Supplementary Information for:

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

Yongqi Wu[1,2], Muyao Yang[1,2], Yongchao Xu[1,2], Li Jia[1,2], Jun Li[1], Xiaohan Wang[1], Jiayu Zhao, Yinli Cai, Yiwen Zhang, Shisheng Sun[1,*].

[1]Laboratory for Disease Glycoproteomics, College of Life Sciences, Northwest University, Xi'an, 710069, P. R. China

[2]These authors contributed equally.

*Correspondence: suns@nwu.edu.cn

Supplementary Information

1. Supplementary Figures

Supplementary Figure 1 Comparison of glycan compositions identified from 24 mouse tissues in this study with existing glycan databases.

Supplementary Figure 2 Reproducibility of glycoproteomic data.

Supplementary Figure 3 Summary of glycoproteomic identification across 24 mouse tissues.

Supplementary Figure 4 LC chromatograph and MS/MS spectra of glycan isoforms identified by StrucGP.

Supplementary Figure 5 Representative MS/MS spectra of site-specific *N*-glycans with different antennas.

Supplementary Figure 6 Sex-specific features of the renal *N*-glycoproteome.

Supplementary Figure 7 Tissue-specific glycosylation of commonly expressed glycoproteins.

Supplementary Figure 8 Examples of site-specific glycan heterogeneity and conserved glycosylation across tissues.

Supplementary Figure 9 Comparative clustering of glycopeptides, glycoproteins, and glycans across tissues.

Supplementary Figure 10 Representative MS/MS spectra of different glycan isoforms identified in different tissues.

Supplementary Figure 11 Structural diversity and tissue-specific distribution of functionally enriched glycan epitopes.

Supplementary Figure 12 Representative examples of glycan co-occurrence patterns at single glycosites.

Supplementary Figure 13 Heatmap showing the co-occurrence preferences across modular structural features of *N*-glycans.

2. Supplementary Datas (separate excel files and a separate PDF file)

Supplementary Data 1. Summary of identified intact *N*-glycopeptides, glycosites, glycans, and glycoproteins across 24 mouse tissues.

Supplementary Data 2. Detailed annotation of the 3,045 *N*-glycan structures.

Supplementary Data 3. Representative MS/MS Spectra for all complete glycans displayed in the main and supplementary figures (a separate PDF file).

Supplementary Data 4. Expression profiles of 209 glycosyltransferase genes from published RNA-seq data.

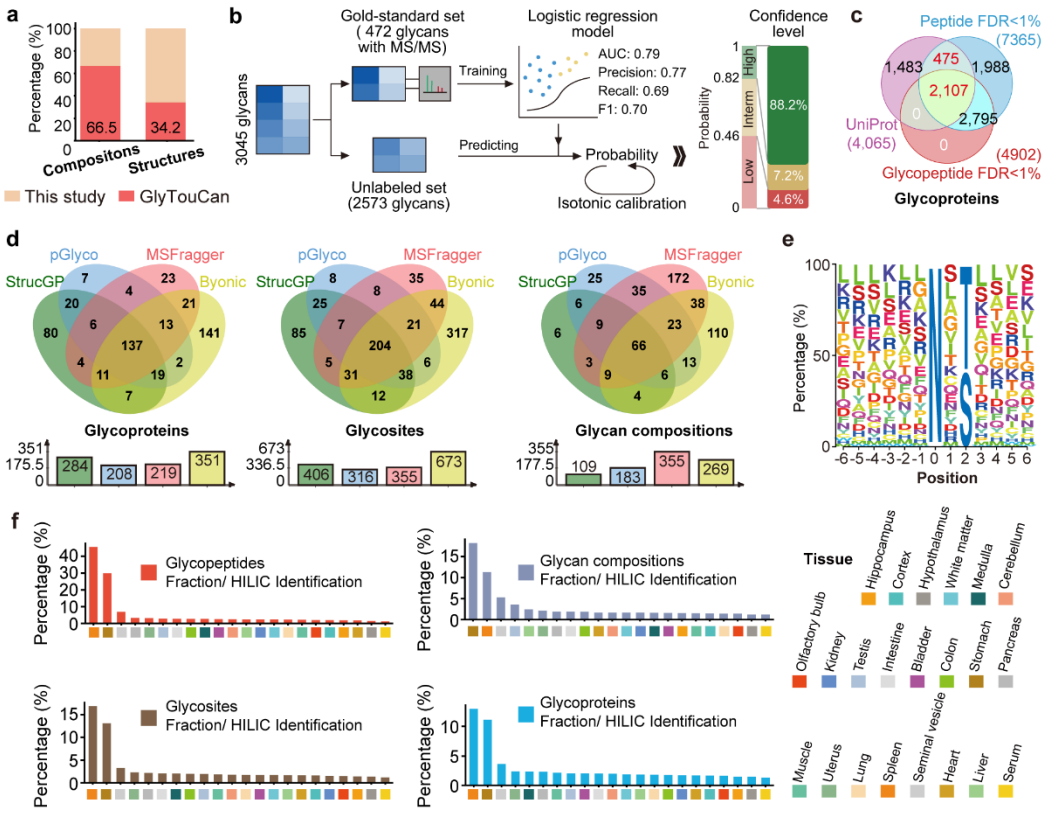
Supplementary Data 5. List of glycoproteins enriched in different subcellular localizations in all different mouse tissues.

Supplementary Data 6. Tissue distribution of ubiquitously and tissue-specifically detected glycoproteins across 24 mouse tissues.

