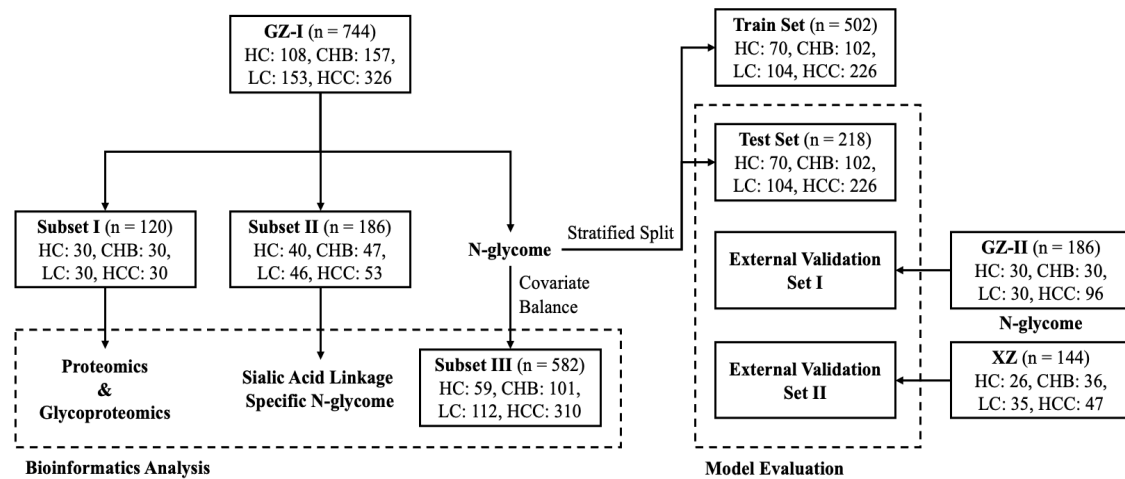
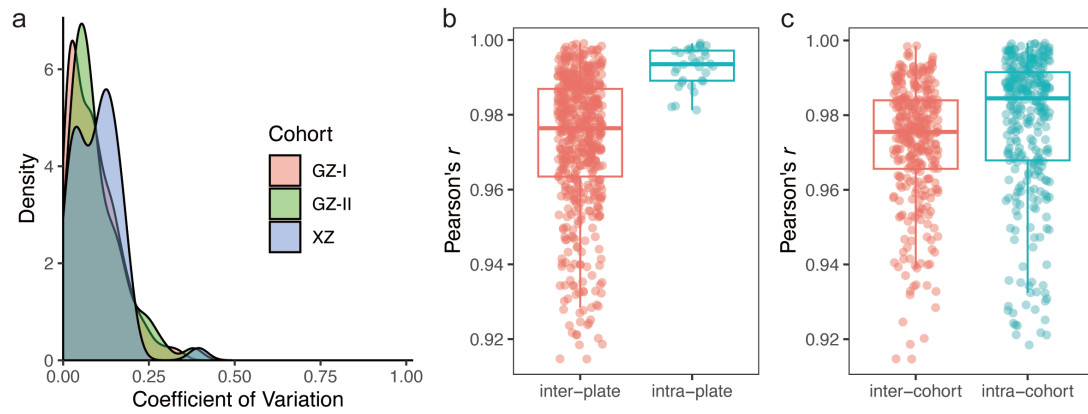
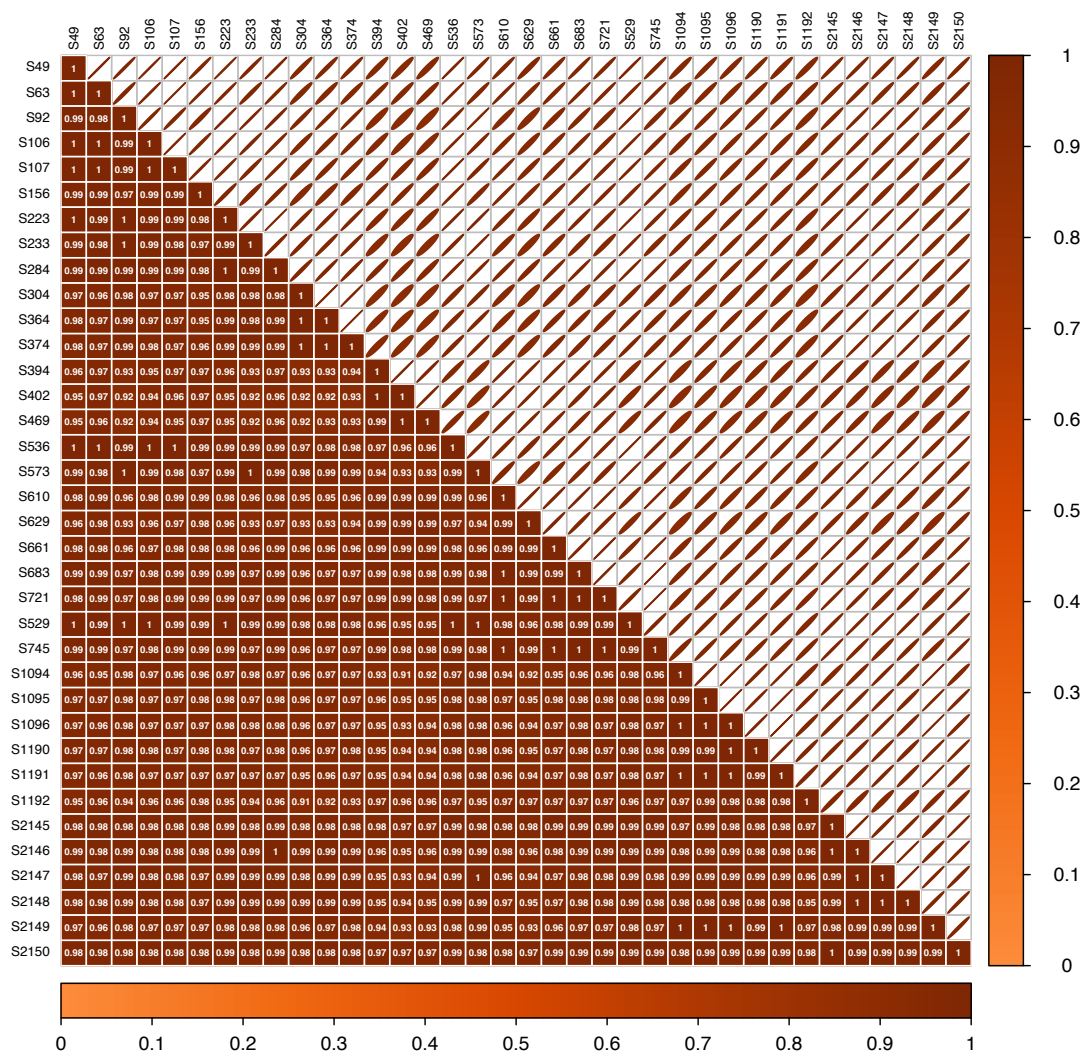




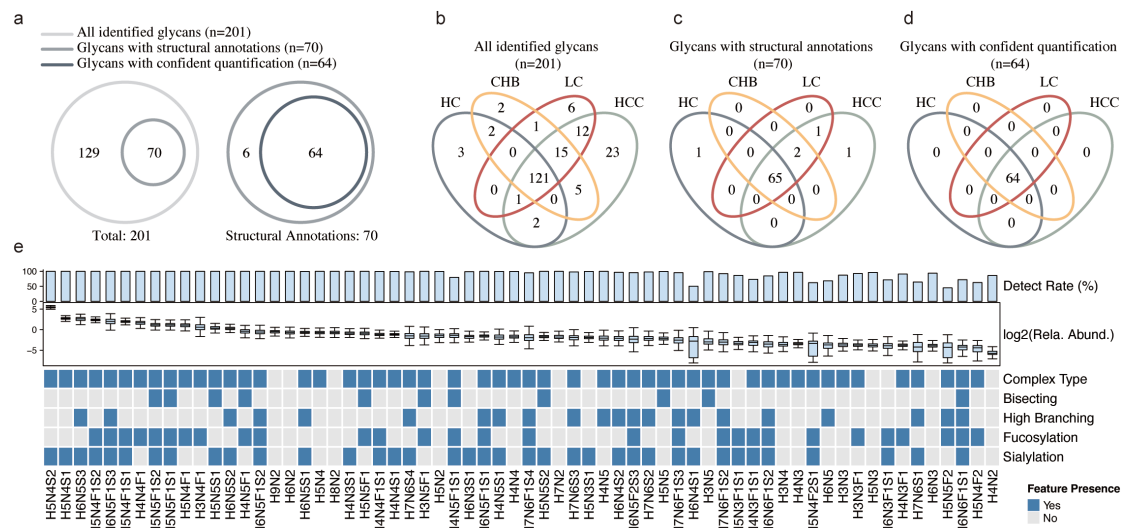
Supplementary Figure 1. Schematic of sample allocation for glycomics, glycoproteomics, and model development and validation. Two subsets of the GZ-I cohort were curated for proteomics and glycoproteomics analysis, and sialic linkage analysis, respectively. Another subset of the GZ-I cohort N-glycomics profilings were used for bioinformatics analysis. GZ-II and XZ cohorts were only used for model evaluation.



