Electronic Supplementary Material (ESI) for ChemComm.

Supplemental Information

Site- and structure-specific characterization of N-glycoprotein markers of MCF-7 cancer stem cells using isotopic-labeling quantitative N-glycoproteomics

Yue Wang[1,#], Feifei Xu[2,#], Kaijie Xiao[1], Yun Chen[2,*], Zhixin Tian[1,*]

[1]School of Chemical Science & Engineering, Shanghai Key Laboratory of Chemical Assessment and Sustainability, Tongji University, Shanghai 200092, China
[2]School of Pharmacy, Nanjing Medical University, Nanjing 211166, China

# These authors contribute equally. *Correspondence to: Zhixin Tian, zhixintian@tongji.edu.cn; Yun Chen, ychen@njmu.edu.cn

Contents

in MCF-7 CSCs relative to MCF-7 obtained from spectrum 19410 of TR1.

Figure S7. Identification of intact N-glycopeptide ANHSGAVVLLKR-N2H8F0S0 from Integrin alpha-6 (P23229, ITA6_HUMAN, N323) in MCF-7 CSCs obtained from spectrum 20996 of TR1.

Figure S8. Identification of intact N-glycopeptide YHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWR-N2H6F0S0 from CD151 antigen (P48509, CD151_HUMAN, N159) in MCF-7 obtained from spectrum 29643 of TR1.

Figure S9. Identification of intact N-glycopeptide YHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWR-N2H7F0S0 from CD151 antigen (P48509, CD151_HUMAN, N159) in MCF-7 obtained from spectrum 29643 of TR1, with a fold of 0.36 down-regulation which quantitated by the left two peaks.

Figure S10. Identification of intact N-glycopeptide LSAVNSIFLSHNNTK-N4H5F1S0 from CD97 antigen (P48960, CD97_HUMAN, N453) in MCF-7 CSCs obtained from spectrum 27516 of TR3.

Figure S11. Identification of intact N-glycopeptide AFNSTLPTMAQMEK-N5H6F1S3 from CD44 antigen (P16070, CD44_HUMAN, N57) in MCF-7 CSCs obtained from spectrum 30374 of TR3.

Figure S12. Identification of intact N-glycopeptide HLNGTITAK-N2H8F0S0 from CD109 antigen (Q6YHK3, CD109_HUMAN, N247) in MCF-7 CSCs obtained from spectrum 20021 of TR3.

Methods

Chemicals and Reagents.

Dithiothreitol (DTT, 3483-12-3), iodoacetamide (IAA, 144-48-9), 2,2,2-trifluoroethanol (TFE, ≥99%, 75-89-8), sodium cyanoborohydride (25895-60-7), acetaldehyde-$^{13}C_2$ (99 atom % ^{13}C, 1632-98-0), ammonium hydroxide solution (28-30% NH_3 basis, 1336-21-6), trifluoroacetic acid (TFA, 99%, 76-05-1), formic acid (FA, 64-18-6), trypsin (Sigma-Aldrich, St.Louis, MO, USA) and all HPLC solvents were purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetaldehyde solution (40% in H_2O, 75-07-0) was obtained from General Reagent (Shanghai). Ultrapure water was produced on site by Millipore Simplicity System (Billerica, MA, USA).

Cell culture of MCF-7 cells and MCF-7 CSCs.

The breast cancer cell line MCF-7 (ATCC, Manassas, VA, USA) was cultured in DMEM (Thermo Scientific Hyclone, MA, USA) supplied with 10% fetal bovine serum, 100 U/mL penicillin and 100 μg/mL streptomycin at 37°C and 5% CO_2. MCF-7 CSCs were kindly donated by Dr. Yongmei Yin[1] (Nanjing Medical University, China) and cultured with six-well plates (Corning, New York, USA) and MammoCult™ Human Medium Kit (Stem cell technologies, Vancouver, Canada) according to the manufacturer's guidelines. MCF-7 CSCs were directly isolated from MCF-7 cultures and both MCF-7 and MCF-7 CSCs were cultured at similar passage numbers. Briefly, cells were split every 5-7 days with 0.25% trypsin and between-splits feeding of fresh medium. In addition, we also examined the characteristics of each MCF-7 CSCs sample, including marker expression and mammosphere-forming efficiency, to confirm their stemness maintenance.[2, 3] In order to preserve the cellular properties of MCF-7 CSCs, the cells were kept up to 5 passages and their characteristics were regularly examined by flow cytometry.

In general, MCF-7 CSCs have a high expression of CD44 and a low expression of CD24.[4-6] Thus, CD44 and CD24 were used as cell-surface markers to isolate MCF-7 CSCs using flow cytometry, as done in previous studies. In detail, 1×10^6 of MCF-7 CSCs were stained with 1 μL anti-CD44-PE antibody (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) and 1 μL anti-CD24-APC antibody (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) in PBS at 4°C for 25 min. Cells were then washed three times with PBS and analyzed on a FACScan flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). The data were analyzed using the Cell Quest software (BD Biosciences, San Jose, CA). The result shows that the characteristics of MCF-7 CSCs have been maintained (Figure S1).

Protein extraction and trypsin digestion.

Cells (MCF-7 or MCF-7 CSCs) were disrupted on ice in 1 mL of lysis buffer (0.1 M Tris/HCl, 4% SDS, pH 8.0) by sonication (Ningbo Scientz Biotechnology CO, .LTD, China) for 15 min. The whole cell lysates were centrifuged at 14,000 rpm and 4 °C for 15 min, and the supernatant protein mixtures were collected. After acetone precipitation, proteins were dissolved in 1 mL of 8 M urea and diluted in 10 mL ultrapure water. Protein concentration was determined by BCA assay (SK3021, Sangon Biotech, Shanghai, China).

One mg of proteins were reduced with 20 mM DTT (20 min, 55 °C), alkylated with 20 mM iodoacetamide (in the dark, 30 min, RT), and digested with trypsin (1:50 w/w, 37 °C, 16 h, stopping reagent 0.5% TFA). The digests were desalted using house-made C18-tip and eluted with 400 μL of 50% ACN and 400 μL of 80% ACN. Desalted peptides were concentrated and stored at -20 °C for further use.

ZIC-HILIC enrichment of intact N-glycopeptides.

Intact N-glycopeptides were enriched using ZIC-HILIC (zwitterionic type of hydrophilic interaction chromatography) particles. Briefly, desalted peptides were redissolved in 80% ACN with 1% TFA (defined as washing buffer) and loaded onto a house-made pipette tip containing 30 mg ZIC-HILIC particles (Merk Millipore, 5 µm, 200 Å) which were pre-equilibrated with 0.1% TFA (defined as elution buffer). The tip was then washed using 800 µL washing buffer. Enriched N-glycopeptides were eluted with 400 µL elution buffer, dried in a vacuum concentrator, and finally resuspended in ultrapure water for further analysis.

Isotopic diethyl labelling of the enriched intact N-glycopeptides.

Stock solution of $NaBH_3CN$ (600 mM), CH_3CHO (20%, w/w), $^{13}CH_3{}^{13}CHO$ (20%, w/w), NH_4OH (4%, v/v) and formic acid (5%, v/v) were freshly made. Diethylation of N-terminal and lysine amino groups with CH_3CHO and $NaBH_3CN$ was carried out using the same protocol as reported for peptides. Two identical aliquots of MCF-7 and MCF-7 CSCs N-glycopeptides were enriched and re-suspended in 100 µL TFE, and 8 µL 20% acetaldehyde or acetaldehyde-$^{13}C_2$ was added. Subsequently, 8 µL freshly prepared 600 mM $NaBH_3CN$ was added and incubated at 37°C for 1 h, and the reaction was quenched with incubation with 8 µL 4% (v/v) NH_4OH for 1 min followed by addition of 6 µL 5% (v/v) FA. After concentrated, the labeled N-glycopeptides were desalted using house-made C18-tip and eluted with 250 µL of 50% ACN and 250 µL of 80% ACN. Desalted peptides were concentrated and re-suspended in ultrapure water for further analysis.

