Probing the contribution of individual polypeptide GalNAc-transferase isoforms to the O-glycoproteome by inducible expression in isogenic cell lines

John Hintze[1], Zilu Ye[1], Yoshiki Narimatsu[1], Thomas Daugbjerg Madsen[1], Hiren Joshi[1], Adam Linstedt[2], Collin Bachert[2], Ulla Mandel[1], Eric P. Bennett[1], Sergey Y. Vakhrushev[1], Katrine T. Schjoldager[1,3]

[1]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark.

[2]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, United States

Running title: O-glycoproteome contribution of individual polypeptide GalNAc-transferases

To whom correspondence should be addressed: Katrine T. Schjoldager; [1]Copenhagen Center for Glycomics, Dept. of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. Tel.: 45-35327797; Fax: 45-35367980; E-mail: schjoldager@sund.ku.dk.

Keywords: Inducible expression, O-glycosylation, Glycoproteomics, GalNAc-transferase, O-GalNAc, Protein glycosylation, Mucin type, Glycan, Glycosyltransferase

---

Contents:

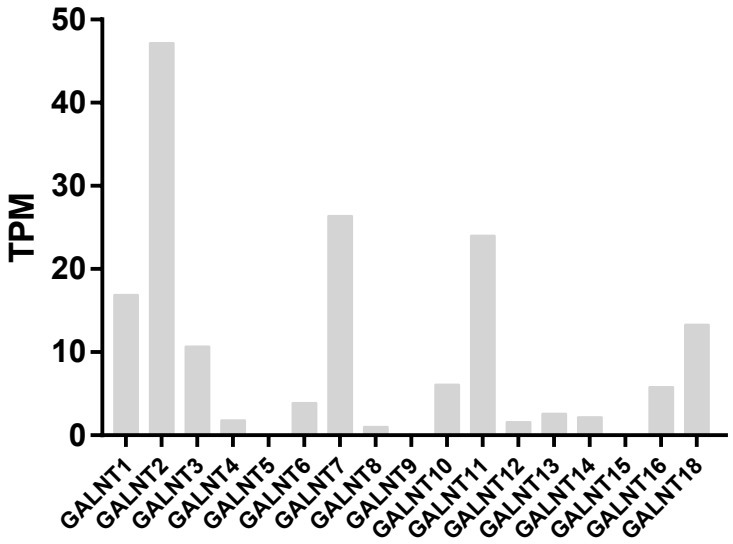
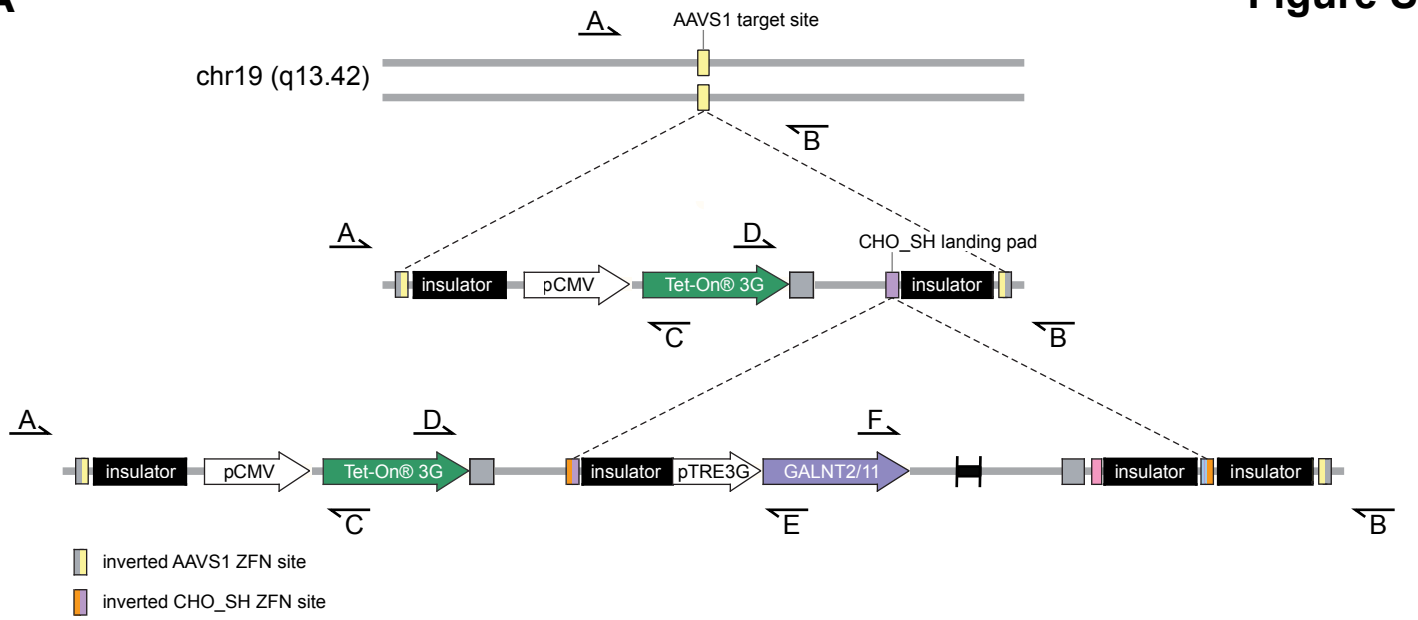


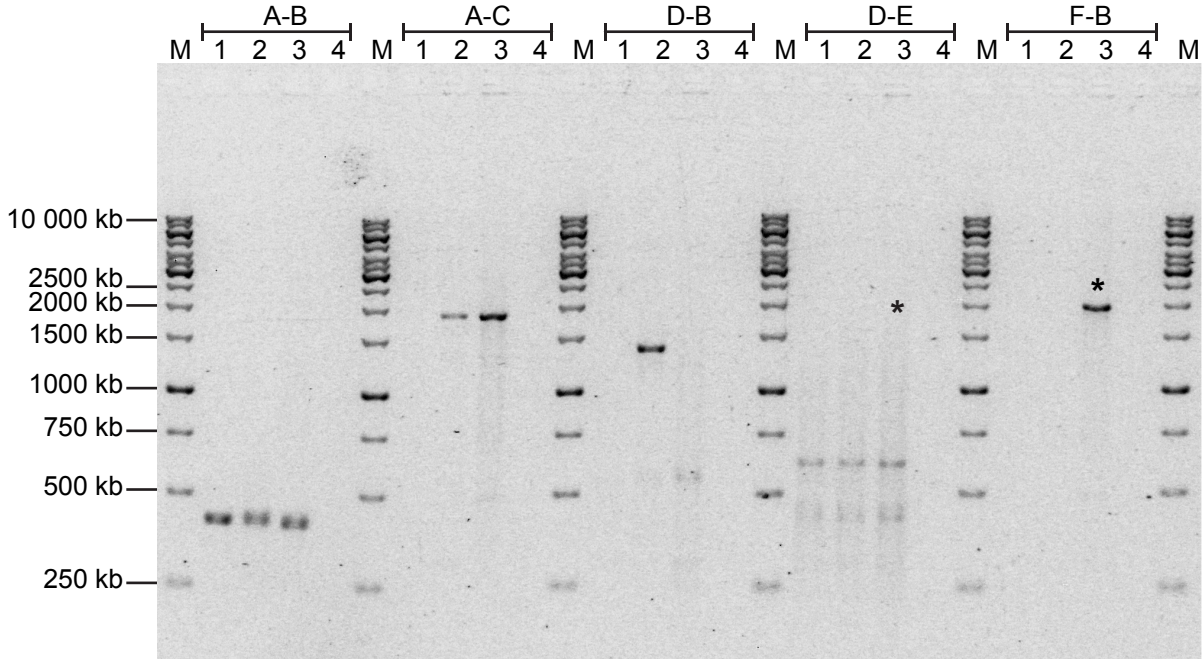
Figure S1. HEK293 *GALNT* **expression profile.** RNA-seq expression data of the human *GALNT* gene family in HEK293 cells. data downloaded from the human protein atlas (1) (https://www.proteinatlas.org/about/download). *GALNT2*, *GALNT7* and *GALNT11* are the most highly expressed isoforms in HEK293.

