

# O-GlcNAc transferase recognizes protein substrates using an asparagine ladder in the TPR superhelix

Zebulon G. Levine, Chenguang Fan, Michael S. Melicher, Marina Orman, Tania Benjamin, Suzanne Walker\*

*Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, United States*

## Supplementary Methods

### *Reagents*

The chemicals used in synthesizing UDP-GlcNAz were either purchased from Sigma Aldrich or Alfa Aesar. UDP-GlcNAz was synthesized according to reported procedures.<sup>1</sup> 2,9-dimethyl-5,6,11,12-tetrahydro-dibenzocyclooctene was purchased from Sigma Aldrich. UDP-GlcNAc and antibody CTD110.6 were purchased from Sigma Aldrich. Immunoblotting reagents were purchased from Life Technologies. Human ProtoArray protein microarrays were purchased from Life Technologies. Microarray four-well incubation plates were purchased from Greiner Bio-One. Lifterslip slides were purchased from Thermo Scientific. Microarray blocking buffer was purchased from ArrayIt. Anti-mouse IgG, Alexa-fluor 647 conjugate and streptavidin-Alexa-fluor 647 were purchased from Life Technologies. Azadibenzocyclooctyne-biotin (ADIBO-biotin) and Azadibenzocyclooctyne-5kDa-PEG (ADIBO-PEG) were purchased from Alfa Aesar. Human cDNA clones were obtained from the National Institutes of Health (NIH) mammalian gene collection (MGC). Recombinant estrogen receptor  $\alpha$  was purchased from Thermo Fisher Scientific. Rabbit reticulocyte lysate *in vitro* transcription/translation system (TNT SP6 or T7) was purchased from Promega. <sup>35</sup>S-Methionine was purchased from Perkin Elmer. All materials for imaging radioactive gels were purchased from GE. UDP-Glo assay reagents, including detection reagents, nucleotides, and nucleotide sugars were purchased from Promega. Bio-Spin P6 size exclusion columns were purchased from BioRad.

### *Generation and use of mutant OGT and purification of recombinant OGT*

All OGT in this research refers to the 1036 amino acid 'ncOGT' isoform. OGT mutant 5N5A expression vector was constructed by PCR-amplifying the ncOGT containing pET24b expression vector<sup>2</sup> using primers 5'-CGTTTGC GACTGGACCGAC-3' and 5'-CGGGCACAGACGCAGAGC-3'. The PCR product was gel purified and the dsDNA fragment shown in table S5 (purchased from Integrated DNA Technologies as a gBlock) was inserted using isothermal assembly.<sup>3</sup> The construct identity was confirmed using Sanger sequencing of the insert before transformation into BL21(DE3) for expression. Purification of both wild-type and mutant ncOGT was carried out as previously described.<sup>2</sup>

Before use, OGT aliquots were thawed on ice, and centrifuged at 15000 rpm at 4 °C for 20 minutes to pellet any aggregated protein. The supernatant was removed and stored on ice until use, with the concentration measured based upon A280 with an extinction coefficient of 117580 M<sup>-1</sup> cm<sup>-1</sup>.

### *Peptides for reactions*

All peptides (sequences shown in Table S5) were purchased from Biomatik, and dissolved as concentrated stocks in ultrapure water. pH of stocks was adjusted to ~7 with dilute sodium hydroxide. Concentration of peptides was determined by UV spectroscopy based upon measuring the A<sub>274</sub> ( $\lambda_{\text{max, tyrosine}}$ ) in water using an extinction coefficient of 1400 M<sup>-1</sup> cm<sup>-1</sup>.

### *General data analysis*

All data was analyzed using the R statistical computing package<sup>4</sup> in the RStudio development environment (2016).<sup>5</sup> Plots were produced using the package ggplot2.<sup>6</sup> Array analysis was performed using the package limma.<sup>7</sup> Robust linear regression for staining correction was performed using the MASS package.<sup>8</sup> General data analysis was aided by the data manipulation packages dplyr<sup>9</sup> and reshape2.<sup>10</sup> The majority of plots were made on the basis of accession number, so that any isoforms that differed on the array were handled separately. All analysis scripts are available on GitHub at <https://github.com/SuzanneWalkerLab/GlcNAzMicroarrayTPR>.