C18-RPLC-MS/MS (HCD) analysis of the 1:1 mixture of the labelled intact N-glycopeptides of MCF-7 CSCs and MCF-7 cells.

For one RPLC-MS/MS analysis, calculated 200 µg proteins from MCF-7 or MCF-7 CSCs were used as starting material (before ZIC-HILIC enrichment). The N-glycopeptides were separated on a 70 cm long analytical column (360 µm o.d. × 75 µm i.d.) packed with C18 particles (300 Å, 5 µm) on a Dionex Ultimate 3000 RSLC nano-HPLC system (Thermo Fisher Scientific) without the trap column. Buffer A is mixture of 99.8% H_2O and 0.2% FA; buffer B is mixture of 95.0% ACN, 4.8% H_2O, and 0.2% FA. Elution at a constant flow of 300 nL/min was conducted at the following gradient. The gradient was 4 h in total for complex samples: 2% buffer B for 25 min for sample-loading and 2-40% B in 135 min, followed by an increase to 95% B in 5 min, held for another 5 min and held for 2% B for the last 65 min for equilibration.

Eluted N-glycopeptides were detected online with nano-ESI tandem mass spectrometry using a Q Exactive mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). MS spectra were acquired in the 700-2000 *m/z* range using a mass resolution 70k (*m/z* 200). For MS/MS spectra, the mass resolution was set at 17.5k. Fragmentation was obtained in a data-dependent mode (Top20) with higher-energy collisional dissociation (HCD). The

automatic gain control (AGC) target value and maximum injection time were placed at 2 × 10^5 and 50 ms for MS and at 5 × 10^5 and 250 ms for MS/MS scans. Isolation window and dynamic exclusion were set at 3.0 *m/z* and 20.0 s. Stepped normalized collision energies was optimally set at 20.0%, 30.0%, and 40.0%. The temperature of the ion transfer capillary was set to 280 °C. The spray voltage was set to 2.8 kV.

With the above setting, three technical replicates were carried out, and three RPLC-MS/MS (HCD) datasets (TR1, TR2, and TR3) were acquired. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE[7] partner repository with the dataset identifier PXD013418.

Database search and identification of intact N-glycopeptides in MCF-7 CSCs and MCF-7 using GPSeeker.

The RPLC-MS/MS (HCD) datasets were searched by DB search engine GPSeeker for intact N-glycopeptide identification with FDR control; the details have been reported elsewhere and only a brief description is given here. Four theoretical customized human intact N-glycopeptides databases of two directions (forward and decoy) and two labels (light and heavy diethylation) were first created, and each dataset was searched against the four databases independently. The search parameters for the precursor and fragment ions are isotopic abundance cutoff (IPACO), isotopic peak *m/z* deviation (IPMD), and isotopic abundance deviation (IPAD); the adopted IPACO, IPMD, IPAD values for both the precursor and the fragment ions are 40%, 20 ppm, and IPAD=50%, respectively. Initial GPSMs were obtained with the following refinement criteria: Y1 ions, Top4; minimal percentage of matched fragment ions of N-glycosite-containing peptides, ≥10%; minimal matched product ions of N-glycan, ≥1; TopN hits, N=2 (top1 hits have the lowest P score); p-MPs (peptides matched product ions), ≥5.

For each dataset, the target and decoy GPSMs from search of the four databases were combined and ranked with increasing P score, and a cutoff P score was then chosen to achieve spectrum-level FDR≤1%. Target GPSMs with P scores lower than the cutoff value were grouped with the criteria of “peptide sequence, N-glycosite, and N-glycan linkage” for removal of duplicates and generation of the final list of intact N-glycopeptide IDs.

Quantitation of differentially expressed intact N-glycopeptides in MCF-7 CSCs relative to MCF-7 using GPSeekerQuan.

Relative quantitation of the identified intact N-glycopeptides was carried out using GPSeekerQuan. A mass tolerance of 20 ppm and mass difference of 4.01344 Da were adopted for the search of the paired isotopic envelopes of the precursor ions in the MS spectra; in each isotopic envelope, top3 isotopic peaks were adopted.

For each intact N-glycopeptide ID, all the six isotopic peaks are required to be observed

for each pair of isotopic envelopes; the peak abundance of the three isotopic peaks in each isotopic envelop was summed to obtain the relative ratio (MCF-7 CSCs/MCF-7). At least two ratios need to be observed among the three technical replicates. For the intact N-glycopeptides quantitated at least twice, the p-value was calculated using t-test[8]; and the intact N-glycopeptides with a fold change of no less than 1.5 and p value no bigger than 0.05 were classified as differentially expressed intact N-glycopeptides.

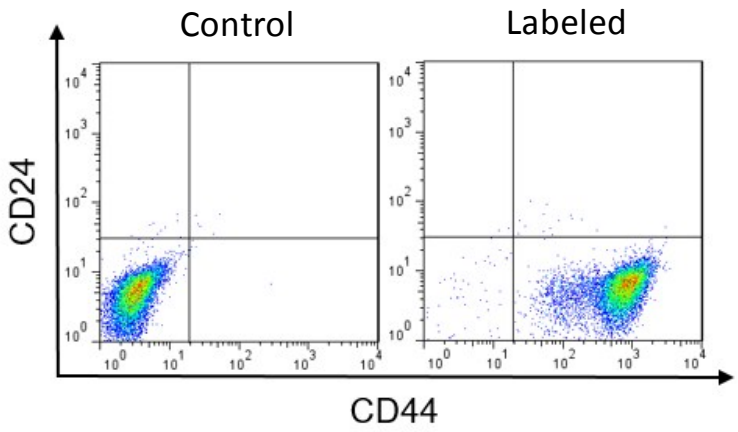

Figure S1. Flow cytometry of anti-CD24-APC and anti-CD44-PE antibodies-labeled MCF-7 CSCs and the control.

