Supplemental Information

Global mapping of glycosylation pathways in human-derived cells

Yi-Fan Huang, Kazuhiro Aoki, Sachiko Akase, Mayumi Ishihara, Yi-Shi Liu, Ganglong Yang, Yasuhiko Kizuka, Shuji Mizumoto, Michael Tiemeyer, Xiao-Dong Gao, Kiyoko F. Aoki-Kinoshita, and Morihisa Fujita

Figure S1

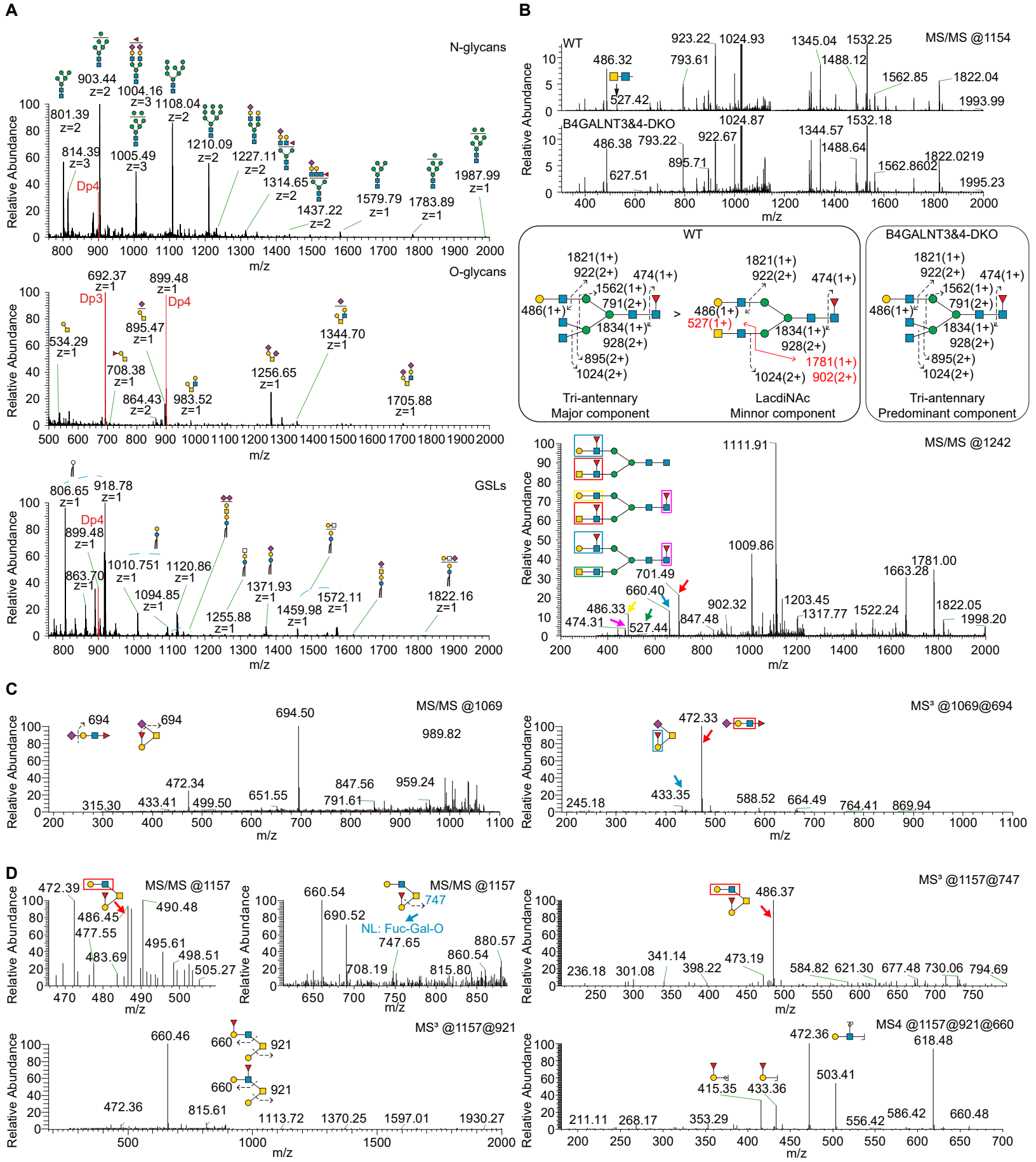

A
N-glycans
Relative Abundance
m/z
Dp4
O-glycans
Dp3
Dp4
GSLs
B
WT
B4GALNT3&4-DKO
MS/MS @1154
MS/MS @1242
Tri-antennary
Major component
LacdiNAc
Minnor component
Tri-antennary
Predominant component
C
MS/MS @1069
MS³ @1069@694
D
MS/MS @1157
MS³ @1157@747
MS³ @1157@921
MS4 @1157@921@660
NL: Fuc-Gal-O

Figure S1. Related to Figure 1: Glycomic signature of HEK293 cells

A, HEK293 cells were subjected to unbiased glycomic profiling for the characterization of N-glycans, O-glycans, and GSLs. All glycan components were permethylated and analyzed by NSI-MS. Permethylated malto-oligosaccharides were used as an external standard for quantification. An equal portion of each glycan species was injected onto NSI-MS in order to quantify their expression level. Top, A series of oligomannose N-glycans were observed as a dominant N-glycan class; middle, mono-, and di-sialo-core 1 structures were the major O-glycan species; bottom, a variety of GSL species were observed. GSLs tends to carry several forms ceramide species (e.g. d18:1 as a sphingoid base with C16:0-C24:1 fatty acids).

B, MS/MS analysis of an N-glycan carrying a LacdiNAc disaccharide and a fucose. Top, MS/MS analysis was carried out for a permethylated N-glycans detected at m/z 1154, demonstrating the presence of LacNAc moiety (Hex-HexNAc) detected at m/z 486 and LacdiNAc structures moiety (HexNAc-HexNAc) detected at m/z 527 respectively. The signal at m/z 527 was disappeared in *B4GALNT3&4*-double-KO; middle, tri-antennary structure was more abundant than LacdiNAc structure at m/z 1154; bottom, MS/MS analysis of an N-glycan detected at m/z 1242 was carried out to determine fucosylation sites. LacNAc moiety (yellow box), fucosylated LacNAc moiety (blue box), LacdiNAc moiety (green box), fucosylated LacdiNAc moiety (red box), and core fucose moiety (purple box) were detected at m/z 486, 660, 527, 701, and 474, respectively.

C, Detection of O-Fuc and SA1+F1+Core1. MS/MS analysis was carried out for the O-glycan at m/z 1069 (left). A fragment at m/z 694 in the MS/MS analysis of the O-glycan at m/z 1069 was subjected to MS^3 analysis (right). A fragment (m/z 472) corresponding Hex-HexNAc derived from O-Fuc structure was major, whereas a trace amount of Hex-dHex moiety (m/z 433) derived from SA1+F1+Core1 was also detected.

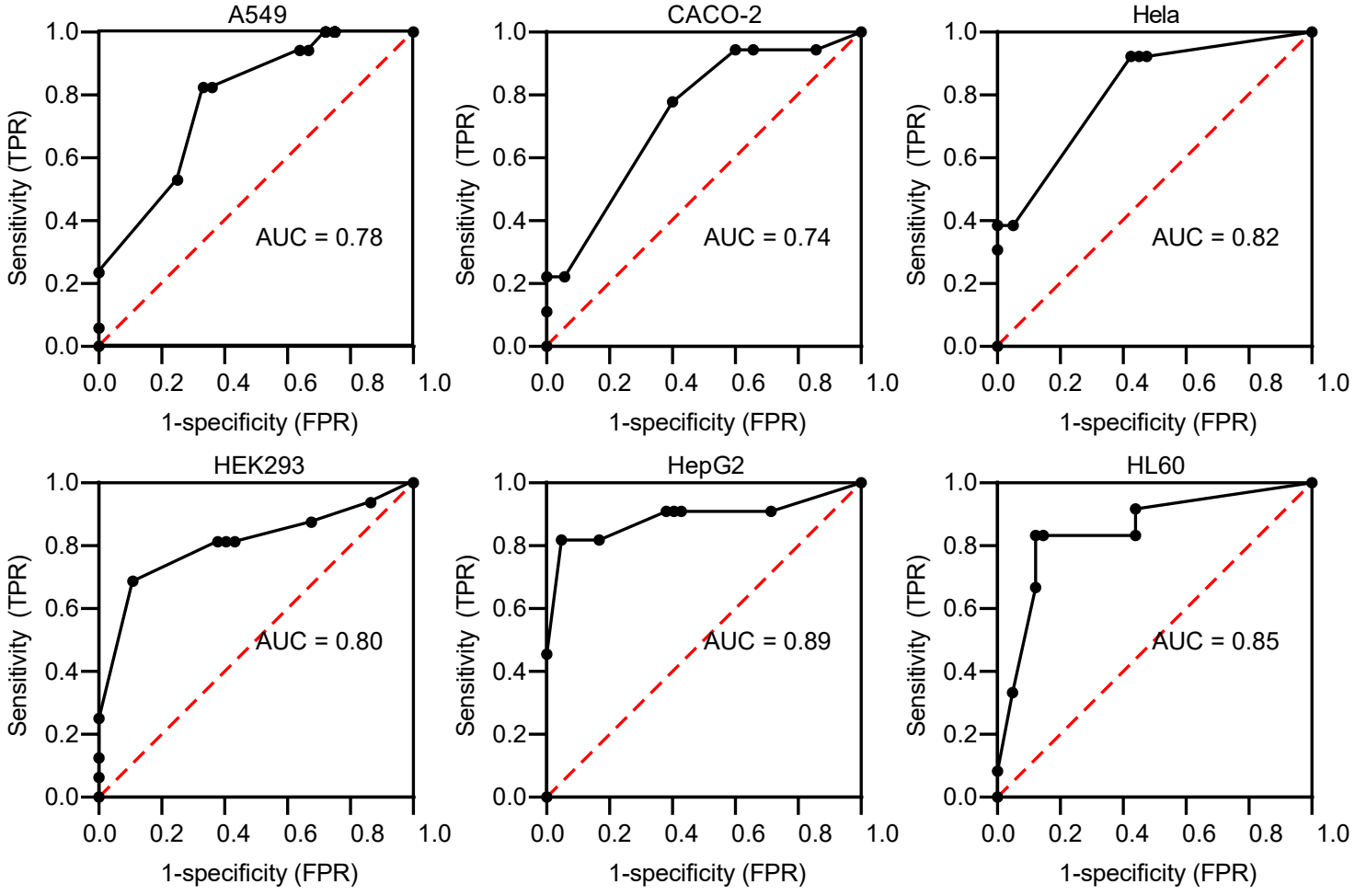
D, Determination of the fucosylated position in an O-glycan structure (F1+Hex1+Core 2) at m/z 1157. MS^n analysis was carried out for the O-glycan carrying fucose at m/z 1157 (upper left). A fragment (m/z 486) represents Hex-HexNAc moiety. A fragment (m/z 747), which represents a structure appeared with a neutral loss of Fuc-Gal-O, was further subjected to MS^3 (lower left), demonstrating the presence of a Hex-HexNAc moiety (m/z 486). To distinguish a fucosylation site on either GlcNAc or Gal, a fragment (m/z 660) in the MS^3 spectra (m/z 921 at m/z 1157) was subjected to MS^4 (right). The MS^4 spectra indicates that both Fuc-Gal and Fuc-GlcNAc structures exist.

A

TPM	0	0.01	0.1	0.2	0.5	1	2	3	4	5	6	8	10
TP	21	21	20	20	20	20	19	13	7	7	2	2	2
FP	32	28	22	19	9	9	9	2	0	0	0	0	0
TN	0	4	10	13	23	23	23	30	32	32	32	32	32
FN	0	0	1	1	1	1	2	8	14	14	19	19	19
Total	53	53	53	53	53	53	53	53	53	53	53	53	53

TPM	0	0.01	0.1	0.2	0.5	1	2	3	4	5	6	8	10
Accuracy	0.40	0.47	0.57	0.62	0.81	0.81	0.79	0.81	0.74	0.74	0.64	0.64	0.64
Precision	0.40	0.43	0.48	0.51	0.69	0.69	0.68	0.87	1.00	1.00	1.00	1.00	1.00
Recall	1.00	1.00	0.95	0.95	0.95	0.95	0.90	0.62	0.33	0.33	0.10	0.10	0.10
Specificity	0.00	0.13	0.31	0.41	0.72	0.72	0.72	0.94	1.00	1.00	1.00	1.00	1.00
F1-score	0.57	0.60	0.63	0.67	0.80	0.80	0.78	0.72	0.50	0.50	0.17	0.17	0.17

B



C

GlycoMaple (HPA's RNA-seq data) vs MS data (Fujitani et al.,*PNAS* (2013))