### *Array Glycosylation and O-GlcNAc Detection via Monoclonal Antibody*

Human ProtoArray microarrays were warmed from -20 °C to 4 °C for 20 min and washed twice (5 mL × 5 min/wash) with tris-buffered saline, pH 7.4, (TBS) containing 0.04% Tween 20 (TBST) in a four-well array incubation plate on a circular shaker at 50 rpm. Microarrays were blocked for 1.5 hours at 4 °C with Microarray Blocking buffer (ArrayIt) and washed again with TBST buffer (5 mL × 5 min/wash) three times. Microarrays were incubated with 200 μL 3 μM OGT and 40 μM UDP-GlcNAc in PBS buffer (pH 7.4) under a coverslip for 2 hours at room temperature. OGT was omitted in the control arrays in the incubation step. Following that, the arrays were washed four times with TBST (5 mL × 5 min/wash) plus a middle wash with 0.5% SDS in TBS. Microarrays were incubated with primary antibody (anti-O-GlcNAc, CTD110.6; 1:250 dilution in TBST) overnight in a moistened ziplock bag at 4 °C. To label modified proteins, an anti-mouse Alexa Fluor 647 conjugated secondary antibody (1:250 dilution in TBST) was incubated for 1 hour at room temperature. The arrays were washed again as above and dipped in water three times, spin-dried (200 ×g, 5 min), and scanned with a GenePix 4000B scanner.

### *Array Glycosylation with UDP-GlcNAz and detection via biotin-streptavidin*

Human ProtoArray microarrays were first washed, blocked, and re-washed similarly to the antibody detection method. The microarrays were then incubated with 200 μL 3 μM OGT and 290 μM UDP-GlcNAz in PBS (pH 7.4) buffer under a coverslip for 2 hours at room temperature. Control arrays lacked OGT, while 5NSA arrays had the mutant enzyme at 3 μM instead of wild-type OGT. Following the two hour incubation, the arrays were washed four times with TBST (5 mL × 5 min/wash) and incubated with 200 μL 270 μM iodoacetamide and 1 μM 2,9-dimethyl-5,6,11,12-tetrahydro-

dibenzocyclooctene in PBS buffer under a coverslip for 30 min. After the incubation, the arrays were washed again and incubated with 200  $\mu$ L 1  $\mu$ M ADIBO-biotin in PBS buffer under a coverslip for 30 min. The arrays were washed with TBS containing 0.5% SDS three times (5 mL  $\times$  5 min/wash) and twice with TBST (5 mL  $\times$  5 min/wash). The arrays were incubated with 200  $\mu$ L 0.5  $\mu$ g/mL streptavidin-Alexa Fluor 647 for 30 min at 4 °C. Finally, the arrays were washed again as above and dipped in water three times, spin-dried (200  $\times$ g, 5 min), and scanned with a GenePix 4000B scanner.

#### *Analysis of array images*

Images of arrays were manually aligned to grid for analysis in GenePix 5.0 software, with the batch-appropriate .gal control files downloaded from the Invitrogen website for each array. Data was saved as a .gpr file, and a .tiff of each channel was saved. Raw .gpr files as well as final normalized data is available on the NIH's gene expression omnibus (GEO) at the accession code GSE107911. Array data was loaded into R using the limma package<sup>7</sup> as median spot values. Background correction was performed by fitting a normal-exponential mixed model as implemented in the "normexp" method of limma command `backgroundCorrect()`.<sup>11</sup>

#### *Data Mining for Known O-GlcNAc Proteins on Microarray*

The O-GlcNAc dataset was downloaded from the PhosphoSitePlus website (accessed June 2017).<sup>12</sup> BioMart<sup>13</sup> was used to convert gene names to human orthologues, and any genes that failed automatic assignment had orthologues manually assigned based on reference to NCBI RefSeq<sup>14</sup> or UniProt.<sup>15</sup> This dataset of human genes, along with the list of genes corresponding to O-GlcNAc modified proteins discovered by Wang et al.<sup>16</sup> and the list of protein contents of the Human ProtoArray (v. 5.0, accessed April 2017) were converted to Human Genome Organization Gene Nomenclature Committee (HGNC) symbols<sup>17</sup> using the limma command `alias2Symbol()`. Any genes not converted by this command or for which there were ambiguities were manually annotated based upon consultation of RefSeq, Uniprot, or BLAST search of the sequence listed for the microarray. The list of HGNC symbols for the protein microarray spots were compared to those found from PhosphoSitePlus<sup>12</sup> or Wang et al.,<sup>16</sup> and these proteins were annotated as known O-GlcNAc Proteins. The list of gene names provided by Invitrogen, Invitrogen descriptions, HGNC symbols, and whether it is present in PhosphoSitePlus,<sup>12</sup> Wang et al.,<sup>16</sup> or both is reported in Table S4.

