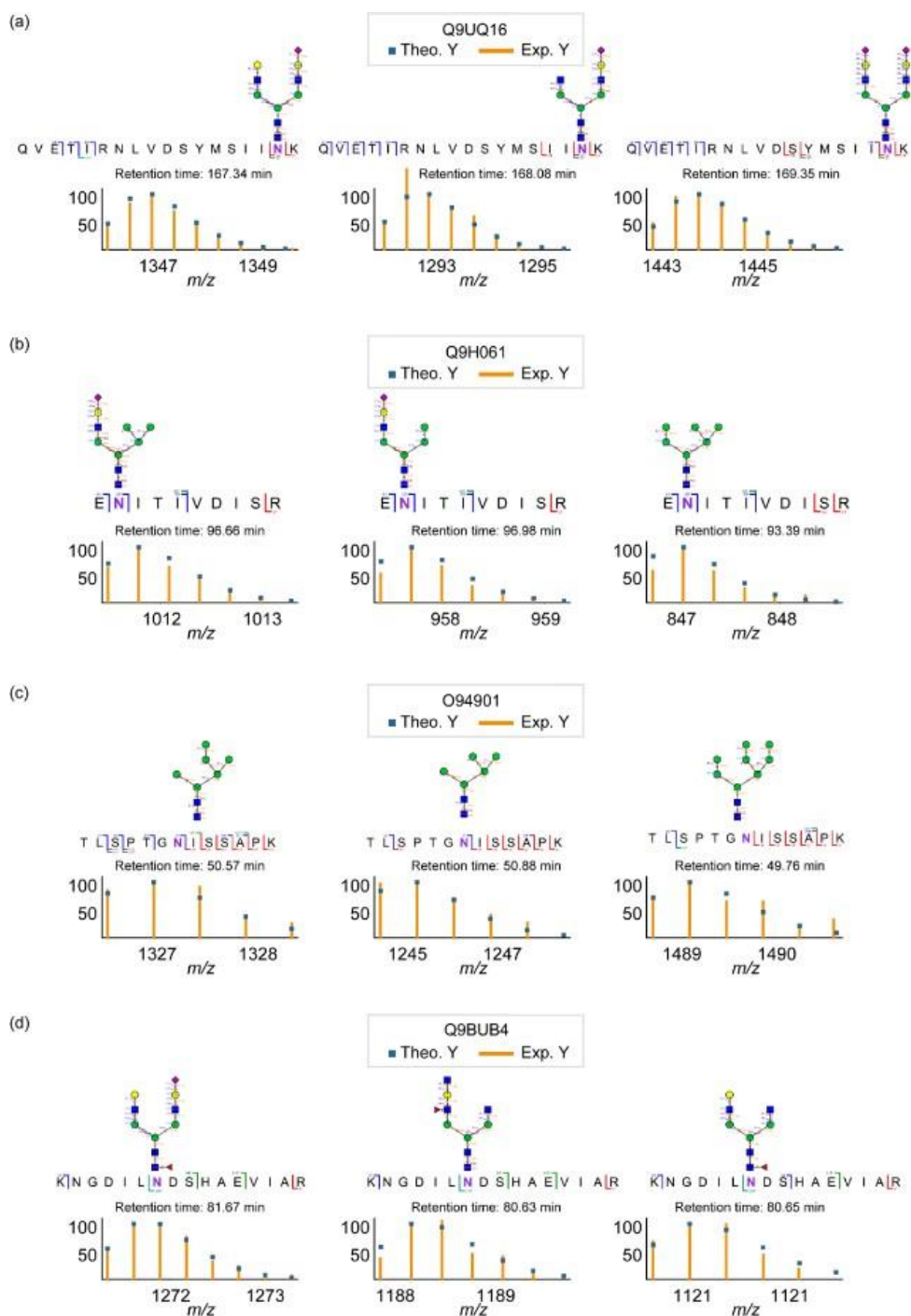
we appreciate the reviewer's emphasis on the importance of Retention Time (RT) for validation. We would like to clarify the data processing presented in Figure 8b. This data represents unique N-glycopeptide identifications grouped by protein, N-glycosite, and monosaccharide composition. To facilitate a clear comparison of software overlap and manage the millions of GPSM entries generated by the four tools, we aggregated the data by retaining only the unique N-glycopeptide IDs with the highest confidence score for each group. We note that the overlap between gPeptide and other software exceeds 50%, and we have updated Supplementary Data 7 to clearly list the unique IDs for each software and added the retention time of each ID from corresponding spectrum.