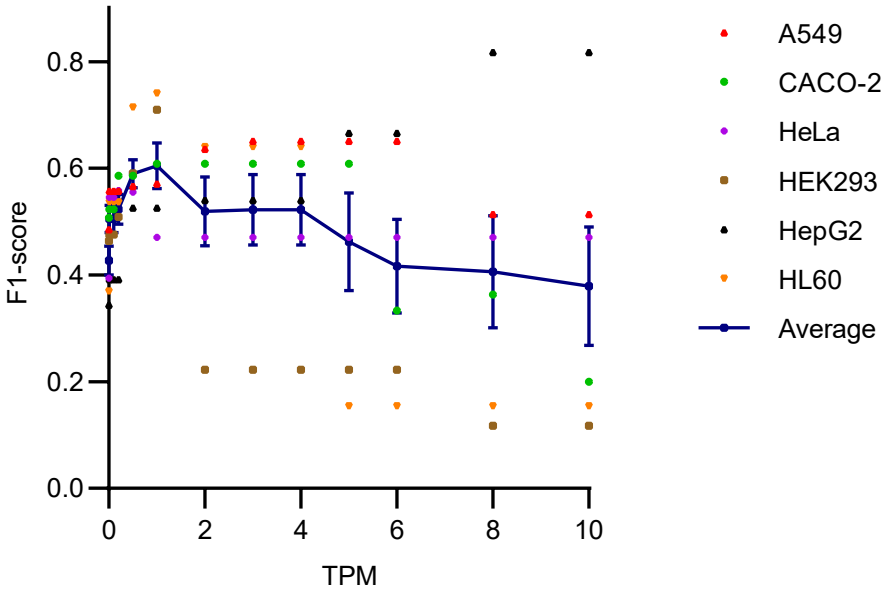
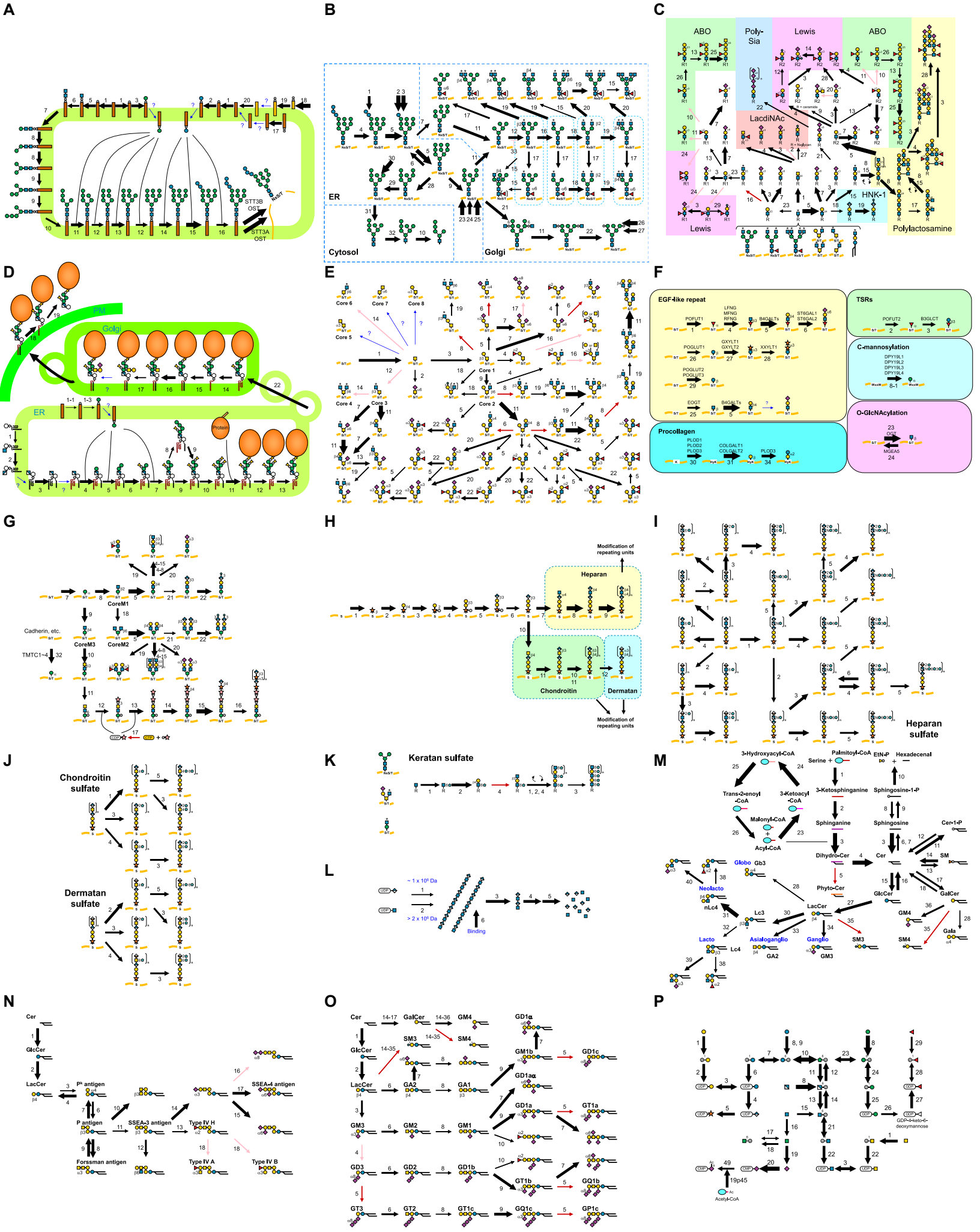


Figure S2. Related to Figure 2: Validation of GlycoMaple in various cell lines

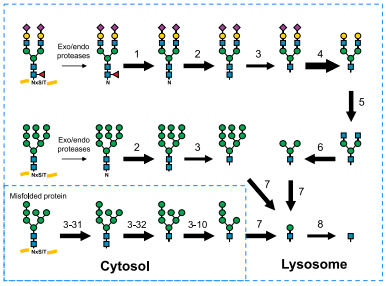
A, Evaluation of GlycoMaple using mucin-type O-glycan structures detected in HEK293. Among the 53 mucin-type O-glycan structures in the GlycoMaple, 21 glycan structures detected in the glycomics data and on recombinant MUC-1 in (A) were used as the actual detected structures. The predicted glycan structures were estimated at various TPM threshold values (0 to 10). The numbers of true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) predictions were calculated. Then, the Accuracy, Precision, Sensitivity (Recall), Specificity, and F1-score under the different thresholds were measured.

B, GlycoMaple analysis was performed using RNA-seq data of 6 cell lines (A549, Caco2, HeLa, HEK293, HepG2, and HL60) that are available in the Human Protein Atlas. As detected O-glycan structures, the data in the published paper (Fujitani et al., 2013) were used. As described in Figure 2, relationships between False Positive Rate (1 – Specificity) and True Positive Rate (Sensitivity) obtained in analyses of 6 human cell lines were plotted as ROC curves. The area under the curve (AUC) values in A549, Caco2, HeLa, HEK293, HepG2, and HL60 cells were 0.78, 0.74, 0.82, 0.80, 0.89, and 0.85, respectively.

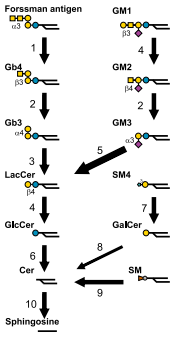
C, F1-scores in 6 human cell lines at various TPM thresholds were plotted as descrived in Figure 2. The data are presented as the mean of the F1-scores in 6 cell lines (blue square) ± SE. It should be noted that the greatest F1-scores were detected at high TPM thresholds in some cells such as HepG2 and A549 cells. There are at least two possible reasons to explain this. First, the RNA-seq and glycomics data used in the analysis were obtained from two independent sources. Second, glycomics analysis still has limitations and cannot detect all glycan structures. Since some estimated glycan structures in GlycoMaple were not observed in the MS analysis of some cell lines (such as HepG2 and A549 cells in the reference paper), the false positive rate was increased at low TPM thresholds.


A
B
ER
Cytosol
Golgi
C
ABO
Poly-
Sia
Lewis
ABO
LacdiNAc
HNK-1
Lewis
Polylactosamine
D
ER
Golgi
Protein
E
F
EGF-like repeat
TSRs
C-mannosylation
O-GlcNAcylation
Procollagen
G
CoreM1
CoreM2
CoreM3
Cadherin, etc.
TMTC1–4
H
Heparan
Chondroitin
Dermatan
Modification of repeating units
I
Heparan sulfate
J
Chondroitin sulfate
Dermatan sulfate
K
Keratan sulfate
L
~ 1 x 10⁶ Da
> 2 x 10⁶ Da
Binding
M
3-Hydroxyacyl-CoA
Palmitoyl-CoA
Serine
EtN-P
Hexadecenal
Trans-2-enoyl -CoA
3-Ketoacyl -CoA
3-Ketosphinganine
Sphingosine-1-P
Malonyl-CoA
Acyl-CoA
Sphinganine
Sphingosine
Cer-1-P
Dihydro-Cer
Cer
SM
Globo
Gb3
Neolacto
nLc4
Phyto-Cer
GlcCer
GalCer
LacCer
Lc3
GM4
Gala
Lacto
Lc4
Asialoganglio
GA2
Ganglio
GM3
SM3
SM4
N
Cer
GlcCer
LacCer
P^k antigen
P antigen
SSEA-3 antigen
Type IV H
SSEA-4 antigen
Forssman antigen
Type IV A
Type IV B
O
GalCer
GM4
SM3
SM4
GM3
GM2
GM1
GD3
GD2
GD1b
GT3
GT2
GT1c
GD1α
GM1b
GD1aα
GD1a
GT1a
GT1b
GQ1b
GQ1c
GP1c
P
GDP-4-keto-6-deoxymannose
19p45
Acetyl-CoA

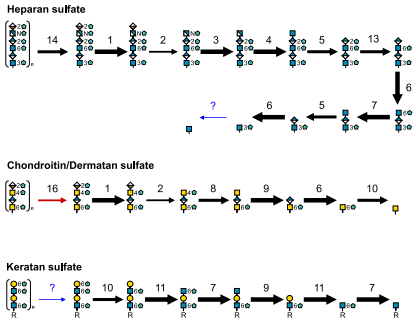
Q



R



S



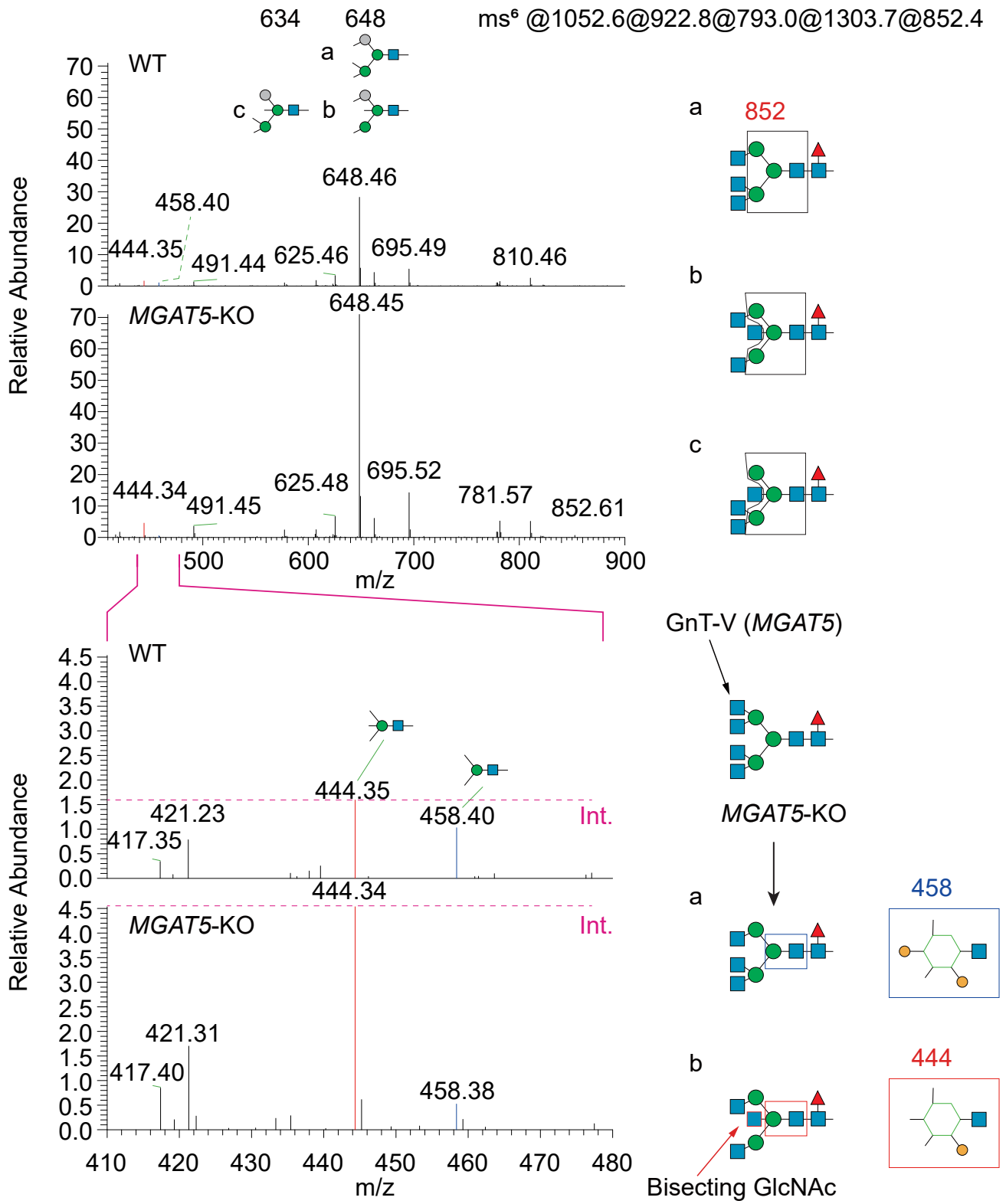
TPM value of genes

x < 0.1	
0.1 ≤ x < 1	
1 ≤ x < 4	
4 ≤ x < 20	
20 ≤ x < 100	
100 ≤ x	

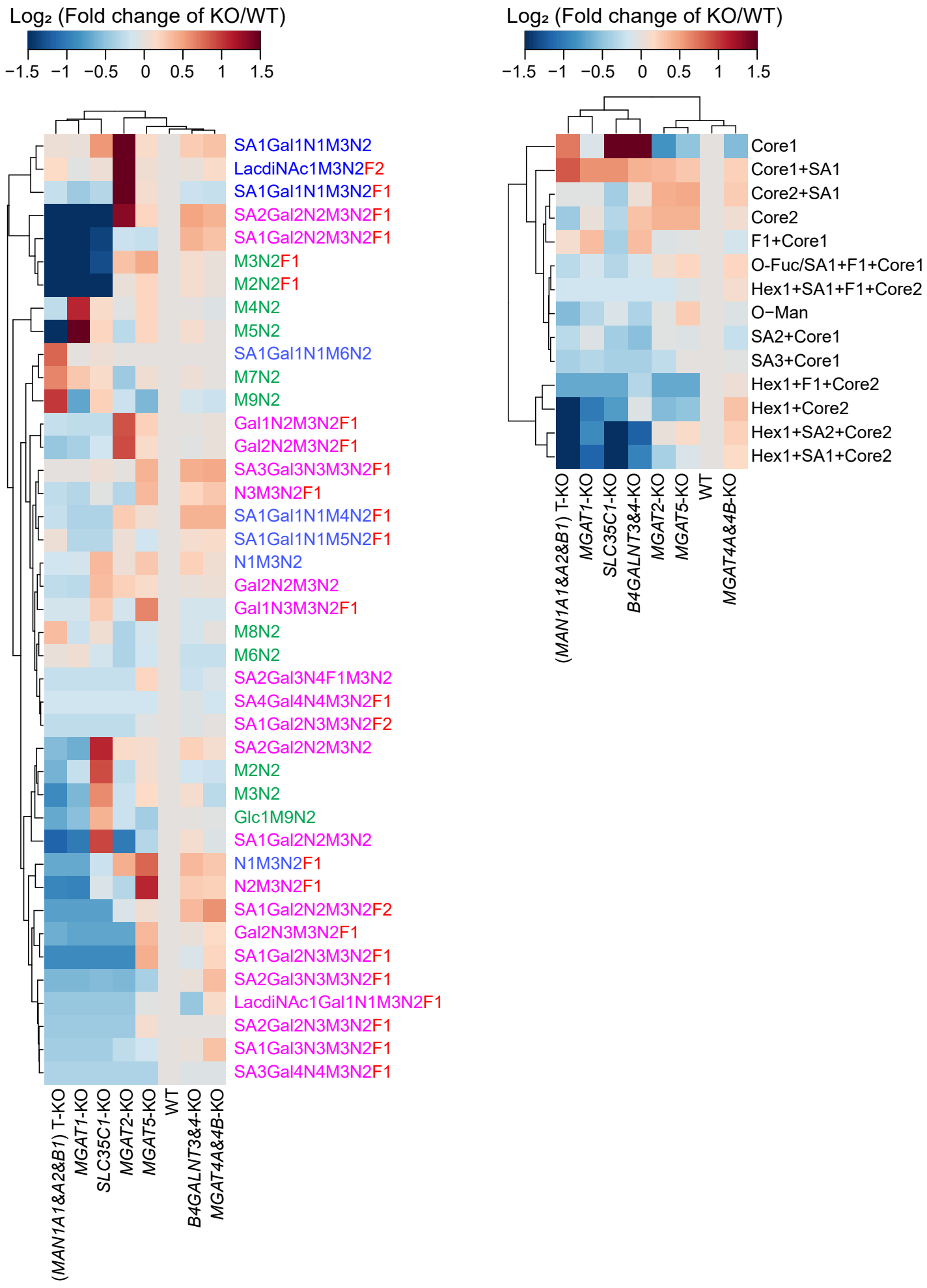
Man	ManNAc	Phoshate	CoA, Ribitol
Gal	GalNAc	Sulfate	Dolichol
Glc	GlcNAc	Fructose	Polyprenol
Fuc	GlcN	Ethanolamine	Diacylglycerol
IdoA	Xyl	Choline	1-alkyl-2-acylglycerol
GlcA	Neu5Ac	Inositol	Acyl-chain