#### *Analysis for comparison of detection methods*

For each of the detection methods (CTD110.6 monoclonal antibody and GlcNAz-biotin labeling method), two arrays were run and analyzed, one that had been treated with wild-type ncOGT and one that had been treated without enzyme as a control (see Table S6 for arrays used).

For the images of arrays shown in Figure 1B and S1, the red channel of each array scan was converted to false color using the same color lookup table for each image using Adobe PhotoShop 2017. The false coloration provided enhanced contrast to allow viewing of the signal, but maintains a similar gradient to that observed in the original channel.

To account for batch-to-batch variation in protein content of the arrays between detection methods, any proteins not present on all arrays were excluded from analysis. For each array, the average intensity value was calculated among all spots representing a single accession number. Note that all control spots

were excluded from this analysis, as were spots corresponding to accession numbers that lacked a corresponding HGNC symbol, had a corresponding HGNC symbol that encoded a long intergenic noncoding RNA, or had a HGNC symbol encoding an antisense RNA. This yielded 7926 unique accession numbers representing 6747 unique HGNC symbols.

The intensity values for OGT-treated arrays were divided by the corresponding control array incubated without OGT to give an intensity ratio. The base-2 logarithmic transform of these ratios was taken, and the corresponding z-score of this logarithm-transformed data was generated by subtracting the mean of all proteins and dividing this subtracted data by the standard deviation (Figure S3B). Based upon a comparison of these z-scores to those expected under a standard normal distribution we determined a cutoff of a z-score of 2 for considering any protein a hit (Figure S3C). The standardized data was raised to the second power to reverse the log transform for plotting (Figure 1C) and reporting in Table S1.

Note that Figure 1D, as well as numbers in the main text, analyze the results on a per-HGNC symbol basis to prevent skewing of the data based upon over-representation of proteins for which there are multiple accession numbers on the array from single genes. If an HGNC symbol is represented by multiple accession numbers on the array, it is considered a hit if any of the accession numbers are a hit. Table S1 lists all proteins on a by accession number basis with corresponding HGNC codes.

#### *In vitro validation of OGT substrates using IVT-mass tagging*

The procedure was based on the previously reported method<sup>18</sup> with modifications. The <sup>35</sup>S-Met labeled proteins were obtained by *in vitro* transcription and translation (IVT) following the manufacturer's instructions. Following the IVT reaction, 10 μM OGT and 100 μM UDP-GlcNAz or UDP-GlcNAz alone were added into the mixture and incubated at 30 °C for 1.5 hours. 500 μM iodoacetamide was added and incubated at room temperature for 30 min. Subsequently, azadibenzocyclooctyne-5 kDa-PEG (ADIBO-PEG) was added to the mixture at a final concentration of 2 mM and the mixture was incubated at room temperature for 30 min. After the incubation, SDS loading buffer was added to each sample, which was subjected to SDS-PAGE. The gel was then dried and exposed to a Phosphor-storage screen for 24 hours and imaged with a Typhoon scanner. The gels are shown in Figure S4.

#### *Comparison of circular dichroism of wild-type and mutant OGT*

Recombinant OGT was buffer exchanged in CD buffer (10 mM sodium phosphate, 1 mM DTT, pH 7.5) using Bio-Spin P6 columns according to the manufacturer's instructions. Protein was diluted to 1 μM for both proteins and a circular dichroism spectrum was measured on a Jasco J-815 spectropolarimeter using a 1 mm path length quartz cuvette (Hellma Analytics, type 110-QS). Buffer-corrected spectra were recorded from 260 nm to 190 nm as the average of 5 accumulations at 20 °C with a 1 nm bandwidth at a scanning speed of 50 nm/min and a data integration time of 1 second. Data were converted to mean residue ellipticity  $[\theta]$  according to the formula:<sup>19</sup>

$$[\theta] = \frac{\theta \text{ (mdeg)}}{10 * C_{\text{protein}} \left(\frac{\text{mol}}{\text{L}}\right) * \text{Length} \text{ (\# amino acids)} * l \text{ (cm)}}$$

using a protein length of 1055 for tagged, recombinant OGT.