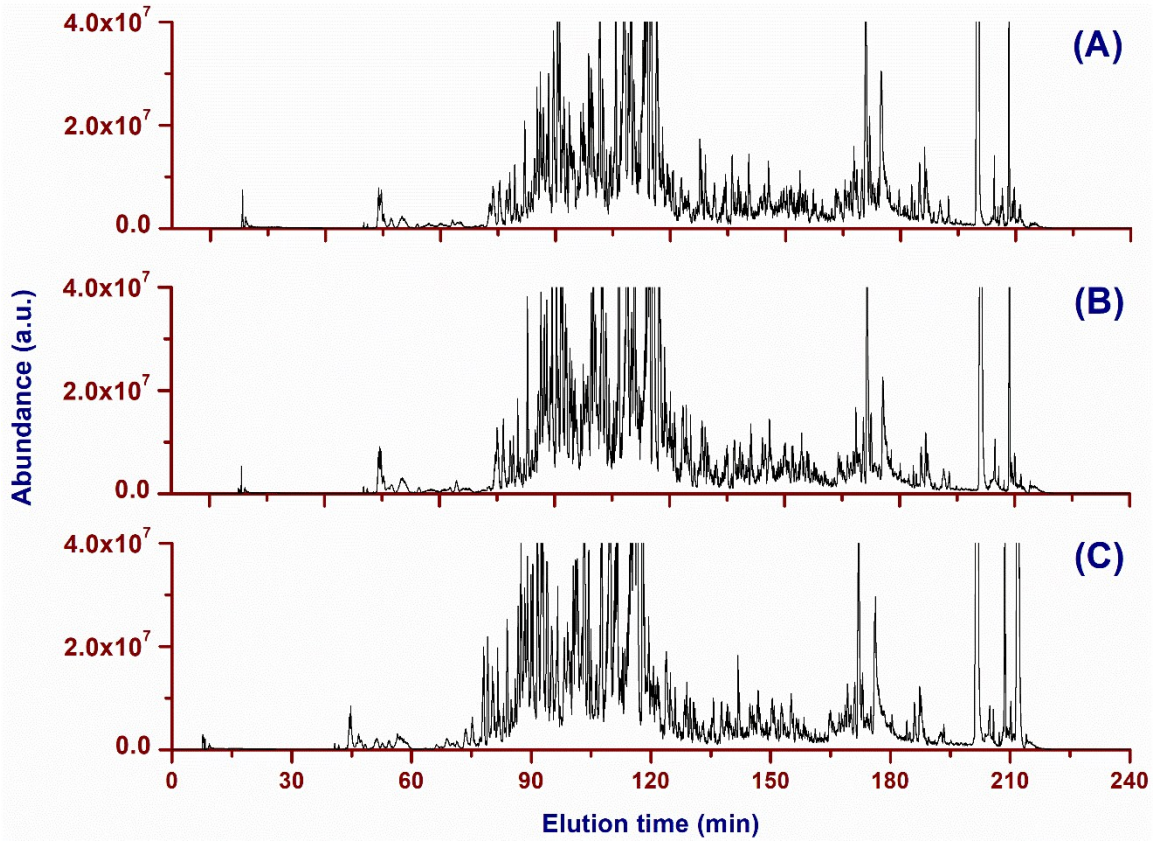

Figure S2. MS-only base-peak chromatograms from RPLC-MS/MS analysis of the 1:1 mixture of the light- and heavy-diethylated intact N-glycopeptides enriched from MCF-7 cells and MCF-7 cancer stem cells; (A, B, C), three technical replicates.

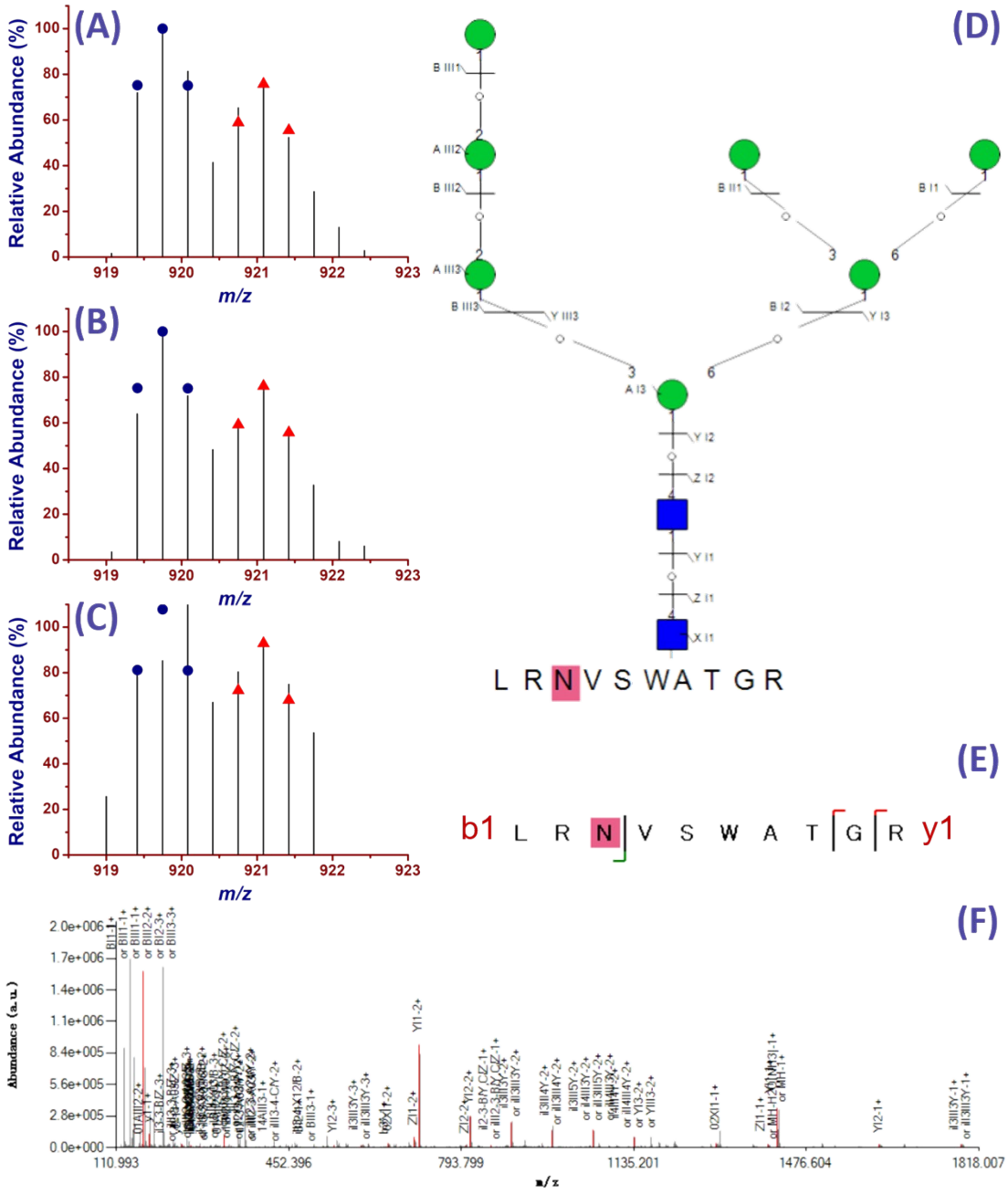

Figure S3. Differential expression with a fold of 0.80±0.03 down-regulation of intact N-glycopeptide LRNVSWATGR-N2H7F0S0 from Monocyte differentiation antigen CD14 (P08571, CD14_HUMAN, N151) in MCF-7 CSCs relative to MCF-7 as quantified from TR1 (A), TR2 (B) and TR3 (C). The isotopic envelope fingerprinting map fingerprinting of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (D) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M21M)61M(31M)61M. (E) Graphical fragmentation map of peptide sequence LRNVSWATGR. (F) Annotated MS/MS spectrum 25452 from TR2 of peptide sequence and N-glycan chain identified using GPSeeker.