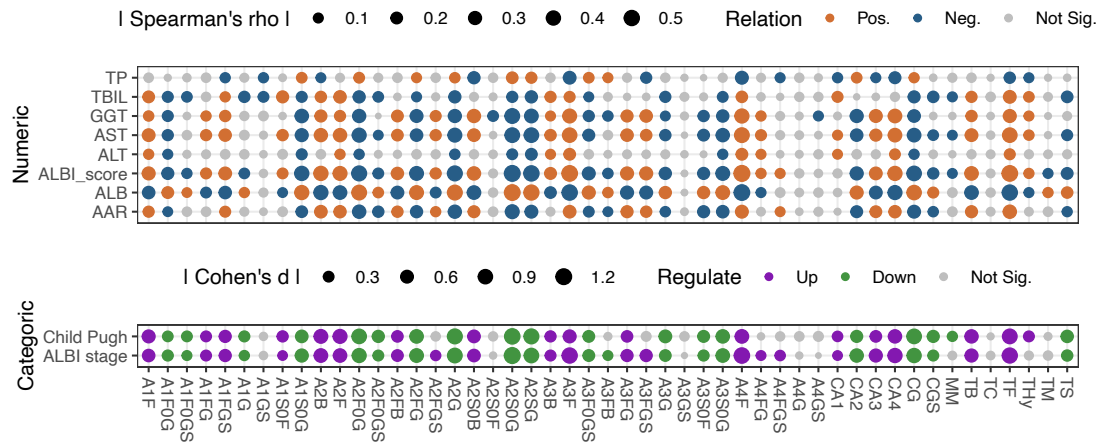
Supplementary Figure 2. a. Coefficient of variances (CVs) distributions of quality control (QC) samples in the GZ-I ($n = 24$), GZ-II ($n = 6$), and XZ ($n = 6$) cohort. **b-c.** Pairwise coefficients of correlation (Pearson's r) of QC samples across plates (b) and across cohorts (c). Source Data are provided with this paper.



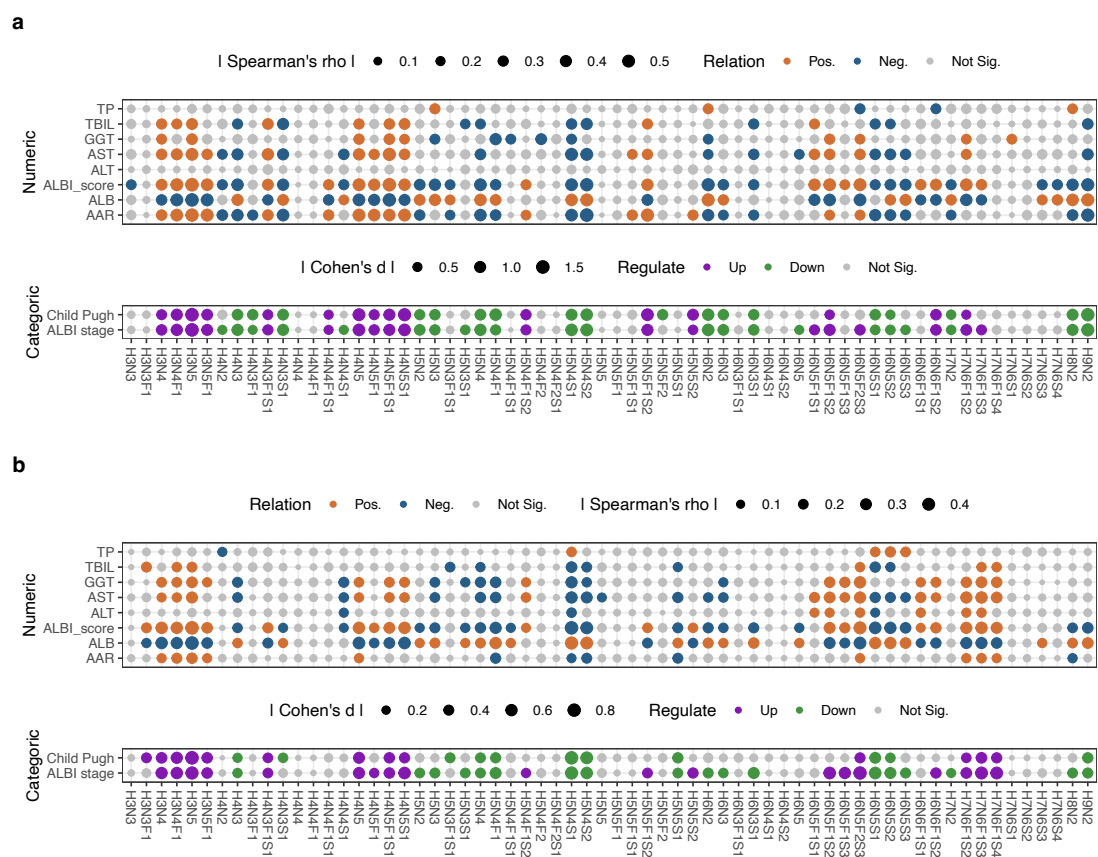
Supplementary Figure 3. Corrplot for QC samples in all three cohorts. “S” represents “QC Sample”. Colors are mapped to correlation coefficients. Number are correlation coefficients. Source Data are provided with this paper.



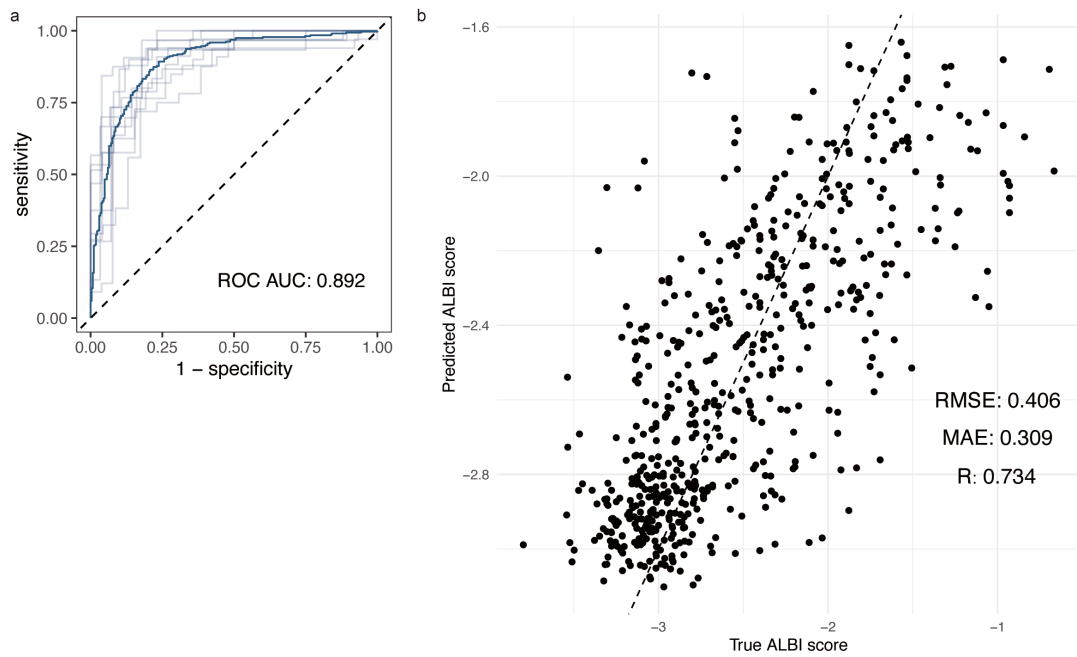
Supplementary Figure 4. N-glycan identification overview. **a.** Number of identified N-glycans. **b-d.** Numbers of identified N-glycans (b), N-glycans with structural annotations (c), and N-glycans with confident quantifications (d) across groups. **e.** Structural characteristics, missing value proportions, and relative abundance distributions of confident quantified N-glycans. Source Data are provided with this paper.