### *In vitro activity assay of OGT against CK2 and HCF-E10S peptide substrates*

Reactions between OGT and peptide substrates CK2 or HCF-E10S were carried out at 37 °C at a variety of peptide concentrations (as shown in Figure 2B–C), with 100 nM OGT in a 20 µL final reaction volume in a buffer containing 1 mM trishydroxypropylphosphine, 20 mM MgCl<sub>2</sub>, 20 mM Tris, and 150 mM NaCl at pH 7.4. Reactions were initiated by addition of OGT and quenched with UDP-Glo detection reagent<sup>20</sup> after 6 minute reactions, a time point at which the reaction was still under initial rate (Figure S5B). After quenching, plates were incubated for 1 hour at room temperature before reading luminescence using a Promega GloMax Explorer plate reader. All points were measured in triplicate, and 5N5A and wild-type OGT were measured simultaneously. Data was fit to a Michaelis-Menten equation to provide the curves shown in Figure 2B and 2C, with parameters as shown in the figure.

For UDP-Glo testing of UDP-GlcNAc and UDP-GlcNAz, reactions were carried out as above, but with 100 µM final concentration of CK2 peptide and sugar donor concentrations reported in Figure S1. Time course data reported against CK2 peptide in Figure S5B was carried out at 200 µM CK2 peptide acceptor concentration as described above, but was quenched at the indicated time rather than 6 minutes, and only single replicates were measured for each time point.

### *In vitro activity assay of OGT against Estrogen Receptor $\alpha$*

Reactions between OGT and estrogen receptor  $\alpha$  (ER) were carried out at 37 °C at a final concentration of 1 µM ER, 75 nM OGT, in a buffer containing 1 mM trishydroxypropylphosphine, 20 mM MgCl<sub>2</sub>, 20 mM Tris, 150 mM NaCl at pH 7.4 in 25 µL final reaction volume. Reactions were initiated by addition of OGT and quenched with UDP-Glo detection reagent every two minutes from 2 to 12 minutes. After quenching, plates were incubated for 1 hour at room temperature before reading luminescence. All points were measured in triplicate, and 5N5A and wild-type OGT were measured simultaneously. A linear fit was used to determine the rate of glycosylation, with the y intercept fixed to be the same for both enzyme variants. The slope for the fit of the 5N5A data was divided by the slope of the fit for wild-type data to obtain the factor for adjusting microarrays treated with 5N5A relative to wild-type OGT. The results of this experiment are shown in Figure S8A.

### *Normalization and comparison of array replicates*

The data analysis process used to compare array replicates is outlined in Figure S6. Arrays were read into R, background-corrected, and together normalized for staining. OGT treated arrays were normalized for activity with 5N5A versus wild-type OGT differences in activity being corrected for based upon *in vitro* kinetics on estrogen receptor  $\alpha$  (ER). For comparisons to control, the OGT-treated arrays were median-normalized for the log<sub>2</sub>-transformed data. A permutation test<sup>21</sup> was carried out to determine statistical significance, and a Storey's q value procedure<sup>22</sup> was used to determine false discovery rate (FDR). Hits for any given comparison were proteins with a less than 5% FDR and a greater than 2-fold change in median values between groups. The details of each step are described below.

### *Normalization of arrays for differences in streptavidin staining*

For each block on microarrays of interest, robust linear regression was used to perform a linear fit between reported concentrations of biotinylated control proteins and the background-corrected fluorescent signal using iteratively re-weighted least squares as implemented in the `rlm()` function in the R MASS package.<sup>8</sup> This robust variation of linear regression was used in order to prevent skewing by outliers. All background-corrected spot fluorescence intensity values on a given block were then divided by the slope obtained from this fit to correct for staining differences between blocks. If the fit either failed to converge for a given block during the iterative re-weighting procedure, indicating poor correlation, or it showed a negative relationship between control concentration and fluorescent signal, then the fit for that block was rejected. To still scale the values on these blocks appropriately, signal values were instead divided by the average of slopes obtained for all other blocks on a given array, thus treating these blocks as the average of all other blocks in terms of degree of staining. This process is repeated for every array analyzed, thus normalizing the slope of all biotinylated control series spots to 1 on every block (Figure S7B–D).