Figure S2

A

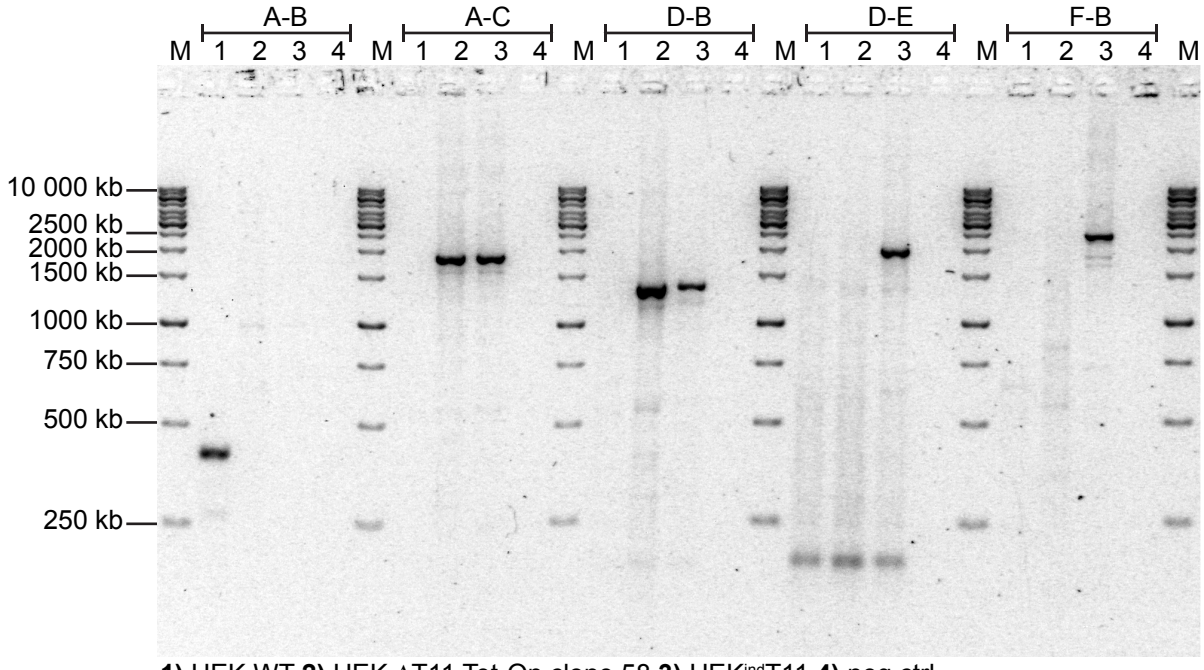


B



1) HEK WT **2)** HEK ΔT2 enriched Tet-On bulk **3)** HEK^{ind}T2 **4)** neg ctrl

C

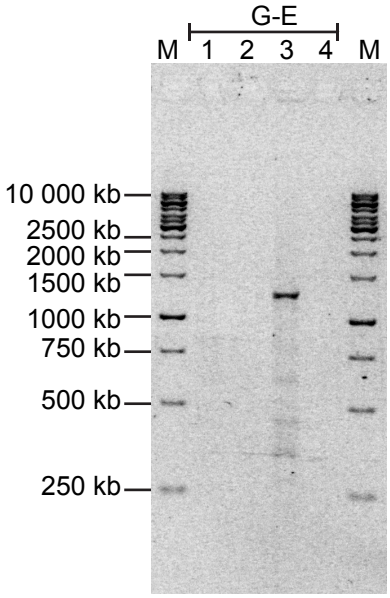


1) HEK WT **2)** HEK ΔT11 Tet-On clone 58 **3)** HEK^{ind}T11 **4)** neg ctrl

Figure S2. Confirmation of KI architecture by junction PCR. A) Graphic depiction of primer (A-E) binding sites. **B)** Junction PCR reactions for HEK WT, HEKΔT2 enriched Tet-On bulk and HEK^{ind}T2 with indicated primer pairs, separated on a 2 % agarose gel. All reactions for HEK^{ind}T2 produced the expected size, except for the internal (D-E) and right (F-B) junction. The expected products are indicated with an asterisk. Positive WT (A-B) and left junction (A-C) for HEK^{ind}T2 reactions indicates mono-allelic integration of the Tet-On cassette. The right junction (F-B) of the inducible *GALNT2* cassette is positive but 500 bp smaller than the expected size. Sanger sequencing reveled it to be due to lack of one insulator element. **C)** Junction PCR reactions for HEK WT, HEK ΔT11 Tet-On clone 58 and HEK^{ind}T11 with indicated primer pairs, separated on a 2 % agarose gel. All reactions produce the expected product sizes. Negative WT (A-B), positive left (A-C) and right (D-B) junction indicate that HEKΔT11 Tet-On clone 58 is bi-allelic for the first KI. Positive left (A-C), right (D-B), internal (D-E) second right (F-B) junction for HEK^{ind}T11, all with the expected sizes, indicate that the clone is mono-allelic for the second KI.

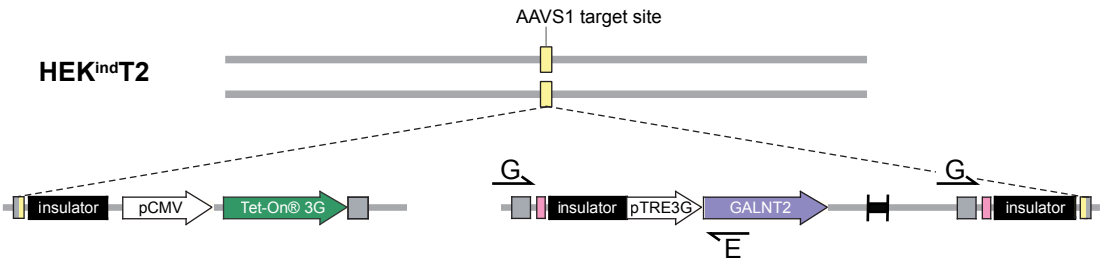
Figure S3

A



1) HEK WT **2)** HEK ΔT2 enriched Tet-On bulk **3)** HEK^{ind}T2 **4)** neg ctrl

B



C

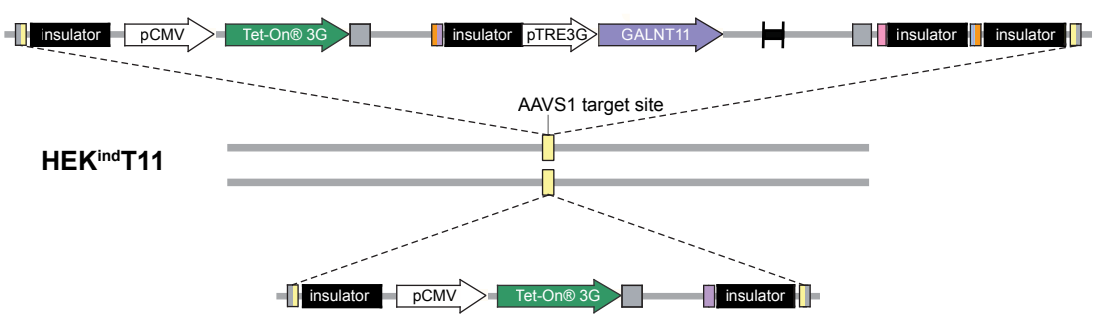
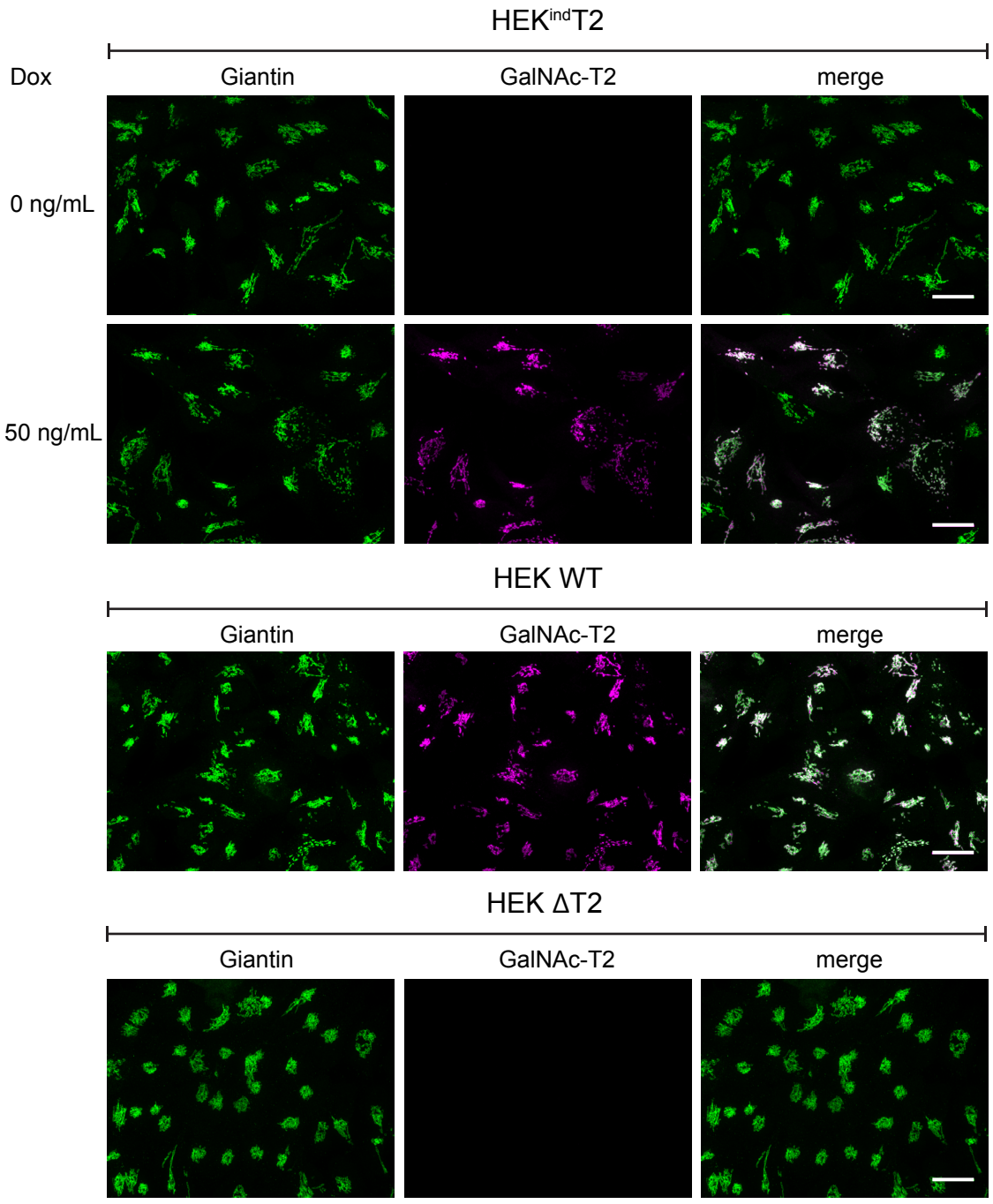


Figure S3

Figure S3. Detailed mapping of HEK^{ind}T2 architecture by junction PCR. A) Agarose gel separation of additional junction PCR for HEK^{ind}T2 using primer pair G-E generated a product of 1300 bp. This PCR product was sequenced confirming a junction between two *GALNT2* cassettes. Further PCR analysis using primers binding 3' of primer G did not result in product. We hypothesize that the unforeseen tandem integration of the *GALNT2* cassette caused truncation of the 3' end of the first cassette. **B)** Suggested architecture of the inducible platform for HEK^{ind}T2. **C)** Suggested architecture of the inducible platform for HEK^{ind}T11.