Figure S3. Related to Figure 2: Mapping of glycan metabolic pathways in HEK293 cells

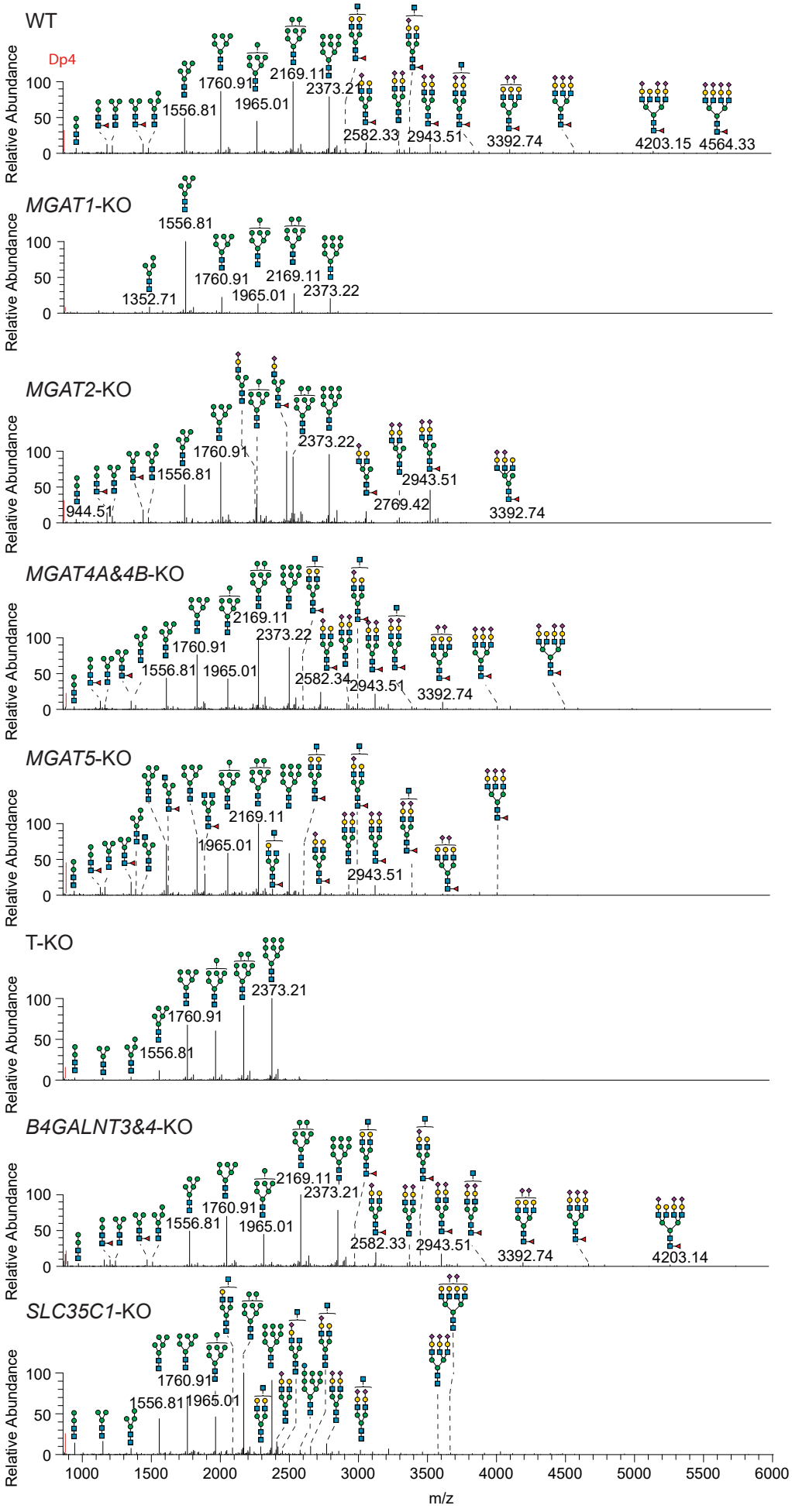
A, Gene expression in HEK293 cells was analyzed using RNA-seq. Based on the gene expression profile of HEK293 cells, pathways for (A) lipid-linked oligosaccharide (LLO) biosynthesis and OST, (B) N-glycan processing and branching , (C) complex capping of N-glycans / O-glycans / GSLs, (D) biosynthesis of GPI-anchored proteins, (E) biosynthesis of mucin-type O-glycans, (F) biosynthesis of O-fucose / O-glucose / collagen-O-galactose / O-GlcNAc / C-Man, (G) biosynthesis of O-Man, (H) biosynthesis of GAGs, (I) biosynthesis of HS, (J) biosynthesis of CS and DS, (K) biosynthesis of keratan sulfate (KS), (L) biosynthesis and catabolism of hyaluronan, (M) biosynthesis of core GSLs, (N) biosynthesis of Gbs, (O) biosynthesis of gangliosides, (P) biosynthesis of sugar nucleotides, (Q) lysosomal degradation of N-glycans, (R) lysosomal degradation of GSLs, and (S) lysosomal degradation of GAGs were visualized. The expression data (TPM values) of glycan-related genes were used for mapping. Arrows indicate the gene expression in each reaction. Thin pink arrows (TPM < 0.1) or solid red arrows ($0.1 \leq$ TPM < 1) indicate that the responsible genes for the pathways are not expressed or rarely expressed in the cells, respectively. The black arrows ($1 \leq$ TPM) indicate that the genes in the pathways are expressed in the cells. The thickness of these arrows shows the expression levels of the genes: thin black arrows, $1 \leq$ TPM < 4; normal black arrows, $4 \leq$ TPM < 20; thick black arrows, $20 \leq$ TPM < 100; very thick black arrows, $100 \leq$ TPM. If several genes overlapped in a reaction, the maximum TPM value among the values of overlapped genes was used as default (The setting could be changed. See Supplementary Information1). When several gene products make a complex for a reaction, the minimum TPM value of the subunit genes was used. Blue arrows indicate the reactions for which the responsible genes were not clear. Each numbered reaction is listed in Table S3.


A
634
648
ms⁶ @1052.6@922.8@793.0@1303.7@852.4
WT
a
b
c
Relative Abundance
70
60
50
40
30
20
10
0
458.40
444.35
491.44
625.46
648.46
695.49
810.46
MGAT5-KO
648.45
444.34
491.45
625.48
695.52
781.57
852.61
500
600
700
800
900
m/z
a
852
b
c
GnT-V (MGAT5)
MGAT5-KO
WT
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
444.35
421.23
417.35
458.40
Int.
444.34
MGAT5-KO
Int.
421.31
417.40
458.38
410
420
430
440
450
460
470
480
m/z
a
458
b
444
Bisecting GlcNAc

B


Log₂ (Fold change of KO/WT)
−1.5 −1 −0.5 0 0.5 1 1.5
SA1Gal1N1M3N2
LacdiNAc1M3N2F2
SA1Gal1N1M3N2F1
SA2Gal2N2M3N2F1
SA1Gal2N2M3N2F1
M3N2F1
M2N2F1
M4N2
M5N2
SA1Gal1N1M6N2
M7N2
M9N2
Gal1N2M3N2F1
Gal2N2M3N2F1
SA3Gal3N3M3N2F1
N3M3N2F1
SA1Gal1N1M4N2F1
SA1Gal1N1M5N2F1
N1M3N2
Gal2N2M3N2
Gal1N3M3N2F1
M8N2
M6N2
SA2Gal3N4F1M3N2
SA4Gal4N4M3N2F1
SA1Gal2N3M3N2F2
SA2Gal2N2M3N2
M2N2
M3N2
Glc1M9N2
SA1Gal2N2M3N2
N1M3N2F1
N2M3N2F1
SA1Gal2N2M3N2F2
Gal2N3M3N2F1
SA1Gal2N3M3N2F1
SA2Gal3N3M3N2F1
LacdiNAc1Gal1N1M3N2F1
SA2Gal2N3M3N2F1
SA1Gal3N3M3N2F1
SA3Gal4N4M3N2F1
(MAN1A1&A2&B1) T-KO
MGAT1-KO
SLC35C1-KO
MGAT2-KO
MGAT5-KO
WT
B4GALNT3&4-KO
MGAT4A&4B-KO
Log₂ (Fold change of KO/WT)
−1.5 −1 −0.5 0 0.5 1 1.5
Core1
Core1+SA1
Core2+SA1
Core2
F1+Core1
O-Fuc/SA1+F1+Core1
Hex1+SA1+F1+Core2
O−Man
SA2+Core1
SA3+Core1
Hex1+F1+Core2
Hex1+Core2
Hex1+SA2+Core2
Hex1+SA1+Core2
(MAN1A1&A2&B1) T-KO
MGAT1-KO
SLC35C1-KO
B4GALNT3&4-KO
MGAT2-KO
MGAT5-KO
WT
MGAT4A&4B-KO

C


WT
Dp4
Relative Abundance
100
50
0
1556.81
1760.91
1965.01
2169.11
2373.21
2582.33
2943.51
3392.74
4203.15
4564.33
MGAT1-KO
1556.81
1352.71
1760.91
1965.01
2169.11
2373.22
MGAT2-KO
1760.91
2373.22
1556.81
2943.51
944.51
2769.42
3392.74
MGAT4A&4B-KO
2169.11
2373.22
1760.91
1556.81
1965.01
2582.34
2943.51
3392.74
MGAT5-KO
2169.11
1965.01
2943.51
T-KO
2373.21
1760.91
1556.81
B4GALNT3&4-KO
2169.11
2373.21
1760.91
1556.81
1965.01
2582.33
2943.51
3392.74
4203.14
SLC35C1-KO
1760.91
1556.81
1965.01
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
m/z

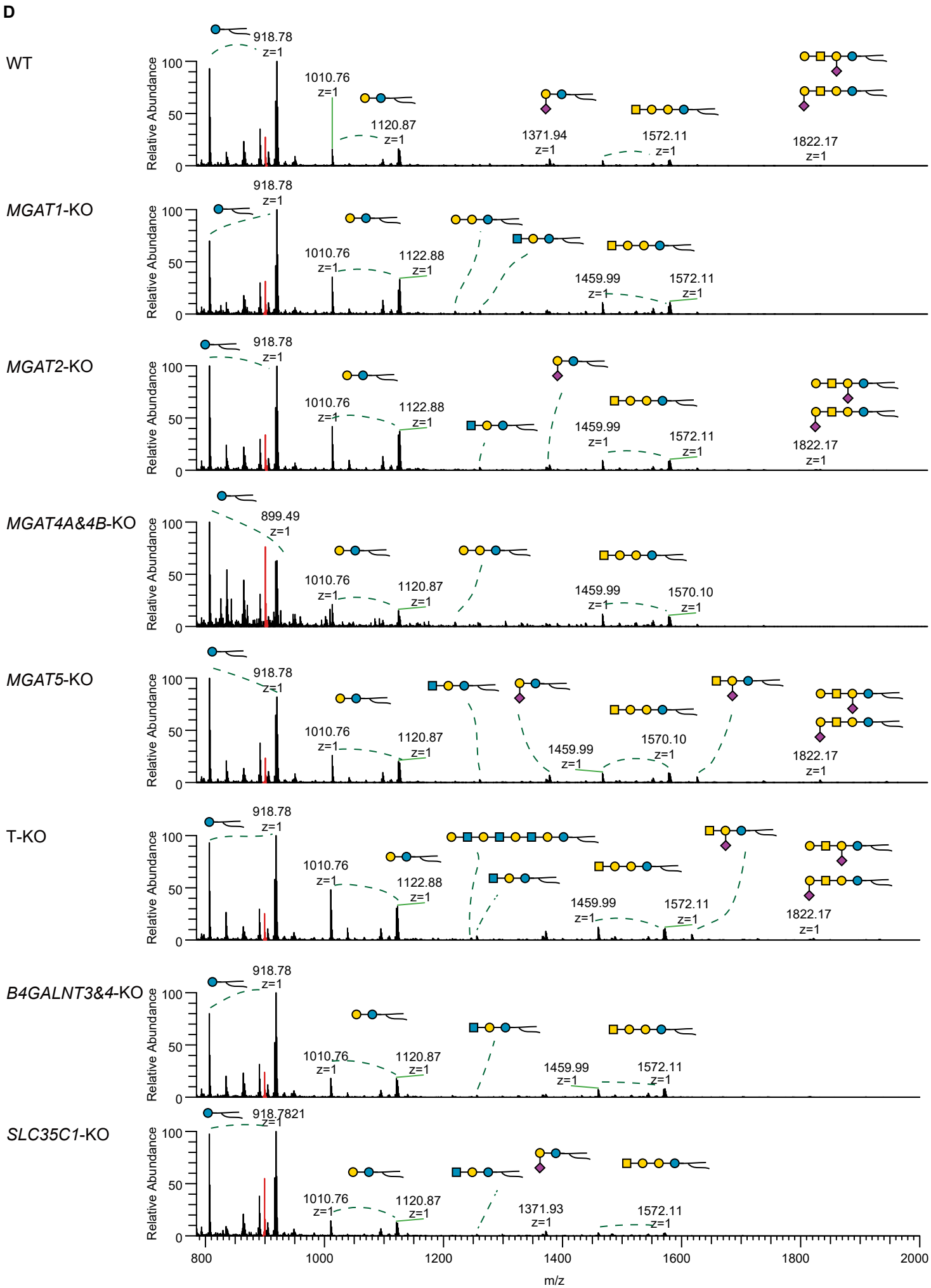

D
WT
MGAT1-KO
MGAT2-KO
MGAT4A&4B-KO
MGAT5-KO
T-KO
B4GALNT3&4-KO
SLC35C1-KO
Relative Abundance
m/z
918.78 z=1
1010.76 z=1
1120.87 z=1
1371.94 z=1
1572.11 z=1
1822.17 z=1
1122.88 z=1
1459.99 z=1
899.49 z=1
1570.10 z=1
918.7821 z=1
1371.93 z=1

Figure S4. Related to Figure 4: Glycosylation is altered in the rationally engineered HEK293 cell lines