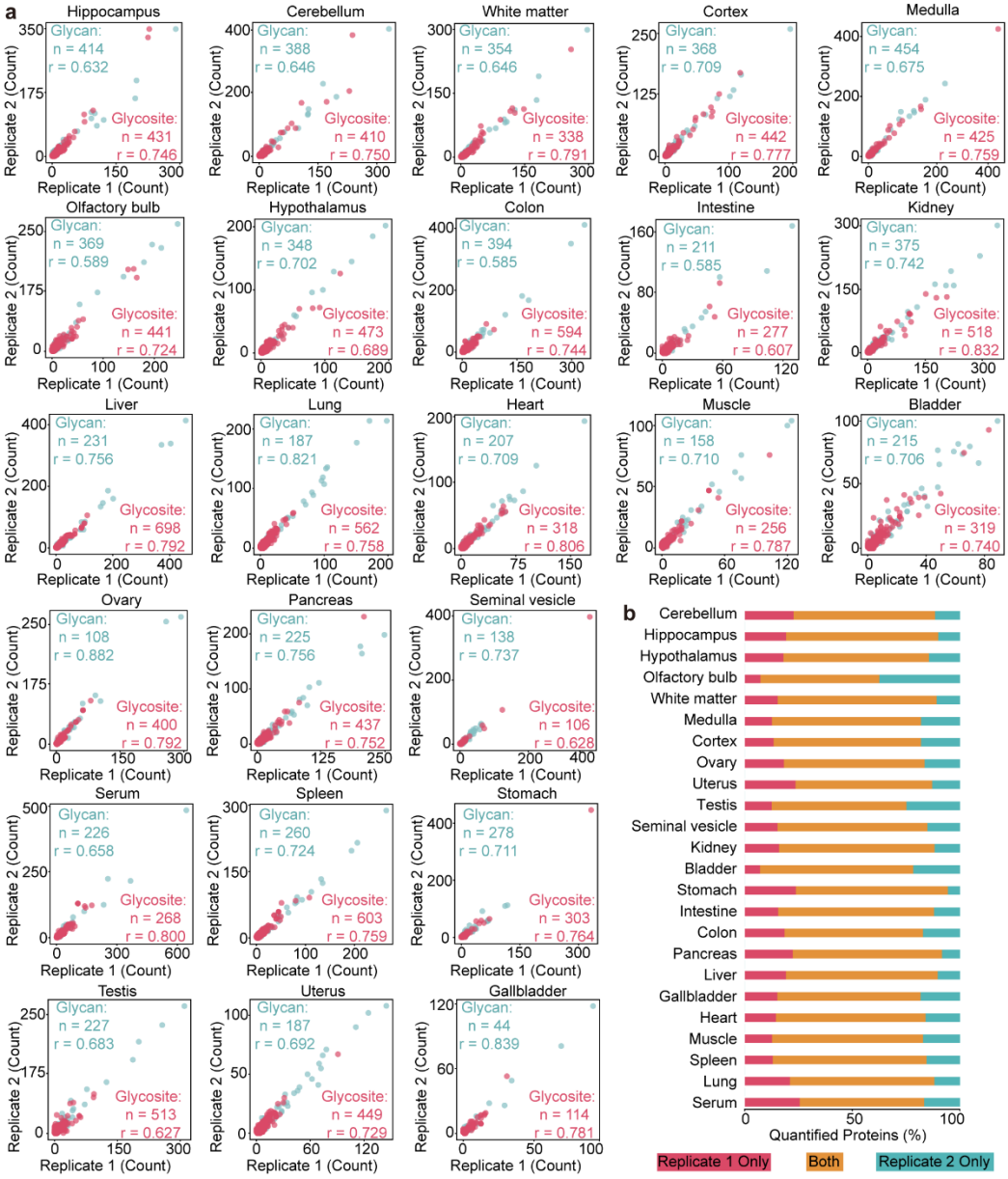
Supplementary Data 7. Site-specific *N*-glycan heterogeneities and conservation across mouse tissues.

Supplementary Data 8. List of the top 20 glycoproteins for each tissue.

Supplementary Data 9. Tissue-specific glycan compositions (a) or glycan structures (b) identified across 24 mouse tissues.



Supplementary Figure 1. Comparison of glycan compositions identified from 24 mouse tissues in this study with existing glycan databases. (a) Comparison of N-linked glycans identified from 24 mouse tissues in this study with the GlyTouCan database. (b) Flowchart illustrating the machine learning–based confidence assignment process and the resulting classification of *N*-glycans into high-, intermediate-, and low-confidence groups (t_high = 0.82, t_low = 0.46) based on calibrated probability outputs from the logistic regression model. (c) Comparison of glycoproteins identified from 24 mouse tissues in this study with UniProtKB annotations. The glycoproteins were identified based on their glycosite-containing peptides with FDR<1% at the peptide level or both peptide and glycan (glycopeptide) levels. (d) Comparison of glycoproteins, glycosites and glycan compositions identified from gallbladder in this study by MSFragger, Byonic, and pGlyco3. (e) Motif analysis of *N*-glycosites based on FDR<1% at both peptide and glycan levels. (f) Comparison of the number of identified glycopeptides, glycoproteins, glycosites, and glycan compositions in matched tissue samples with and without fractionation. Panels C-F present results derived exclusively from high-confidence glycans (probability ≥ 0.82). Source data are provided as a Source Data file.

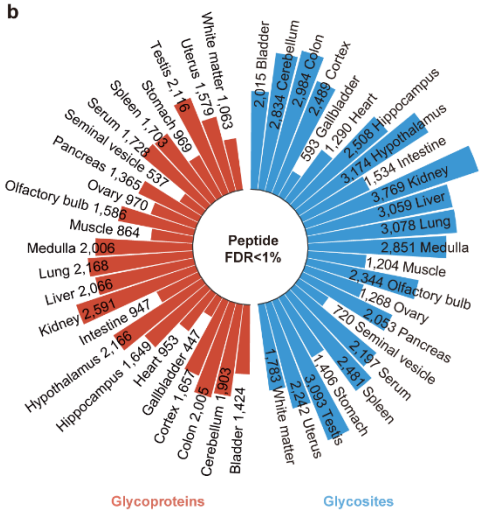


Supplementary Figure 2. Reproducibility of glycoproteomic data. (a) Scatter plots comparing the glycan structure and glycosite identifications between duplicate LC-MS runs across 24 mouse tissues. (b) Overlapped and unique glycoproteins identified by duplicate mass spectrometry analyses of the same glycopeptide sample. Source data are provided as a Source Data file.