We agree that unique IDs found by a single tool require careful interpretation. In order to evaluate the unique IDs, we conducted verification in two aspects. First, after performing the oxidation modification with Met and Trp, we conducted a new search for adrenal gland2-1. While adding these variable modifications increased the search space, 95.3% IDs we reported were successfully replicated (Ref. Fig. 2). The remaining 59 IDs that were not replicated might be related to the update of our software version. It is also possible that the existence of dynamic modifications has led to higher matching scores for GPSMs that contain dynamic modifications. Besides, we have performed oxidation modification with Met and Trp, with overalkylation with Met in Fragpipe. There were 131 carrying such dynamic modifications or their combinations in 4,479 GPSMs, only counting 2.9%. Although it is impossible to rule out the possibility of these modifications, it can be estimated that the proportion of such IDs is relatively low.



Second, taking adrenal gland2-1 as an example, the matches of the top several N-glycoproteins were manually verified in the MS1 and MS2 spectra. Three random N-glycopeptides separately from Q9UQ16, Q9H061, O94901 and Q9BUB4 were showed in Ref. Fig. 3. Among these 12 glycopeptides, 11 of them have been identified to contain site-diagnostic ions from MS2. Besides, the retention time of each N-glycopeptide was reasonable. Therefore, we believe that these IDs have some value.

In summary, while we acknowledge that no software can completely eliminate all chemical noise, the combination of strict FDR control and the manual confirmation of parent ions and fragment ions provides credibility for the IDs. We believe that the credibility of the majority of these IDs is relatively high. It is also important to note that each software can produce multiple candidates for a given spectrum; due to differing scoring mechanisms, each also generates a unique ID. Besides, future version of gPeptide will incorporate more extensive dynamic modification modeling to further automate this validation. We thank the reviewer for these valuable suggestions.



Ref. Fig. 3 Manual verification of unique IDs from adrenal gland2-1.

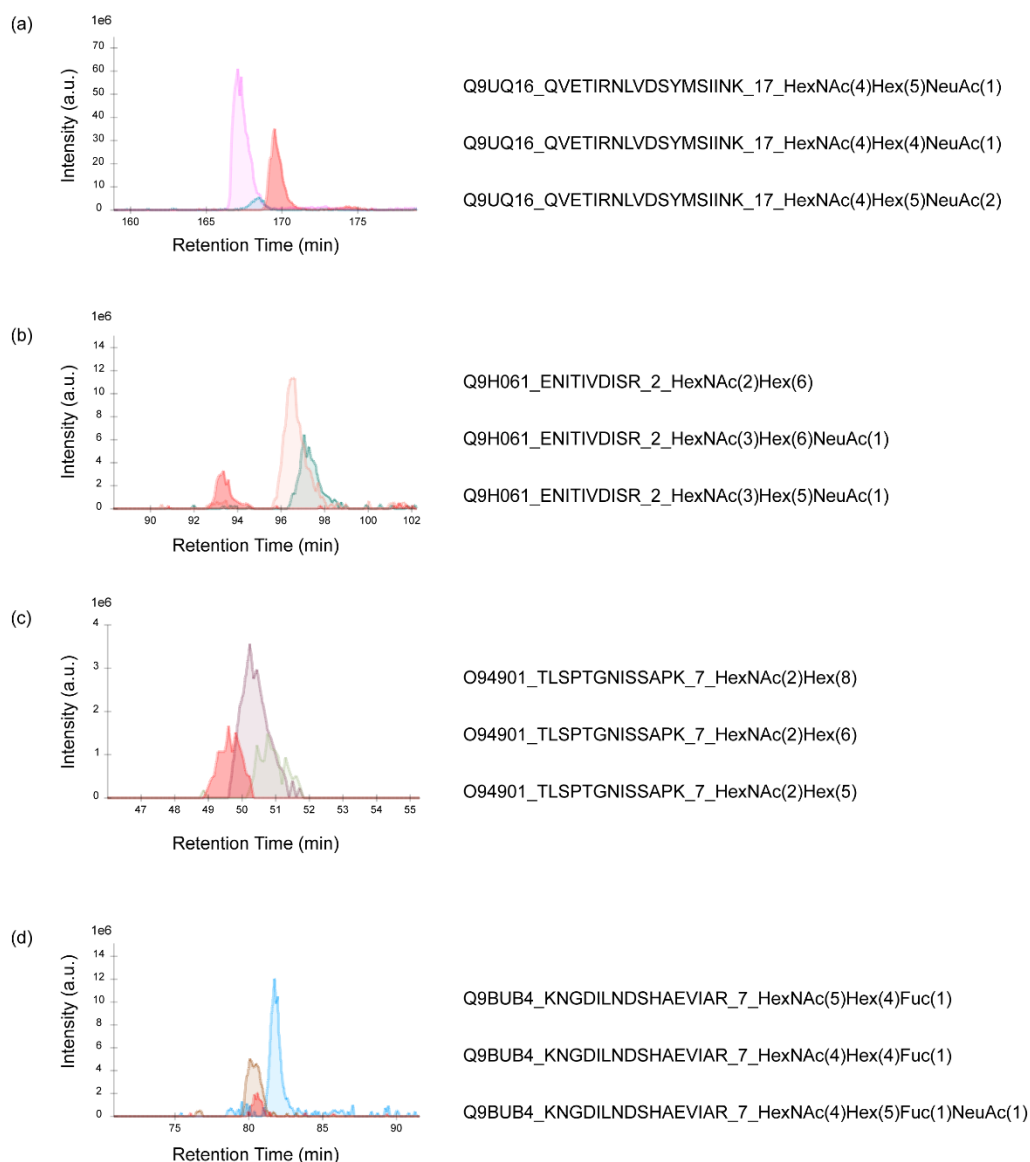
Evaluation:

The authors have revised the supplementary data provided for all glycopeptide spectrum matches where retention time, file, and scan number are available and can be used to evaluate externally. This information would also be easily accessible in the EICs generated with Skyline. However, the authors have not provided the Skyline analysis documents that they claimed to have created.

Response:

Thank you for your comments. After reorganizing these precursor ions, we input them into Skyline. The EIC of each parent ions were correctly extracted. This .zip file has also been uploaded to panoramaweb together with all the skyline files of each subject

(<https://panoramaweb.org/Tongji%20University%20-%20Tian-BAMS/N-glycoproteome%20of%2018%20human%20organs/project-begin.view>).



Ref.Fig.1

The secondary analysis performed with Met and Trp oxidation on “adrenal gland2-1” is interesting and provides slightly more confidence in the results, but it does not address all other modifications pointed out by Reviewer #1 that give rise to false positive identifications.

Response:

We thank the reviewer for the evaluation. Our current version of gPeptide does not support an exhaustive dynamic modification search for all possible artifacts (such as comprehensive overalkylation on N-terminal/Met, or ammonia adducts) simultaneously, as this would cause an unmanageable explosion in the search space and compromise FDR estimation. Our previous test based FragPipe was intended as a proxy to estimate the frequency of oxidation modification on Met and Trp and overalkylation on Met, but we agree it does not completely rule out the presence of

specific mass overlaps in the gPeptide output.

As the reviewer correctly mentioned, retention time is a critical factor in validating glycoforms on the same peptide backbone. As demonstrated in the newly provided Skyline EICs in Ref. Fig. 1, our annotated glycoforms exhibit logical and highly predictable elution profiles.

We have added in the discussion of the revised manuscript acknowledging this limitation of the software: “The unique IDs from each software should be interpreted with caution, as not all possibilities have been taken into account, such as excessive acetylation of the N-terminal/methionine, amide adducts, and so on”. We will continue to enhance the performance of the software in the future so that we can achieve these cross-validation.

The use of multiple search engines to identify glycopeptides is of high interest and the endeavor is appreciated. By using multiple search strategies, the false negative rate is most likely reduced as more true IDs could be captured, but the rate is not assessed. Additionally, just using multiple search engines is no proof of “reliability of the IDs output by this study” as claimed by the authors. Ideally, the output of all tools would be used to recalculate the confidence of identification of each PSM as implemented in other proteomics applications (DOI: 10.1038/nbt.3109). However, this approach has not been implemented in this study as far as I understand. Alternatively, the conclusions could be based on the most reliable IDs which all search engines identified as the same conclusion (i.e. glycopeptide ID) was reached from the same spectrum after testing with multiple approaches. However, this approach has not been used either since the larger number of glycopeptides reported correspond to the output of all search engines. The proportion of true IDs are unknown since not all were manually validated and the low overlap of “21.1% of intact N-glycopeptides being consistently identified across all platforms” raises more questions than reassurance of the unique IDs. That said, presenting the results from all search engines is valuable and could be used in future studies to perform the aforementioned FDR-rescoring strategy. Nonetheless, the tone of the results and discussion with regards to the multi-software approach should be adapted recognizing the concerns presented by Reviewer #1 and here.