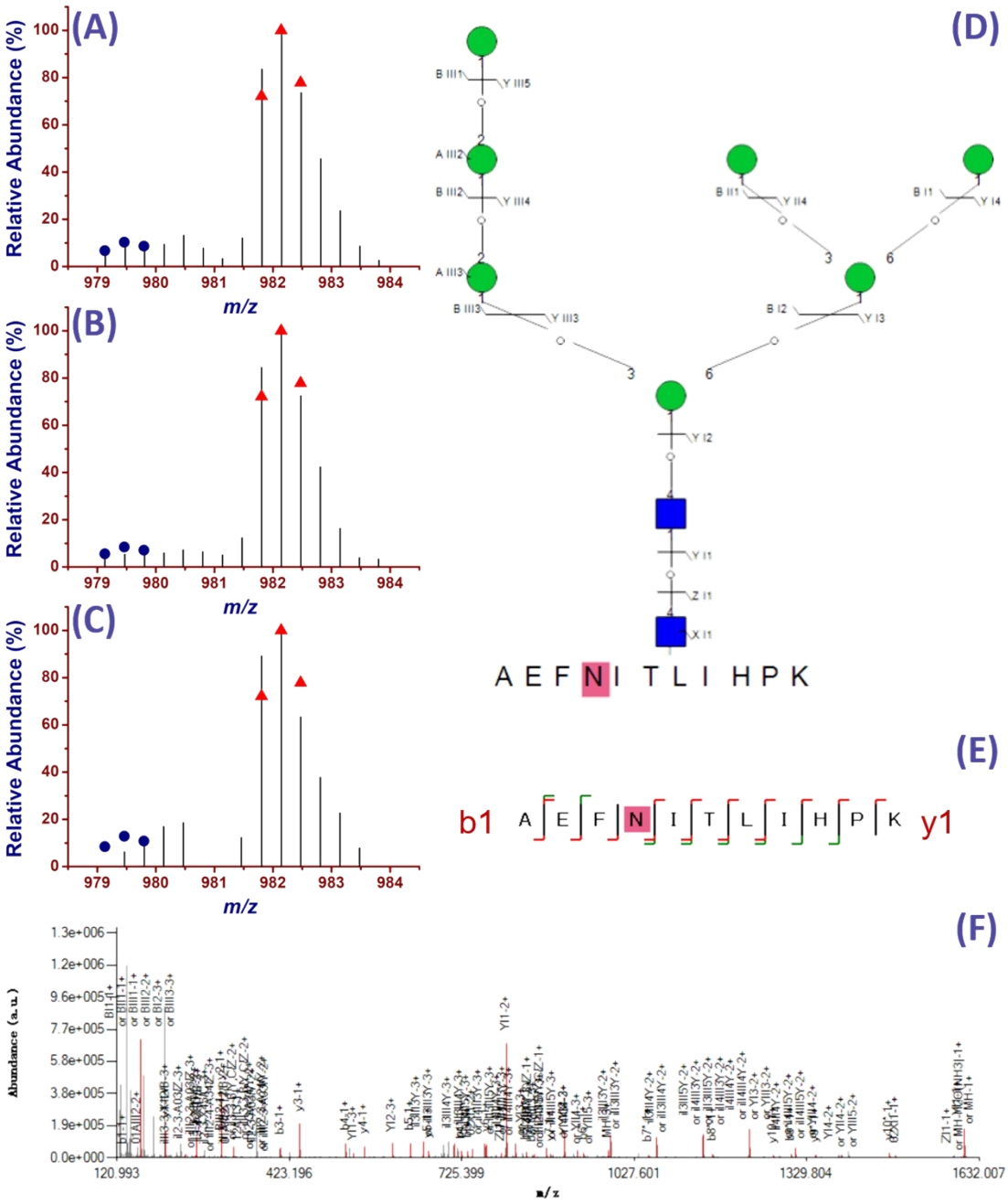

Figure S4. Differential expression with a fold of 12.03±1.07 up-regulation of intact N-glycopeptide AEFNITLIHPK-N2H7F0S0 from Aminopeptidase N (P15144, AMPN_HUMAN, N234) in MCF-7 CSCs relative to MCF-7 as quantified from TR1 (A), TR2 (B) and TR3 (C). The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (D) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M21M)61M(31M)61M. (E) Graphical fragmentation map of peptide sequence AEFNITLIHPK. (F) Annotated MS/MS spectrum 28897 from TR1 of peptide sequence and N-glycan chain identified using GPSeeker.

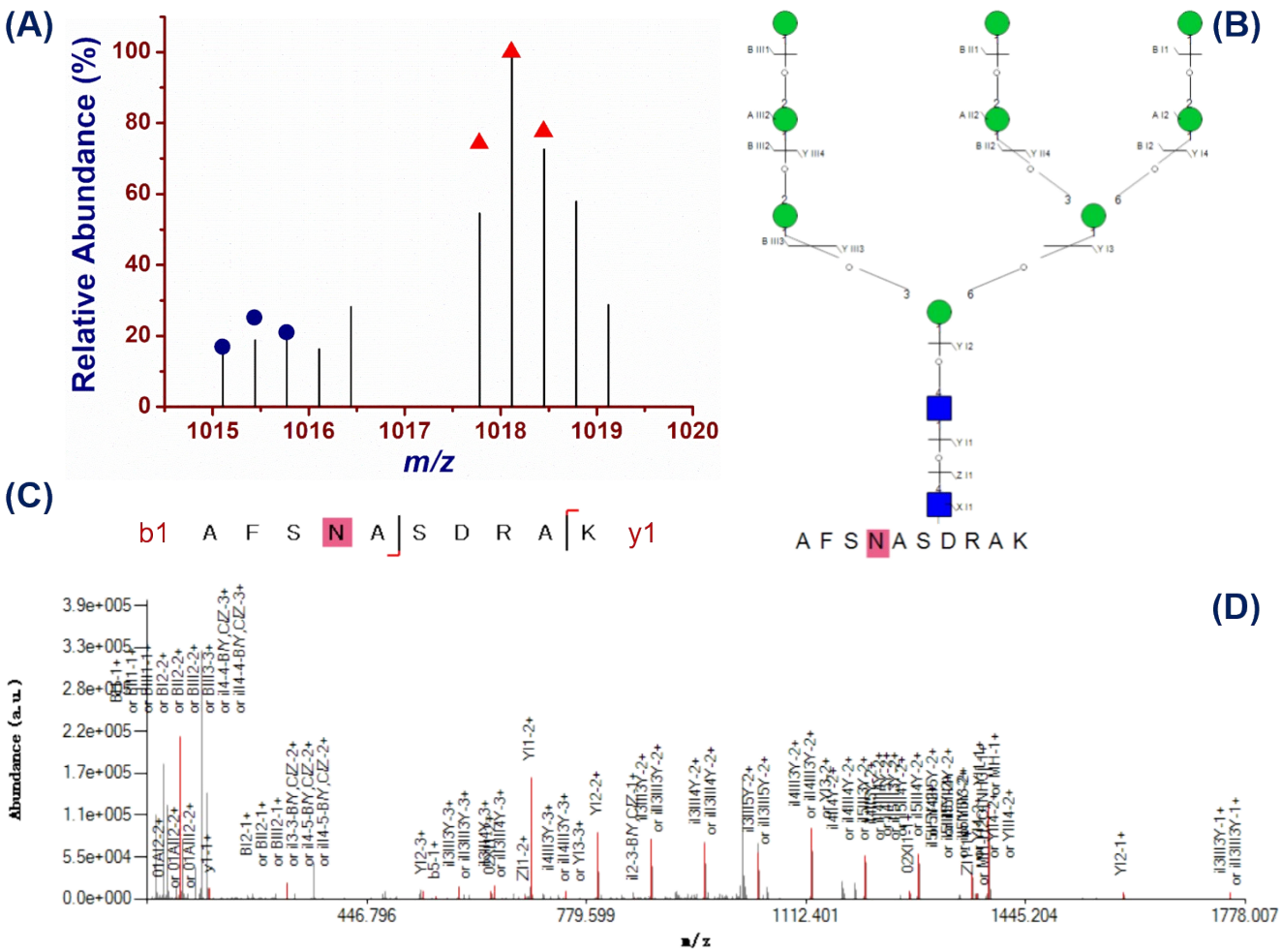

Figure S5. Differential expression with a fold of 4.03 up-regulation of intact N-glycopeptide AFSNASDRAK-N2H9F0S0 from Zinc finger protein GLI1 (P08151, GLI1_HUMAN, N344) in MCF-7 CSCs relative to MCF-7 obtained from spectrum 19407 of TR3. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M21M)61M(31M21M)61M21M. (C) Graphical fragmentation map of peptide sequence AFSNASDRAK. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