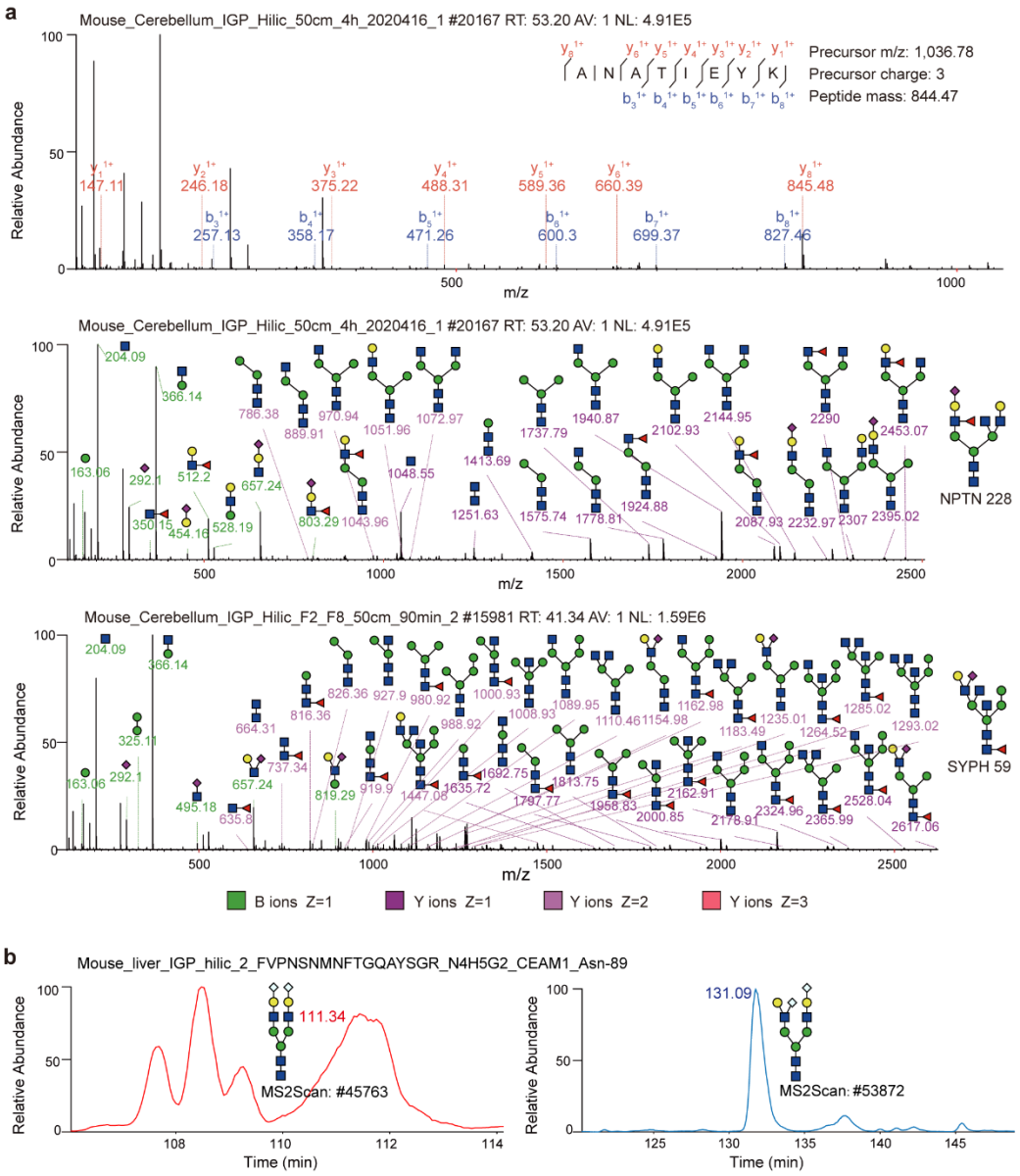
a

Glycopeptide Enrichment							
Tissue	MAX	HILIC	Fraction	PSMs	Unique IGPs	High PSMs	High Unique IGPs
Bladder	3	3	13	36,365	6,006	35,691	5,803
Cerebellum	3	4	12	65,481	8,469	63,526	8,145
Colon	3	5	12	59,739	10,146	58,059	9,727
Cortex	3	3	12	63,212	7,803	61,249	7,494
Gallbladder	2	3	0	6,375	1,304	6,306	1,275
Heart	1	0	12	16,271	4,642	15,872	4,508
Hippocampus	3	4	12	63,733	7,623	61,409	7,246
Hypothalamus	3	3	13	67,233	8,523	65,320	8,192
Intestine	3	3	12	33,411	6,206	32,465	5,971
Kidney	10	10	15	127,201	12,208	121,626	11,723
Liver	3	3	12	70,026	10,862	66,546	10,215
Lung	5	3	12	58,255	8,139	57,104	7,893
Medulla	3	3	12	58,472	7,954	56,551	7,603
Muscle	3	2	13	19,835	3,909	19,330	3,768
Olfactory bulb	3	3	12	46,985	6,575	45,297	6,221
Ovary	6	5	0	22,140	2,264	21,626	2,154
Pancreas	3	4	12	53,027	7,082	50,038	6,676
Seminal vesicle	3	3	12	24,742	1,660	23,694	1,559
Serum	3	5	18	124,034	6,903	123,735	6,806
Spleen	3	3	12	38,013	8,246	36,947	7,961
Stomach	4	5	12	22,774	4,980	22,098	4,788
Testis	5	7	12	79,149	8,392	75,503	7,866
Uterus	3	5	16	41,174	6,581	40,565	6,395
Whitematter	3	3	12	60,890	7,531	58,842	7,164