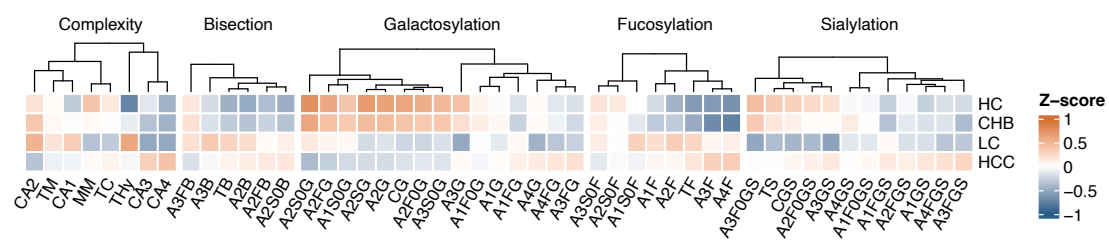
Supplementary Figure 5. Heatmap showing the effect sizes of statistical analyses between derived traits and liver function tests (LFTs). Numeric variables are colored blue and orange. Categorical variables are colored green and purple. Colored points represent significant (BH-adjusted p -value < 0.05) associations. Pos: positive, Neg: negative, Not Sig: Not significant. Source Data are provided with this paper.



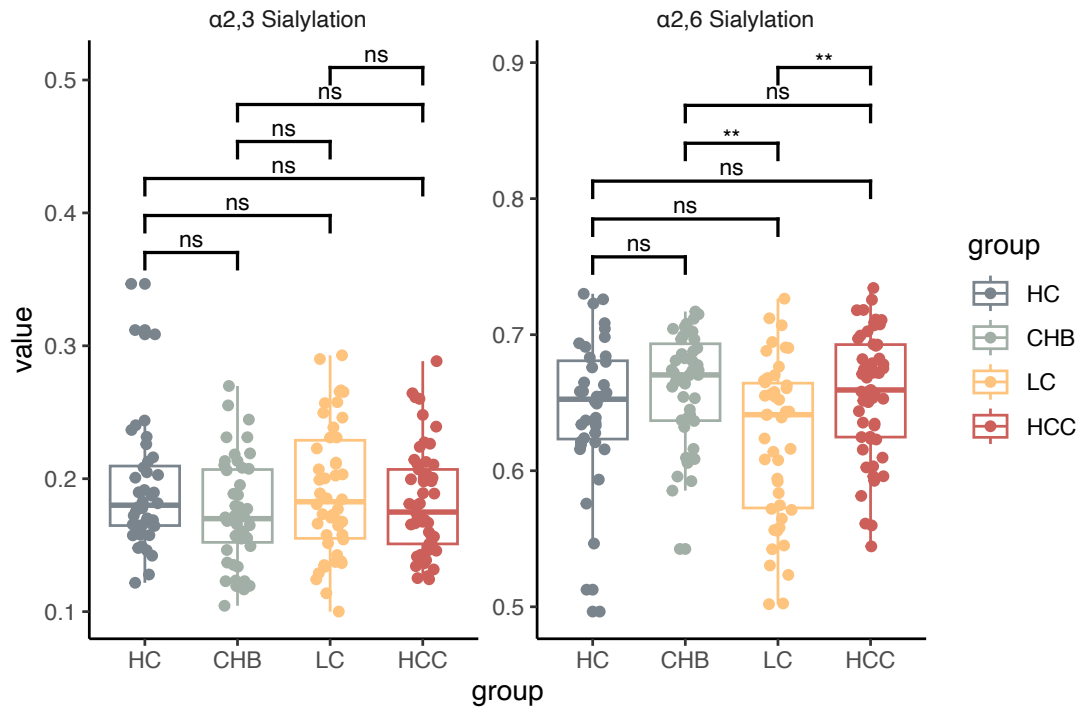
Supplementary Figure 6. Heatmaps showing the effect sizes of statistical analyses between glycans and liver function tests (LFTs) within LC samples (a) and HCC samples (b). Numeric variables are colored blue and orange. Categorical variables are colored green and purple. Colored points represent significant (BH-adjusted p -value < 0.05) associations. Pos: positive, Neg: negative, Not Sig: Not significant. Source Data are provided with this paper.



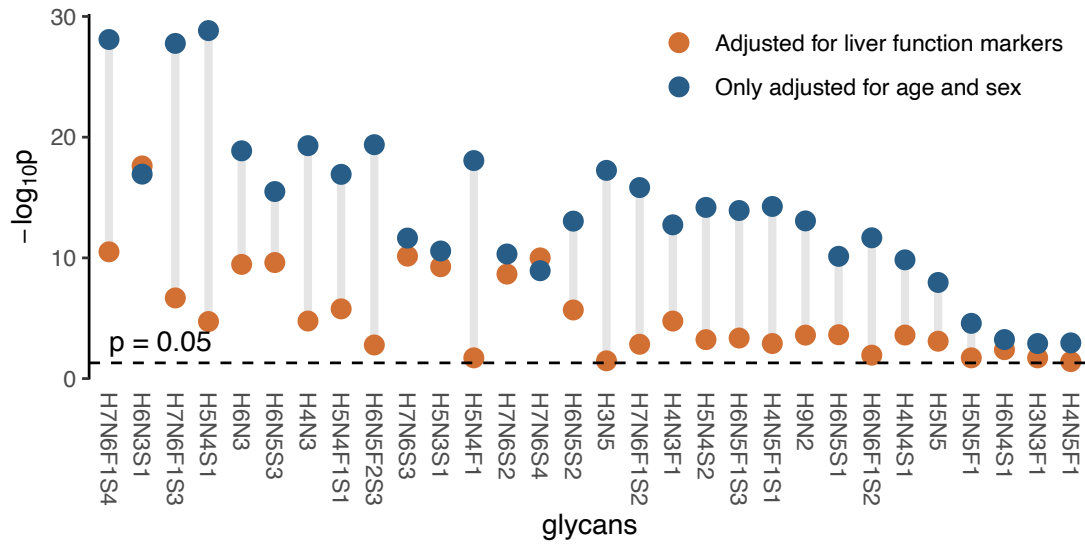
Supplementary Figure 7. a. ROC curves of random forest models predicting ALBI stage from N-glycome. Ten-fold cross-validation was used to evaluate the classification performance. Solid curve is the ROC curve using cross-validated predictions. Translucent curves are ROC curves of different folds in cross-validation. **b.** Predicted ALBI scores from N-glycome mimic the real ALBI score. Each point represents an individual. The dotted line is the fitted line between true and predicted ALBI score. RMSE: root-mean-square error, MAE: mean absolute error, R: correlation coefficient. Source Data are provided with this paper.



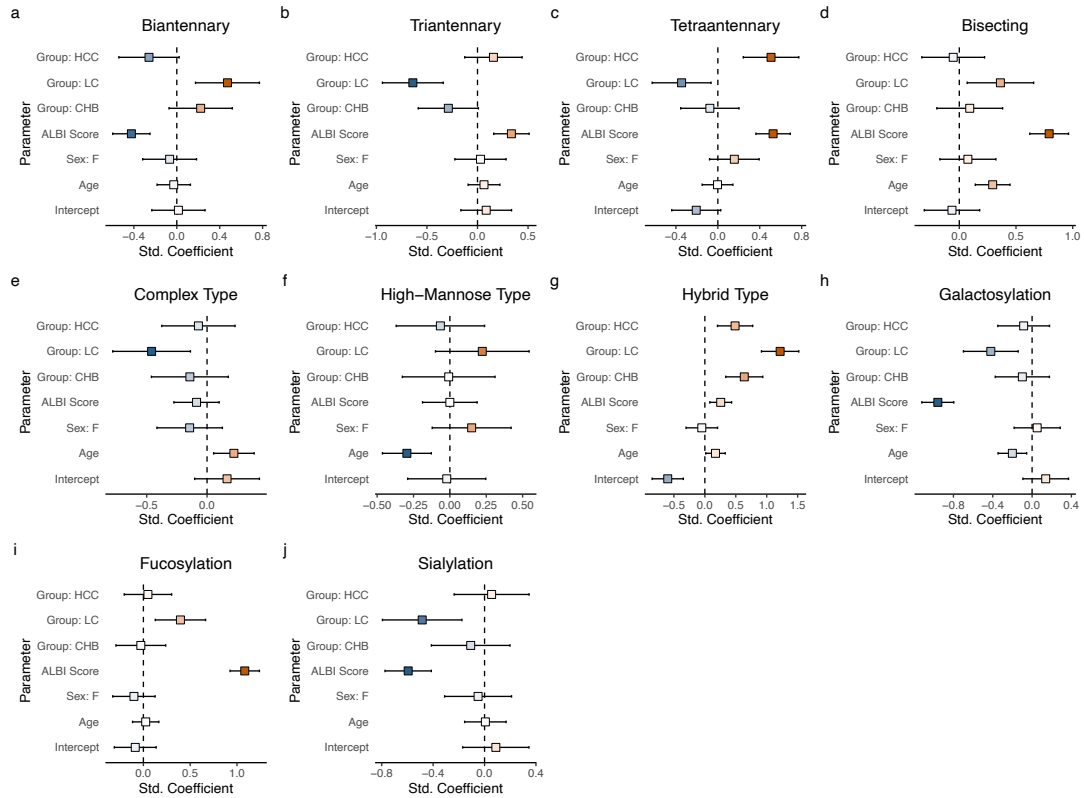
Supplementary Figure 8. Heatmap of derived traits across four groups (HC, CHB, LC, HCC). Z-scores were calculated before being averaged. Source Data are provided with this paper.