Response:

We sincerely appreciate your comments, which are of great significance to our work.

1) First, the manuscript has been revised in the results and discussion. We removed the description related to cross-validation and replaced it with “to capture a broader landscape of the human N-glycoproteome” and so on.

2) Besides, we have updated the credibility level of the ID in the Supplementary Data 7, based on how many softwares have identified them. The IDs that have been verified by all four software packages are the most trustworthy. This will improve the readability of the results. The corresponding code has also been updated to GitHub (<https://github.com/HuXiaoyu-TJU/A-comprehensive-landscape-of-human-organ-N-glycoproteome.git>) simultaneously.

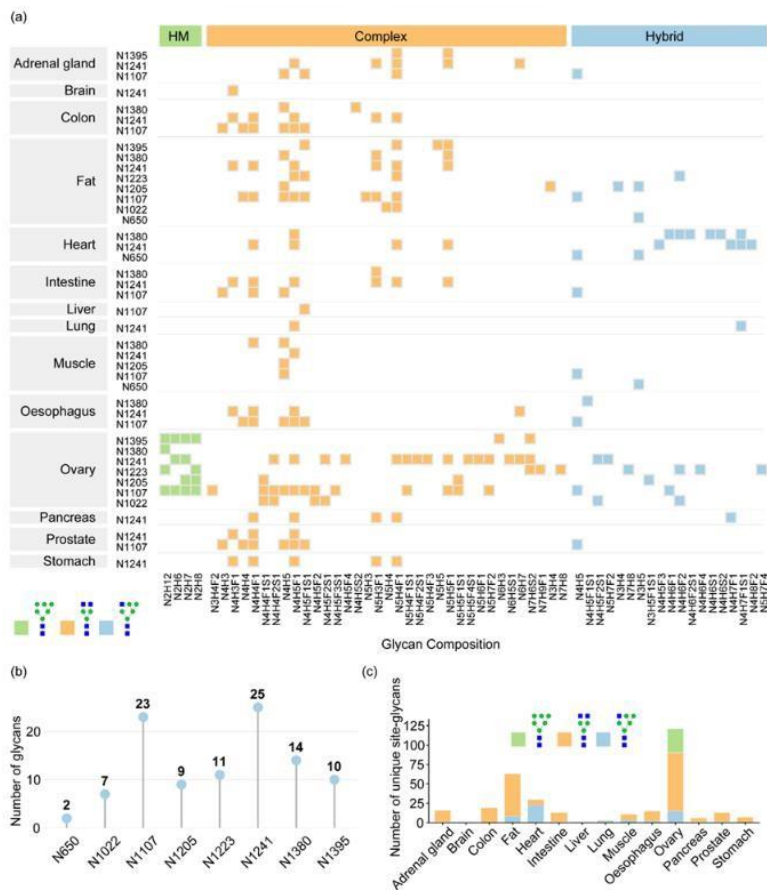
Comment #12:

Reviewer #1:

While glycoproteomics data is presented, the main focus of the data analysis is on the glycomics aspect of this data: what glycan features can be found in what organs. – In principle this information could also be obtained with, much simpler, glycomics studies. The data also contain valuable information on the proteins carrying the glycans, this part of the data could be further leveraged by reporting organ-specific glycan profiles for shared glycoproteins and protein-specific glycan profiles within one organ.

Author's Response:

in addition to the existing protein-related analysis including the upset analysis, heat map of co-identification analysis, GO and KEGG and the machine learning models, we have added some new analysis of the shared N-glycoproteins. The TOP3 glycoproteins that were identified with the most of IDs were CD91, lumican and laminin subunit gamma-1. Since the identification results of the former two have been discussed separately in the results section, we have included the site-specific analysis of laminin subunit gamma-1 in Supplementary Figure 36.



Supplementary Figure 36.

Evaluation:

The authors have taken seriously the recommendation from Reviewer #1 and produced an excellent representation of glycosylation site diversity across sites in multiple tissues.