Figure S4


HEK^indT2
Dox
Giantin
GalNAc-T2
merge
0 ng/mL
50 ng/mL
HEK WT
Giantin
GalNAc-T2
merge
HEK ΔT2
Giantin
GalNAc-T2
merge

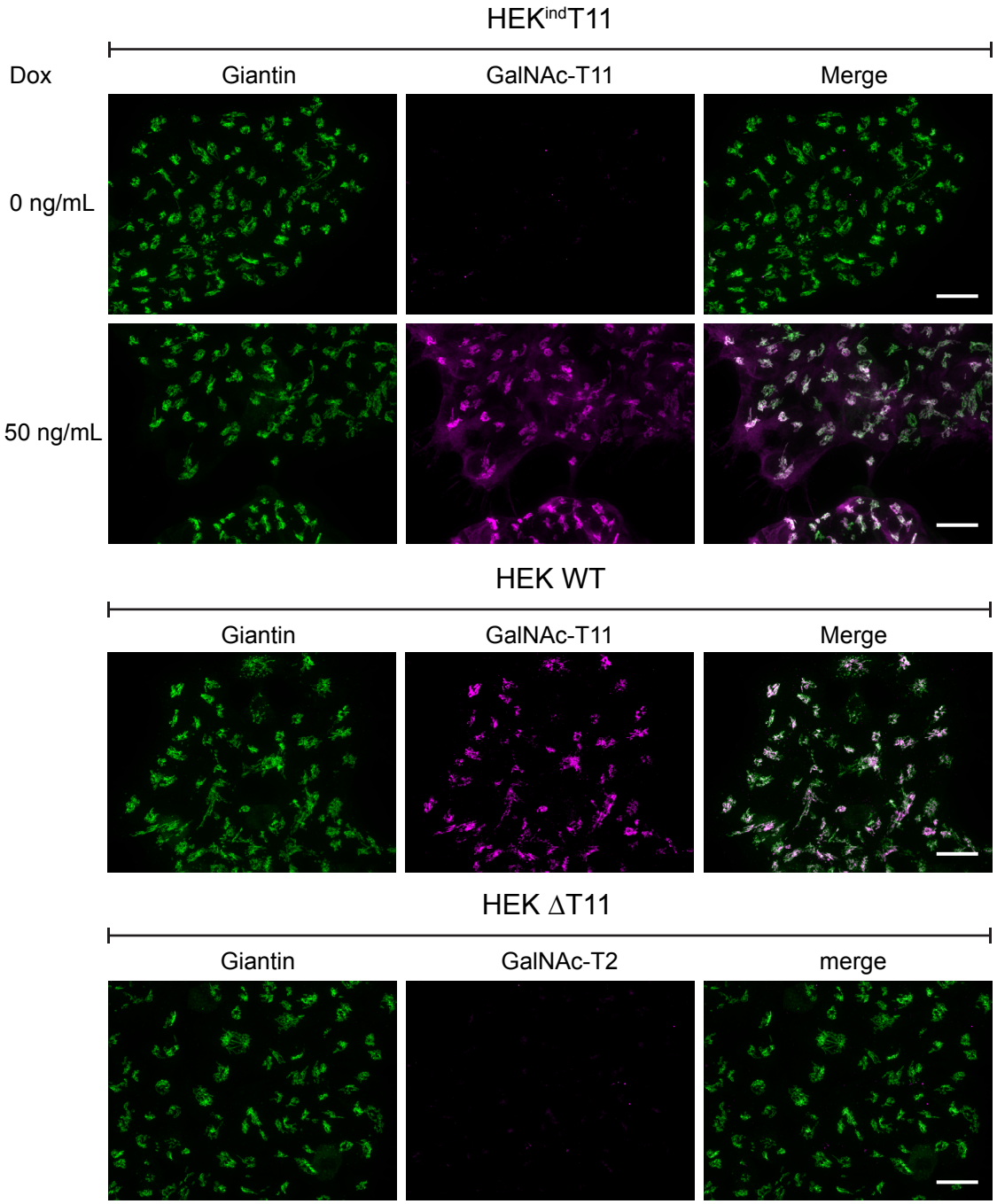


Figure S4. Sub-cellular localization of induced GalNAc-T2 and T11. Immunofluorescence stainings of HEK^{ind}T2 **(A)** or HEK^{ind}T11 **(B)** cells grown on coverslips stained with antibodies against Golgi marker Giantin, GalNAc-T2 or GalNAc-T11, and subsequently imaged by confocal microscopy. Induced GalNAc-T2 and GalNAc-T11 colocalizes with Giantin as in HEK WT cells. Scale bar = 20 µm.

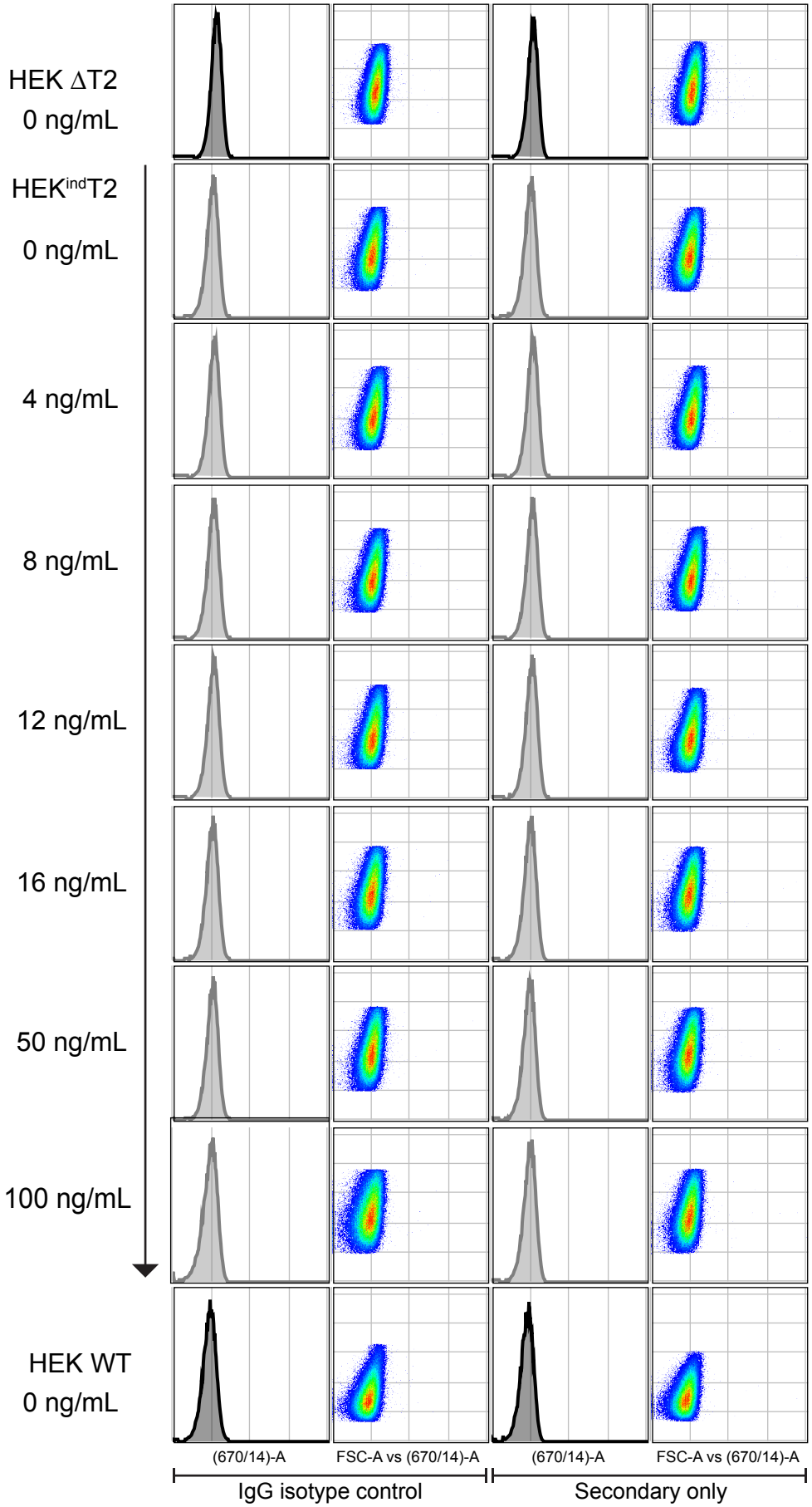


Figure S5. Control stainings for flow cytometric quantification of induced GalNAc-T2. HEK ΔT2, HEK^{ind}T2 and HEK WT cells were seeded, cultured for 24 h and then induced with doxycycline at indicated concentration for 48 h before fixation, permeabilization and staining with anti-IgG isotype control in combintation with Alexa Fluor 647 secondary antibody or Alexa Fluor 647 secondary antibody only. Neither control show any variation between samples.