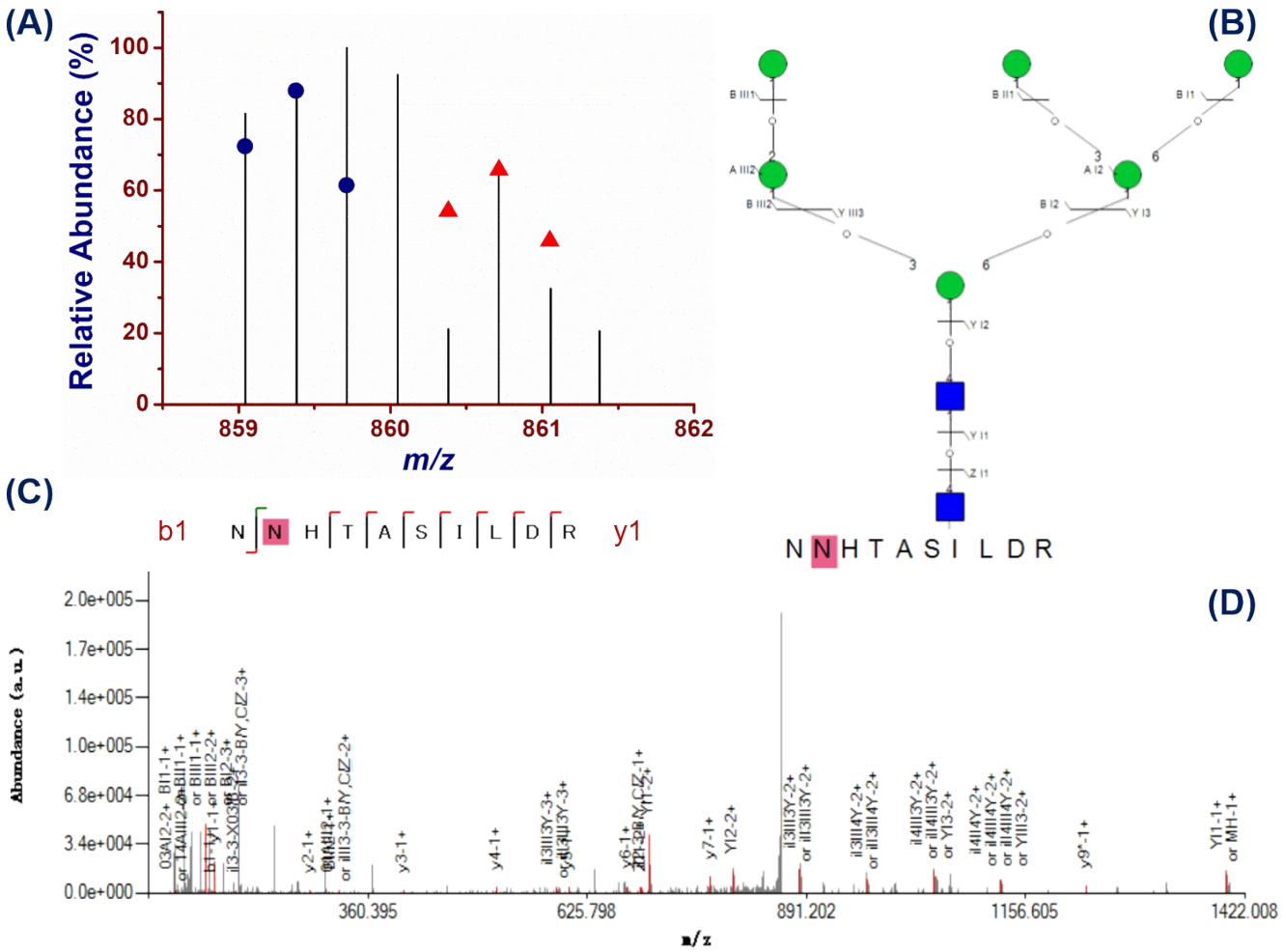

Figure S6. Differential expression with a fold of 0.44 down-regulation of intact N-glycopeptide NNHTASILDR-N2H6F0S0 from CD63 antigen (P08962, CD63_HUMAN, N130) in MCF-7 CSCs relative to MCF-7 obtained from spectrum 19410 of TR1. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M)61M(31M)61M. (C) Graphical fragmentation map of peptide sequence NNHTASILDR. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

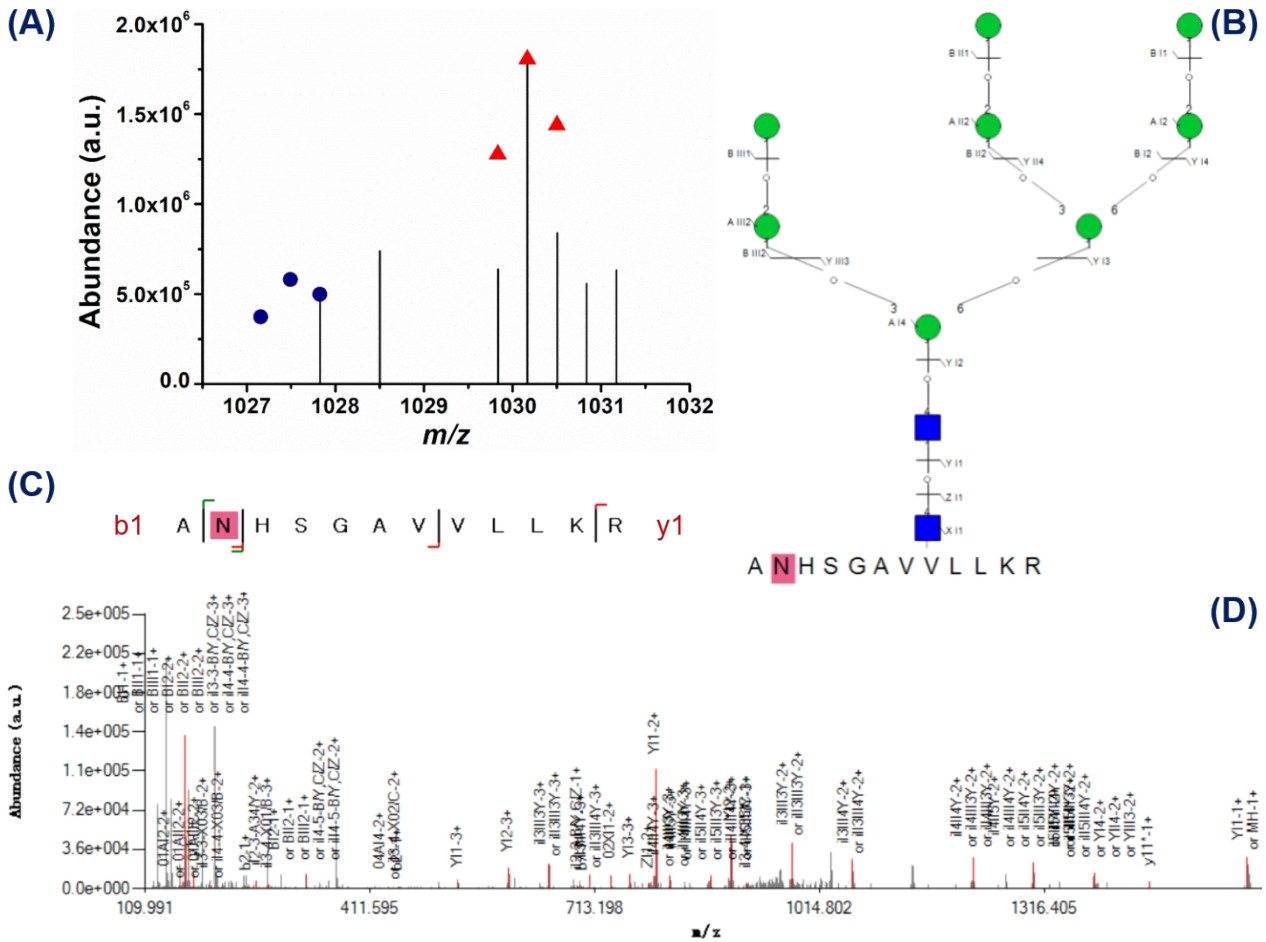

Figure S7. Identification of intact N-glycopeptide ANHSGAVVLLKR-N2H8F0S0 from Integrin alpha-6 (P23229, ITA6_HUMAN, N323) in MCF-7 CSCs obtained from spectrum 20996 of TR1, with a fold of 1.69 up-regulation which quantitated by the single right peak. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M)61M(31M21M)61M21M. (C) Graphical fragmentation map of peptide sequence ANHSGAVVLLKR. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