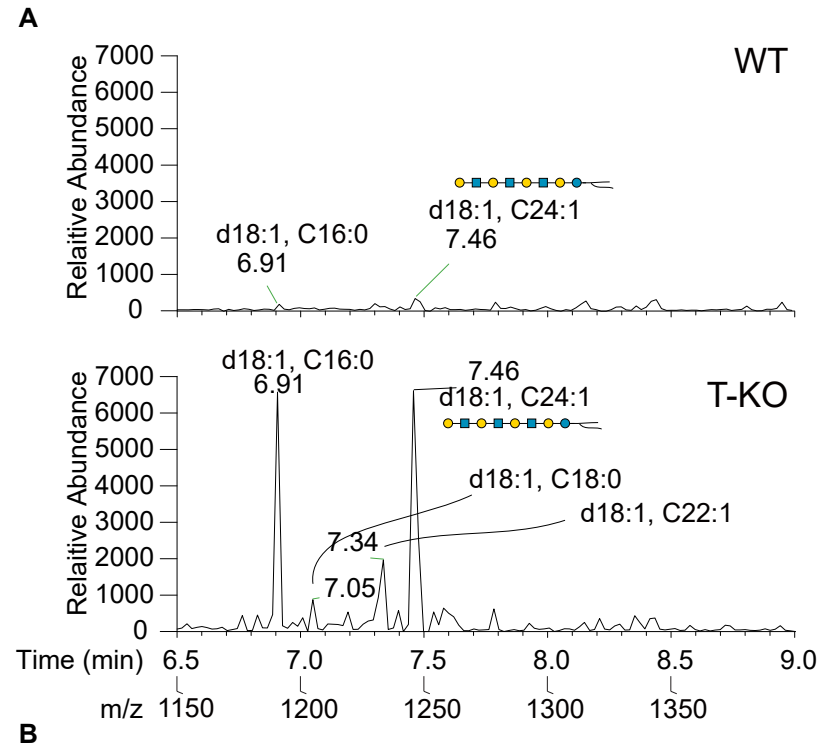
A, Determination and quantification of N-glycan carrying bisecting GlcNAc in HEK293 WT and *MGAT5*-KO cells. MS^n analyses were carried out for several permethylated N-glycans, which potentially carrying bisecting GlcNAc epitope. For example, the molecular ion detected at m/z 1052 (2+) was subjected to MS^6 analysis demonstrating fragment ions produced from m/z 852 which composed of 3 Hex and 1 HexNAc in black open box that produces signature ions m/z 648 and 634 (2 Hex and 1 HexNAc), and m/z 444 and 458 (1 Hex and 1 HexNAc). The ion of m/z 634 was barely detected (less than 1% of total ion intensity), indicating the structure c is a minor glycan component (upper figure). Based on the number of methyl group (orange circle) on monosaccharide residue of permethylated N-glycan, it shows different m/z, e.g. m/z 458 carries 2 Me groups (blue open box) on Man, whereas m/z 444 carries 1 Me group (red open box) on Man, indicating that one of hydroxy group on the first Man attached on the second GlcNAc on chitobiose core is occupied by a GlcNAc residue, forming a bisecting GlcNAc epitope. The ion intensity detected at m/z 444 is significantly increased in *MGAT5*-KO cells. Although WT expresses a reasonable amount of bisecting GlcNAc epitope (the ratio of 444:458=3:2), *MGAT5*-KO cells express more bisecting-GlcNAc type N-glycans than that of WT due to a lack of *MGAT5*. Incapability of synthesizing higher complex N-glycans by knocking out *MGAT5* gene alters the cellular glycosylation pathway leading to an increase of production of bisecting GlcNAc epitope (lower figure).

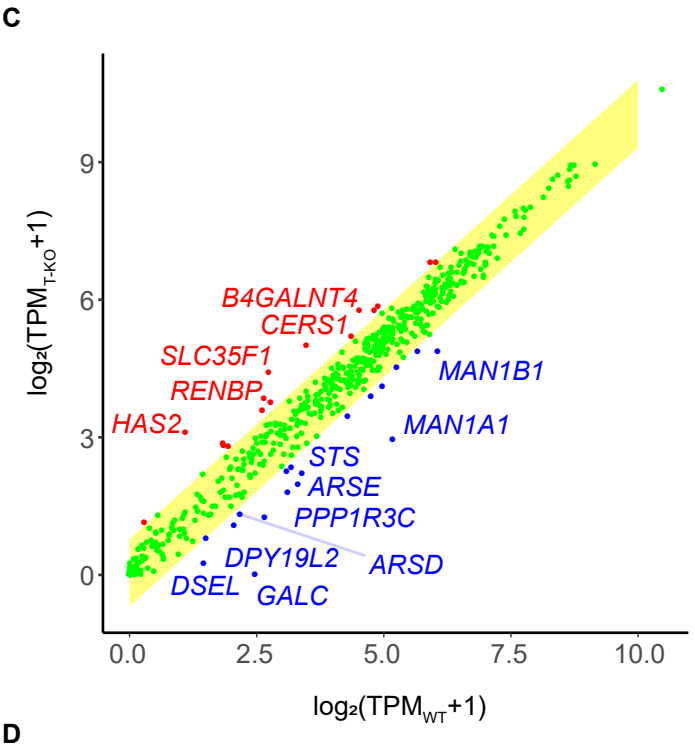
B, Comparative glycomic analysis of seven KO cells. Relative abundance of N-glycans (left panel) and O-glycans (right panel) in HEK293 *MAN1A1*-, *A2*-, and *B1*-triple-KO (T-KO), *MGAT1*-KO, *MGAT2*-KO, *MGAT4A&4B*-KO, *MGAT5*-KO, *B4GALNT3*- and *4*-KO, and *SLC35C1*-KO cells. The data are visualized as clustered heatmaps. The relative amounts of glycan structures in each cell type were calculated, and were compared with those of HEK293 cells. The $\log_2$ values of the data are visualized as heatmaps. The data used the mean value from two independent experiments.

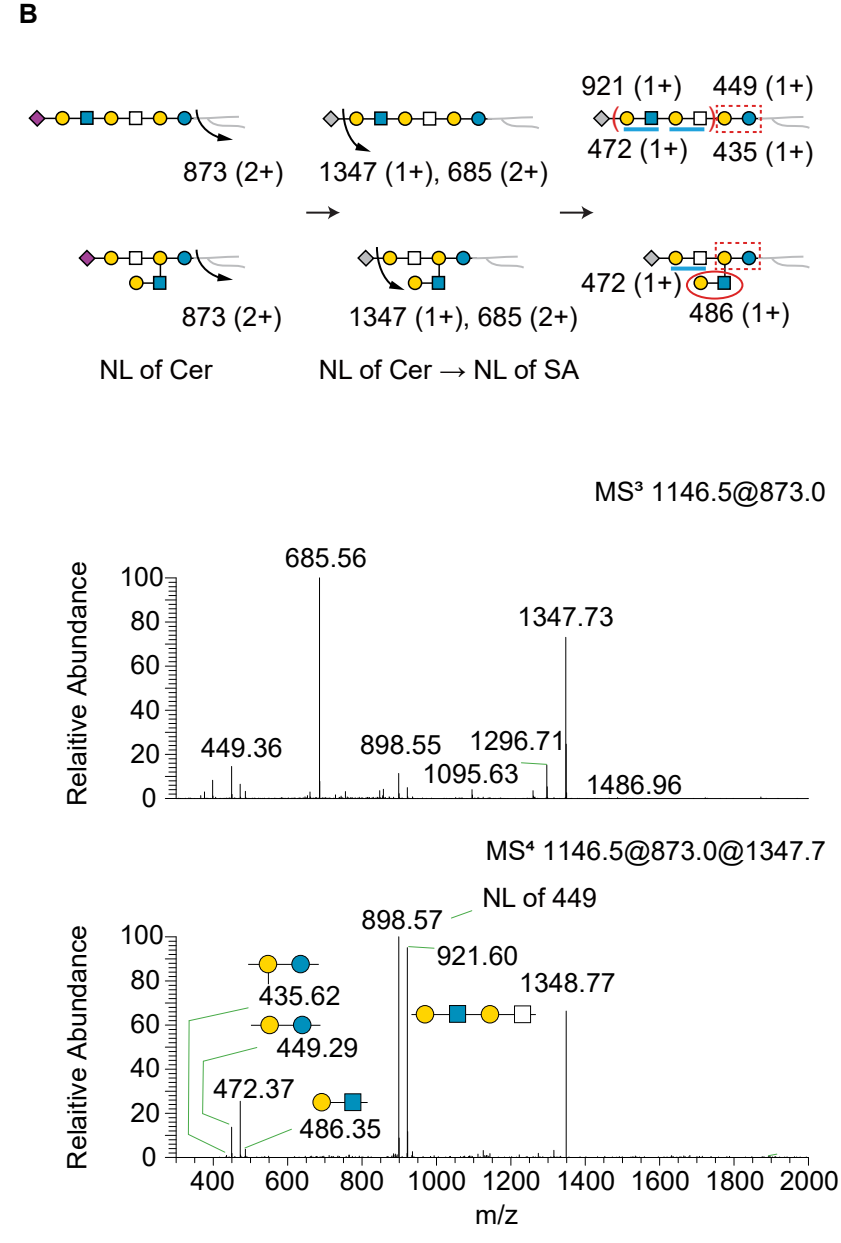
C, N-linked glycosylation is altered in the rationally engineered HEK293 cell lines. A selected set of glycan-related gene KO cells were subjected to glycomics analysis. N-glycans from equivalent numbers of cells were analyzed for each cell type shown here. The full MS profiles of N-glycans were deconvoluted by Xtract software. Briefly, N-linked glycans were released from glycoproteins harvested from the indicated cell types. MS glycan profiles were obtained following permethylation. An internal standard (Dp4, red line) was added in equal amount to each glycan preparation to facilitate quantification of glycan abundances.

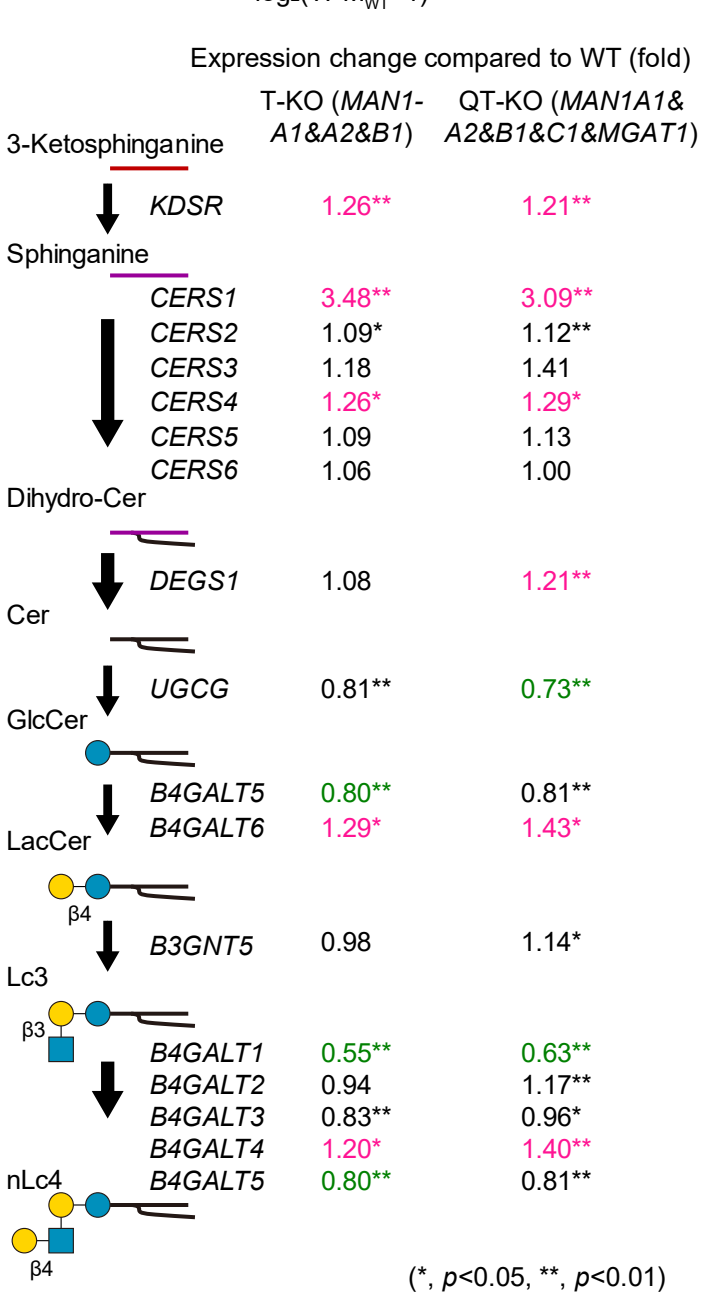
D, GSLs glycosylation is altered in the rationally engineered HEK293 cell lines. GSLs from equivalent numbers of cells were analyzed for each cell type. The full MS profiles of GSLs were deconvoluted

by Xtract software. Intact GSLs were permethylated with 12C-methyliodide prior to MS analysis. An internal standard (Dp4, red line) was added in equal amount to each glycan preparation to facilitate quantification of glycan abundances. In the T-KO (*MAN1A1&A2&B1*) cells, a significant signal of polylactosamine chain was detected compared to other type cells.


A
WT
Relative Abundance
d18:1, C16:0
6.91
d18:1, C24:1
7.46
T-KO
d18:1, C18:0
d18:1, C22:1
7.34
7.05
Time (min)
m/z


C
log₂(TPM_T-KO+1)
log₂(TPM_WT+1)
B4GALNT4
CERS1
SLC35F1
RENBP
HAS2
MAN1B1
MAN1A1
STS
ARSE
PPP1R3C
DPY19L2
ARSD
DSEL
GALC


B
921 (1+)
449 (1+)
472 (1+)
435 (1+)
873 (2+)
1347 (1+), 685 (2+)
486 (1+)
NL of Cer
NL of Cer → NL of SA
MS³ 1146.5@873.0
685.56
1347.73
449.36
898.55
1296.71
1095.63
1486.96
MS⁴ 1146.5@873.0@1347.7
NL of 449
898.57
921.60
1348.77
435.62
449.29
472.37
486.35
Relative Abundance
m/z


D
Expression change compared to WT (fold)
T-KO (MAN1-A1&A2&B1)
QT-KO (MAN1A1&A2&B1&C1&MGAT1)
3-Ketosphinganine
KDSR 1.26** 1.21**
Sphinganine
CERS1 3.48** 3.09**
CERS2 1.09* 1.12**
CERS3 1.18 1.41
CERS4 1.26* 1.29*
CERS5 1.09 1.13
CERS6 1.06 1.00
Dihydro-Cer
DEGS1 1.08 1.21**
Cer
UGCG 0.81** 0.73**
GlcCer
B4GALT5 0.80** 0.81**
B4GALT6 1.29* 1.43*
LacCer
β4
B3GNT5 0.98 1.14*
Lc3
β3
B4GALT1 0.55** 0.63**
B4GALT2 0.94 1.17**
B4GALT3 0.83** 0.96*
B4GALT4 1.20* 1.40**
B4GALT5 0.80** 0.81**
nLc4
β4
(*, p<0.05, **, p<0.01)