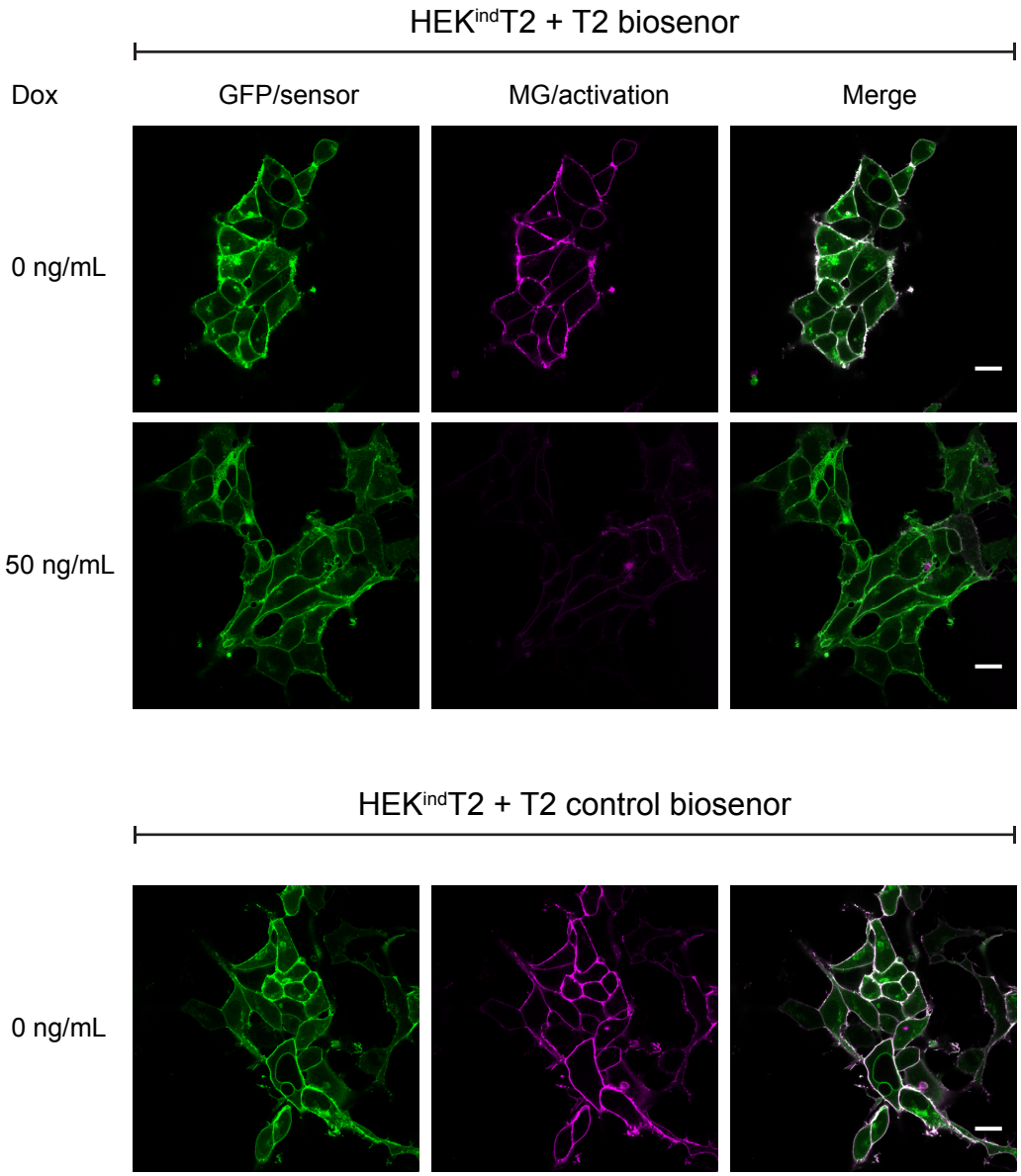
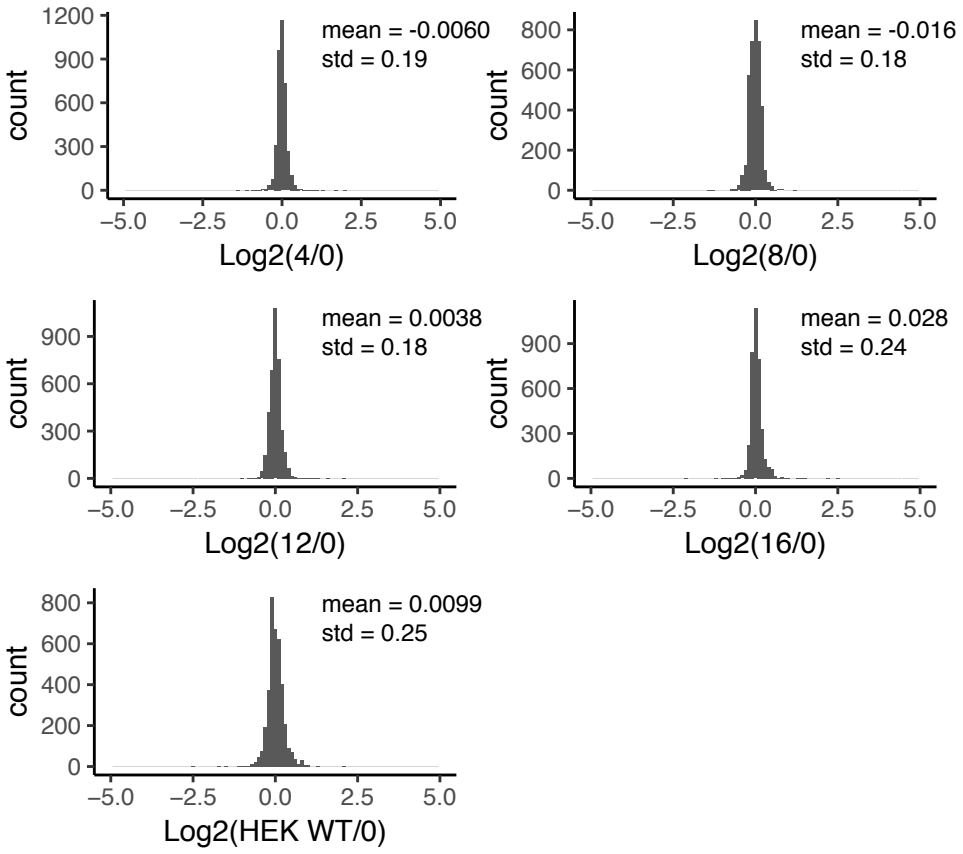


Figure S6. Live cell imaging of GalNAc-T2 specific biosensor activation. HEK^{ind}T2 cells stably expressing T2 biosensor or control T2 biosensor (Δgly) were grown on coverslips and induced for 48 h before being incubated with cell impermeable malachite green (MG) and imaged by fluorescence microscopy. The sensor traffics to the cell surface and glycosylation by induced GalNAc-T2 leads to decrease in MG signal. Non-induced cells display strong sensor activation and MG signal, comparable to that of cells expressing the control biosensor. Scale bar = 20 μm.