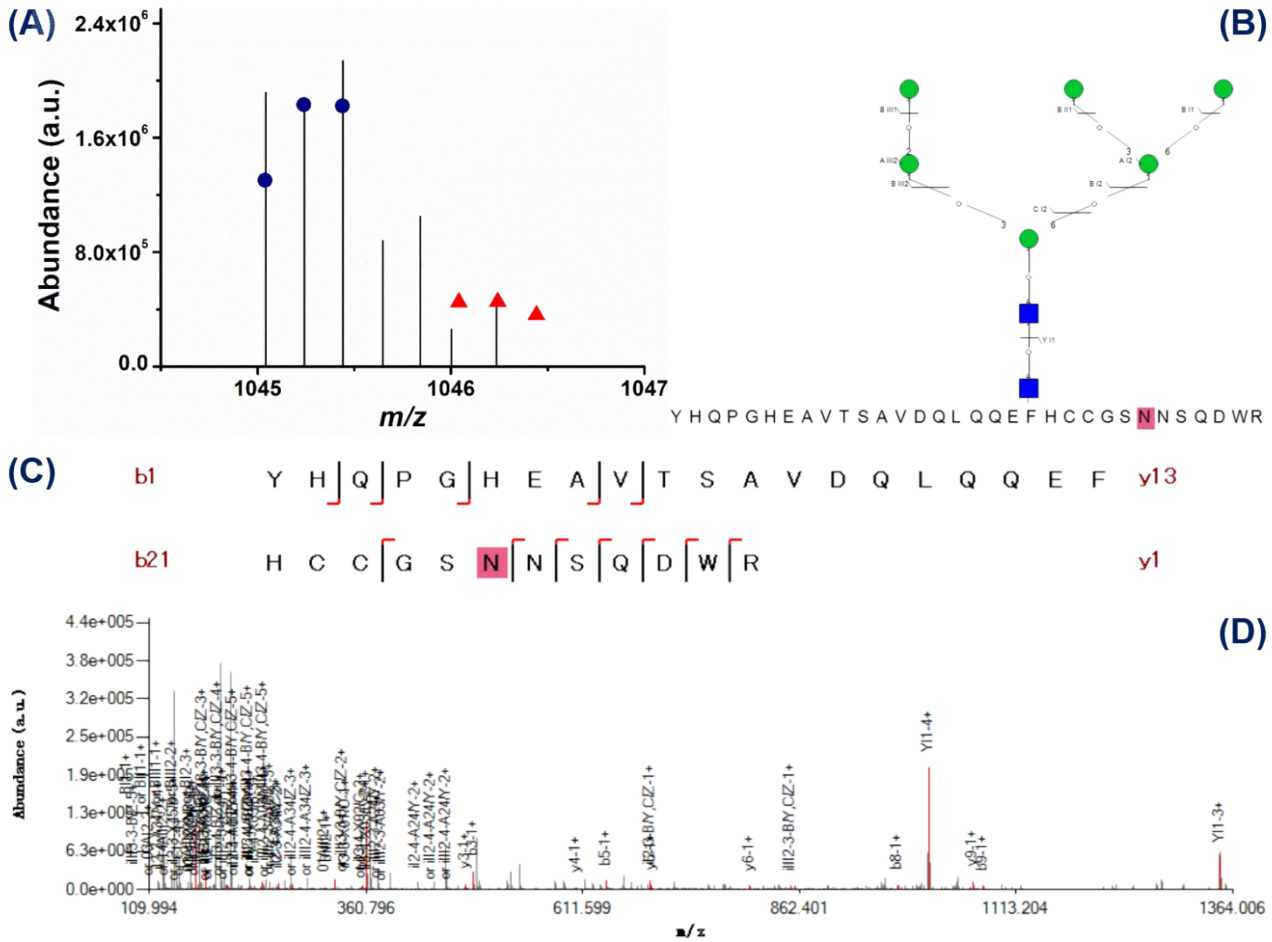

Figure S8. Identification of intact N-glycopeptide YHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWR-N2H6F0S0 from CD151 antigen (P48509, CD151_HUMAN, N159) in MCF-7 obtained from spectrum 29643 of TR1, with a fold of 0.19 down-regulation which quantitated by the left two peaks. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M)61M(31M)61M. (C) Graphical fragmentation map of peptide sequence YHQPGHEAVTSAVDQLQQ EFHCCGSNNSQDWR. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

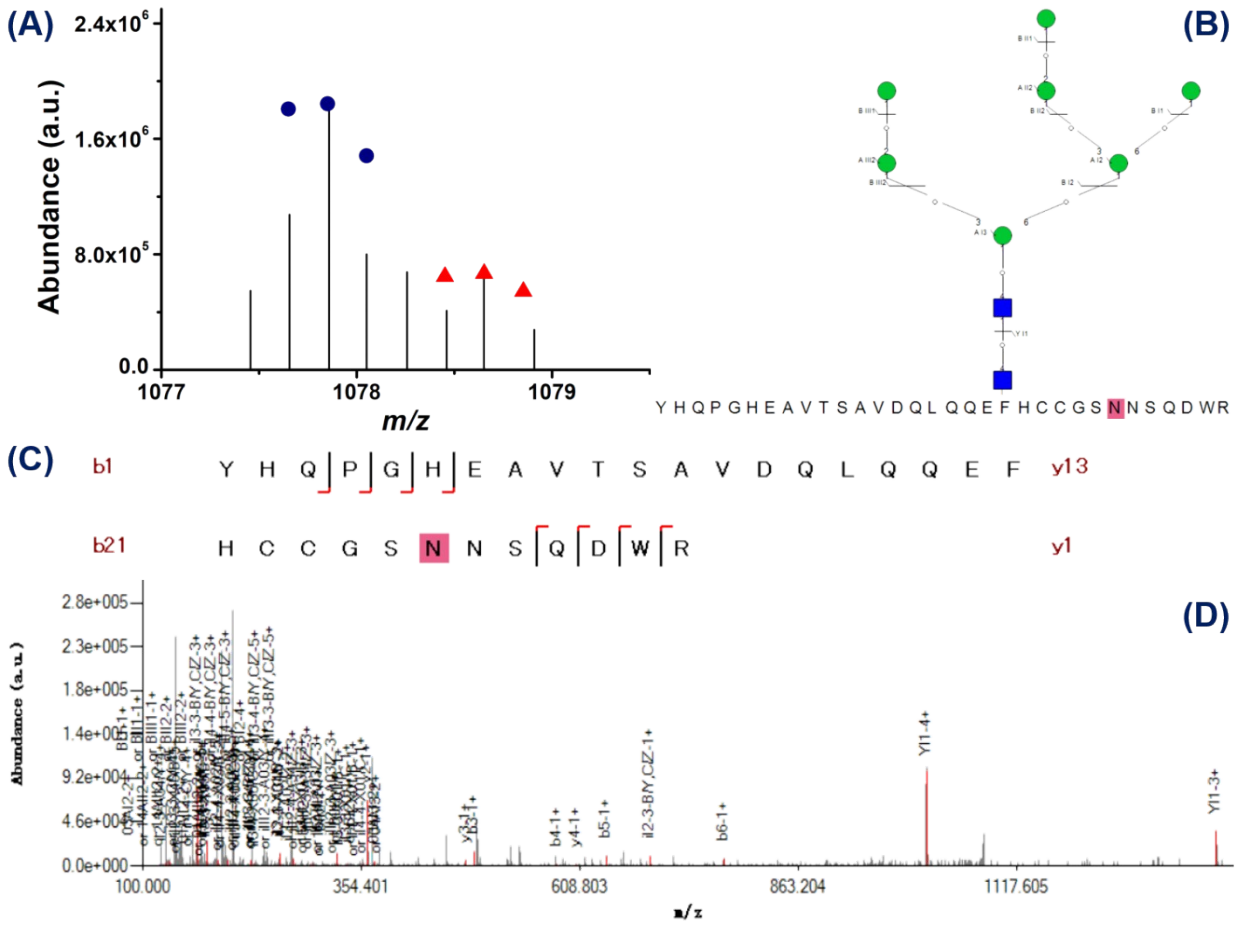

Figure S9. Identification of intact N-glycopeptide YHQPGHEAVTSAVDQLQQEFHCCGSNNSQDWR-N2H7F0S0 from CD151 antigen (P48509, CD151_HUMAN, N159) in MCF-7 obtained from spectrum 29643 of TR1, with a fold of 0.36 down-regulation which quantitated by the left two peaks. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M)61M(31M21M)61M. (C) Graphical fragmentation map of peptide sequence YHQPGHEAVTSAVDQLQQ EFHCCGSNNSQDWR. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