I would suggest promoting Supplementary Figure 36 to a main figure in the manuscript and expand the discussion around it if possible around glycosylation-related enzymes that might give a specific tissue the observed glycosylation profile. This example of laminin subunit gamma-1 presents a visual summary of the richness of the data contained in this study presented and it would be a shame its hidden within the supplementary data.

Response:

Thanks for your comments. We have added it to the manuscript as a main figure. Additionally, in order to maintain the overall consistency of the chart, a minor modification was made to Figure 6c. Instead of counting the glycans at the linkage level, it now counts the glycans at the compositional level.

Besides, we reviewed the proteins that had been widely identified across 18 different organs. They are indeed structural components of the extracellular matrix rather than enzymes. However, we agree that their diverse glycosylation profiles are driven by glycosylation-related enzymes expressed in a tissue-specific

manner. Following the reviewer's advice, we expanded the discussion to bridge the observed glycan patterns with the expression of key enzymes (e.g., MGAT series) reported in these tissues.

Major comments:

Comment #1: ambiguity of glycosylation sites with HCD data for glycopeptides with multiple glycosylation sites

One of the examples chosen for presenting a novel glycosylation site (Q07954 - N1767) with the N-X-C motif was an unfortunate choice since this peptide has two glycosylation sites.

I tried to visualize this spectrum externally, but I could not find any glycopeptide spectrum match for LYWISSGNHTINR_12_Hex(9)HexNAc(2) in Supplementary Data 7. Why is this PSM not reported by any of the search engines used, but still annotated in the main manuscript?

Response:

These two glycopeptides were identified from liver1, TR2. And, the data of each organ have been updated on GitHub (<https://github.com/HuXiaoyu-TJU/A-comprehensive-landscape-of-human-organ-N-glycoproteome.git>). There is one tsv file for each organ. The NGPs we shown in Figure 2e and f in the last version of the manuscript are two IDs from the same MS2 23880 of liver 1, TR2.

The Supplementary Data 7 shown all search results which were filtered into one ID per spectrum rather than each ID of each raw. For the same spectrum, there may be two or more identifications in gPeptide and pGlyco. Only the ID with the highest score will be retained. We have carried out this processing for all the search results of four software packages to ensure a unified standard. During this process, this ID was filtered out. However, it is presented in the tsv file of liver tissues.

A clear reference to the spectrum file and scan number is required for independent validation. Reporting of universal spectrum identifiers is recommended: <https://www.psidev.info/usi>.

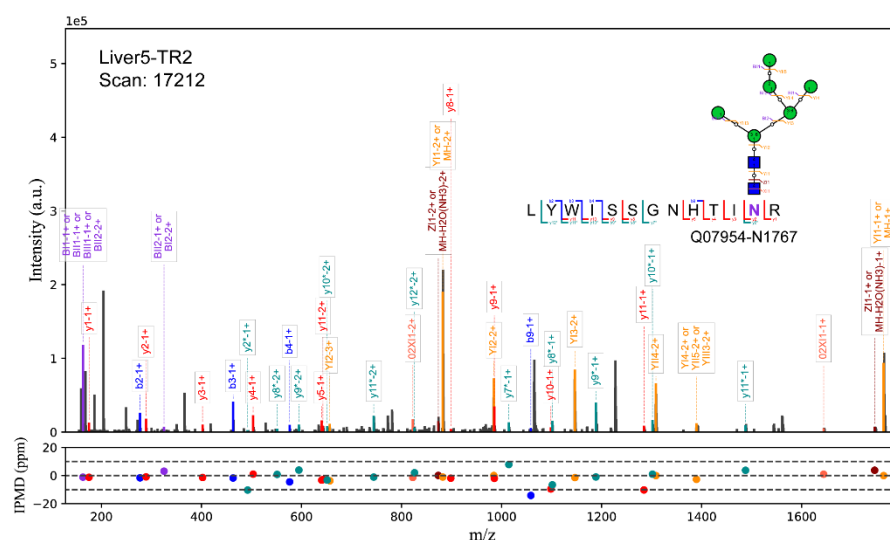
Response:

In this manuscript, the raw files and scan numbers were added to all the individual spectra. Thanks for your comments.

It is not clear from Supplementary Figure 26 if the fragment y4* is a real fragment or noise annotation. Were isotopologues observed for this ion? All other subsequent y-ions (i.e. y6 and higher) carrying a HexNAc mass could be generated by the alternative glycosylation site N1763. Similarly, the intensity of the b9* ion cannot be confirmed with the provided information.