A



B

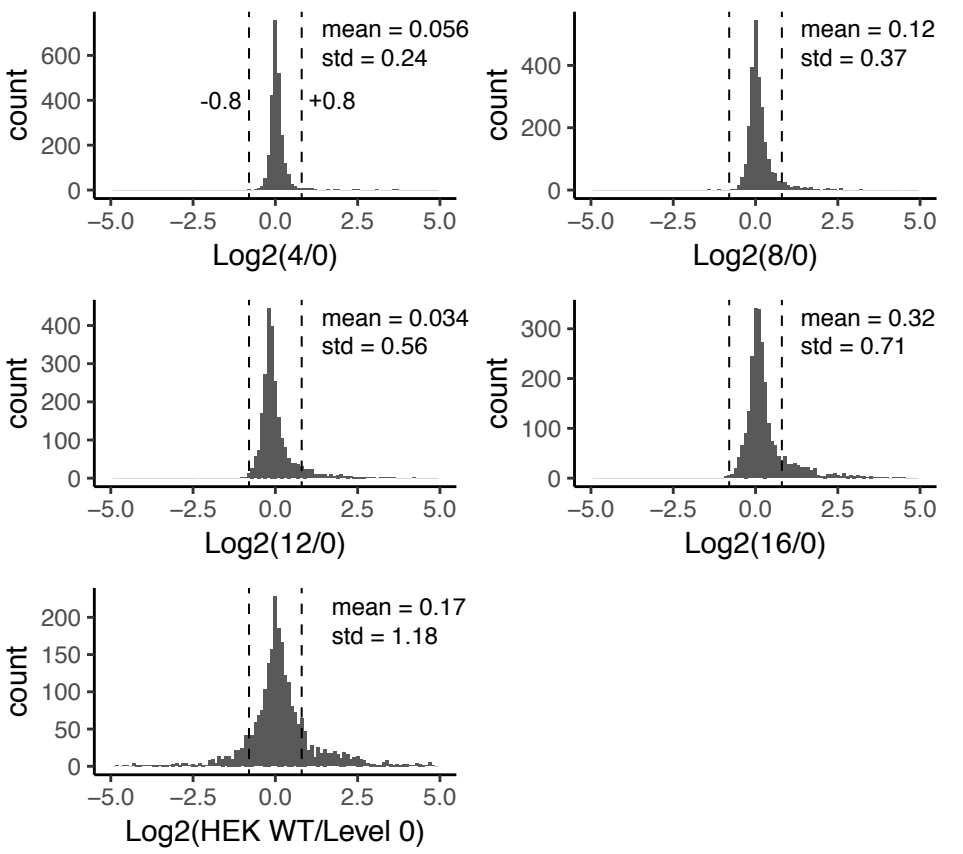
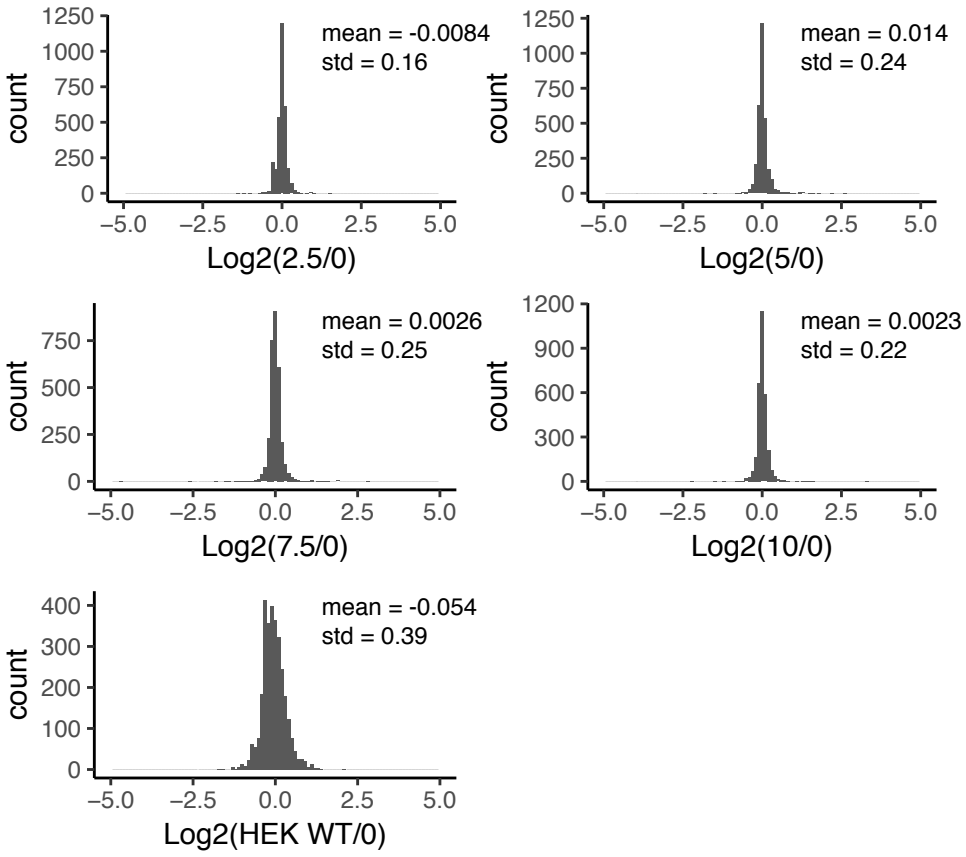


Figure S7. Histograms of quantified peptides and O-glycopeptides from HEK^{ind}T2. A) LC-MS3 quantified peptides in the ratio check were normalized to HEK^{ind}T2 cultured in absence of doxycycline (0) and log2 transformed. Mean and standard deviation, assuming normal distribution, showed for each histogram. **B)** Quantified O-glycopeptides were normalized to HEK^{ind}T2 cultured in absence of doxycycline (0) and log2 transformed. Dashed vertical lines at +/- 0.8 where 0.8 (approximately 3x standard deviation at the lowest induction level) was chosen as a threshold for grouping of induced O-glycopeptides.

A



B

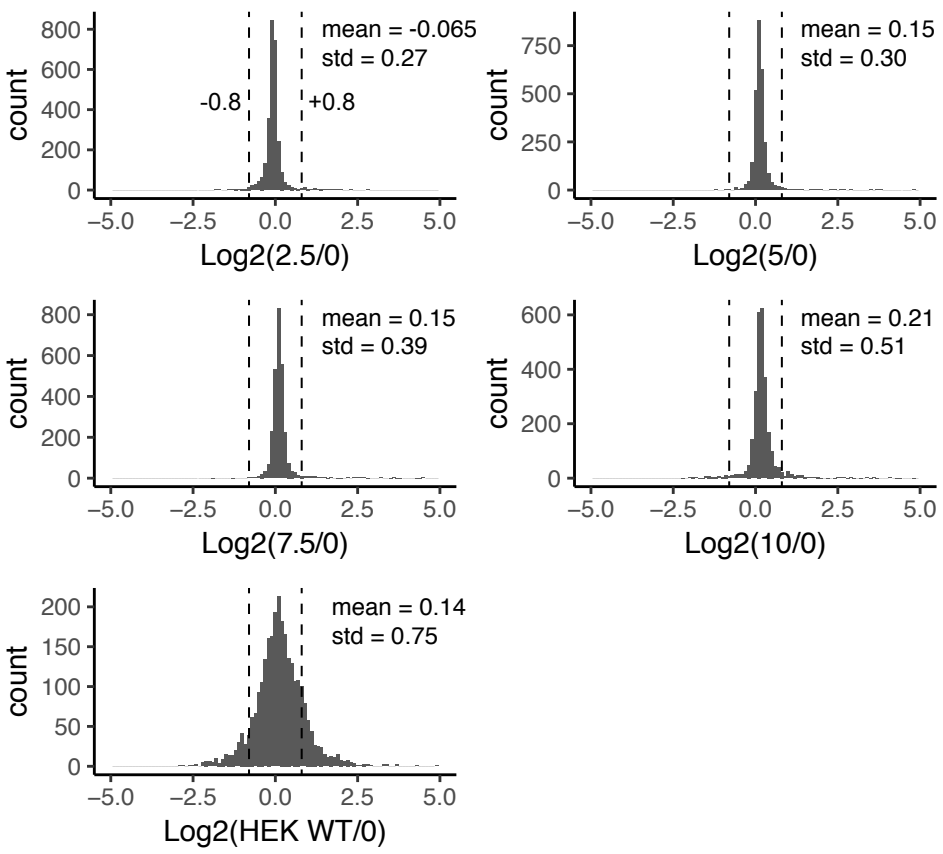


Figure S8. Histograms of quantified peptides and O-glycopeptides from HEK^{ind}T11. A) LC-MS3 quantified peptides in the ratio check were normalized to HEK^{ind}T11 cultured in absence of doxycycline (0) and log2 transformed. Mean and standard deviation, assuming normal distribution, showed for each histogram. **B)** Quantified O-glycopeptides were normalized to HEK^{ind}T11 cultured in absence of doxycycline (0) and log2 transformed. Dashed vertical lines at +/- 0.8 where 0.8 (approximately 3x standard deviation at the lowest induction level) was chosen as a threshold for grouping of induced O-glycopeptides.

Figure S9

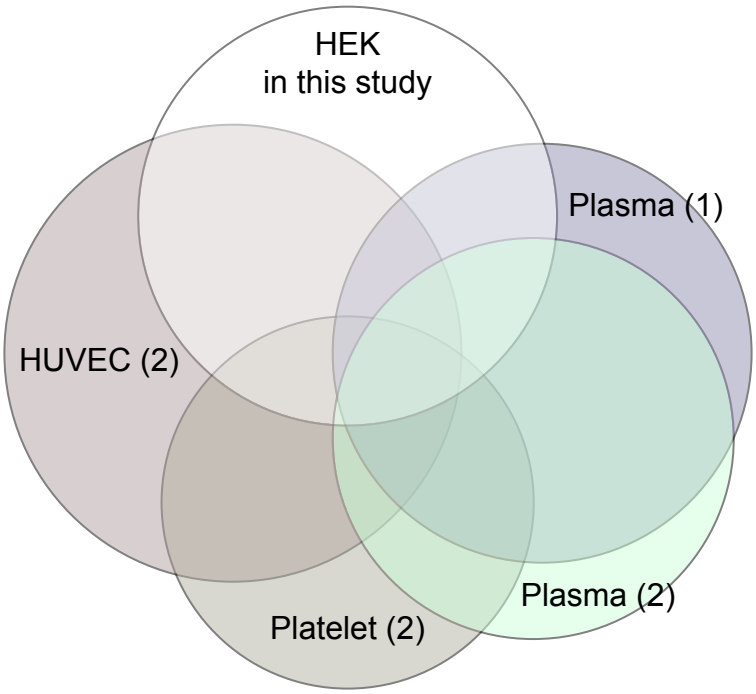


Figure S9. Overlap of O-glycoproteins identified in this and previous studies. Venn diagram illustration of sialyl-T (ST) O-glycoproteins identified in this study compared to those identified in Khetarpal et al. (2) and King et al. (3) from plasma, platelets and HUVEC cells.

.

Figure S10

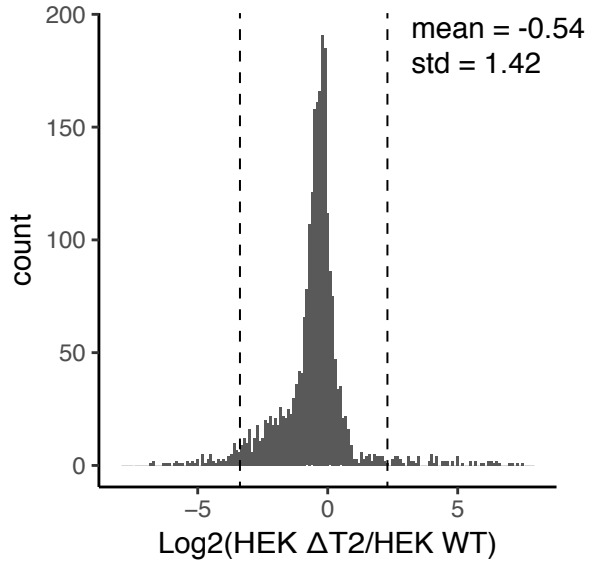


Figure S10. Histogram of quantified O-glycopeptides from HEKΔT2. Quantified O-glycopeptides normalized to HEK WT. Dashed vertical lines at +/- two standard deviations from the mean. As this experiments compares WT with loss-of-function T2-specific glycopeptides, minus two standard deviations from the mean was used as a cut-off (-3.38) to define O-glycopeptides significantly affected by *GALNT2* KO.