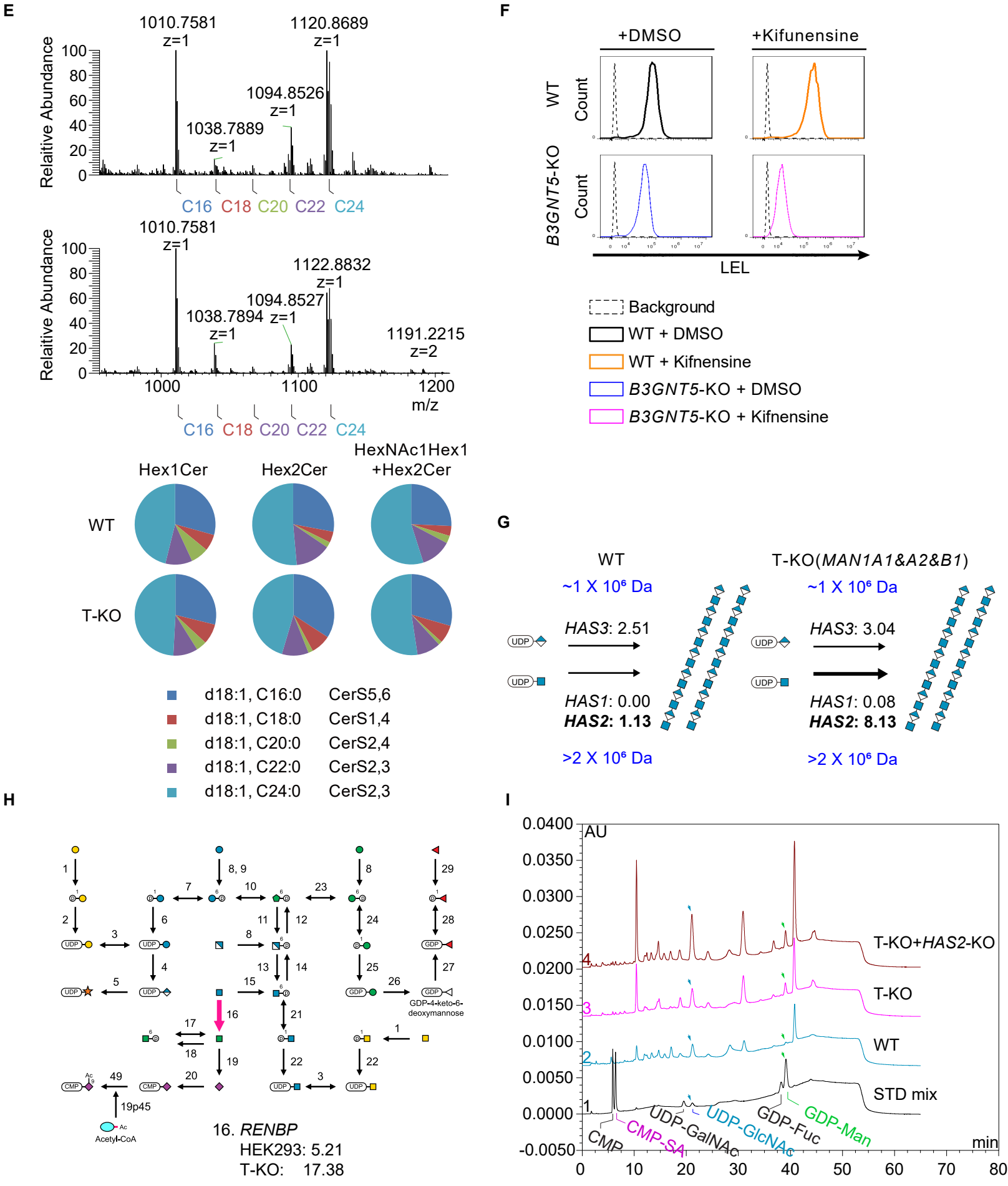

E
1010.7581
z=1
1120.8689
z=1
1094.8526
z=1
1038.7889
z=1
Relative Abundance
C16 C18 C20 C22 C24
1010.7581
z=1
1122.8832
z=1
1094.8527
z=1
1038.7894
z=1
1191.2215
z=2
1000
1100
1200
m/z
Hex1Cer
Hex2Cer
HexNAc1Hex1
+Hex2Cer
WT
T-KO
d18:1, C16:0 CerS5,6
d18:1, C18:0 CerS1,4
d18:1, C20:0 CerS2,4
d18:1, C22:0 CerS2,3
d18:1, C24:0 CerS2,3
F
+DMSO
+Kifunensine
WT
B3GNT5-KO
Count
LEL
Background
WT + DMSO
WT + Kifnensine
B3GNT5-KO + DMSO
B3GNT5-KO + Kifnensine
G
WT
T-KO(MAN1A1&A2&B1)
~1 X 10⁶ Da
HAS3: 2.51
HAS1: 0.00
HAS2: 1.13
>2 X 10⁶ Da
~1 X 10⁶ Da
HAS3: 3.04
HAS1: 0.08
HAS2: 8.13
>2 X 10⁶ Da
UDP
H
GDP-4-keto-6-deoxymannose
Acetyl-CoA
19p45
16. RENBP
HEK293: 5.21
T-KO: 17.38
I
AU
T-KO+HAS2-KO
T-KO
WT
STD mix
CMP
CMP-SA
UDP-GalNAc
UDP-GlcNAc
GDP-Fuc
GDP-Man
min

Figure S5. Related to Figure 5: Increase of LacNAc-containing GSLs and hyaluronan in *MAN1A1&A2&B1*-T-KO cells.

A, Increase of LacNAc-containing GSLs in T-KO cells. GSLs extracted from WT and T-KO cells were permethylated and subjected to NSI-MS analysis by total ion monitoring (TIM). The resulting TIM profiles were filtered for detection of the signature fragment ion corresponding to the loss of terminal Hex-HexNAc disaccharide epitope in order to report the presence of LacNAc-containing GSLs. Theoretical mass of the B2-ion resulting from a neutral loss of doubly charge LacNAc disaccharide was used to assess the relative expression of LacNAc-containing GSLs in WT and T-KO cells.

B, MS^n analysis of a GSL carrying LacNAc disaccharide in T-KO cells. MS^n analysis was carried out for the GSL detected at m/z 1146.5. Neutral loss of Cer was triggered to obtain fragment ions derived from GSL-glycan moiety (upper), showing a loss of sialic acid from the oligosaccharide backbone. MS^4 analysis was carried out to determine the hexasaccharide structure, demonstrating the presence of a linear type internal tetrasaccharide moiety Hex-HexNAc-Hex-HexNAc) detected at m/z 921 and a linear and a branched form of internal disaccharide (Lactose) detected at m/z 449 and 435, respectively. By comparison of ion intensities observed at m/z 449 and 435 as well as m/z 921 and 486, a linear type GSL shows higher abundance than that of branched type. A blank box indicates a HexNAc (GalNAc or GlcNAc) residue due to a difficulty to determine the monosaccharide residue by MS analysis.

C, Comparison of the expression of glycan-related genes in WT versus *MAN1A1*/*A2*/*B1*/*C1*/*MGAT1*-quintuple knockout (QT-KO) cells. TPM values (averages of triplicated data) were calculated and plotted as $\log_2$(TPM + 1) values. The yellow area represents the predicted interval expression in WT or QT-KO cells. Representative examples of genes with higher expressions in WT or QT-KO cells are indicated by blue or red text, respectively.

D, Expression changes of genes involved in the biosynthetic pathway of neolacto-series GSLs. Gene expression of T-KO and QT-KO cells were compared with HEK293 WT cells and illustrated as the fold change. The values of fold change that were more than 1.2 and less than 0.83 were shown as color in red and blue, respectively. The data represent means of triplicated experiments. *, $p < 0.05$; **, $p < 0.01$. (two-tailed Student's t test).

E, Lipid composition of LacCer. Representative zoom spectra for the mass range containing LacCer ceramide forms are presented for a WT and T-KO (*MAN1A1*, *A2*, and *B1*). The m/z values at which LacCer species were detected are shown along the x-axis (left). Summary of the mean relative abundances (n = 2) of HexCer, LacCer, and Gb4 ceramide forms in WT controls and T-KO were shown as pie charts (right). Ceramide consisting of d18:1-C18:0, which is mainly synthesized by

CERS1 and CERS4 ceramide synthase, was increased in T-KO cells, which is consistent with upregulated *CERS1* and *CERS4* genes in T-KO.

F, Flow cytometric analysis of cells stained by fluorescent-conjugated LEL. HEK293 wild-type (WT) and *B3GNT5*-KO cells treated with DMSO or a mannosidae-I inhibitor, kifunensine, were stained by fluorescent-conjugated LEL, and analyzed by flow cytometry. Background, without lectin staining.

G, Expression values (TPM) of genes required for hyaluronan biosynthesis in WT and T-KO cells.

H, Comparison of sugar nucleotide biosynthetic pathway between HEK293 WT and T-KO cells. The TPM value of RENBP (step 16) was 3-fold increased in T-KO cells compared to WT cells. Pink arrow represents expression (TPM value) of the gene responsible for the reaction was increased more than 2 times in T-KO compared to WT cells. Each numbered reaction is listed in Table S3.

I, Amounts of UDP-GlcNAc and GDP-Man were analyzed by a Dionex high-performance anion-exchange chromatography. UDP-GlcNAc (blue arrow) and UDP-Man (green arrow) were significantly increased in T-KO and T-KO+HAS2-KO cells, respectively, whereas the expression of CMP and CMP-sialic acid (SA) among the three sample populations did not alter and stay on a quite similar level. Due to a lack of ability to synthesize higher complex N-glycans by knocking of mannosidase-I genes, both KO cells tends to accumulate UDP-GlcNAc, which are important donors to build matured glycoprotein-glycan structures. Interestingly, T-KO+HAS2-KO cell tends to express higher abundance of UDP-GlcNAc than that of T-KO cells, indicating that knocking out *HAS2* gene in T-KO cell leading to an accumulation of cellular UDP-GlcNAc due to a disruption of hyaluronan biosynthesis.

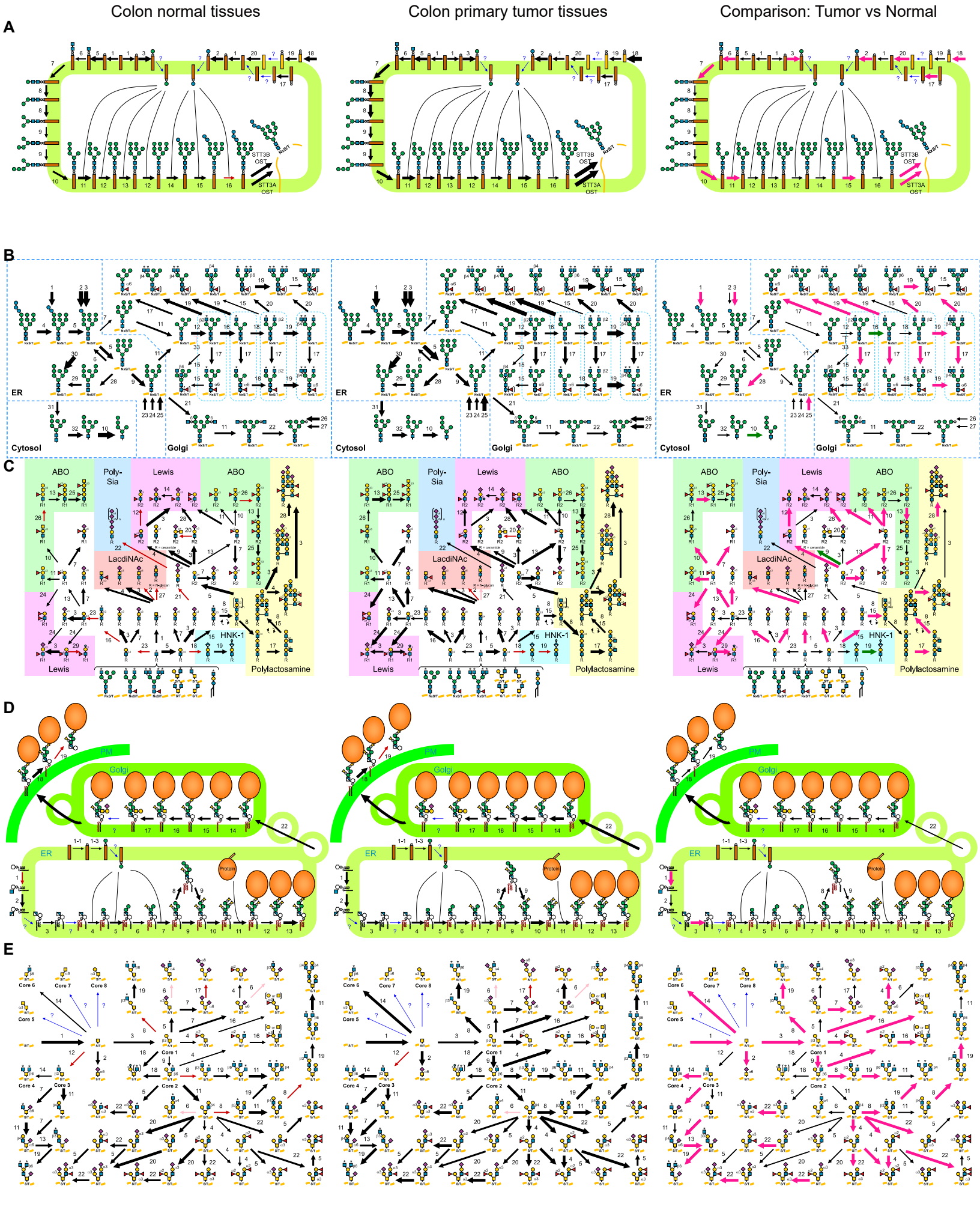
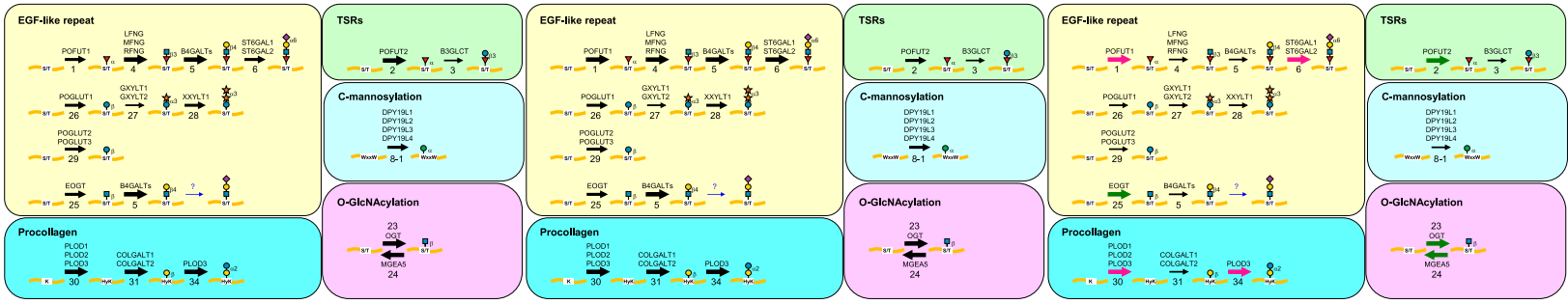

Colon normal tissues
Colon primary tumor tissues
Comparison: Tumor vs Normal
A
B
ER
Cytosol
Golgi
C
ABO
Poly-Sia
Lewis
LacdiNAc
HNK-1
Polylactosamine
D
PM
Golgi
ER
Protein
E
Core 1
Core 2
Core 3
Core 4
Core 5
Core 6
Core 7
Core 8