Response:

In this raw file, no isotopic peaks of y_4^* from N1767 were detected. The original intensity of y_4^* is 2.5E3, while the ion with the highest abundance is 8.8E5. However, both b_9^* and y_6^* have an isotope peak from N1763. We found that in the TR2 of liver5, there was also identification of N1767. However, similarly, the intensity of the site-specific ions y_2^{*1+} was found without isotope peaks detected. The absence of detected isotope peaks might be due to that the theoretical calculated isotope peaks usually have very low intensities. But we do indeed agree that eliminating such a result would be a better approach for the readers.



Ref. Fig. 2

Generally, HCD spectra is not the best way to confirm glycosylation site of peptides with multiple possible glycosylation sites. Particularly, for less frequent glycosylation sites such as the N-X-C sites proposed for glycopeptide LYWISSGNHTINR. In these cases a targeted electron based fragmentation spectrum or combined with HCD (e.g. ETHcd) is more informative for site localization.

Supplementary Data 3 does not provide the annotated spectrum file or scan number to allow independent evaluation. This should be made accessible more easily for revision and readers. Ideally, chromatographically resolved EICs of the glycoform isomers LYWISSGNHTINR_8_Hex(9)HexNAc(2) and LYWISSGNHTINR_12_Hex(9)HexNAc(2) should be included to further confirm these assignments. Alternatively, re-analysis of the sample using ETHcd could also provide the necessary information.

Response:

We sincerely appreciate the reviewer's professional insight regarding the site localization for non-consensus motifs (N-X-C). We agree that accurately assigning glycosylation sites within a peptide containing multiple potential sequons (e.g., LYWISSGNHTINR) is challenging using HCD fragmentation alone, as it often lacks sufficient fragment ions for unambiguous site determination.

Given that we currently lack ETHcd data to provide definitive site-specific evidence,

and considering that these site isomers cannot be baseline-resolved under our current RPLC conditions, we have decided to remove the identification results related to the two potential N-X-C glycosites (N1767 and N3858) from the revised manuscript. This ensures the highest level of stringency and reliability for our human N-glycoproteome atlas.

Excluding the annotation of all glycopeptides with more than one glycosylation site might remove some of this ambiguity. I further evaluated in Annotator (DOI: 10.1101/2025.01.18.633732) the glycopeptide Q07954_LYWISSGNHTINR_12_HexNAc(4)Hex(7)Fuc(2) proposed by pGlyco3 with scan ID “File: Liver2_NGP_TR2_250529 – Scan#: 17401”. Although the glycosylation site was annotated at the N12 amino acid residue, when comparing both annotation options, the annotation of site localization at N8 provides a better sequence coverage and annotation of larger proportion of the fragment ion spectrum total ion current. This seems like an incorrect glycosylation site assignment due to ambiguous HCD spectra.



Response:

Thank you for your insightful comment regarding the multiple glycosites from same glycopeptides. After carefully re-examining our data and analysis, we acknowledge your concern. To avoid any potential confusion for our readers, we have removed the N1767 and N3858 from the manuscript.

However, it is worth emphasizing that while this specific finding remains inconclusive, we cannot entirely rule out its validity. Most importantly, regardless of whether this result is ultimately existed or not, it does not affect the core conclusions of the manuscript. We have changed the description and corresponding figures. Besides, the Supplementary Data 2 and 3 has been revised. The glycosites on the same peptide are combined and presented, and the corresponding counts also change to increase by only one. Because in over 200 raw files, the scores for the different sites of same peptide may vary, and the retained sites may not be the same, we adopted this strategy.

Thank you again for your valuable feedback, which has helped us improve the clarity and rigor of our work.

Comment #2: missing Skyline results and standardized peptide identification results

Although all the mass spectrometry data was provided, the identification results were not provided in the standardized formats suggested by the proteomics standard initiative, e.g. .mzIdentML (mzID). This complicated data revision for this manuscript and will be a barrier for interested readers. I understand that multiple search engines were used of which some might not support peptide results reporting in .mzID, but at least the universal spectrum identifier (DOI: 10.1038/s41592-021-01184-6) should be provided for all spectra interpreted. Specially for those provided in all figures of the manuscript.