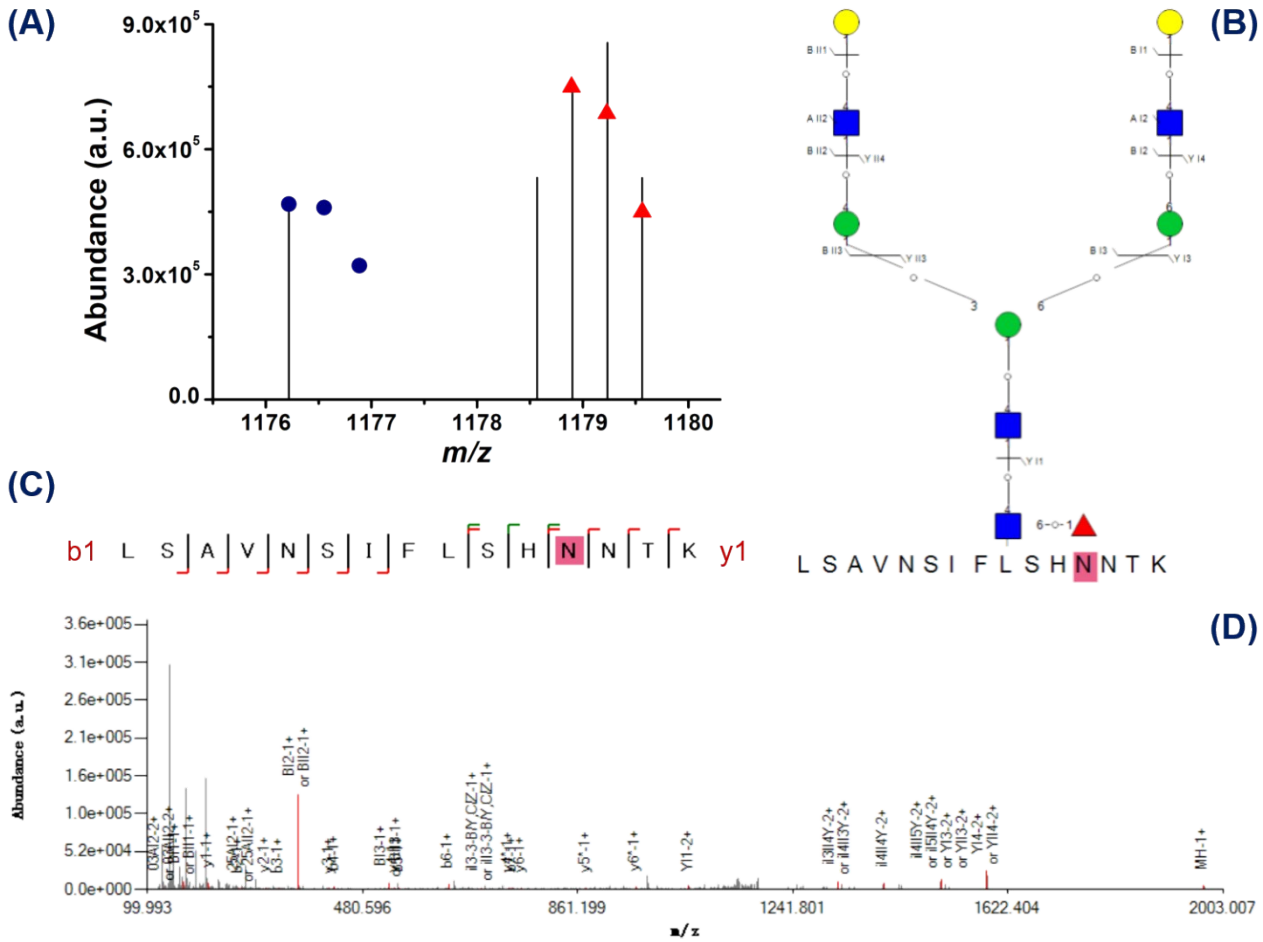

Figure S10. Identification of intact N-glycopeptide LSAVNSIFLSHNNTK-N4H5F1S0 from CD97 antigen (P48960, CD97_HUMAN, N453) in MCF-7 CSCs obtained from spectrum 27516 of TR3, with a fold of 1.60 up-regulation which quantitated by the single left peak. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y(61F)41Y41M(31M41Y41L) 61M61Y41L. (C) Graphical fragmentation map of peptide sequence LSAVNSIFLSHNNTK. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

