Supplementary data for site-specific *N*-glycoproteomic analysis of four M2 macrophage subtypes derived from human THP-1 cells

Pengfei Li[ab], Chen Ma[a], Lin Chen[a], Zexuan Chen[a], Shanshan You[ac], Jun Li[a], Shisheng Sun*[a]

*College of Life Sciences, Northwest University, 229 North Taibai Road, Life Science Building, Room 221, Xi'an, Shaanxi Province 710069, P. R. China

Email: suns@nwu.edu.cn

Phone: 86-29-88302411

Experimental section

Sample Source

Human monocytic cell line THP-1 was obtained from the Cell Bank of Chinese Academy of Sciences (Shanghai, China). THP-1 monocytes were cultured in RPMI 1640 medium (VivaCell, Shanghai, China), supplemented with 10% fetal bovine serum (Biolnd, Israel; VivaCell, Shanghai, China) and 1% penicillin–streptomycin (Solarbio, China) at 37 ℃ and 5% CO_2. The peptide samples of THP-1 cells-derived M2a, M2b, M2c, and M2d macrophage were obtained from our previous experiments, and the successful polarization of four M2 subtypes had been confirmed, as described previously [10].

Cell lysis and protein extraction

The denaturing buffer that contained 8 M urea (Merck, Darmstadt, Germany) and 1 M ammonium bicarbonate (Merck, Darmstadt, Germany) was added to cells for cell lysis. The denatured proteins were reduced by 5 mmol/L dithiothreitol at 37 ℃ for 1 h, and then alkylated by 15 mmol/L iodoacetamide (Merck, Darmstadt, Germany) in the dark for 30 min at room temperature. The remaining iodoacetamide was quenched by adding 2.5 mmol/L dithiothreitol and incubated for 10 min at room temperature. The protein solutions were first diluted two-fold with deionized water and then the proteins were digested by sequencing grade trypsin (Promega, Madison, WI, USA) with the ratio of 1:100 (trypsin to total protein, w/w) at 37 ℃ for 2 h. The solutions were further diluted four-fold with deionized water, and additional trypsin (trypsin to total protein, 1:100, w/w) was added with overnight incubation at 37 ℃ overnight. The sample solutions were acidified to pH<2 with trifluoroacetic acid (TFA), centrifuged at 15 000*g* for 15 min, and then the digested peptides in the supernatant were purified using a C18 column. The peptides were eluted from the column by 60% acetonitrile (ACN)/0.1% TFA. ACN and TFA were purchased from Thermo Fisher Scientific (Waltham, MA, USA).

Enrichment of N-Linked Intact Glycopeptides

The intact glycopeptides were enriched using hydrophilic interaction chromatography (HILIC) micro-column (Cat: VH950010-0, Agela Technologies, China), and the method has been described in detail in previous publications [11]. Briefly, peptides were diluted to final solvent composition of 80% ACN/1% TFA for HILIC enrichment. Before sample loading, the columns were washed twice each by 0.1% TFA and 80% ACN/0.2% TFA. The columns were washed three times with 80% ACN/0.2% TFA after samples were loaded. Finally, the glycopeptides were eluted in 0.1% formic acid (FA) solution (Thermo Fisher Scientific, USA). To reduce the biological variance, the samples from three biological replicates were pooled into one sample for further analyses.

Mass Spectrometry Analysis

Intact glycopeptides in each samples underwent three LC-MS/MS runs on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, Germany). About 1 μg

intact glycopeptides were separated by an Easy-nLC 1200 system with a 75 μm × 50 cm Acclaim PepMap-100 C18 separating column protected by a 75 μm × 2 cm guarding column. Mobile phase flow rate was 200 nL/min and consisted of 0.1% FA in water (A) and 80% ACN/0.1% FA (B). The gradient profile for LC separation was set as follows: 3–7% B for 3 min, 7–40% B for 200 min, 40–68% B for 20 min, 68–99% B for 4 min and 99% B for 13 min. The spray voltage (+) was set at 2,400 V. Orbitrap spectra with automatic gain control (AGC) of 4×10^5 were collected from 375–2,000 *m/z* at a resolution of 120,000 followed by oxonium ions (138.055 *m/z* and 204.087 *m/z*)-dependent HCD (33% collision energy) triggered 20% collision energy HCD MS/MS at a resolution of 30,000 using an isolation width of 2 *m/z*. Charge states from 2 to 7 were selected for MS/MS acquisition. Unassigned and singly charged ions were rejected. A dynamic exclusion time of 20 s was set for each precursor ion after being selected once. A representative total ion chromatogram (TIC) was provided to demonstrate the quality of the LC separation, the TIC showed well-dispersed peaks characteristic with uniform line shapes and intensities (Fig. S4).