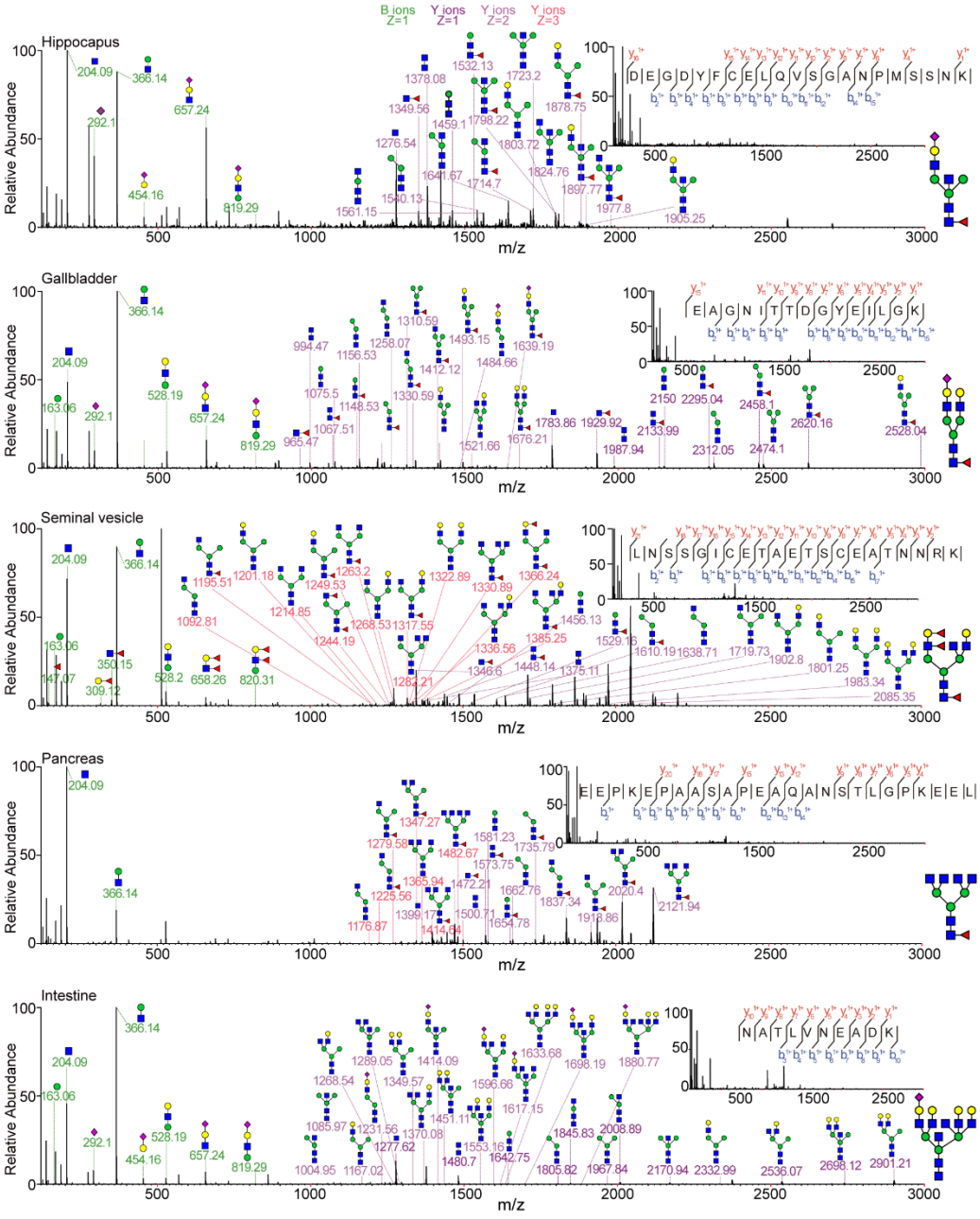
b



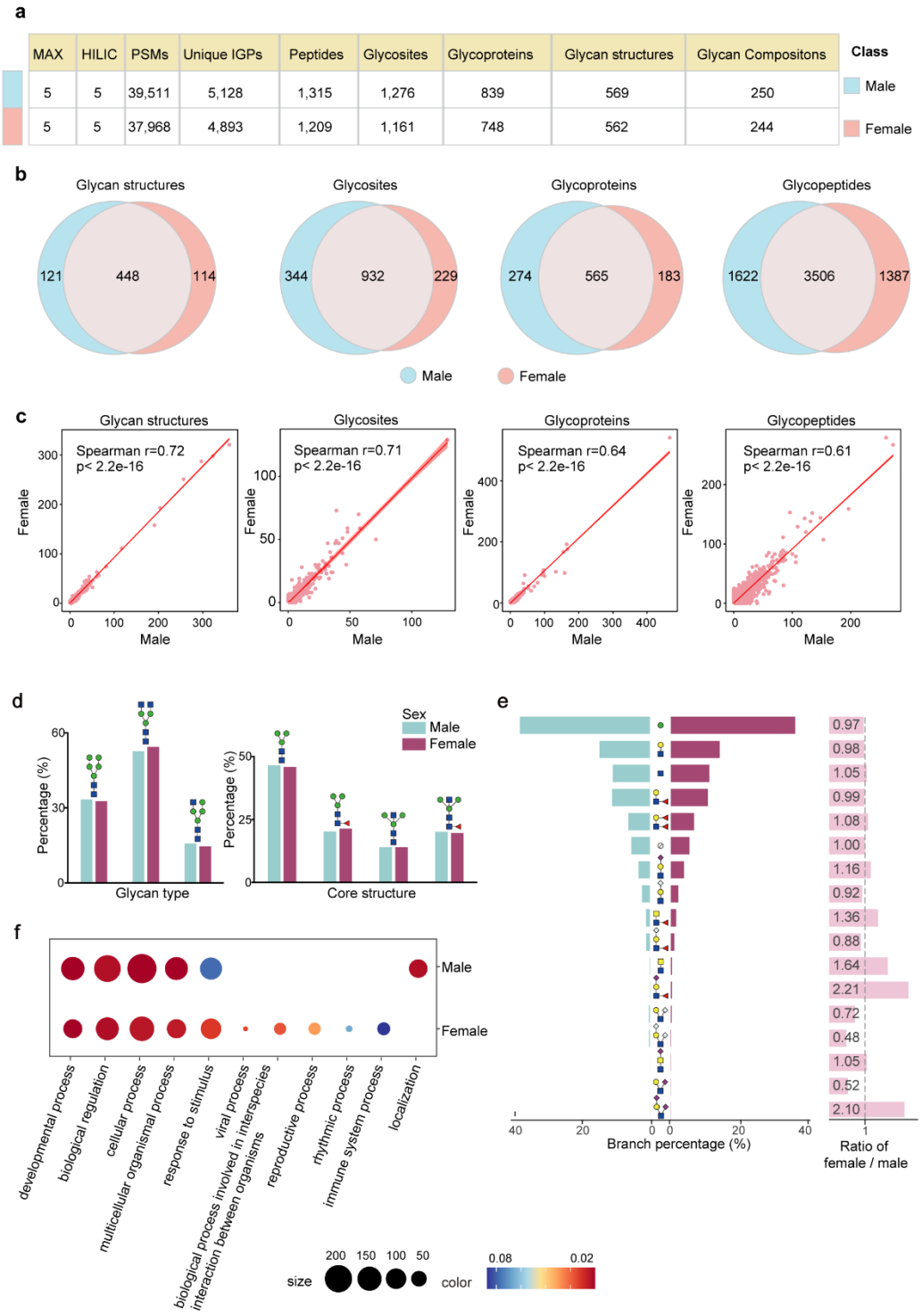
Supplementary Figure 3. Summary of glycoproteomic identification across 24 mouse tissues. (a) Number of identified PSMs and unique IGPs per tissue based on FDR<1% at both peptide and glycopeptide levels. (b) Number of glycoproteins from 24 mouse tissues based on FDR<1% at the peptide level. Source data are provided as a Source Data file.



Supplementary Figure 4. LC chromatograph and MS/MS spectra of glycan isoforms identified by StrucGP. (a) Representative MS/MS spectra of 12 structural isomers with the same composition (N5H5F1S1), with only 2 examples shown here, distinguished by characteristic B and Y fragment ions. (b) Chromatographic separation of selected N4H5G2 isomers by C18 chromatograph.