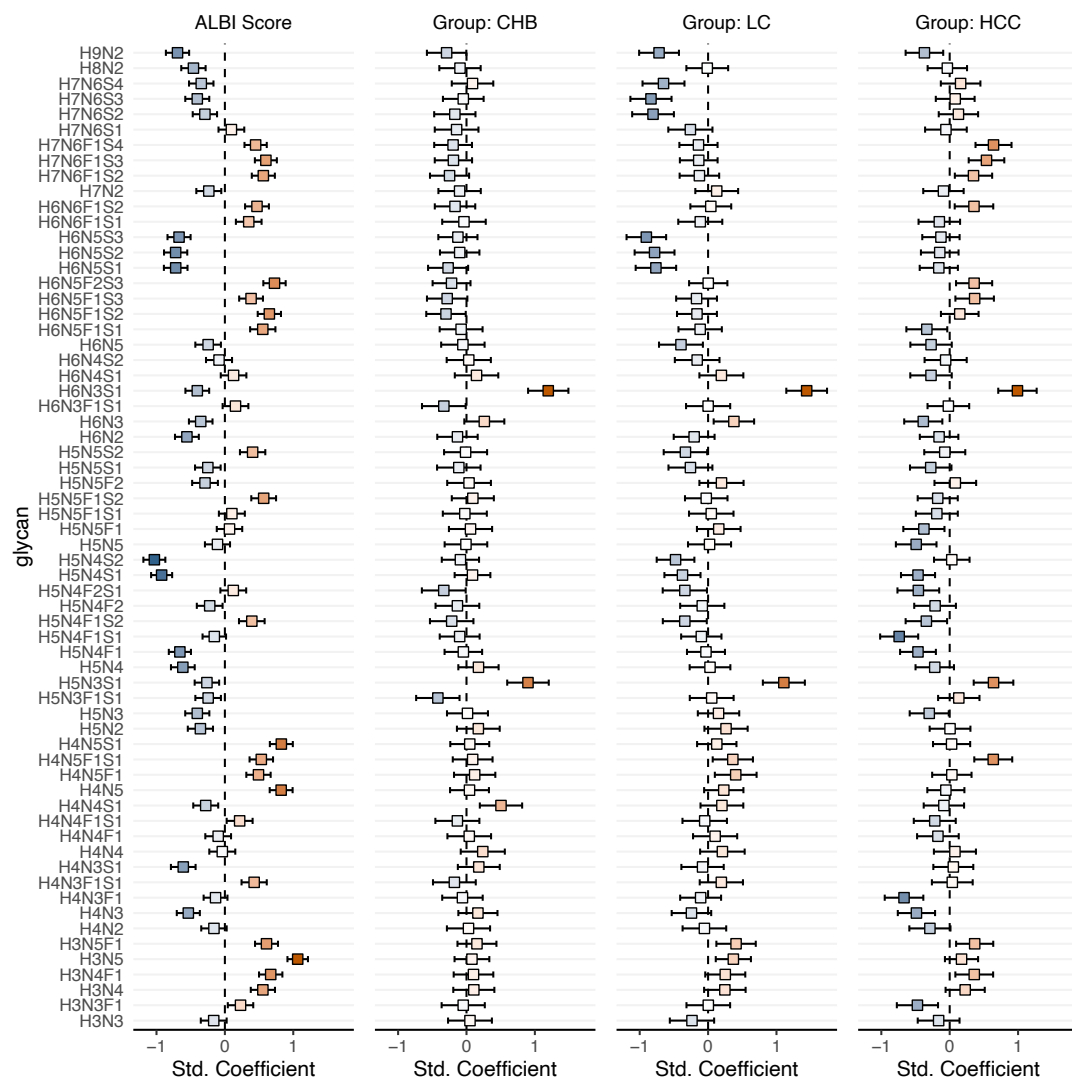
Supplementary Figure 9. Sialylation level of different linkages in four groups. It is defined as the average degree of sialylation for all galactoses. **: $p < 0.01$, ns: not significant. Source Data are provided with this paper.



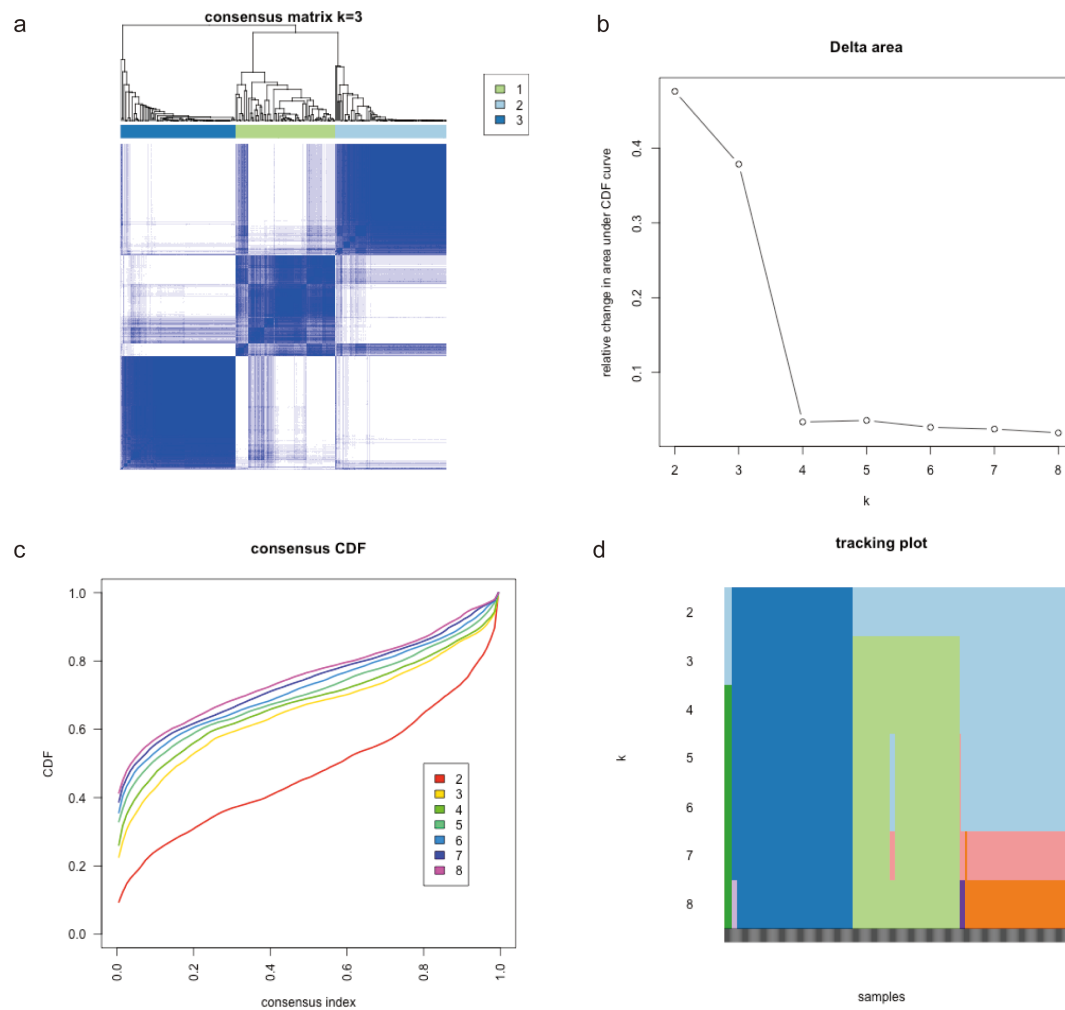
Supplementary Figure 10. Sensitivity test results. After setting all LFTs as covariates in Analysis of Covariate (ANCOVA), 30 of the 48 significant glycans still had BH-adjusted p -values less than 0.05. Source Data are provided with this paper.