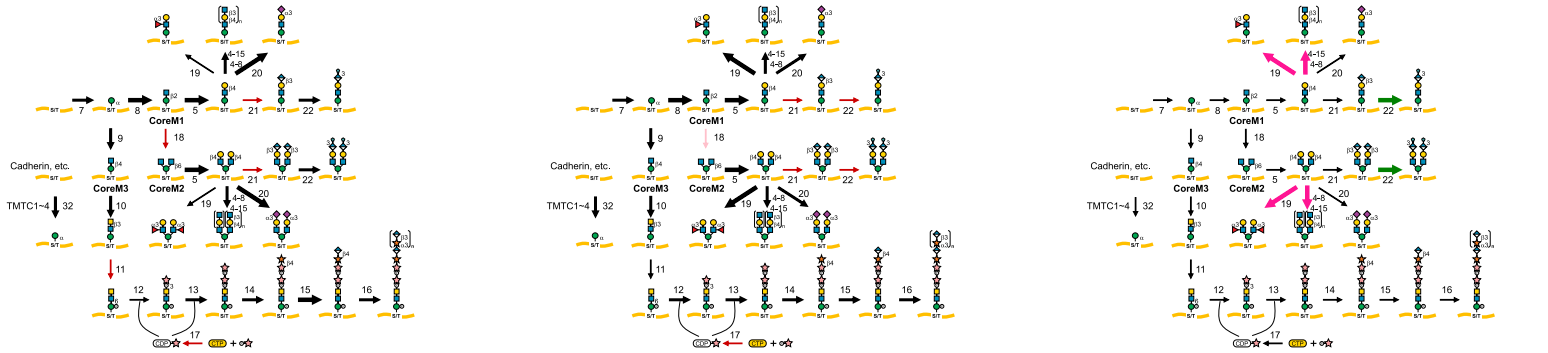
Figure S6

Colon normal tissues | Colon primary tumor tissues | Comparison: Tumor vs Normal

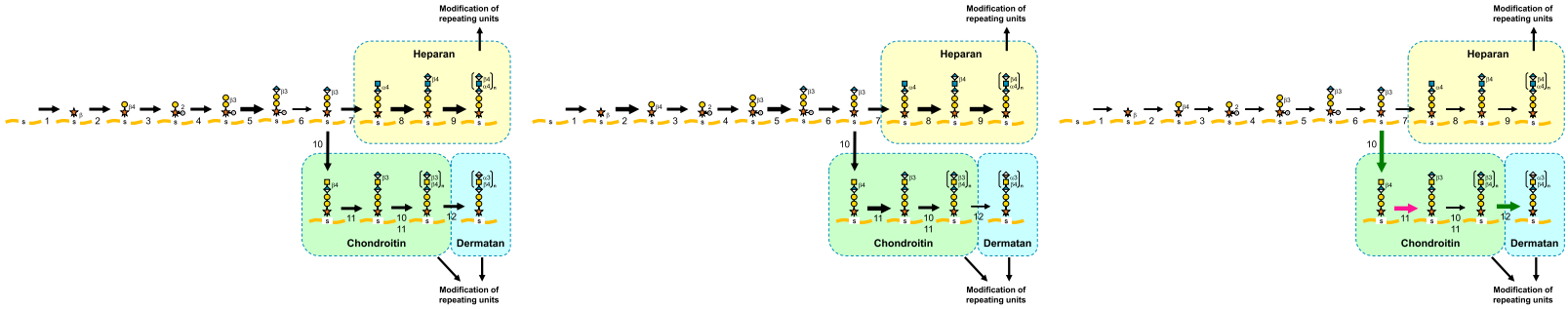
F



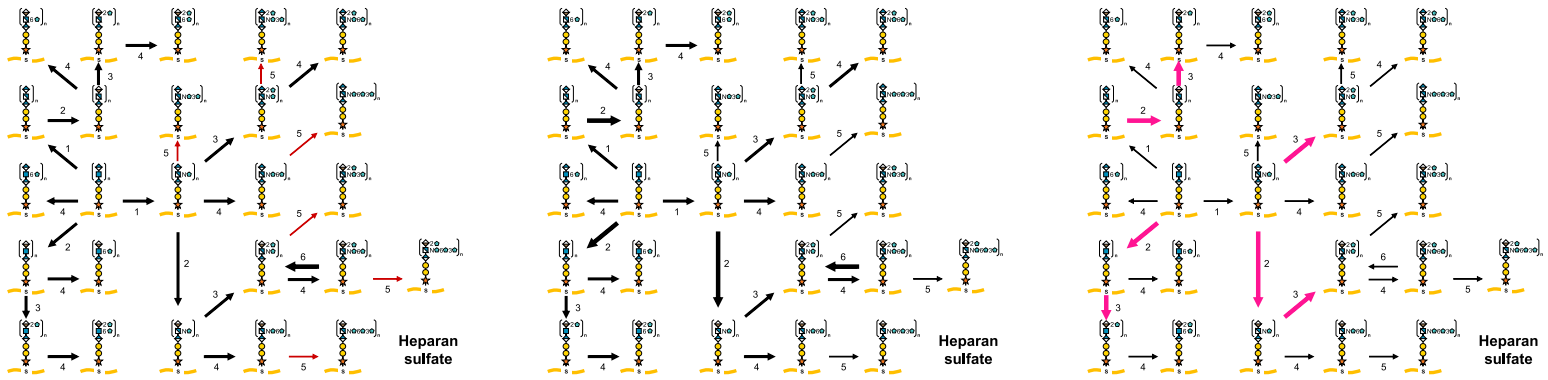
G



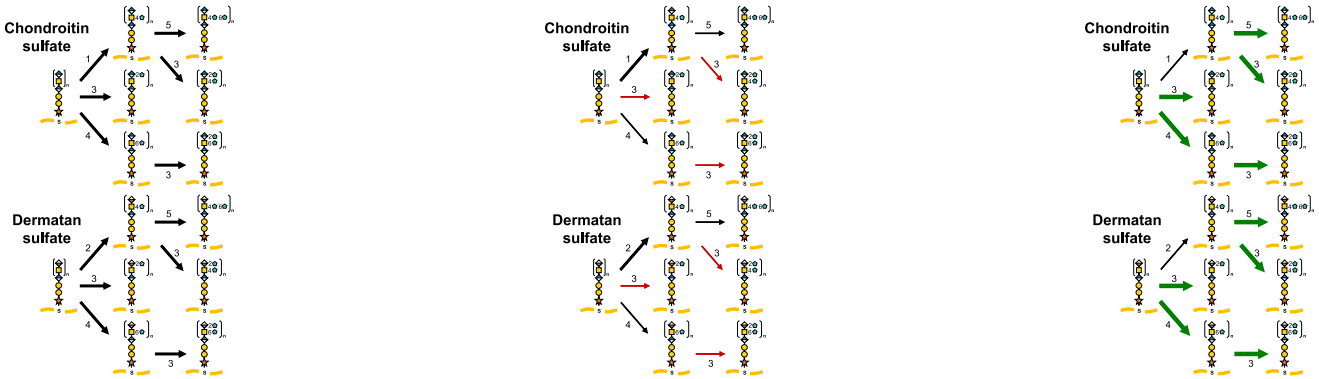
H



I



J

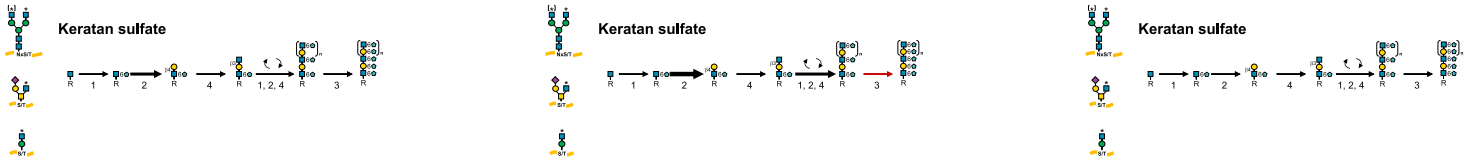


K

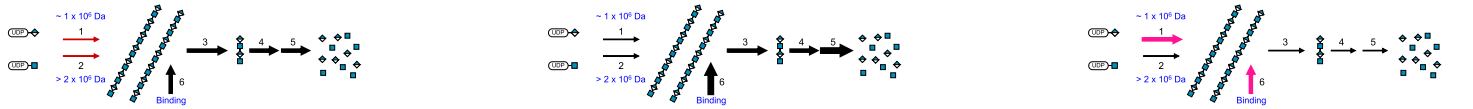
Colon normal tissues

Colon primary tumor tissues

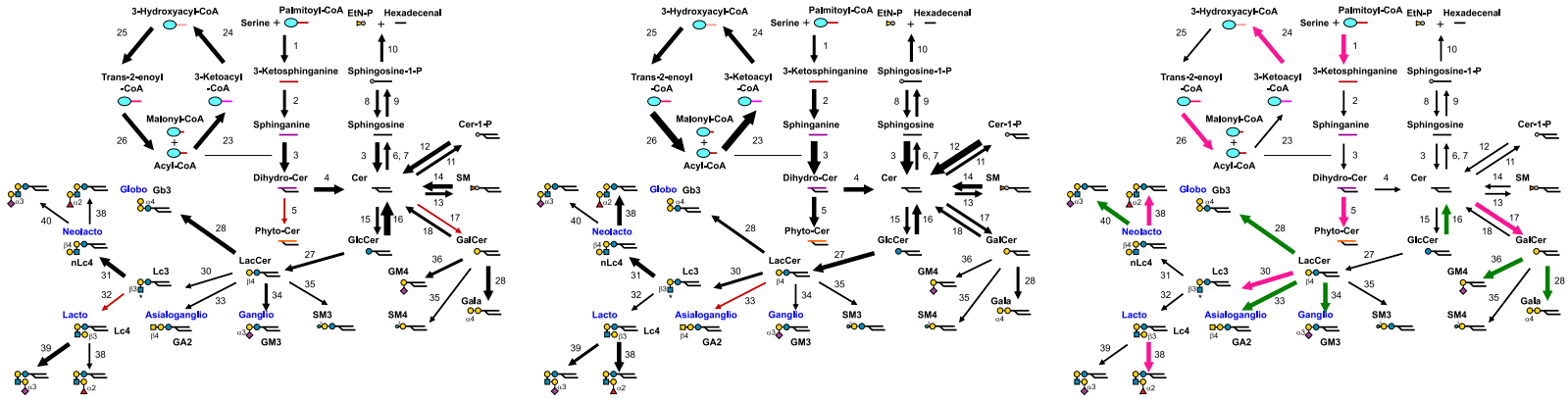
Comparison: Tumor vs Normal



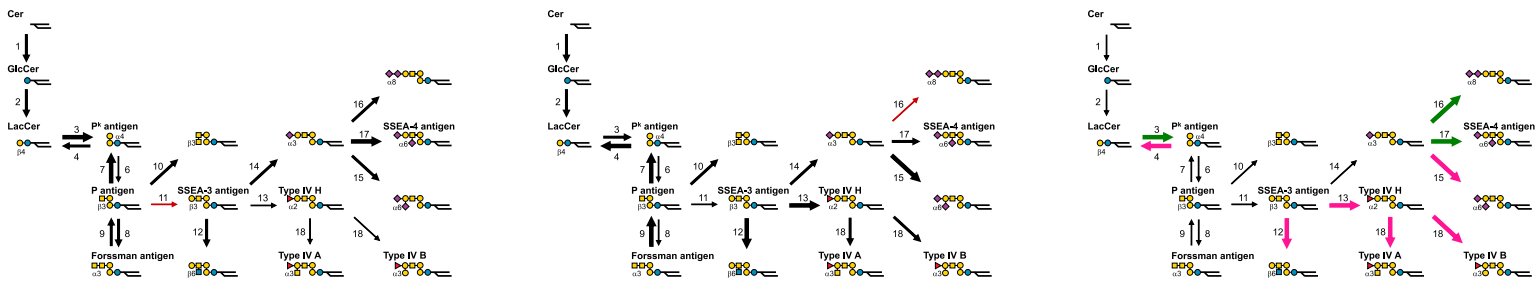
L



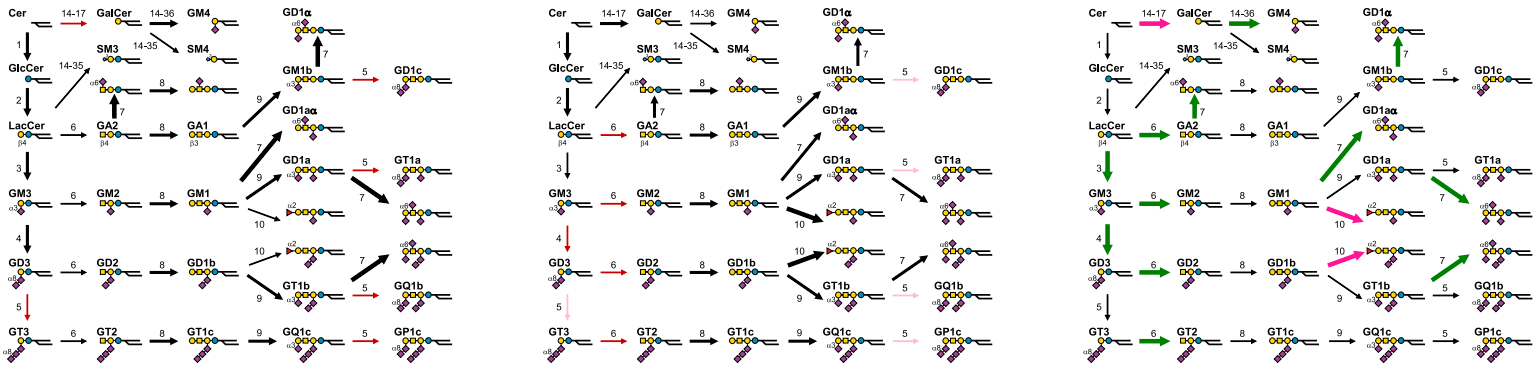
M



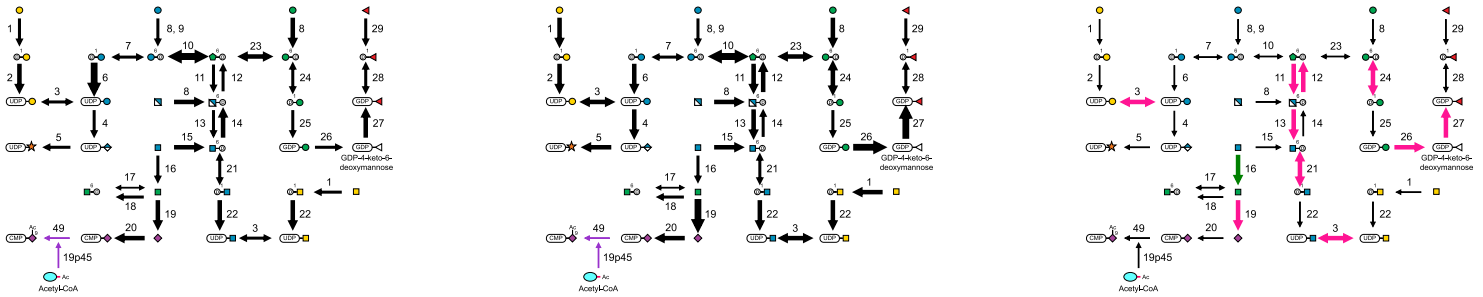
N



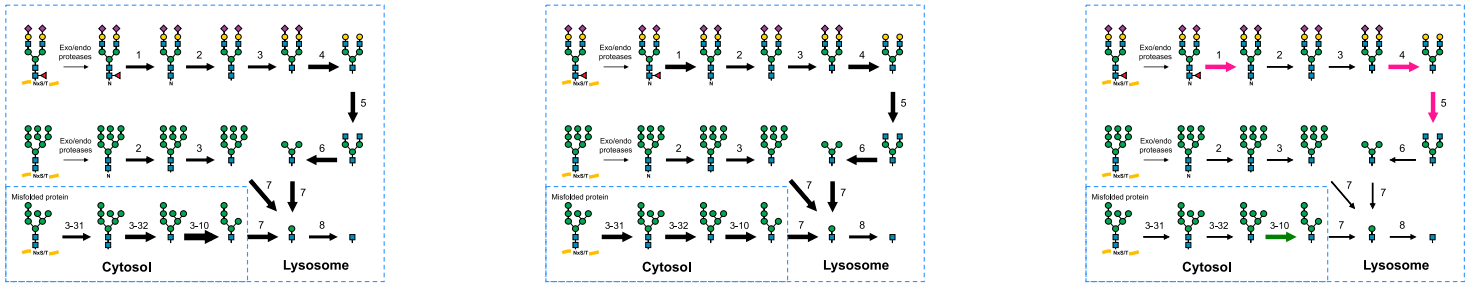
O



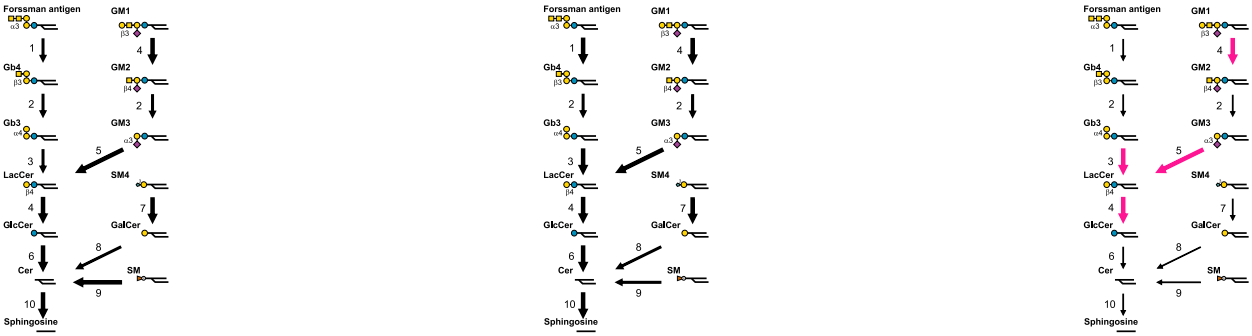
P



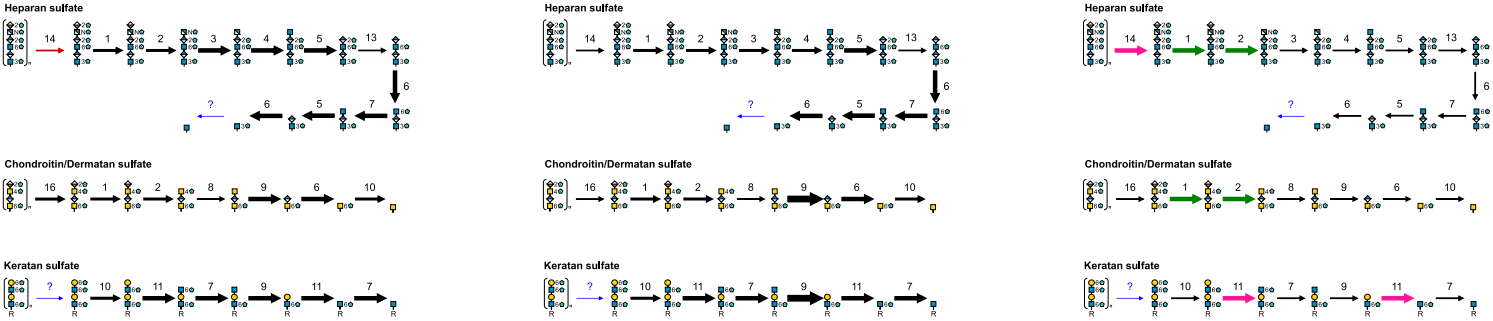
Q



R



S



TPM value of genes		TPM change of genes	
x < 0.1		0.5 ≤ fold < 2	
0.1 ≤ x < 1		fold ≥ 2	
1 ≤ x < 4		0.5 > fold	
4 ≤ x < 20			
20 ≤ x < 100			
100 ≤ x			

Man		ManNAc		Phoshate		CoA	Ribitol
Gal		GalNAc		Sulfate		Dolichol	
Glc		GlcNAc		Fructose		Polyprenol	
Fuc		GlcN		Ethanolamine		Diacylglycerol	
IdoA		Xyl		Choline		1-alkyl-2-acylglycerol	
GlcA		Neu5Ac		Inositol		Acyl-chain	

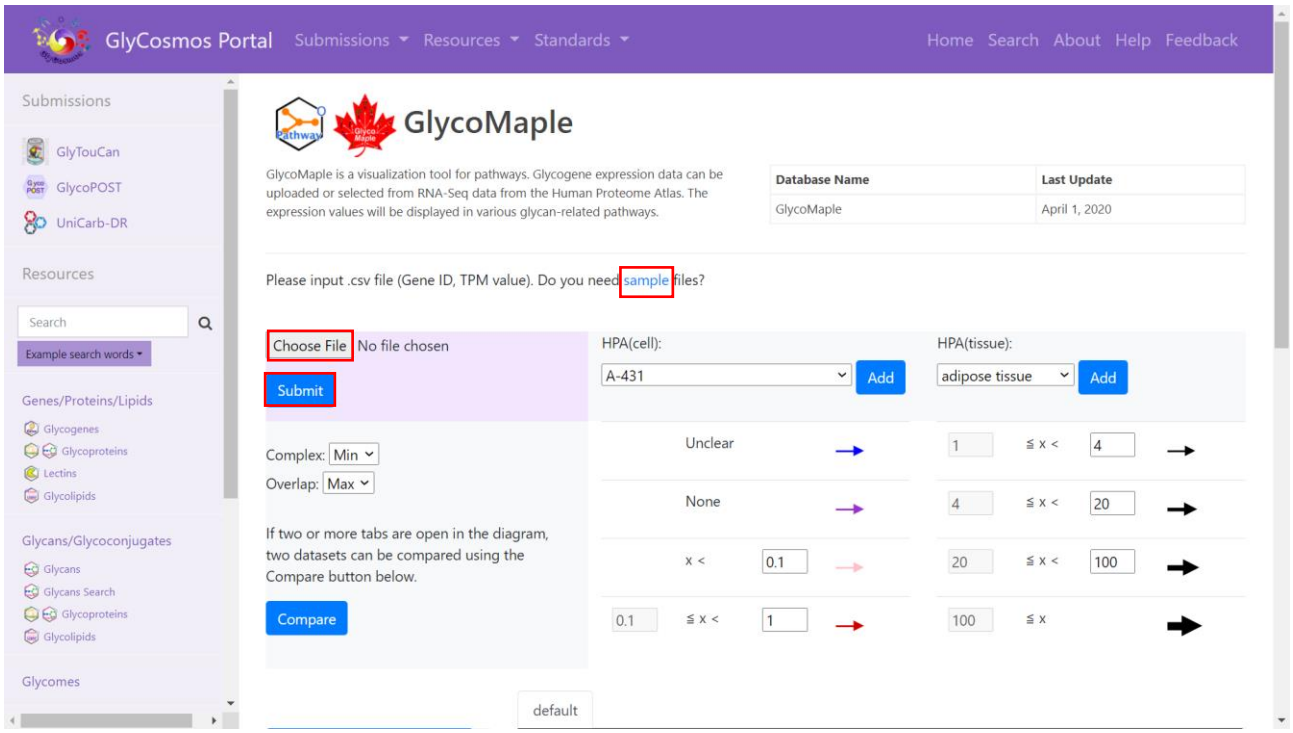
Figure S6. Related to Figure 6: Comparison of pathways for glycosylation pathways between colon normal and primary tumor tissue

Prediction of biosynthesis and catabolic pathways of glycans including (A) lipid-linked oligosaccharide (LLO) biosynthesis and OST, (B) N-glycan processing and branching , (C) complex capping of N-glycans / O-glycans / GSLs, (D) biosynthesis of GPI-anchored proteins, (E) biosynthesis of mucin-type O-glycans, (F) biosynthesis of O-fucose / O-glucose / collagen-O-galactose / O-GlcNAc / C-Man, (G) biosynthesis of O-Man, (H) biosynthesis of GAGs, (I) biosynthesis of HS, (J) biosynthesis of CS and DS, (K) biosynthesis of keratan sulfate (KS), (L) biosynthesis and catabolism of hyaluronan, (M) biosynthesis of core GSLs, (N) biosynthesis of Gbs, (O) biosynthesis of gangliosides, (P) biosynthesis of sugar nucleotides, (Q) lysosomal degradation of N-glycans, (R) lysosomal degradation of GSLs, and (S) lysosomal degradation of GAGs in colon normal (N = 304) (left), primary tumor tissue (N = 288) (middle) and comparison (right). The expression data (TPM values) of glycan-related genes were used for mapping. Arrows indicate the gene expression in each reaction. Thin pink arrows (TPM < 0.1) or solid red arrows (0.1 ≤ TPM < 1) indicate that the responsible genes for the pathways are not expressed or rarely expressed in the cells, respectively. The black arrows (1 ≤ TPM) indicate that the genes in the pathways are expressed in the cells. The thickness of these arrows shows the expression levels of the genes: thin black arrows, 1 ≤ TPM < 4; normal black arrows, 4 ≤ TPM < 20; thick black arrows, 20 ≤ TPM < 100; very thick black arrows, 100 ≤ TPM. If several genes overlapped in a reaction, the maximum TPM value among the overlapped genes were used. When several gene products make a complex for a reaction, the minimum TPM value of the subunit genes was used. Blue arrows indicate the reactions for which the responsible genes were not clear. In the comparison images (right), fold changes that were > 2 and < 0.5 are shown as pink and green arrows, respectively. Each numbered reaction is listed in Table S3.