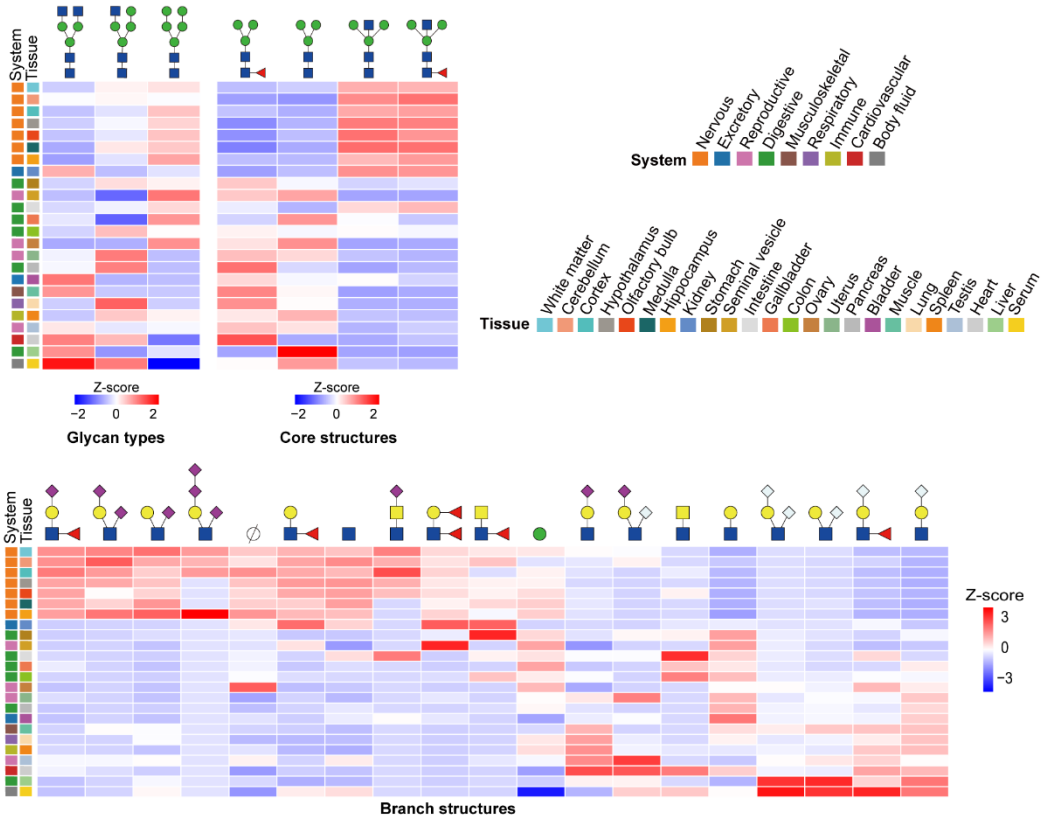


Supplementary Figure 5. Representative MS/MS spectra of site-specific *N*-glycans with different antennas.

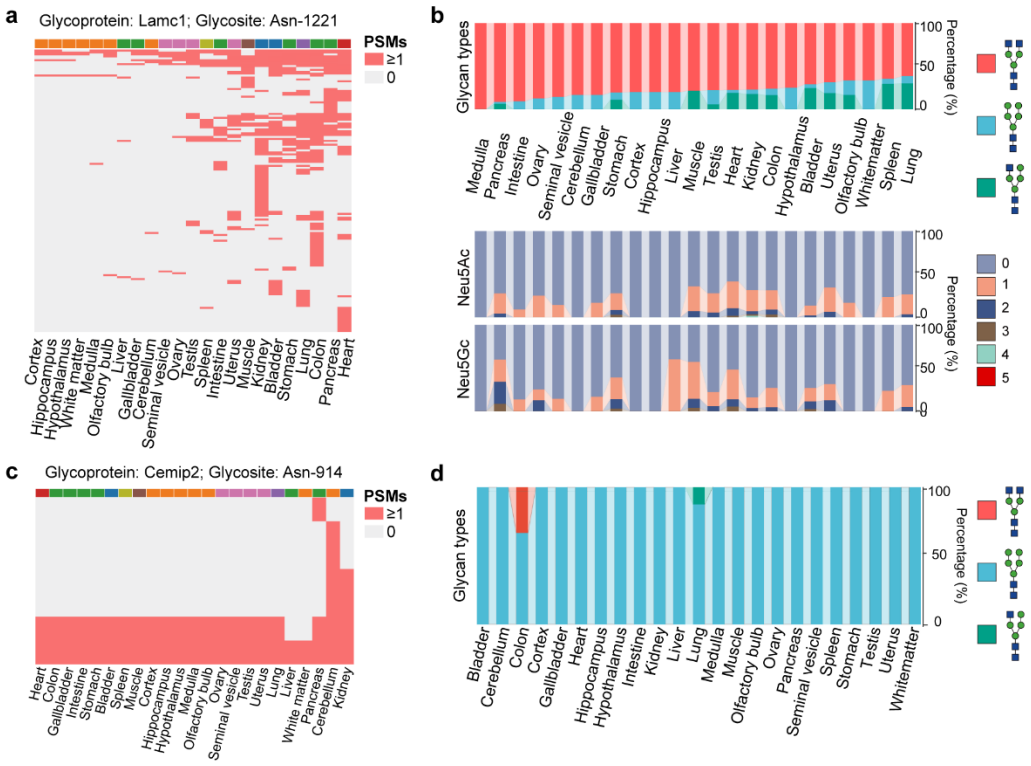


Supplementary Figure 6. Sex-specific features of the renal *N*-glycoproteome. (a) Numbers of intact *N*-glycopeptides (IGPs) identified in male and female kidneys based on FDR <1% at the glycan levels. PSMs: peptide-spectrum matches; IGPs: intact glycopeptides. (b) Overlap of identified glycan structures, glycosites, glycoproteins and glycopeptides between sexes (c) Scatter plots showing strong quantitative correlations