Figure S11

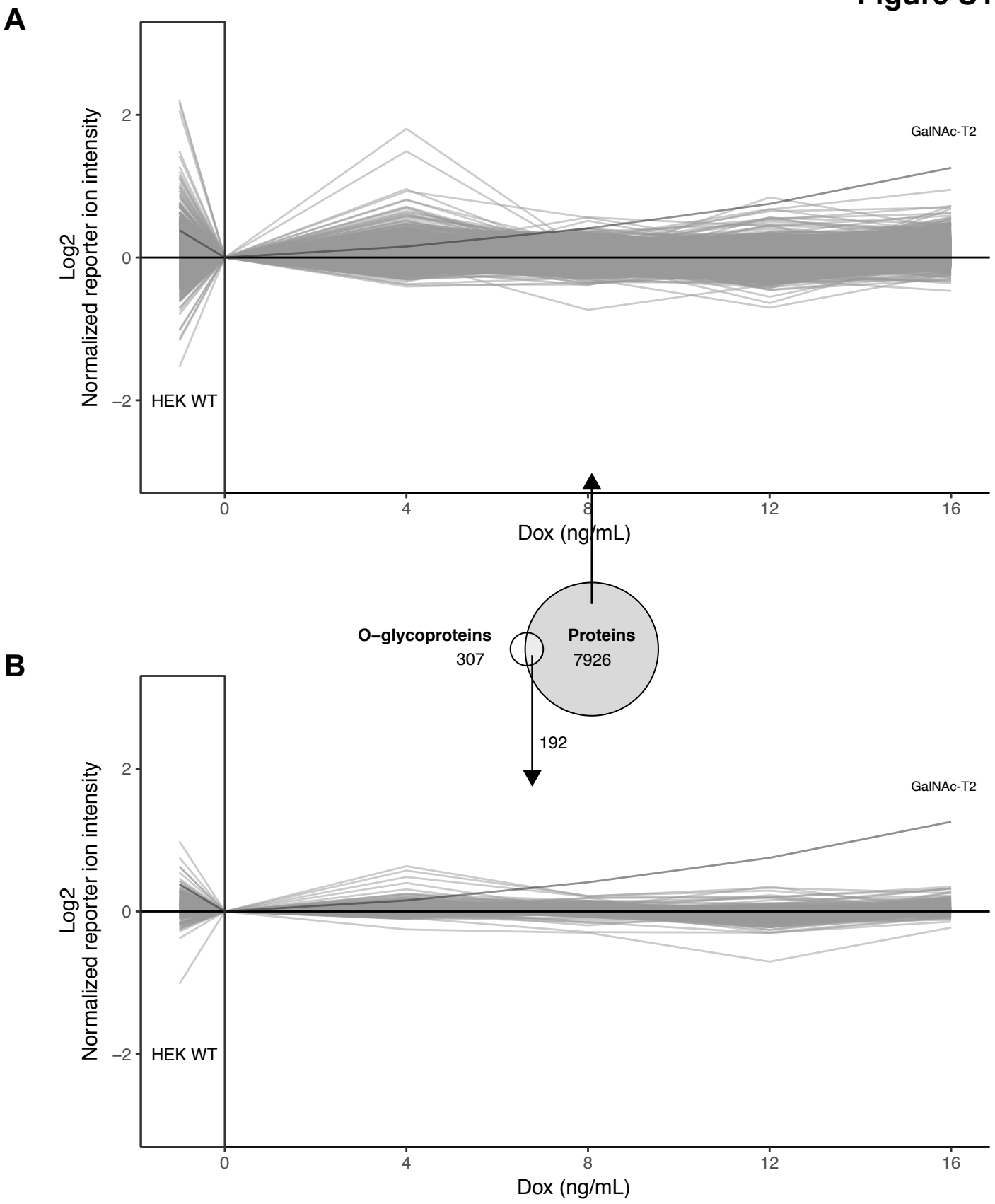


Figure S11. Deep proteome analysis of HEK^{ind}T2 LWAC flow-through. A) In the LWAC FT peptides from 7926 proteins were quantified across all cultures. Data was normalized to HEK^{ind}T2 cultured in absence of doxycycline, log2 transformed and subsequently plotted. Each line represents a protein. GalNAc-T2 (shown in black) display a clear trend of increased quantification with increasing doxycycline concentration. **B)** Of the 307 O-glycoproteins identified in the HEK^{ind}T2 O-glycoproteome 192 were also identified in the LWAC FT.

Figure S12

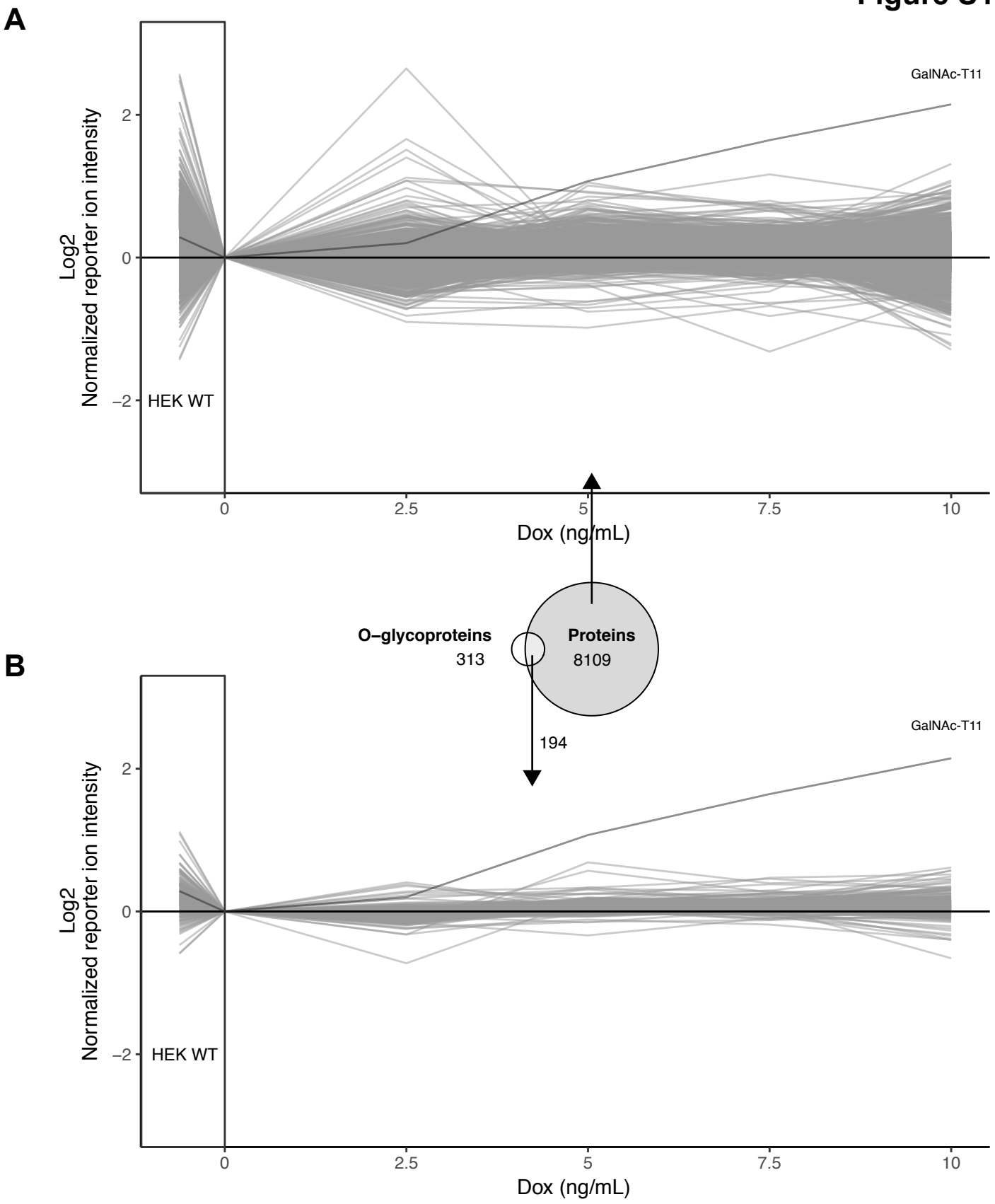


Figure S12. Deep proteome analysis of HEK^{ind}T11 LWAC flow-through. A) In the LWAC FT peptides from 8109 proteins were quantified across all cultures. Data was normalized to HEK^{ind}T11 cultured in absence of doxycycline, log2 transformed and subsequently plotted. Each line represents a protein. GalNAc-T11 (shown in black) display a clear trend of increased quantification with increasing doxycycline concentration. **B)** Of the 313 O-glycoproteins identified in the HEK^{ind}T11 O-glycoproteome 194 were also identified in the LWAC FT.