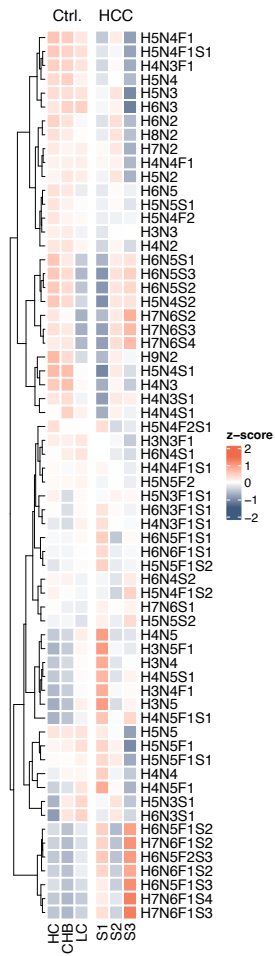
Supplementary Figure 11. Forest plots of key derived traits in multivariate regression analysis. **a.** Proportion of bi-antennary glycans within complex glycans. **b.** Proportion of tri-antennary glycans within complex glycans. **c.** Proportion of tetra-antennary glycans within complex glycans. **d.** Proportion of bisecting glycans within complex glycans. **e.** Proportion of complex glycans within all glycans. **f.** Proportion of high-mannose glycans within all glycans. **g.** Proportion of hybrid glycans within all glycans. **h.** Average degree of galactosylation of branching GlcNAc in complex glycans. **i.** Average number of fucoses. **j.** Average degree of sialylation of galactose in complex glycans. Source Data are provided with this paper.



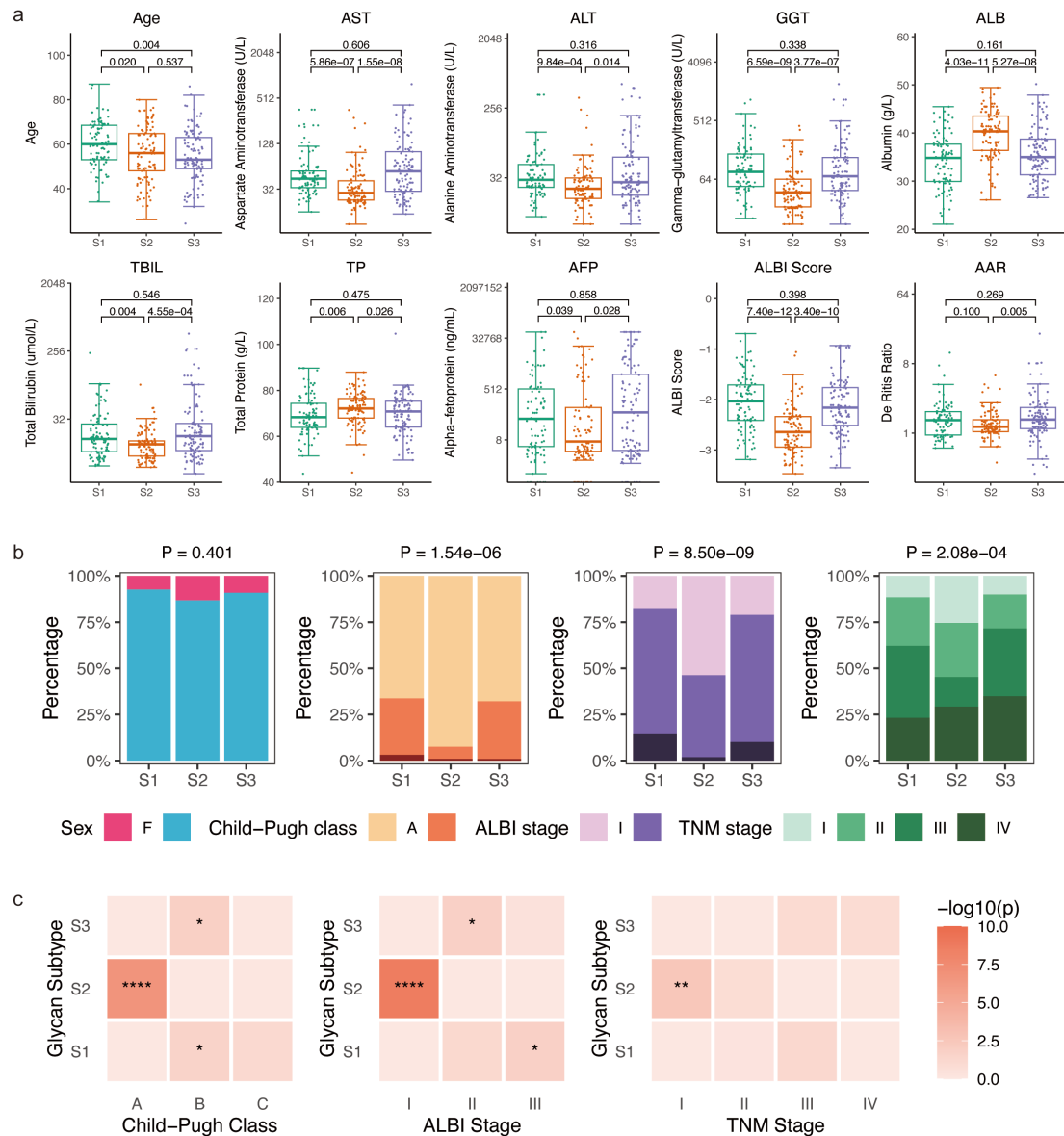
Supplementary Figure 12. Forest plots of glycans in multivariate regression analysis. Source Data are provided with this paper.

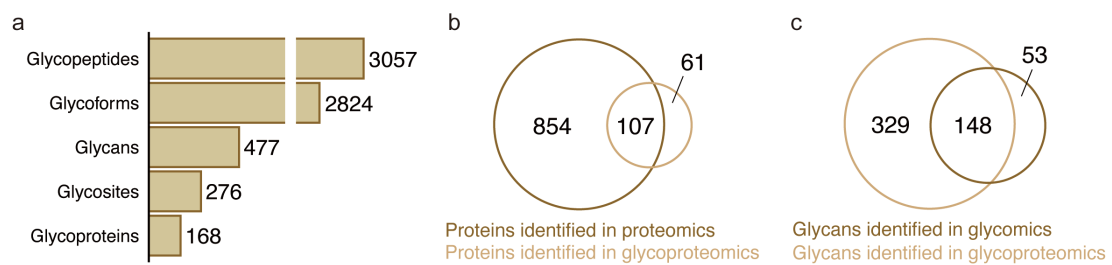


Supplementary Figure 13. Consensus clustering results for N-glycome-based molecular subtyping. a. Consensus matrix when $k = 3$. **b.** Delta area curve. **c.** Consensus CDF curve. **d.** The tracking plot. The figures were generated directly by the ConsensusClusterPlus (v1.68.0) package.

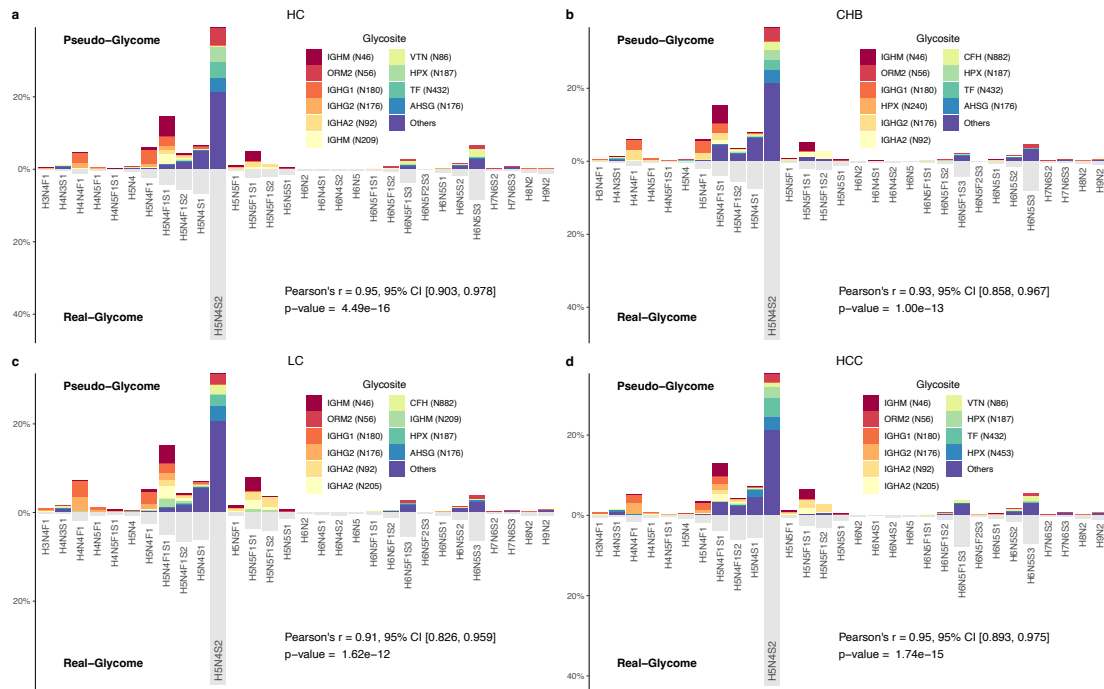


Supplementary Figure 14. N-glycome profiling of three HCC subtypes and three control groups (HC, CHB, and LC). Source Data are provided with this paper.