Intact Glycopeptide Identification

As previously described [5], intact glycopeptides analyses were performed using the built-in glycan branch structure database from StrucGP and the human protein databases (UP000005640). The protein enzymatic digestion was set as trypsin with a maximum of two missed cleavage sites and the potential glycosite-containing peptides were screened with the N-X-S/T motif (X is any amino acid except Proline). The carbamidomethylation (C, +57.02 Da) was set as a fixed modification, and oxidization (M, +15.9949 Da) as a dynamic modification. The mass tolerances for MS1 and MS2 were set at 10 and 20 ppm, respectively. For the Y ions determination, an optional mass shift of ±1 or ±2 Da was allowed in addition to the 20-ppm mass tolerance in MS2. Finally, both peptide and glycan portions required<1% false discovery rates (FDR) for the intact glycopeptide identification.

Analysis and Visualization of *N*-Glycan Structures

The average number of GPSMs for each *N*-glycan structure was calculated from the three replicates and then normalized against the total number of GPSMs in each sample. GlycoVisual tool (https://zenodo.org) was used to visualize the results of StrucGP by the corresponding MS/MS spectra to demonstrate the principle of glycan type and structure determination in intact glycopeptide analysis [5].

Bioinformatics Analysis

These pathway enrichments, including Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), reactome pathway and protein–protein interaction (PPI) analyses (interaction score>0.4) were performed to search for the associated important pathway information and key glycoproteins using STRING database (https://string-db.org) [12]. Whole human genome as background, and $p<0.05$. The disease correlation analysis was performed by STRING database and ClueGO of Cytoscape software [13]

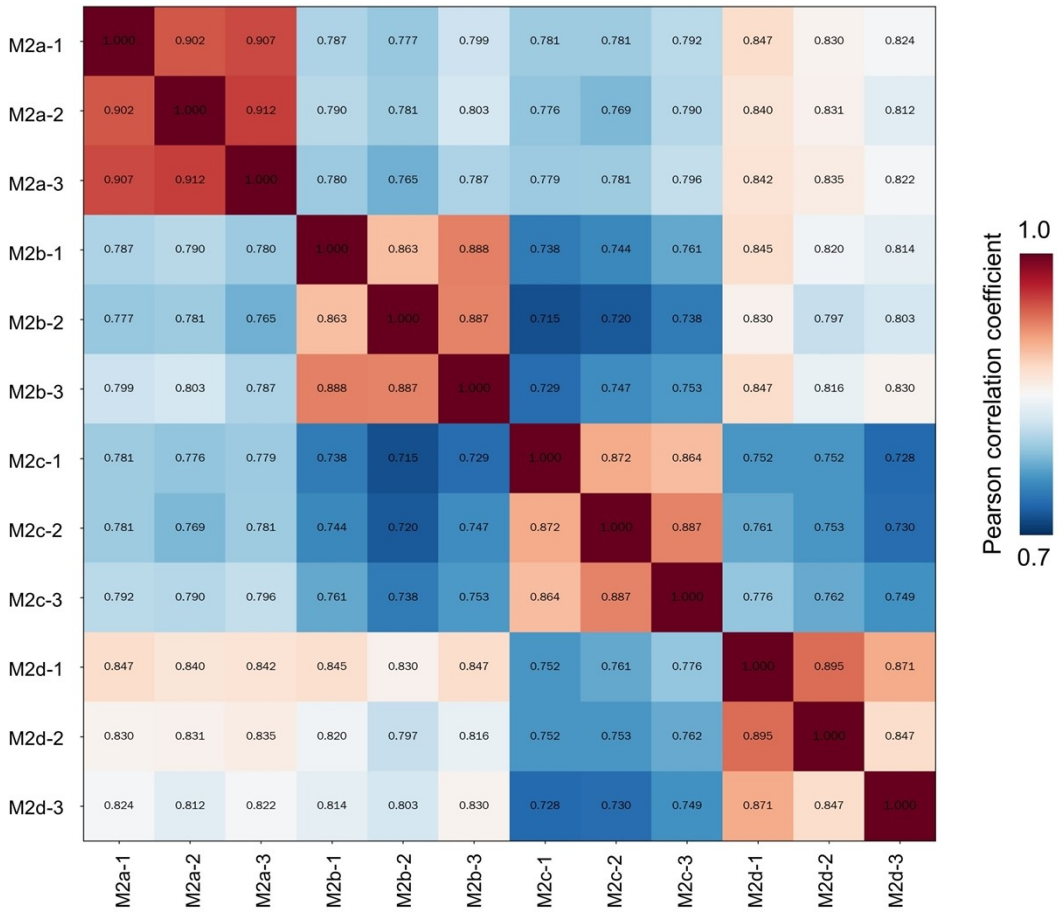

Fig. S1. Repeatability analysis was assessed through Pearson correlation coefficient. All aforementioned correlations were statistically significant (p<0.001).