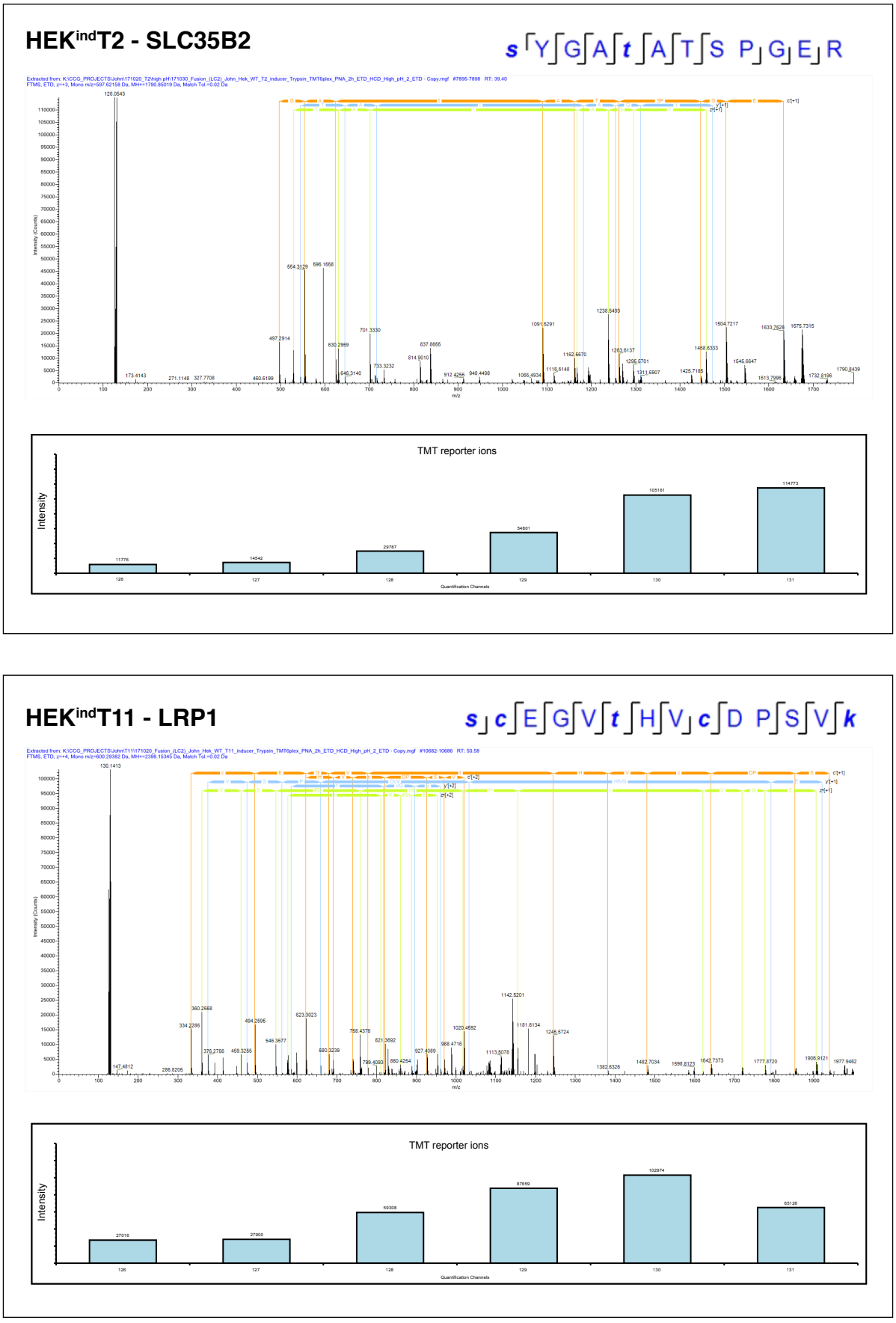


Figure S13. Mass spectra of two induced O-glycopeptides. Representative spectra showing O-glycopeptide identification and quantification of SLC35B2 ^{137}SYGA**T**ATSPGER^{148} in HEK^{ind}T2 and LRP1 LA8-9 ^{1096}SCEGV**T**HVCDPSVK^{1109} in HEK^{ind}T11 (glycosite shown in bold). For the given O-glycopeptides, the ETD spectrum is shown together with the TMT quantification channel values. The ETD spectrum provides unambiguous assignment of the O-glycan composition. TMT channel values were extracted from the corresponding HCD spectrum.

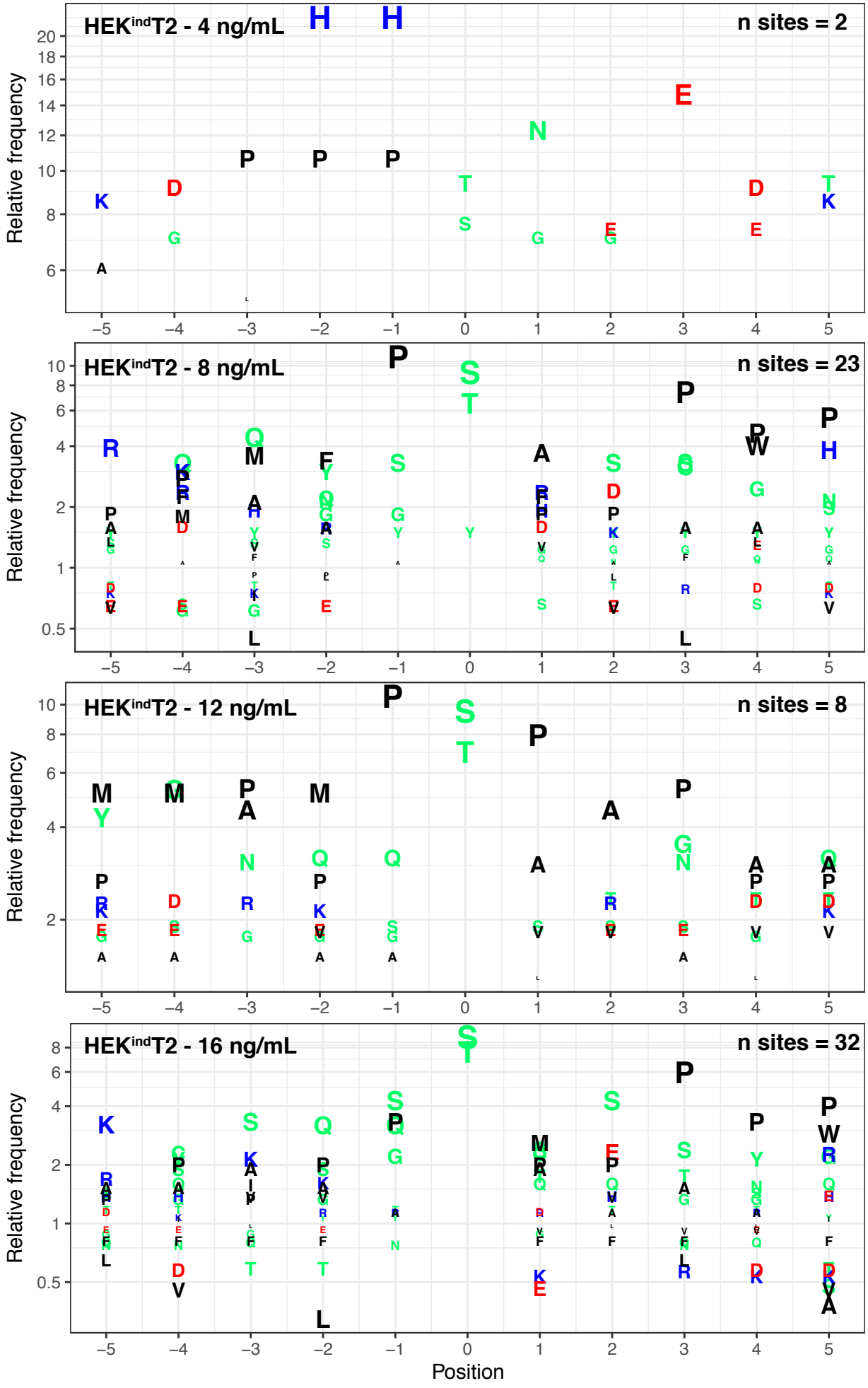


Figure S14. Alignment of induced O-glycosites from HEK^{ind}T2. O-glycosites were extracted from unambiguous single-site O-glycopeptides of each group and aligned in a sequence window spanning +/- 5 amino acids from the O-glycosite. The amino acid frequency is relative to the frequency of amino acids in the UniProt human proteome database (December 2013). GalNAc-T2's known preference for proline at position -1 is evident, and prolines are frequent at position +1 and +3 for glycosites induced at a doxycycline concentration of 8 ng/mL or more.