Response:

Thanks for your comments. In order to verify whether the peptide with glycans can be accurately retrieved in the PRIDE ([Universal Spectrum Identifier // ProteomeXchange](https://www.ebi.ac.uk/pride/)) using USI, we conducted a test using previously published data (<https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00191>). Although the USI syntax is valid, the spectrum remains unresolvable in PRIDE, likely due to current limitations in repository support for the 'Glycan:' prefix.

Universal Spectrum Identifier

what's this? Example USIs Recent USIs Look Up USI Valid USI Skip species

mzspec:PX012424.L02_GP_C18_t1_f1_pentaHILIC_75um_150min_20181116.scan:8292[UNIMOD:36]DIYTFDGLNKGlycan:HexNAc2Hex7K[UNIMOD:36]2

VALID SYNTAX
The syntax of this USI is valid, although it may not be resolvable at any resource

USI

mzspec:

mzspec:PX012424.L02_GP_C18_t1_f1_pentaHILIC_75um_150min_20181116.scan:8292[UNIMOD:36]DIYTFDGLNKGlycan:HexNAc2Hex7K[UNIMOD:36]2	
charge	2
collection_id	2
collection_name	PX012424
collection_type	PXD
error	0
identifier_type	UPPM
index	8292
index_type	scan
interpretation	[UNIMOD:36]DIYTFDGLNKGlycan:HexNAc2Hex7K[UNIMOD:36]2
is_valid	true
ms_run_name	L02_GP_C18_t1_f1_pentaHILIC_75um_150min_20181116
peptide_name	
peptide_charge	2
peptide_is_valid	true
peptide_neutral_mass	1311.671067945
peptide_sequence	DIYTFDGLNKG
peptide_peptide_string	[UNIMOD:36]DIYTFDGLNKGlycan:HexNAc2Hex7K[UNIMOD:36]
peptide_residue_modifications	N
peptide_residue_modifications_index	10
peptide_residue_modifications_name	Glycan:HexNAc2Hex7
peptide_residue_modifications_string	Glycan:HexNAc2Hex7
peptide_residue_modifications_type	unsupported
peptide_residue_modifications_warnings	N0Glycan:HexNAc2Hex7
peptide_residue_modifications_warnings_index	0

The 'Glycan:' prefix is recognized and legal but is not yet supported by this system

The raw spectrum can be found using the correct peptide sequence with the fixed modification (i.e., dimethylation in the reported data), except with the glycan

Universal Spectrum Identifier

what's this?

Example USIs

Recent USIs

Look Up URI

Validate URI

strip sp

mzspec:PXD012424:LO2_GP_C18_tr1_f1_pentaHILIC_75um_150min_20181116:scan:8292:[UNIMOD:36]:DIYTFDGLNKG[UNIMOD:36]/2

show / hide details

mzspec:PXD012424:LO2_GP_C18_tr1_f1_pentaHILIC_75um_150min_20181116:scan:8292:[UNIMOD:36]:DIYTFDGLNKG[UNIMOD:36]/2

Provider	Response	Spectrum / PSM (click to view)	Attributes	Dev Info	Model + Params
iProX	Incomplete spectrum data	n/a		[JSON]	
jPOST	Not Found	n/a	-	-	
MassIVE	Not Found	n/a	-	-	
PeptideAtlas	DatasetNotHere	n/a			
PRIDE	OK	[UNIMOD:36]-DIYTFDGLNKG[UNIMOD:36]	[VIEW]	[JSON]	[A]
ProteomeCentral	DatasetNotHere	n/a	-	-	
MS2PIP	OK	[UNIMOD:36]-DIYTFDGLNKG[UNIMOD:36]	[VIEW]	[JSON]	HCD
KOINA	OK	[UNIMOD:36]-DIYTFDGLNKG[UNIMOD:36]	[VIEW]	[JSON]	ms2pip_HCD2021

Observed spectrum from PRIDE

Ions:

☐ 1⁺
☐ 2⁺
☒ 3⁺

b

☒ 1⁺
☒ 2⁺
☐ 3⁺

c

☐ 1⁺
☐ 2⁺
☐ 3⁺

x

☐ 1⁺
☐ 2⁺
☐ 3⁺

y

☒ 1⁺
☒ 2⁺
☐ 3⁺

z

☐ 1⁺
☐ 2⁺
☐ 3⁺

(deselect All)

Neutral Loss:

☐ NH₃ (*)
 ☐ H₂O (o)
 ☒ Immonium ions
 ☐ Reporter ions
 ☒ Precursor ions

Frags. Mass Type:

☒ Mono
 ☐ Avg

Mass Tol: 20

☐ Th
 ☒ ppm

Update

Peak Assignment:

☒ Most Intense
 ☐ Nearest Match
 ☐ Peak Detect

Peak Label:

☒ Ion
 ☐ m/z
 ☐ None

Width: 650

Height: 400

DIYTFDGLNKG, NH₃⁺ 1312.6783, m/z 656.8428

File: PRIDE spectrum, Exp. m/z: 1427.1168, Charge: 2

Click and drag in the plot to zoom X: ☒ Y: ☐ ☐ Enable tooltip ☒ Plot mass error

a+	b+	b2+	#	Seq #	y+	y2+
116.0706	144.0655	72.5364	1	D	11	
229.1547	257.1496	129.0784	2	I	10	1169.6201
392.2180	430.2429	210.6101	3	Y	9	1056.5360
493.2657	521.2606	261.1339	4	T	8	895.4727
640.3341	668.3290	334.6681	5	F	7	792.4250
755.3610	783.3559	392.1816	6	D	6	645.3566
812.3825	840.3774	420.6923	7	G	5	530.3297
883.4196	911.4145	456.2109	8	A	4	473.3082
996.5037	1024.4986	512.7829	9	L	3	402.2711
1110.5466	1138.5415	569.7744	10	N	2	289.1870
			11	K	1	175.1444

(Click) to move table

Annotated Ion Current: 117206.5244 (11.20%)

Add to N-terms: 28.0313

Variable Modifications: K: 28.0313 [11]

The authors claimed to have evaluated the data in Skyline, but no Skyline documents were deposited in the PRIDE repository. These can easily be shared through [www. https://panoramaweb.org](https://panoramaweb.org) or as zipped folders in PRIDE. Zipped folders can be created from the Skyline GUI with “File -> Share...”. I also recommend using the following publication as the base reference for this tool: DOI: 10.1007/s13361-018-2122-8. If the authors will not provide this information then I would not recommend any of their conclusions based on the output of Skyline nor report it in the manuscript. Data analysis in Skyline implies additional rigorous LC-MS/MS data validation or, at best, transparency in reporting results which should be shared.

Response:

Thanks for your comments. The .sky files were uploaded to <https://panoramaweb.org/Tongji%20University%20-%20Tian-BAMS/N-glycoproteome%20of%2018%20human%20organs/project-begin.view>. There was one .sky file for one subject. There were 74 files shown. This paper has also been cited.

Overall, the overall expression matrix is based on the combination of the information from qualitative information from gPeptide and quantitative information from skyline. It has all been supplemented. Thanks again for your comment.

Minor comments:

-Please revise English of final version. Multiple instances of incorrect grammar were observed in the text

Response:

Thanks for your comments. We have revised it carefully.

-Make sure “spectrogram” is not used in the manuscript as this is incorrect terminology in the mass spectrometry field. It should be “spectrum” or “spectra”

Response:

Thanks for your comments. We have already revised the entire manuscript.

-Supplementary Figure 3a: Legend figures and color should be accompanied by text. Legend of Fucose and Sialic acids are missing

Response:

Thanks for your detailed comments. Supplementary Figure 3a and 3b were both been revised.

-Supplementary Figure 30: The x-axis is not mass. It should be m/z if a fragment ion spectrum is displayed.

Response:

Thanks for your detailed comments. The x-axis has been revised to m/z .

-Glycan nomenclature such as “01Y(61F) 41Y41M(31M)61M” is not standard in the field and might not provide any information about the structure referred to. The authors should consider finding the corresponding GlyTouCan ID so that readers can look up the structure and quickly visualize it in glycan databases such as www.glygen.org

Response:

Thanks for your detailed comments. We have made modifications to the files of

each organ and Supplementary Table 7. The new information has been added to the sheets. These glycans were matched with the accession numbers in glycoTouCan with composition information. Besides, we used the glycowork tool (<https://bojarlab.github.io/glycowork/>) to convert linkages from gPeptide into the standard IUPAC-condensed format, making it easier for readers to understand. The readers can easily visualize glycans structures in GlyTouCan (<https://glytoucan.org/>) or GlyCosmos (<https://glycosmos.org/glycans/text>).