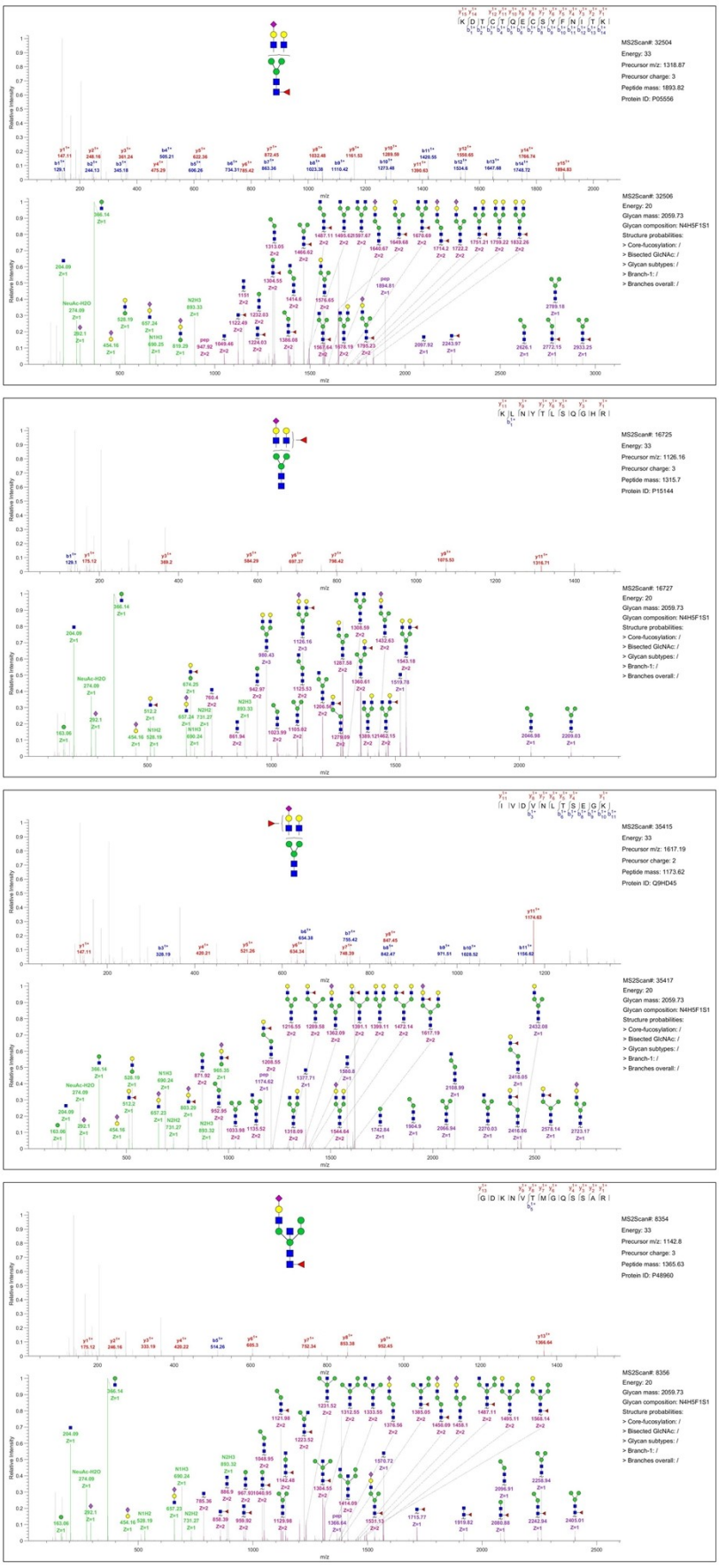

Fig. S2. Representative MS/MS spectra of four different isomers of N4H5F1S1.