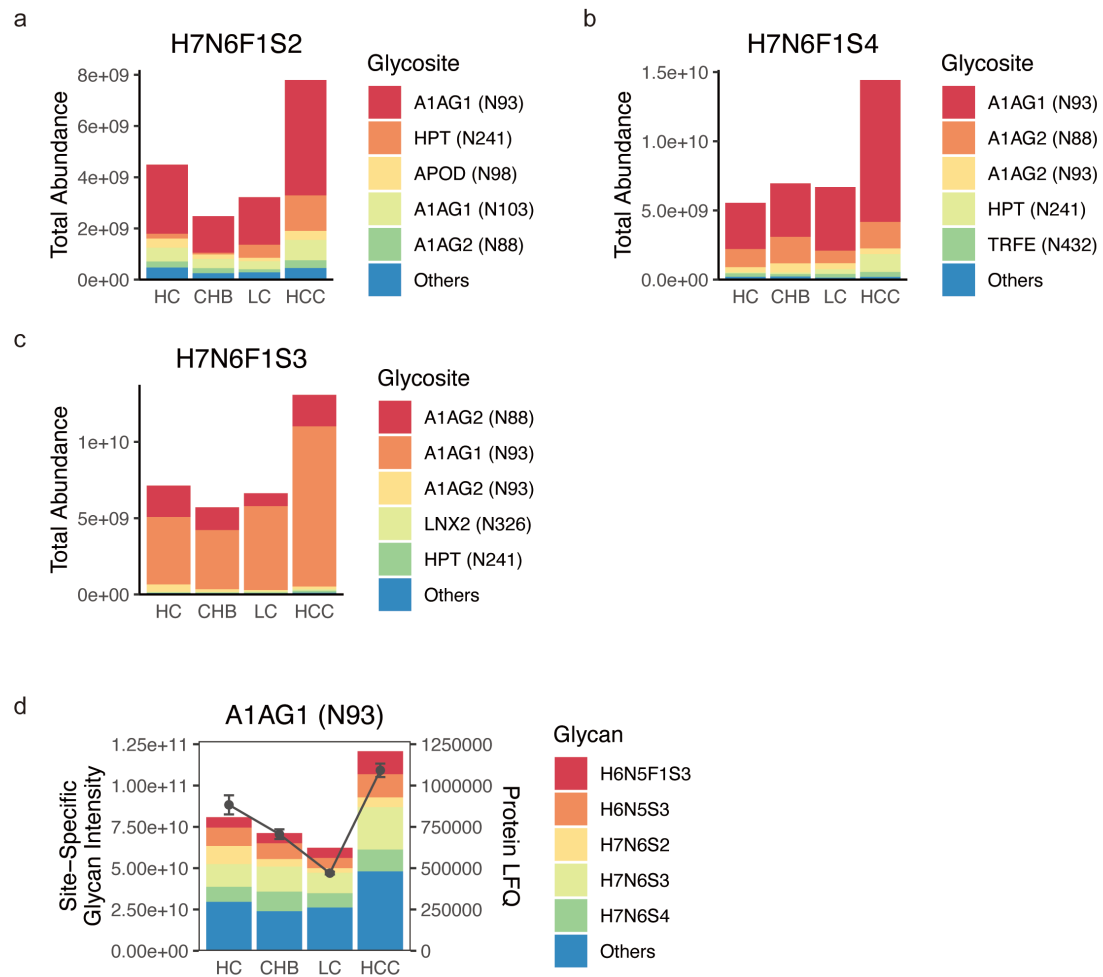




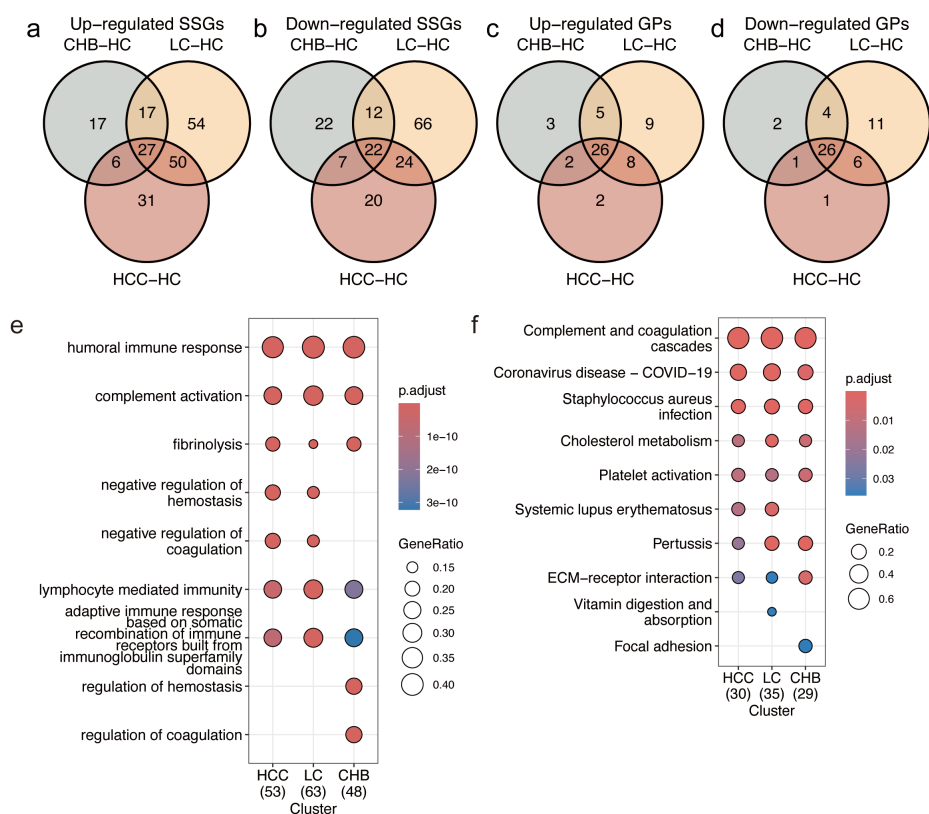
Supplementary Figure 16. Glycoproteomics result overview. **a.** Bar chart showing the total number of identified glycopeptides, glycoforms, glycans, glycosites, and glycoproteins in the glycoproteomics analysis. **b.** Venn diagram illustrating the overlap of proteins identified in the proteomics and glycoproteomics datasets. **c.** Venn diagram depicting the overlap of glycans identified in the glycomics and glycoproteomics datasets. Source Data are provided with this paper.