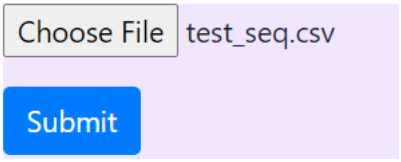
Methods S1. Related to STAR Methods: Instruction of GlycoMaple

Visualization of glycosylation pathways using RNA-seq data 1.
Access https://glycosmos.org/glycomaples/index in a browser.

Download a sample file by clicking the sample [sample]. A sample file including median TPM values of the primary tumors (N = 288) and normal colon tissues (N = 304) is available on the top of the main page.

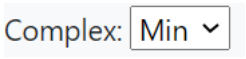


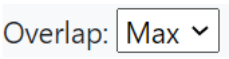
2. Click Choose File [Choose File] to choose the sample file downloaded (test_seq.csv).



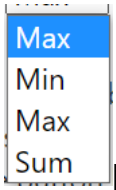
Alternatively, prepare RNA-seq data by yourself. The file should contain “Gene ID” in the first column, and followed by “TPM value in sample1”, “TPM value in sample2”,, from the second column. The data should be saved as a csv file

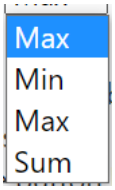
	A	B	C	D	E
1	Gene ID	HEK293_AVE	TKO_AVE		
2	A1BG	2.51	2.48		
3	A1CF	0.03	0		
4	A2M	1.05	2.85		
5	A2ML1	0.05	0.05		
6	A3GALT2	0.09	0.58		
7	A4GALT	3.32	2.39		
8	A4GNT	0.16	0.02		
9	AAAS	120.79	145.64		
10	AACS	33.98	32.14		
11	AADAC	0.06	0		
12	AADACL2	0	0		
13	AADACL3	0	0		
14	AADACL4	0.01	0		
15	AADAT	14.72	9.35		

3. Select the mode of 

[complex] and 

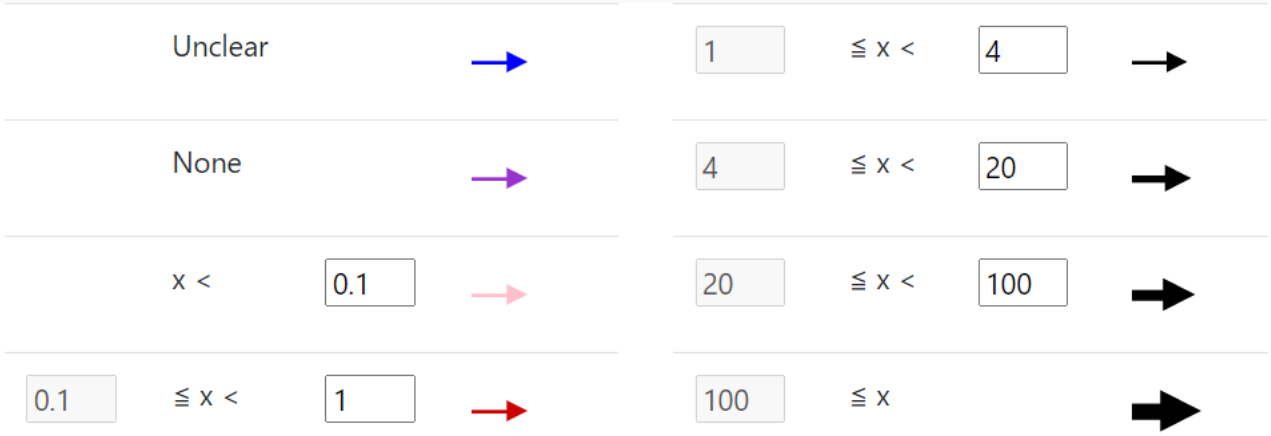
[overlap].

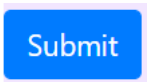
The GlycoMaple allows users to choose different modes to represent the value. Users can select 

[minimum] (Min), [maximum] (Max), [average] (Ave) or [summary] (Sum) in the [complex] and [overlap] parts. In the [complex] part, the default setting is [Min], which means when several gene products make a complex for a reaction, the minimum TPM value of the subunit genes is used to represent the arrow. In the [overlap] part, the default setting is [Max]. This means if several isoenzyme genes overlapped in a reaction, the maximum TPM value of these genes is used to show the arrow. [Ave] uses the average value of the TPM values and [Sum] uses the summary of the TPM values to represent the arrow, respectively. Before submitting this file, users need to select 

"minimum, maximum, average, summary" in the [complex] and [overlap] parts.

4. Select threshold values to show the thickness of arrows with different TPM values.



5. Click 

[submit], then start the processing.

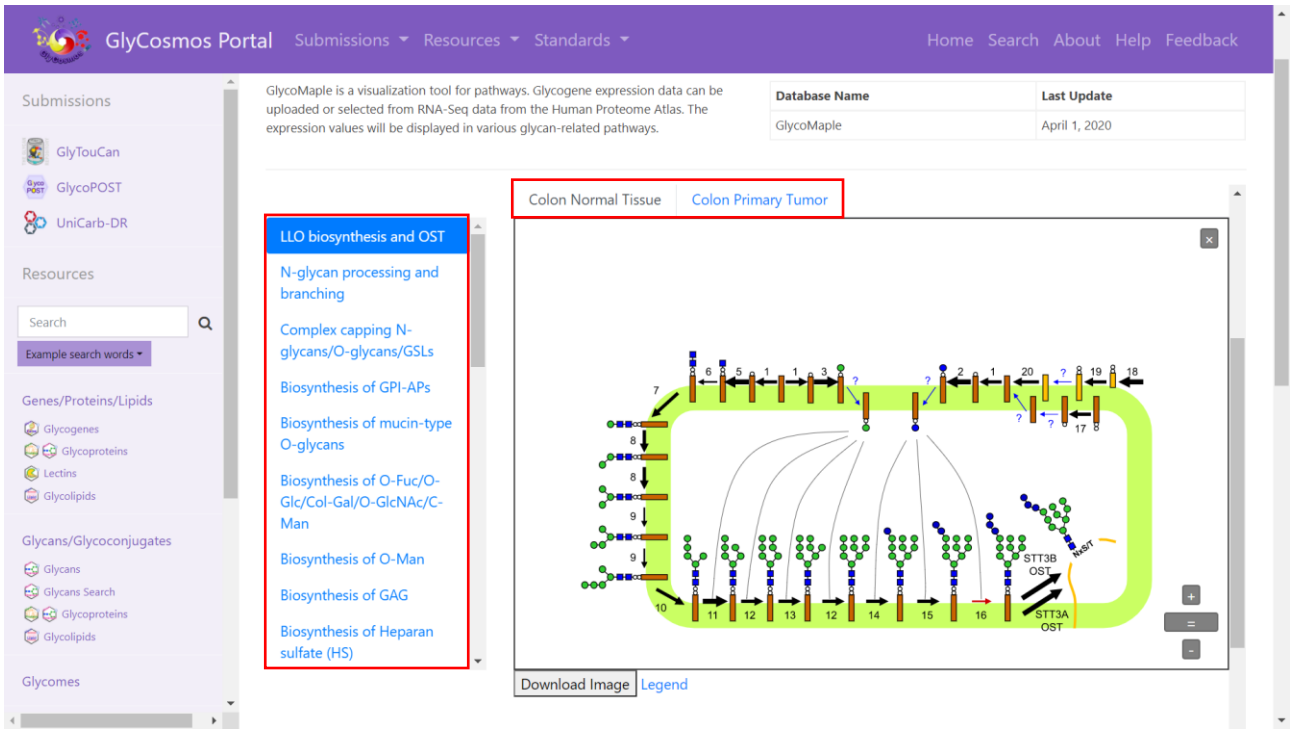
After confirming the setting, click the "Submit" bottom to submit RNA-seq data. The processing will be finished within a few second.