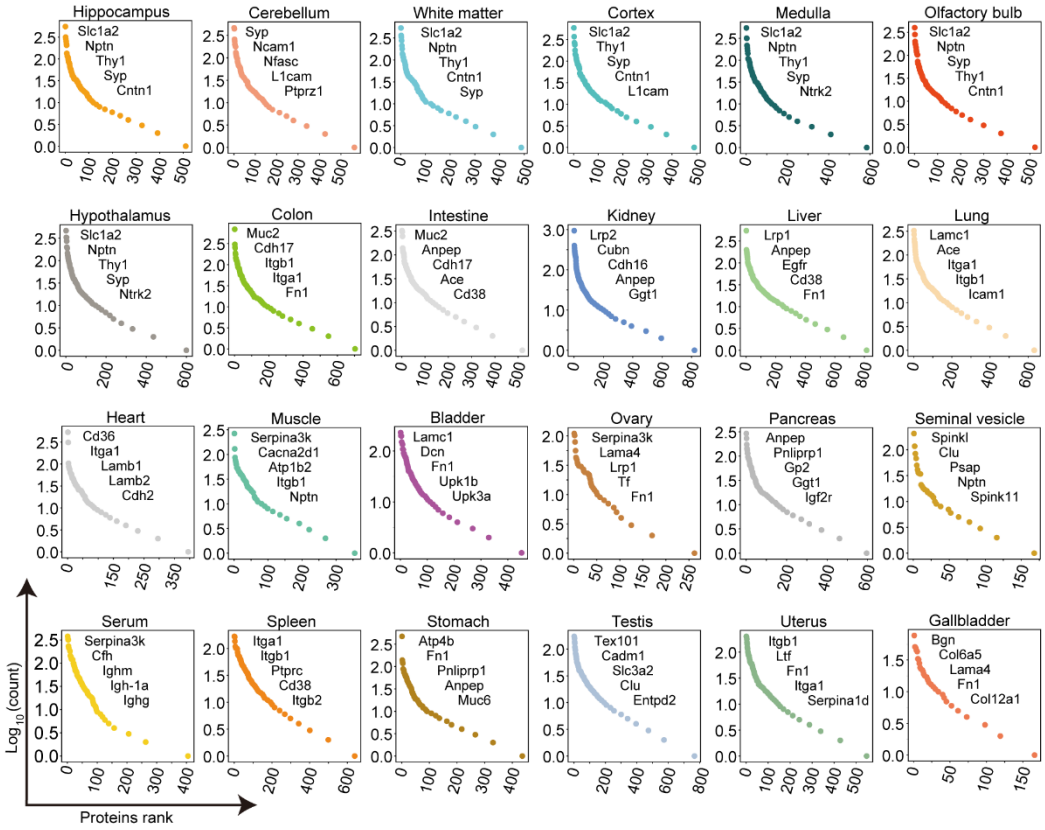
of glycan structures, glycosites, glycoproteins, and glycopeptides between male and female kidneys. (d-e) Distribution of site-specific *N*-glycans in male and female kidneys, including three glycan types and four core structures(d), and 17 branch structures(e). Percentages represent the proportion of unique glycopeptides containing each feature relative to the total glycopeptides identified in the tissue. Female-to-male ratios were calculated based on these percentages to assess sex-biased glycan features. (f) Gene Ontology enrichment analysis of sex-specific renal glycoproteins. Source data are provided as a Source Data file.



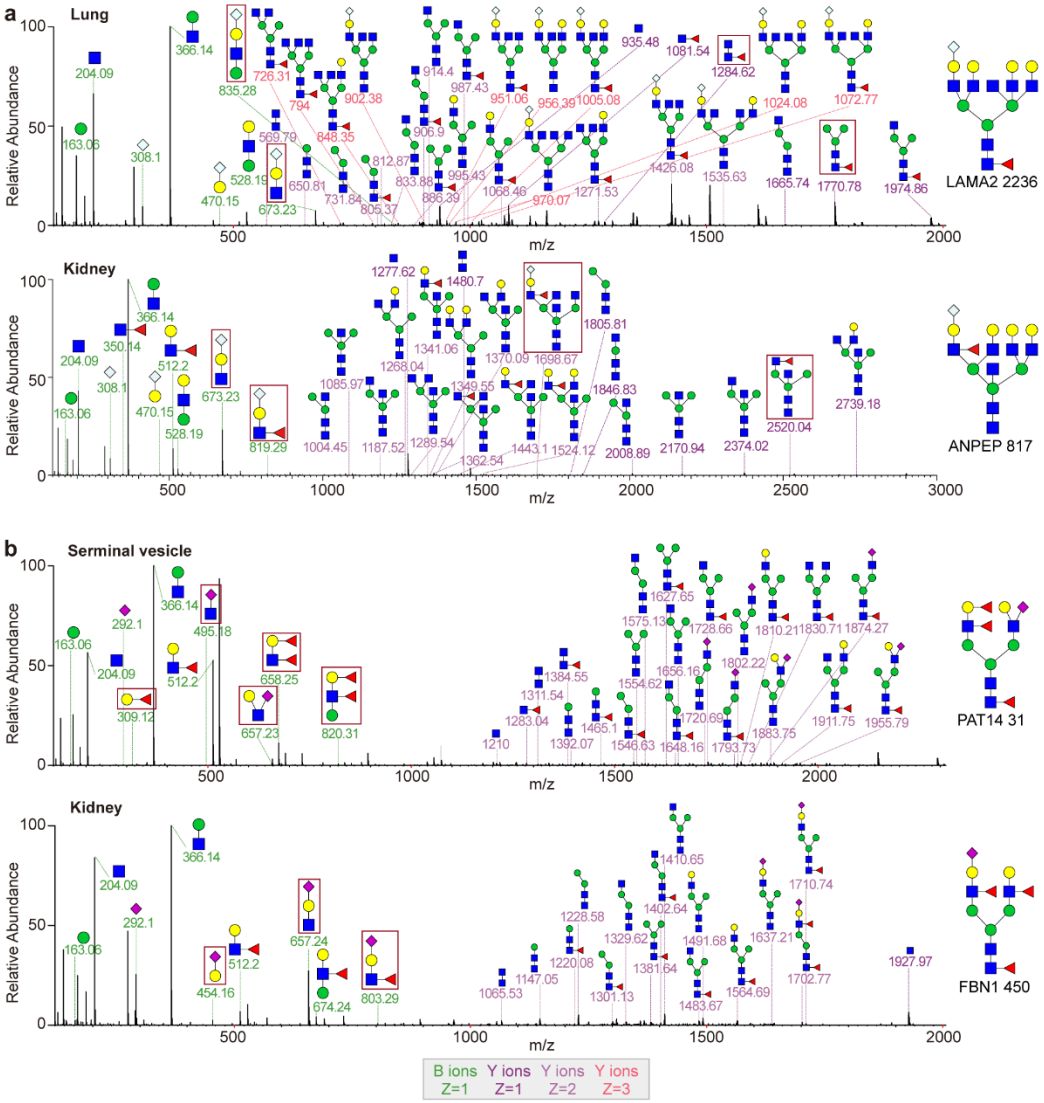
Supplementary Figure 7. Tissue-specific glycosylation of commonly expressed glycoproteins. Heatmap showing site-specific *N*-glycan profiles of 24 glycoproteins commonly expressed in all 24 tissues confirm tissue-dependent structural differences. Source data are provided as a Source Data file.