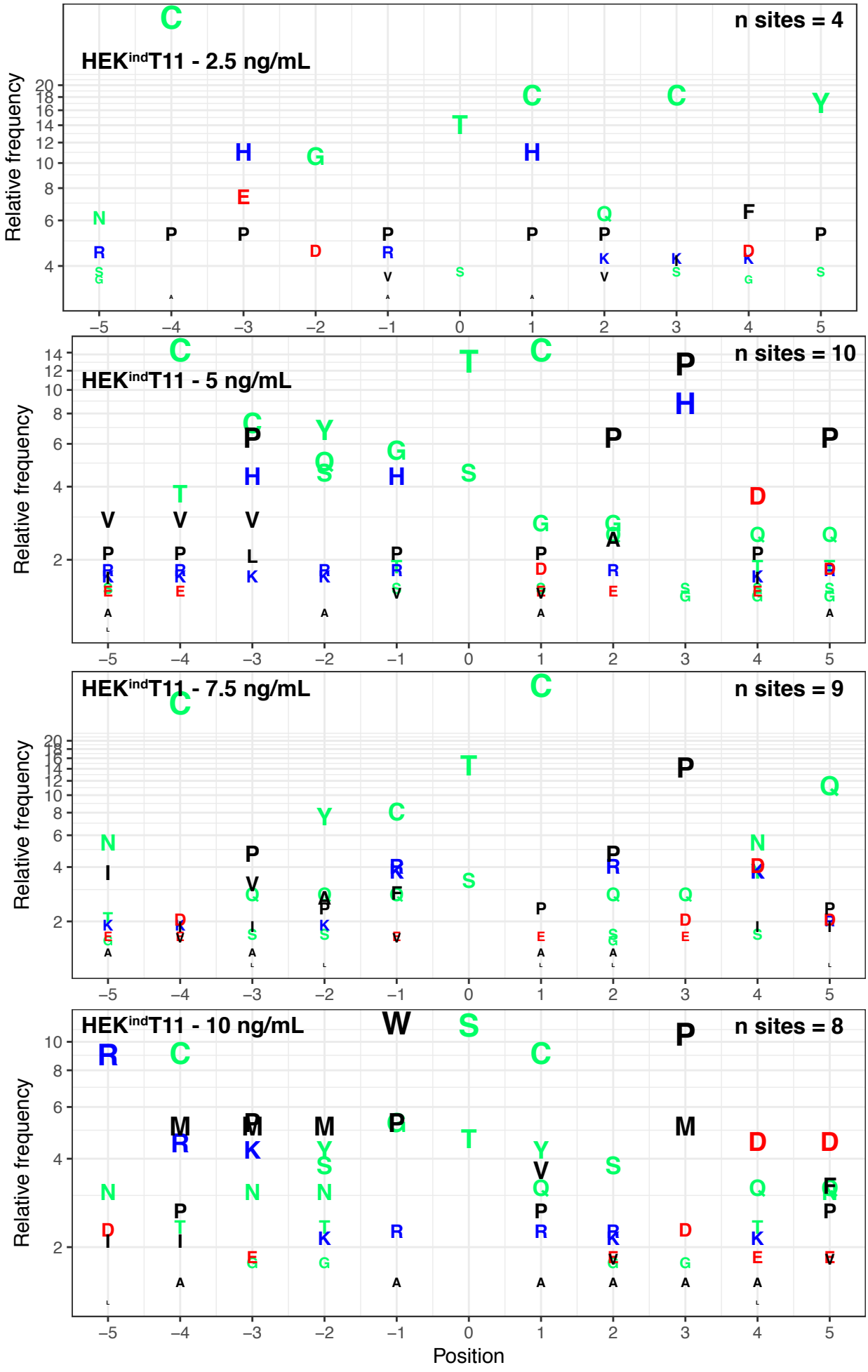


Figure S15. Alignment of induced O-glycopeptides from HEK^{ind}T11. O-glycosites were extracted from unambiguous single-site O-glycopeptides of each group and aligned in a sequence window spanning +/- 5 amino acids from the O-glycosite. The amino acid frequency is relative to the frequency of amino acids in the UniProt human proteome database (December 2013). The LA linker motif XXC^{6}XXX**T**C^{1}-XX is highly enriched at early induction levels, 2.5 to 7.5 ng/mL of doxycycline. At 10 ng/mL a motif of Pro at postion -1 and +3, similar to HEK^{ind}T2 at 16 ng/mL, is enriched. However, the size of the underlying data is limited.

Table S1.

In vitro glycosylation of "low-dose" responding T2-sites

Gene ID	UniProt ID	Identified tryptic glycopeptide	Glycan	Synthetic peptide sequence	GalNAc-transferase								
					T1	T2	T3	T4	T5	T11	T12	T14	T16
CPD	O75976	SFPDQFSTGEPPALDEVPEVR	T8(Hex(1)HexNAc(1))	GRDLNR**SFPDQFSTGEPPAL**	++	+++	++	-	-	++	+	-	+
GALNT10	Q86SR1	ERQPDGTPGGSGAAVAPAAGQGSHSR	2xT	Y**RERQPDGTPGGSGAAVAPA**	+	+++	++	NA	-	-	-	NA	-
B4GALT3	O60512	DQGPTFDYSHPR	1xT	SALFGR**DQGPTFDYSHPR**DV	-	+	-	NA	-	-	-	-	-
CANX	P27824	APVPTGEVYFADSFDR	1xT	PSSPKVTYTYK**APVPTGEVYFA**	-	++++	1+/2++	-	NA	-	-	-	(+)
HSPA5	P11021	LYGSAGPPPTGEEDTAEKDEL	S4(Hex(1)HexNAc(1)) and or T10(HexNAc)	SK**LYGSAGPPPTGEEDTAEK**	-	+	-	-*	-	-	-	-	-
SLC35B2	Q8TB61	SYGATATSPGER	1xT	VMTR**SYGATATSPGER**FTDS	(+)	++	1++/2(+)	NA	NA	-	-	-	NA
ITGA5	P08648	EAPSRSSASSGPQILK	S7(Hex(1)HexNAc(1))	QQKR**EAPSRSSASSGPQILK**	+	1++++/2++	+	NA	-	-	-	-	-

Table S2.

In vitro glycosylation of "high-dose" responding T2-sites

Gene ID	UniProt ID	Identified tryptic glycopeptide	Glycan	Synthetic peptide sequence	GalNAc-transferase								
					T1	T2	T3	T4	T5	T11	T12	T14	T16
TFRC	P02786	LAGTESPVREEPGEDFPAAR	1xTn_1xT	LAGTESPVREEPGEDFPAAR	++	-	-	-	-	-	-	-	-
GOLIM4	O00461	EKPTREVQEVSR	T4(Hex(1)HexNAc(1));S11(Hex(1)HexNAc(1))	TAR**EKPTREVQEVS**RNNDVW	-	-	-	-	-	-	-	-	-
SDF4	Q9BRK5	YSEFFTGSK	S2(Hex(1)HexNAc(1))	PEEVLK**YSEFFTGSK**LVDYA	-	-	-	-	-	-	-	-	-
F2RL1	P55085	VDGTSHVTGK	T4(Hex(1)HexNAc(1));T8(Hex(1)HexNAc(1))	GRSLIGK**VDGTSHVTGK**GVT	-	-	1+/2+	-	-	NA	NA	NA	NA
NUCB2	P80303	KLQQGIPPSGPAGELK	1xT	**KLQQGIPPSGPAGELK**FEPH	-	-	-	-	-	-	-	-	-
GLCE	O94923	AAASESNNYMNHVAK	2xTn_1xT	GFEKR**AAASESNNYMNHVAK**	-	-	-	-	-	-	-	-	-
SEMA4D	Q92854	VVPKPVVAPTLSVVQTEGSR	2xT	KP**VVAPTLSVVQTEGSR**IAT	-	+	-	-	-	-	-	-	-
SDC4	P31431	AGSGSQVPTEPK	1xT	PER**AGSGSQVPTEPK**KLEEN	-	-	-	-	-	-	-	-	-
SDC2	P34741	IPAQTK	1xT	LNIQNK**IPAQTK**SPEETDKE	+	-	++	+ (ON)	-	-	-	-	-
ERGIC2	Q96RQ1	STSTALPPREDDSSQSPNACR	T4(Hex(1)HexNAc(1))	SAFK**STSTALPPREDDSSQS**	1++++/2+	1+/2++++	1++++/2+	NA	+	+	1++/2+	+	HOLE

Table S2

Table S2.
Primers used for junction PCR

Code in Fig. S2	Name	Sequence 5' - 3'
A	AAVS1 F	CCTTACCTCTCTAGTCTGTGCTAG
B	AAVS1 R	CGTAAGCAAACCTTAGAGGTTCTGG
C	Tet3G R	ACTTGGCGTTGTTCCGCAGAAAG
D	Tet3G F	GATATGCTGCCTGCTGACGCTC
E T2	GALNT2 R	TTTCCCTGGAGGGAGGGTCTCC
F T2	GALNT2 F	ATGACAGCAGACAGAAATGGGAAC
E T11	GALNT11 R	CTCGTTGAAGATCATGCCCAGCTC
F T11	GALNT11 F	CAGAAAGGCTCCGTGGCTATGG
G	SSODN F	AGGCCGGCGGATAACTAGCTGATCGCGG

Table S3

Table S3
Touch down PCR program

Temp.	Time	Cycles
95ºC	15 min	
95ºC	30 s	
72ºC*	30 s	X15
72ºC	2 min 45 s	
95ºC	30 s	
58ºC	30 s	X25
72ºC	2 min 45 s	
72ºC	10 min	
12ºC	forever	

*Touch down 72ºC-1ºC per cycle start from cycle one

Supplementary references

1. Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Ponten, F. (2015) Proteomics. Tissue-based map of the human proteome. Science **347**, 1260419

2. Khetarpal, S. A., Schjoldager, K. T., Christoffersen, C., Raghavan, A., Edmondson, A. C., Reutter, H. M., Ahmed, B., Ouazzani, R., Peloso, G. M., Vitali, C., Zhao, W., Somasundara, A. V., Millar, J. S., Park, Y., Fernando, G., Livanov, V., Choi, S., Noe, E., Patel, P., Ho, S. P., Myocardial Infarction Exome Sequencing, S., Kirchgessner, T. G., Wandall, H. H., Hansen, L., Bennett, E. P., Vakhrushev, S. Y., Saleheen, D., Kathiresan, S., Brown, C. D., Abou Jamra, R., LeGuern, E., Clausen, H., and Rader, D. J. (2016) Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans, Nonhuman Primates, and Rodents. *Cell Metab* **24**, 234-245

3. King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., Woetmann, A., Vakhrushev, S. Y., and Wandall, H. H. (2017) Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Advances* **1**, 429-442