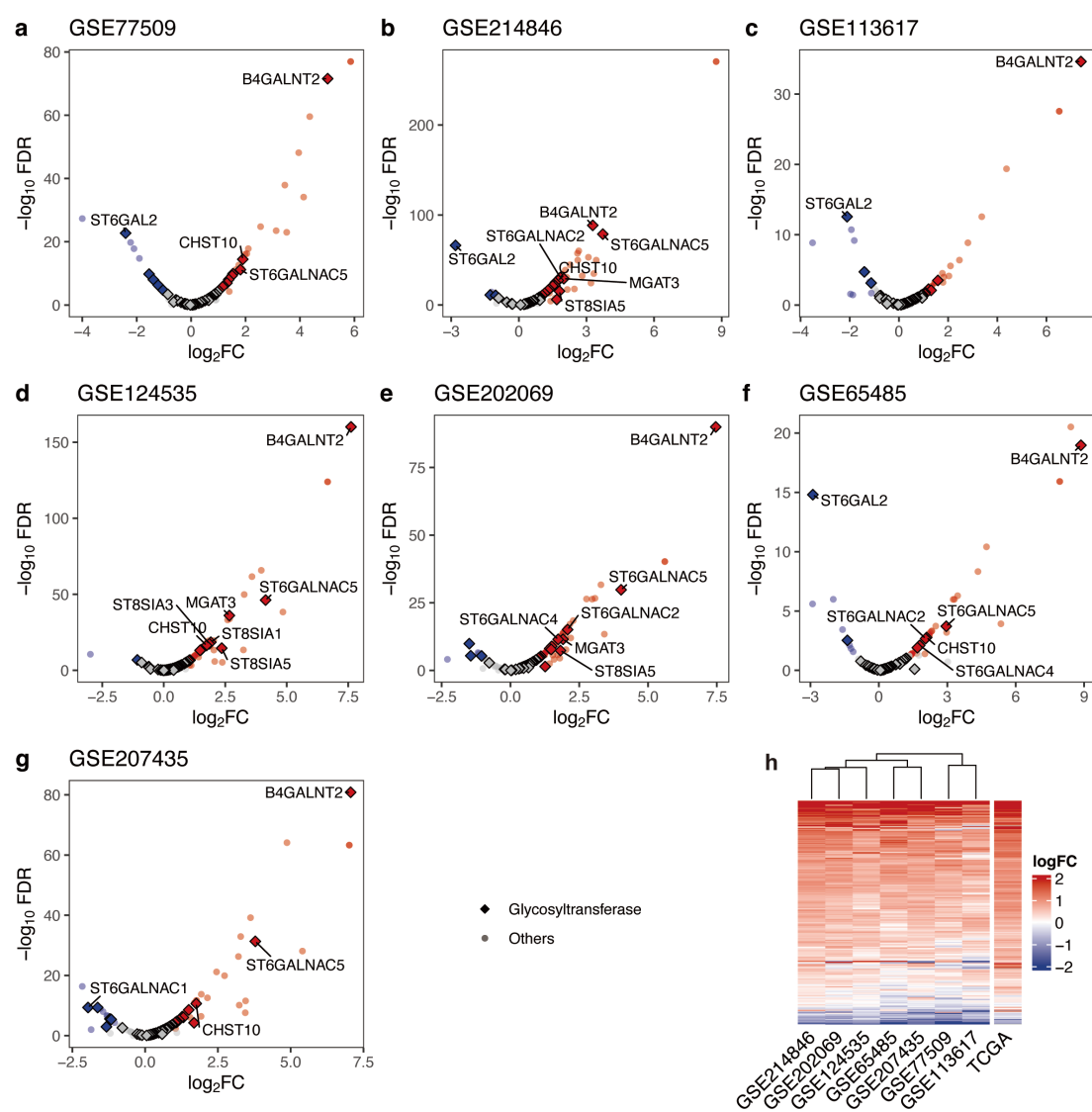
Supplementary Figure 17. Comparison of pseudo-glycome and real glycome for HC (a), CHB (b), LC (c), and HCC (d) pooled samples ($n = 30$ each). Source Data are provided with this paper.



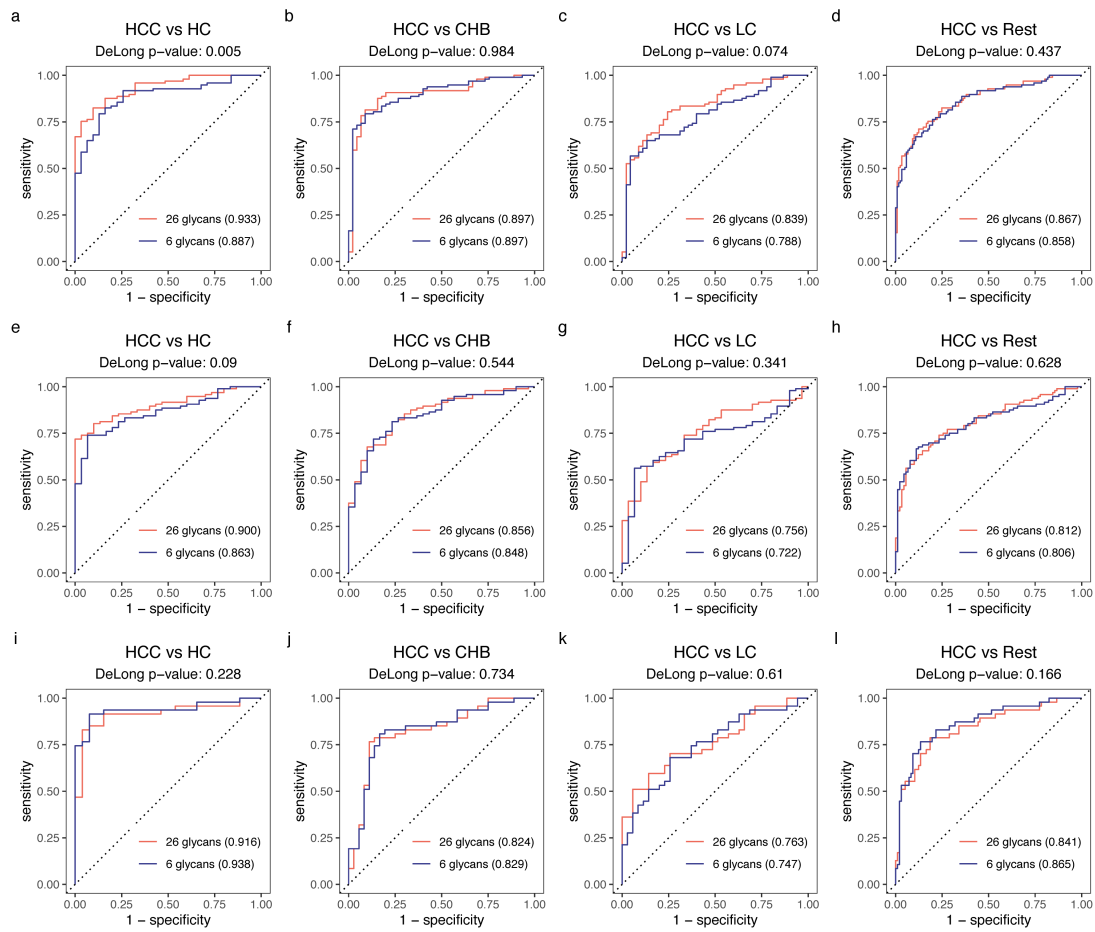
Supplementary Figure 18. The increase of tetra-antennary fucosylation multi-sialylated glycans being mainly a result of the elevation of A1AG1. a-c. Site contribution of H7N6F1S2 (a), H7N6F1S4 (b), and H7N6F1S3 (c). **d.** Average site-specific glycan abundance on A1AG1 N93 (stacked bar plot, left y-axis) and label-free quantification (LFQ) of A1AG1 (line plot, right y-axis). Error bars represent standard errors of protein LFQ. Source Data are provided with this paper.



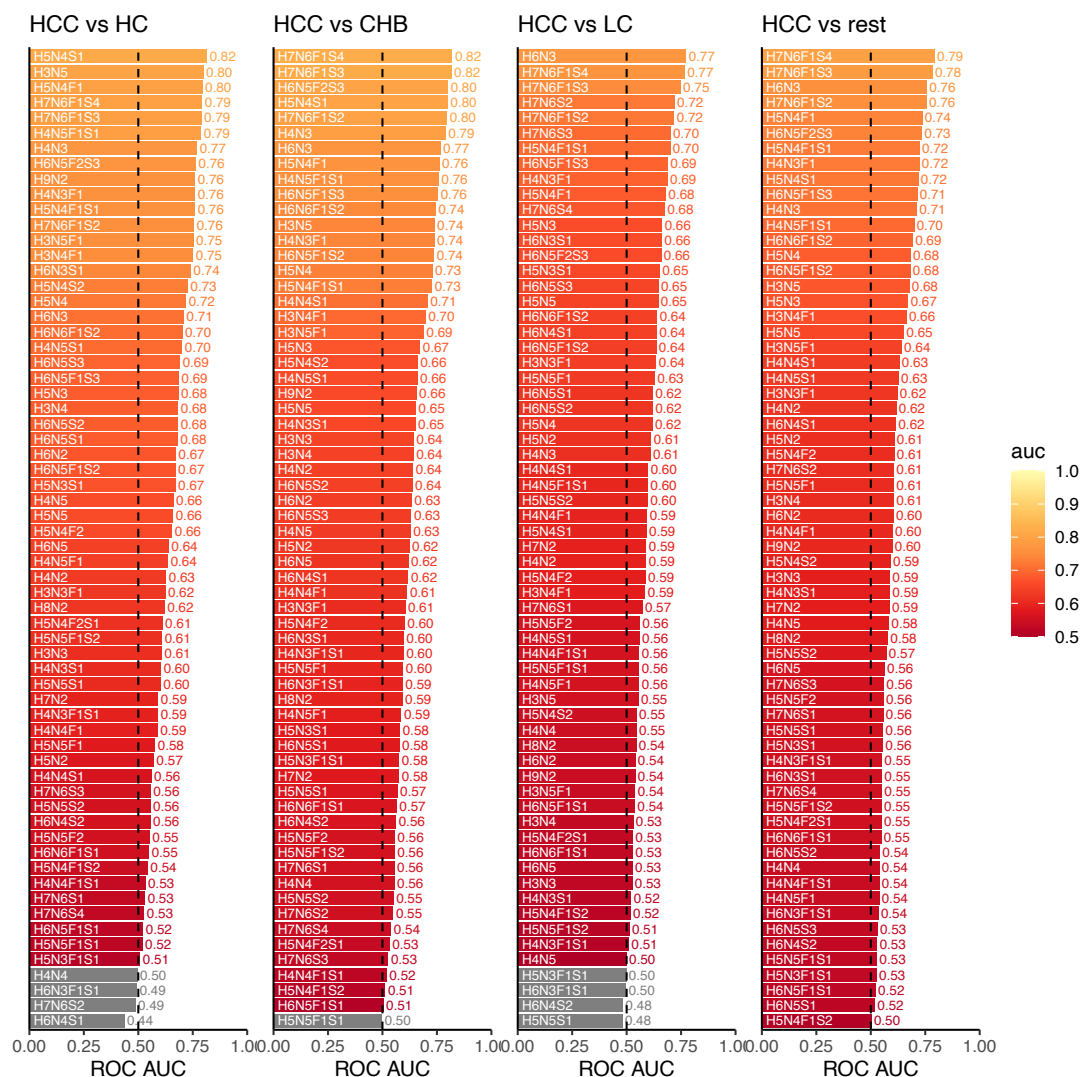
Supplementary Figure 19. a-d. Venn plots summarizing dysregulated site-specific glycans (SSGs) (a, b) and glycoproteins (GPs) (c, d). **e-f.** GO-BP (a) and KEGG (b) enrichment results for dysregulated glycoproteins. Source Data are provided with this paper.