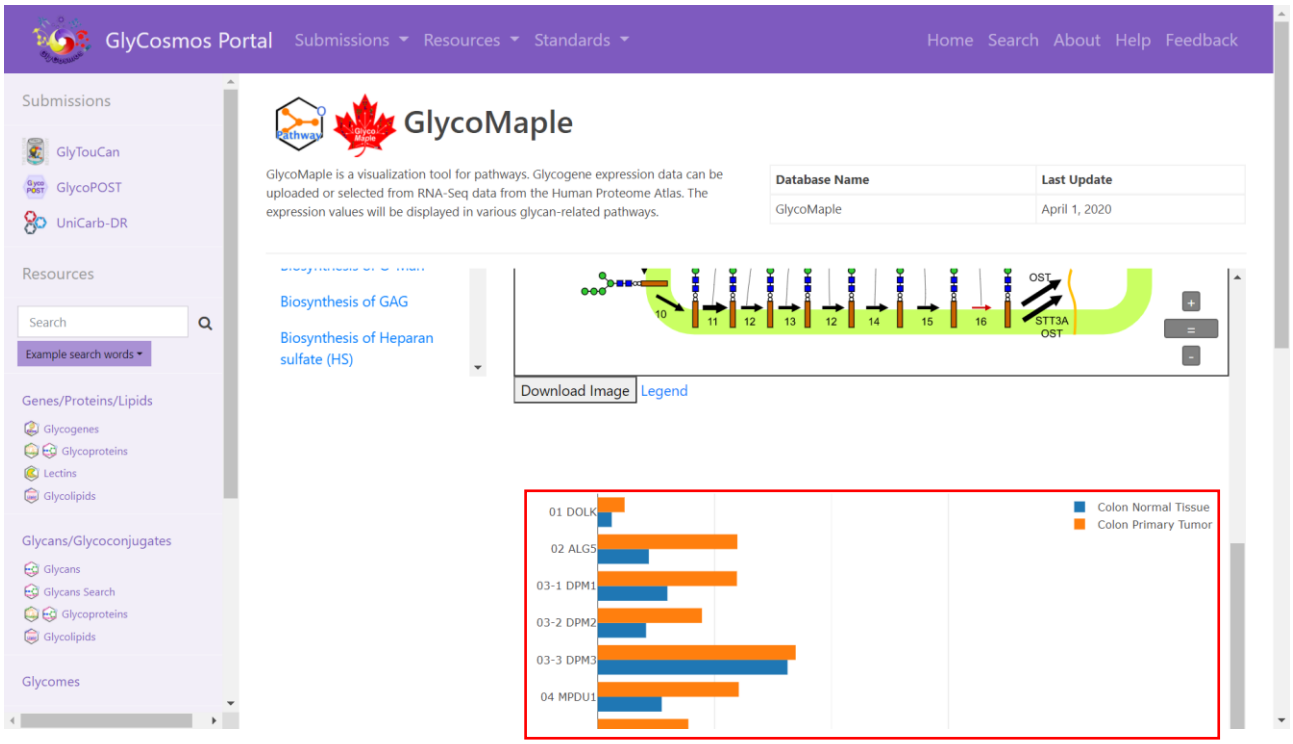
6. Check glycosylation maps and bar plots

The thickness of arrows in glycosylation pathway maps is changed depending on the TPM values of responsible genes. If there is a reaction called A->B->C->D and there is a defect in the gene involved in the reaction B->C, which shows a red or pink arrow, it is assumed that products C and D will not be synthesized no matter how the genes involved in the C->D reaction is highly expressed.

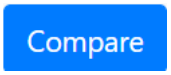
Each arrows links to the human gene nomenclature website, and show information of the genes responsible to the reactions. The TPM values are shown at the bottom of the page as a bar plot. The number of the gene name in the plot are corresponded to the number on the map.

By uploading your own RNA-seq data, 19 glycan metabolic pathways and expression profiles of about 950 glycan-related genes can be visualized.

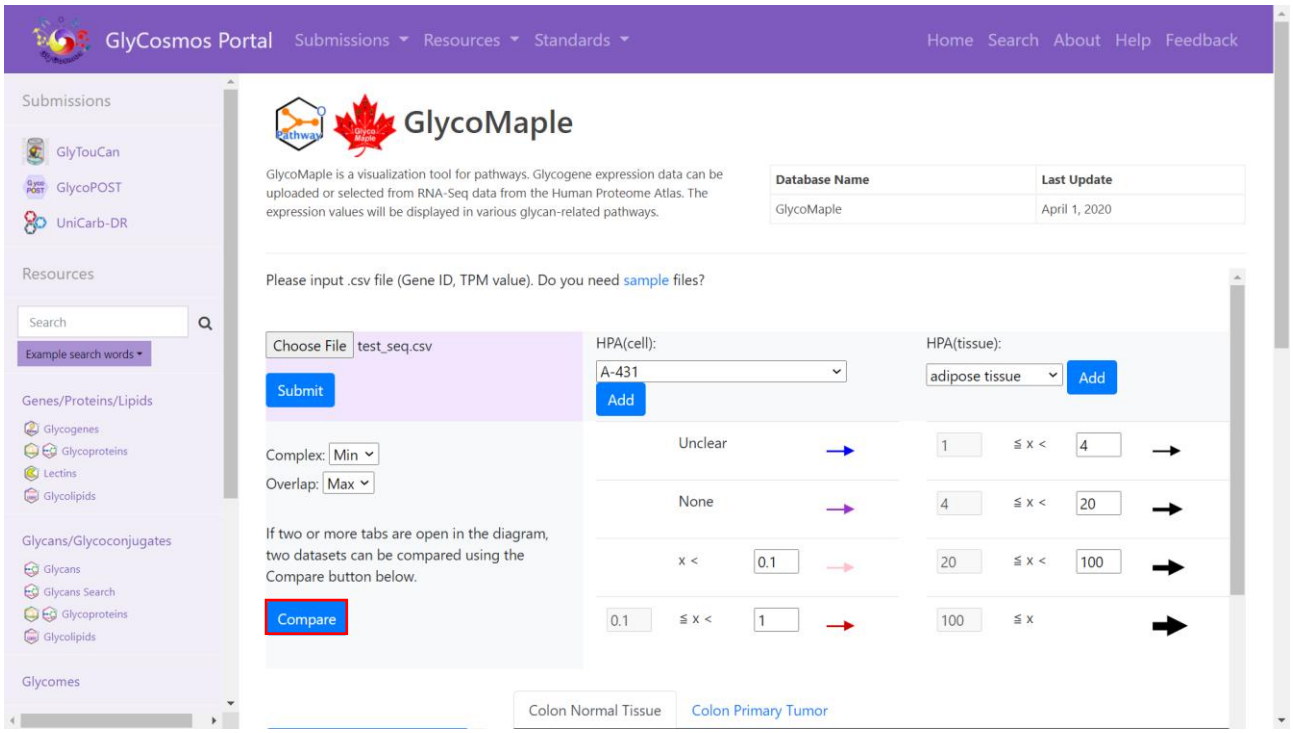


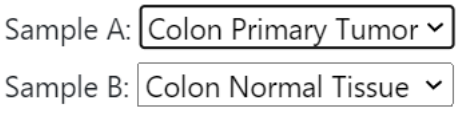


Comparison of glycosylation pathways in different cells/tissues data.

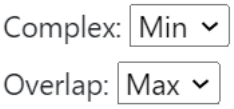
7. Click 

[compare].



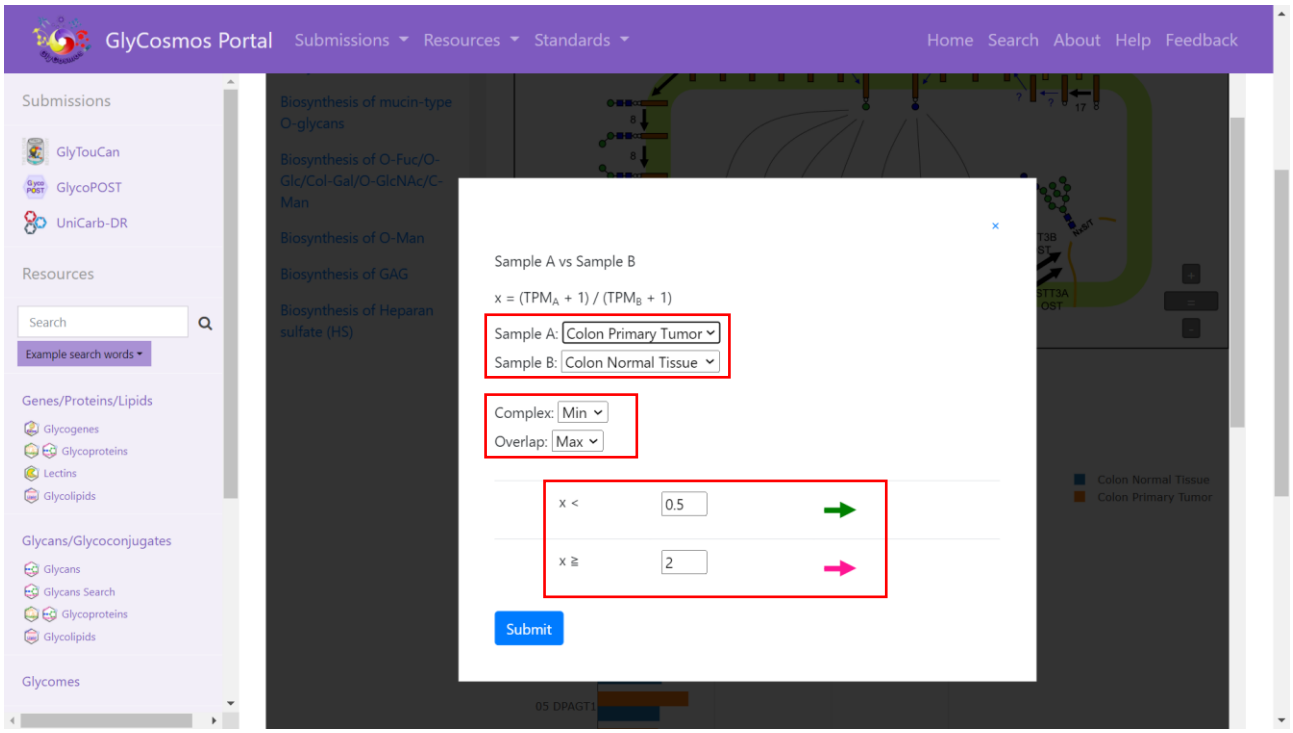
8. Pick two data that you want to compare. 

Two sets of data can be compared in GlycoMaple. Click [compare] and choose two samples that you want to compare.

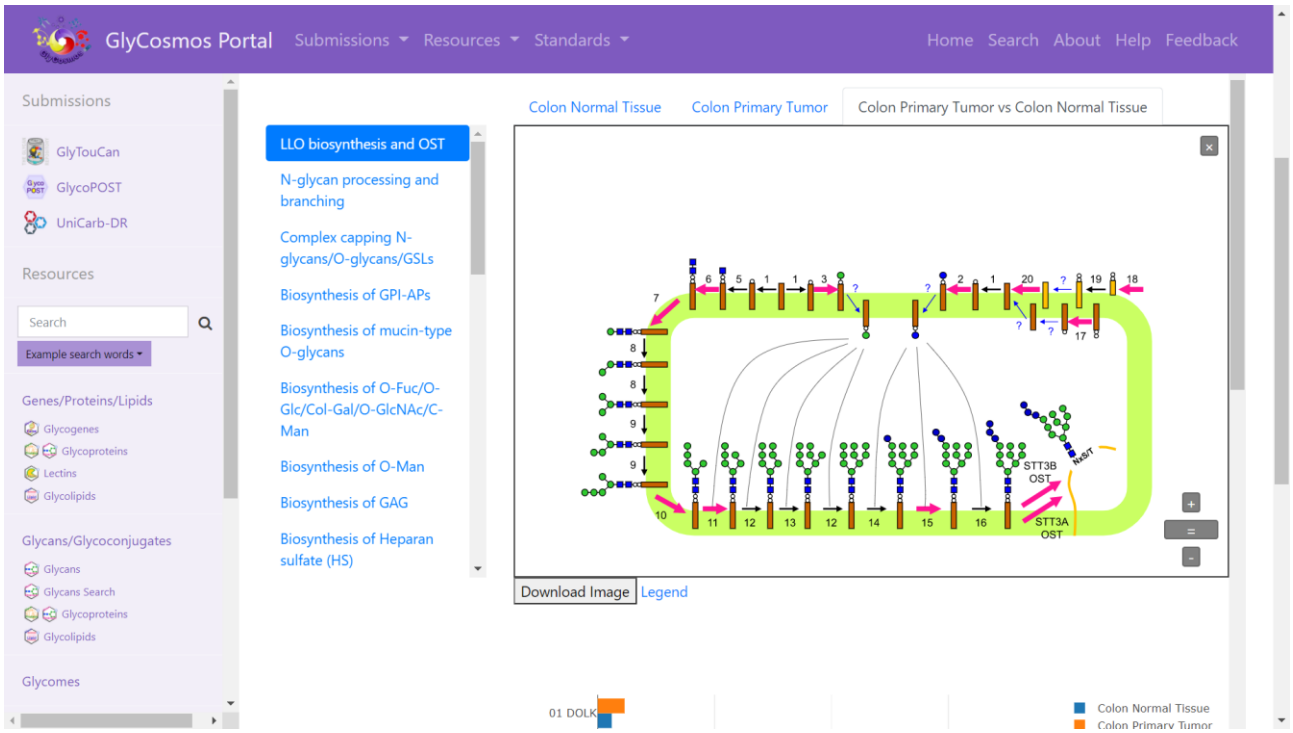
9. Confirm the setting of 

[complex] and [overlap].

10. Confirm the threshold of fold changes to show green and pink arrows.



11. Click the Submit [submit].



Visualization of glycosylation pathways in human cells/tissues using HPA data.

The RNA-seq data of 64 human cell lines and 37 human tissues in the Human Protein Atlas (HPA) have been already deposited on the website. When you select the cells or tissues of your interest, the glycosylation pathway maps in the cells/tissues can be presented.

12. Choose cells/tissues of interest.
13. Click “Add”.