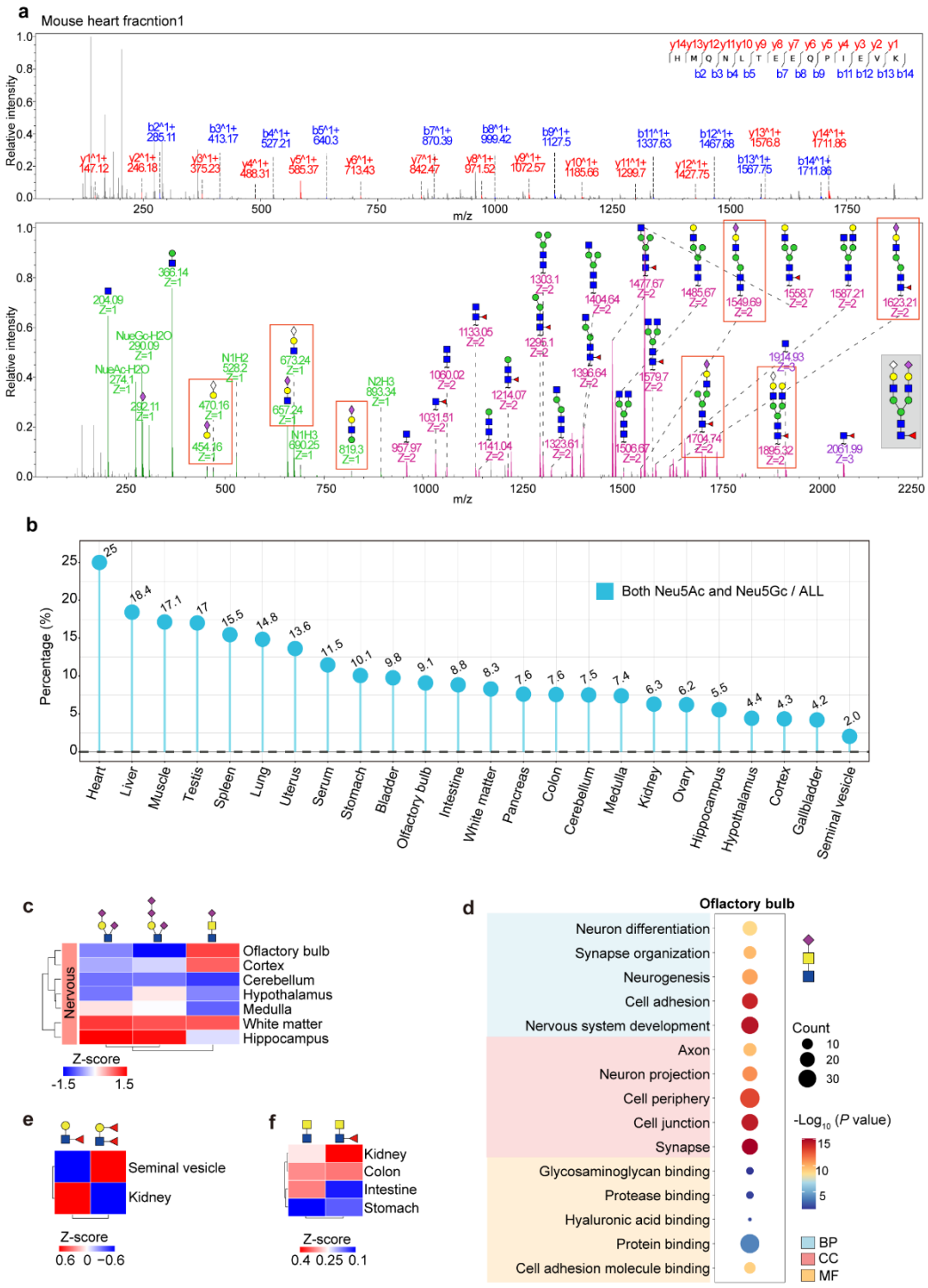
Supplementary Figure 8 Examples of site-specific glycan heterogeneity and conserved glycosylation across tissues. (a) Tissue-specific glycosylation patterns at the glycosite Asn-1221 of Laminin subunit gamma-1 (lamc1), illustrating substantial structural diversity across tissues. (b) Representative glycan types and the number of sialic acid (Neu5Gc and Neu5Ac) residues at the glycosite Asn-1221 of lamc1 across tissues. (c) Tissue-specific glycosylation patterns at the glycosite Asn-914 of Cell surface hyaluronidase (Cemip2), illustrating highly conserved glycan structures across tissues. (d) Representative glycan types at the glycosite Asn-914 of Cemip2 across tissues. Panels a-d present results derived exclusively from high-confidence glycans (probability ≥ 0.82). Source data are provided as a Source Data file.



Supplementary Figure 9. Comparative clustering of glycopeptides, glycoproteins, and glycans across tissues. Dynamic range of the intensity-ranked glycoproteins of 24 tissues. Five of the high abundant glycoproteins that relate to the functional specialization of the corresponding tissue are listed in descending order. Source data are provided as a Source Data file.