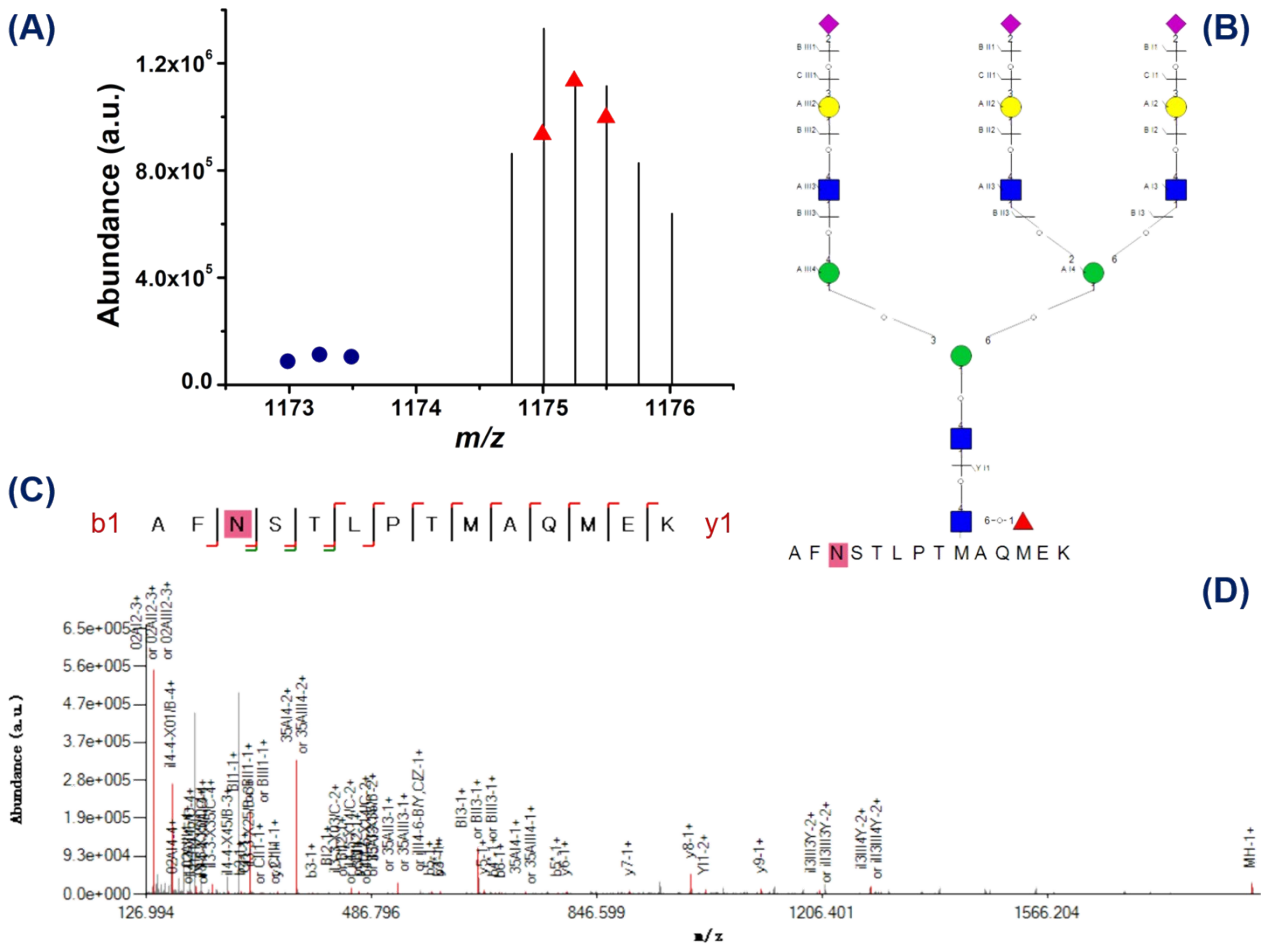

Figure S11. Identification of intact N-glycopeptide AFNSTLPTMAQMEK-N5H6F1S3 from CD44 antigen (P16070, CD44_HUMAN, N57) in MCF-7 CSCs obtained from spectrum 30374 of TR3, corresponding N-glycopeptide in MCF-7 not detected. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y(61F)41Y41M(31M41Y41L32S)61M (21Y41L32S)61Y41L32S. (C) Graphical fragmentation map of peptide sequence AFNSTLPTMAQMEK. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

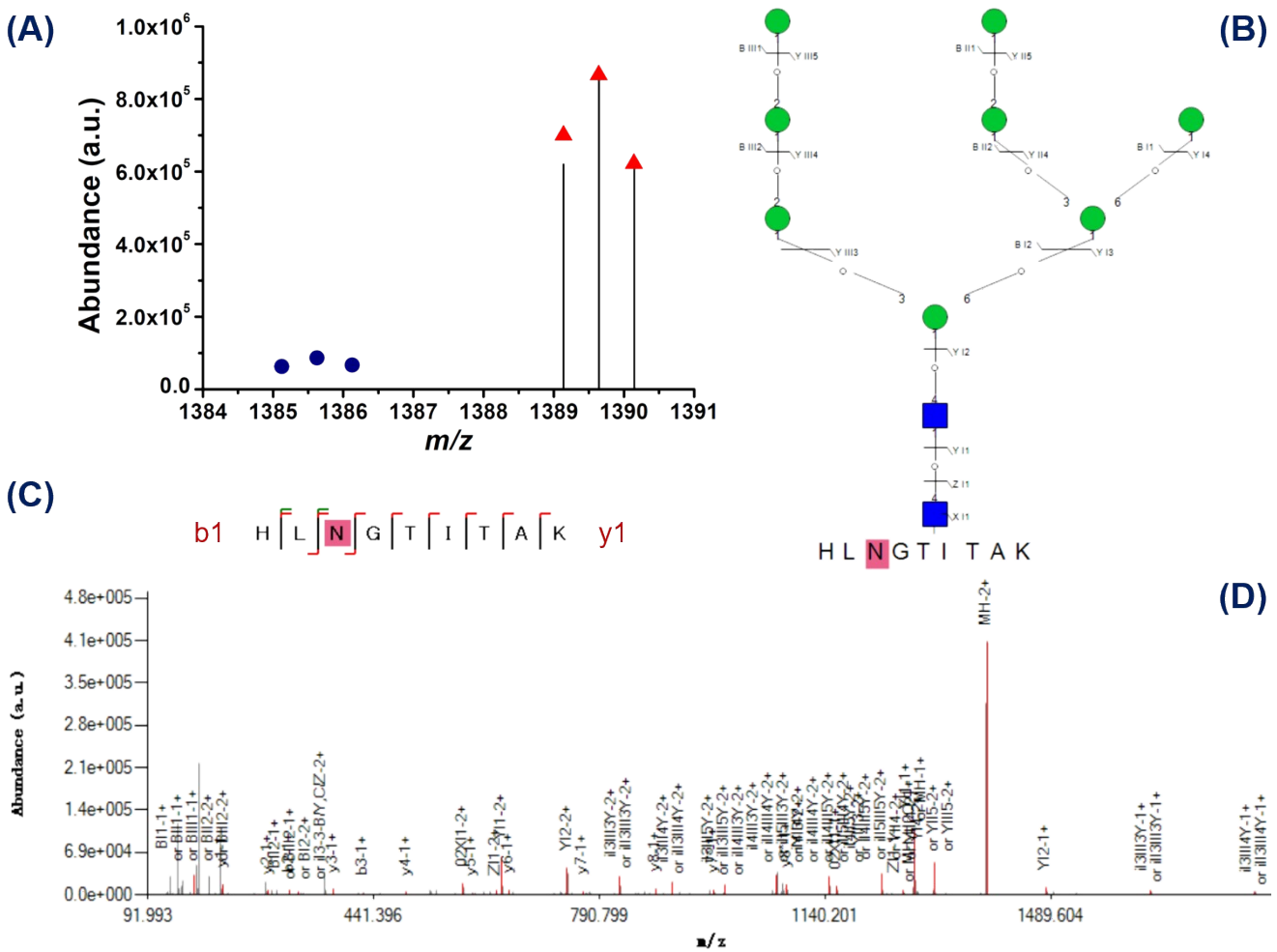

Figure S12. Identification of intact N-glycopeptide HLNGTITAK-N2H8F0S0 from CD109 antigen (Q6YHK3, CD109_HUMAN, N247) in MCF-7 CSCs obtained from spectrum 20021 of TR3, corresponding N-glycopeptide in MCF-7 not detected. (A) The isotopic envelope fingerprinting map of the experimental isotopic envelope (vertical lines) on the corresponding theoretical isotopic envelope (quantitative peaks, navy and red scatters). (B) Graphical fragmentation map of N-glycan chain 01Y41Y41M(31M21M21M)61M(31M21M)61M. (C) Graphical fragmentation map of peptide sequence HLNGTITAK. (D) Annotated MS/MS spectrum of peptide sequence and N-glycan chain identified using GPSeeker.

References

1. J. H. Chu, Y. F. Li, X. M. Fan, J. J. Ma, J. Li, G. P. Lu, Y. H. Zhang, Y. Huang, W. Li, X. Huang, Z. Y. Fu, Y. M. Yin and H. Y. Yuan, *Cell Physiol Biochem*, 2018, **48**, 593-604.
2. X. T. Lin, W. Y. Chen, F. Q. Wei, B. P. Zhou, M. C. Hung and X. M. Xie, *Int J Cancer*, 2017, **140**, 2517-2525.
3. F. Diomede, T. S. Rajan, V. Gatta, M. D'Aurora, I. Merciaro, M. Marchisio, A. Muttini, S. Caputi, P. Bramanti, E. Mazzon and O. Trubiani, *Stem Cells Int*, 2017, DOI: 10.1155/2017/5651287.
4. M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison and M. F. Clarke, *P Natl Acad Sci USA*, 2003, **100**, 3983-3988.
5. J. E. Visvader and G. J. Lindeman, *Nature Reviews Cancer*, 2008, **8**, 755-768.
6. J. E. Visvader and G. J. Lindeman, *Cell Stem Cell*, 2012, **10**, 717-728.
7. Y. Perez-Riverol, A. Csordas, J. W. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma and J. A. Vizcaino, *Nucleic Acids Res*, 2019, **47**, D442-D450.
8. Z. Chen, Q. Yu, L. Hao, F. Liu, J. Johnson, Z. Tian, W. J. Kao, W. Xu and L. Li, *The Analyst*, 2018, **143**, 2508-2519.