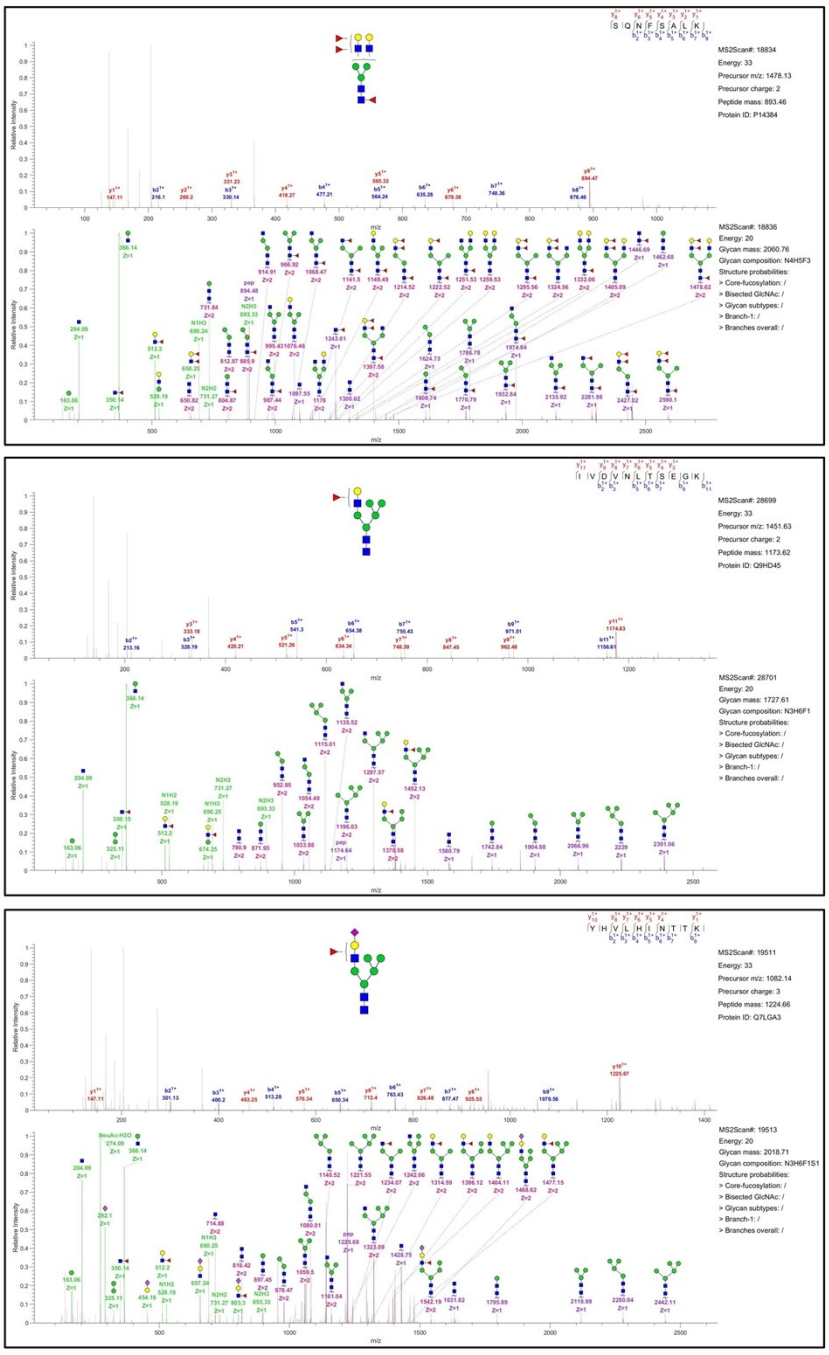

Fig. S3. Representative MS/MS spectra of Lewis structures.

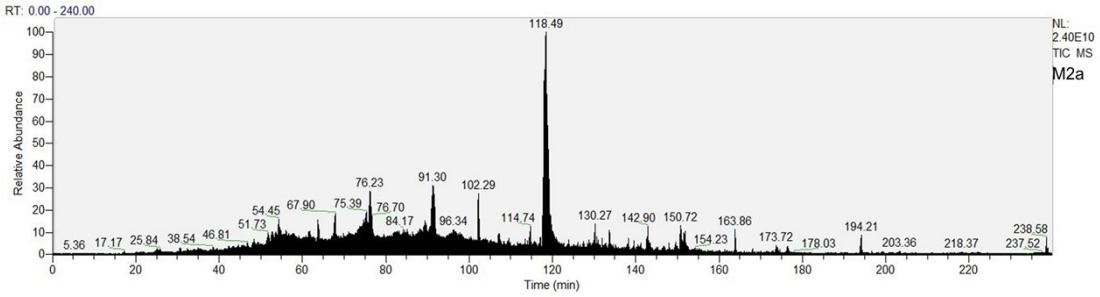

Fig. S4. TIC of M2a macrophage-derived glycopeptide samples.