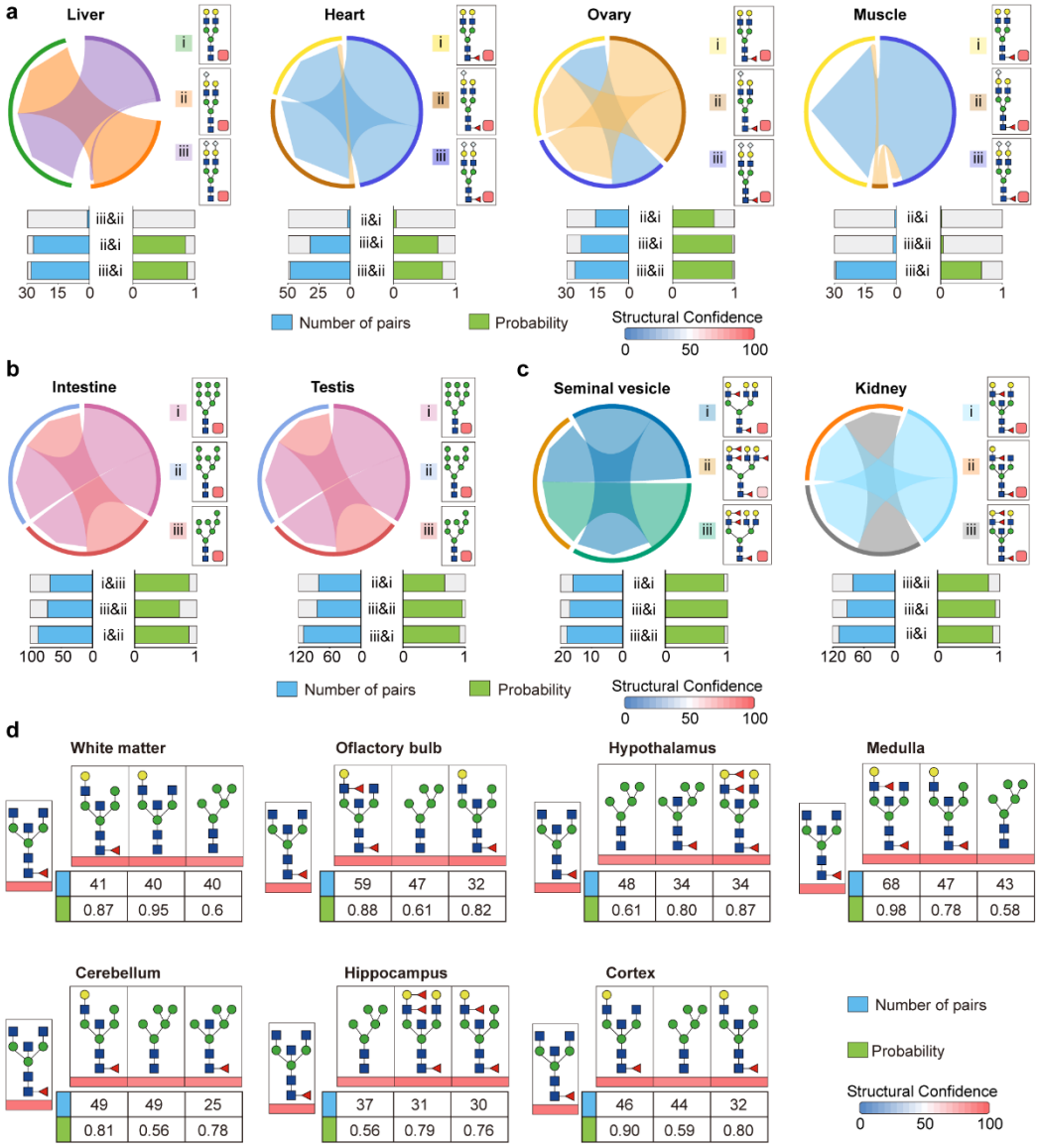


Supplementary Figure 10. Representative MS/MS spectra of different glycan isoforms identified in different tissues.

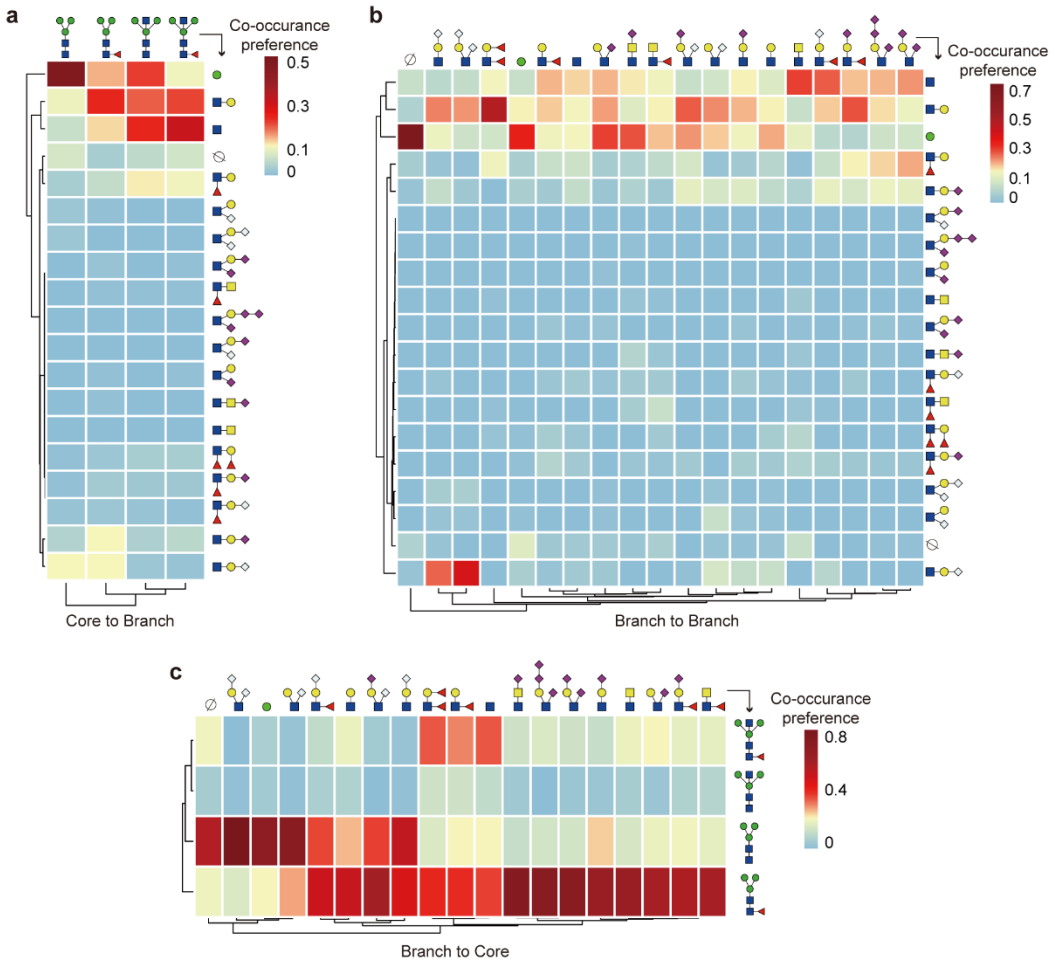


Supplementary Figure 11. Structural diversity and tissue-specific distribution of functionally enriched glycan epitopes. (a) Representative MS/MS spectra of *N*-glycans containing both Neu5Ac and Neu5Gc. (b) Proportion of glycans simultaneously carrying Neu5Ac and Neu5Gc among all identified glycans across tissues. Statistics were derived exclusively from high-confidence glycans (probability

≥ 0.82). (c) Heatmap showing distribution of sialylated *N*-glycans across seven brain regions. (d) Gene Ontology (GO) enrichment of glycoproteins modified by sialylated LacdiNAc glycans in the olfactory bulb. (e) Distribution of Lewis$^{x/a}$ and Lewis$^{y/b}$ epitopes in the seminal vesicle and kidney. (f) Distribution of LacdiNAc and fucosylated LacdiNAc glycans in the kidney, colon, intestine, and stomach. Source data are provided as a Source Data file.



Supplementary Figure 12. Representative examples of glycan co-occurrence patterns at single glycosites. (a) The triplet co-occurrence of sialylated glycans in liver, heart, ovary, and muscle, showing high conditional probabilities. (b) Co-occurrence of oligo-mannose structures (e.g., Man7–Man9) in intestine and testis. (c) Fucosylation-dependent co-occurrence patterns in kidney and seminal vesicle. (d) Co-occurrence patterns of glycans in brain regions. Color encodes structural confidence (0–1). High-confidence: probability ≥ 0.82; Intermediate-confidence: 0.46 ≤ probability < 0.82); Low-confidence: probability < 0.46. Source data are provided as a Source Data file.



Supplementary Figure 13. Heatmap showing the co-occurrence preferences across modular structural features of *N*-glycans. (a) Core-to-branch associations depict the likelihood of specific branch structures occurring alongside distinct core structures. (b) Branch-to-branch associations show how frequently different branch structures co-occur on the same glycan. (c) Branch-to-core associations indicate the probability of specific core structures appearing with given branch structures. Source data are provided as a Source Data file.