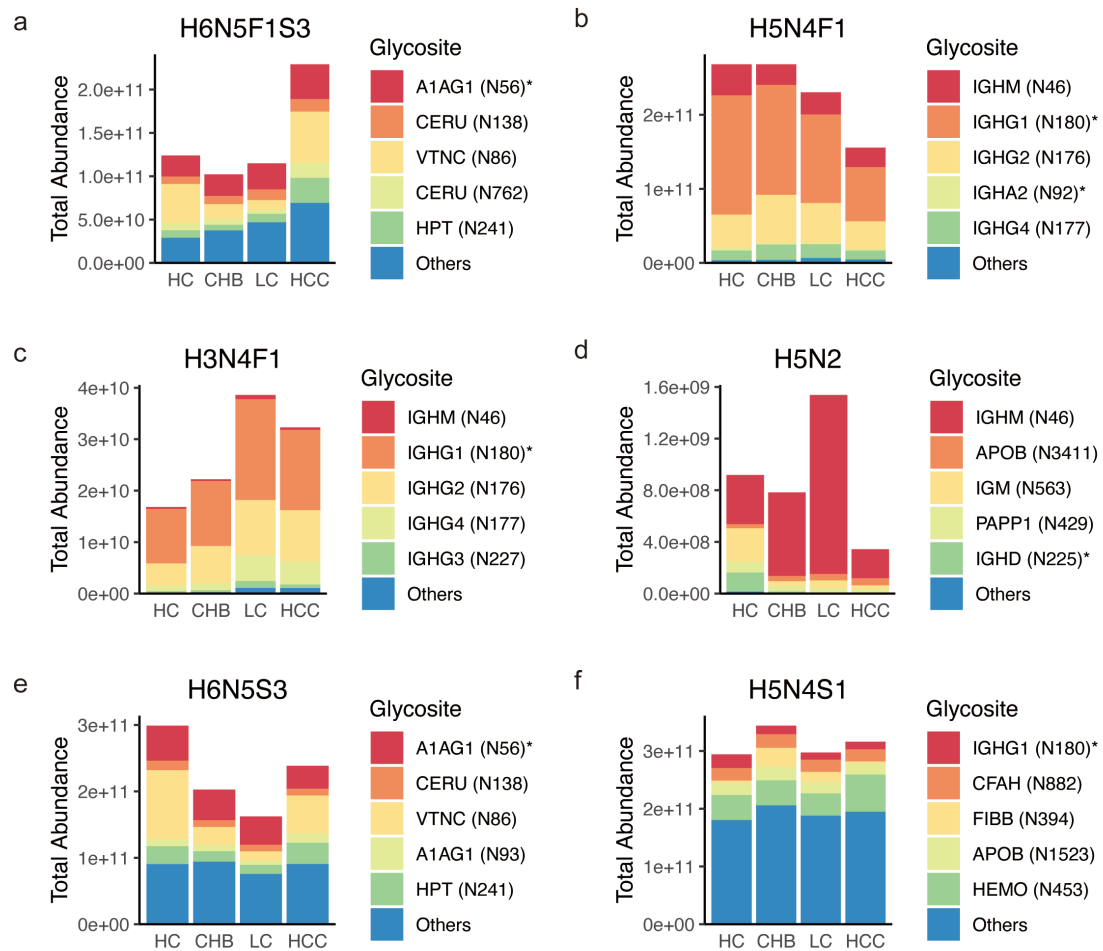
Supplementary Figure 20. a-g. Volcano plots of glycosyltransferases in the seven GEO datasets. **h.** Fold changes of 307 glycosyltransferases between tumor and non-tumor tissues across multiple GEO datasets and the TCGA-LIHC dataset. Source Data are provided with this paper.



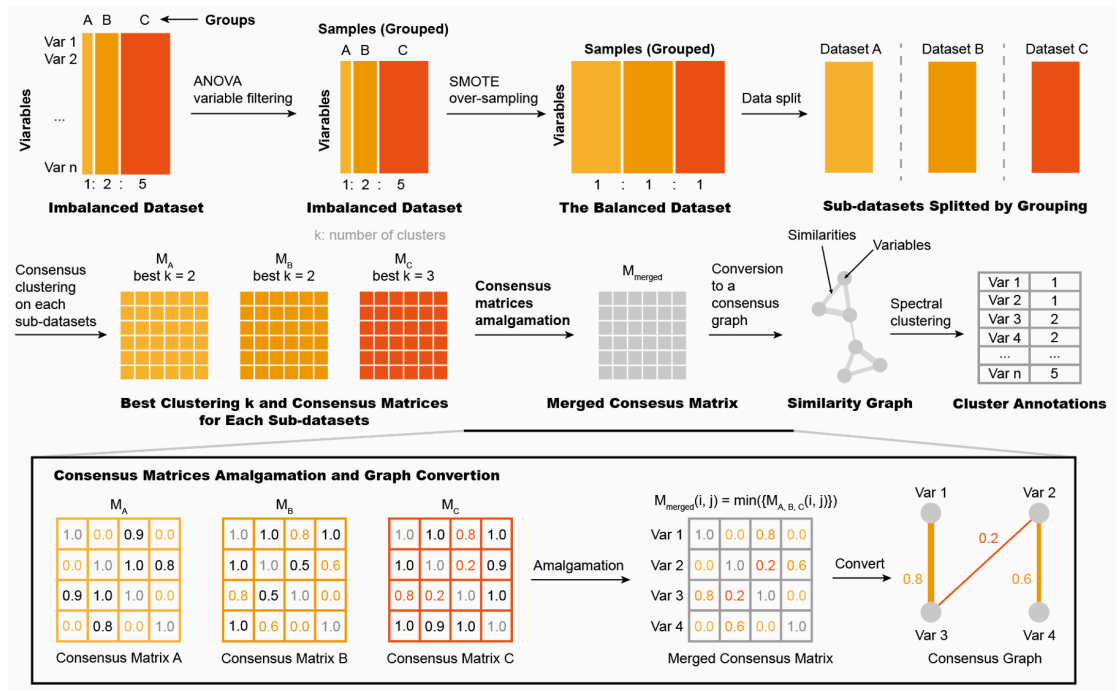
Supplementary Figure 22. ROC curves comparing models trained with all 26 glycans or the six most important glycans. Models were compared through in all three cohorts: (a-d) the test cohort, (e-h) the external validation cohort 1, and (i-l) the external validation cohort 2. ROC curves were compared using DeLong test. Source Data are provided with this paper.



Supplementary Figure 23. ROC AUC values for each glycan in diagnosing HCC. Source Data are provided with this paper.



Supplementary Figure 24. Site contribution of the most consistently important six glycan identified by SHAP algorithm. Source Data are provided with this paper.



Supplementary Figure 25. Workflow of the Minimum Consensus Spectral Clustering (MCSC) algorithm.