### *Normalization of arrays for differences in activity*

Overall, the activity-normalization procedure scales the difference between a given protein's signal on an OGT-treated microarray and the corresponding signal on untreated control microarrays based upon the degree of OGT activity seen on spots of the control OGT substrate estrogen receptor  $\alpha$  (ER) on the same microarray block as any given protein. The activity scaling takes into account both spatial and array-to-array differences in OGT activity based upon ER signal and scales the different OGT mutant activities based upon the difference in biochemical activity against recombinant ER measured as described above. In order to do so, the following procedure was carried out:

First, the median staining-corrected signal of all control ER spots across all microarrays not treated with OGT, termed  $\bar{C}_{ER}$ , was found. The median staining-corrected signal level  $\bar{C}_j$  for every non-control protein  $j$  on the untreated arrays was also calculated. The median value of all ER spots across all of the microarrays treated with wild-type OGT,  $\bar{W}_{ER}$ , was calculated, and the difference in these two values,  $\bar{S}_{ER}$ , was used as a scaling factor for activity:

$$\bar{S}_{ER} = \bar{W}_{ER} - \bar{C}_{ER}$$

In addition, the following correction  $V_{ER,m}$  based on which OGT form,  $m$ , was used to treat an a given microarray was defined as follow:

$$V_{ER,m} = \begin{cases} 1 & \text{where } m = \text{wild-type} \\ 0.434 & \text{where } m = 5N5A \end{cases}$$

The value of  $V_{ER,m}$  is set based upon the difference in activity measured in biochemical tests of recombinant ER.

For each block  $i$  on a microarray treated with an OGT variant, the median value  $\bar{E}_{ER,i}$  was calculated based upon staining-corrected signal for all six individual ER spots in block  $i$ . Block median  $\bar{E}_{ER,i}$  was

then compared to  $\bar{C}_{ER}$ , and only those blocks  $i$  where  $\bar{E}_{ER,i} > \bar{C}_{ER}$  were corrected for activity. For blocks where  $\bar{E}_{ER,i} > \bar{C}_{ER}$ , the difference

$$D_{ER,i} = \bar{E}_{ER,i} - \bar{C}_{ER}$$

is calculated, as are the differences

$$D_{j,i} = \bar{E}_{j,i} - \bar{C}_j$$

for all proteins  $j$  on block  $i$  for which  $\bar{E}_{j,i} > \bar{C}_j$ , where  $\bar{E}_{j,i}$  is the average of staining-corrected signal for spots of protein  $j$  in block  $i$ .

For all calculated  $D_{j,i}$  values, the following correction is carried out:

$$D_{j,i}^{\text{corrected}} = \frac{D_{j,i}}{D_{ER,i}} * \bar{S}_{ER} * V_{ER,m}$$

whereby the difference is scaled based upon the block-specific activity against ER to a constant value ( $\bar{S}_{ER}$ ) and then scaled by whether the microarray was treated with wild-type or mutant enzyme.

The staining-corrected value of every individual spot on blocks where  $\bar{E}_{ER,i} > \bar{C}_{ER}$  were adjust as follows:

$$\text{Spot Signal}_{j,i}^{\text{activity-corrected}} = \text{Spot Signal}_{j,i}^{\text{staining-corrected}} - D_{j,i} + D_{j,i}^{\text{corrected}}$$

where  $\text{Spot Signal}_{j,i}^{\text{staining-corrected}}$  and  $\text{Spot Signal}_{j,i}^{\text{activity-corrected}}$  represent the signal at each individual spot for protein  $j$  on block  $i$  before and after activity correction, respectively.

After activity correction, all data is converted from being on a per-spot basis to a per-accession number basis, meaning that for every array, the average activity-corrected signal was taken for each individual non-control protein, leading to one value per array per accession number being used in all further analysis. As with method comparison, all accession numbers corresponding to antisense RNAs, LINC RNAs or accession numbers that lack an associated HGNC symbol were removed from further analysis, yielding 8968 unique proteins corresponding to 7427 HGNC symbols. Note that this list is longer than the list involved in comparison between detection methods due to greater overlap in array contents between batches of arrays used for these studies versus studies of detection methods.

### *Median normalization of different array treatments*

To compare between untreated control microarrays and OGT-treated arrays to determine degree of specific glycosylation, the OGT-treated arrays needed to be corrected for nonspecific glycosylation. To do so, all arrays of a given type (ie, not treated with OGT, treated with wild-type OGT, or treated with 5N5A mutant) were separately grouped, and the logarithm base 2 of all averaged signal values was taken. The median of these log2-transformed signals was taken for each group, and each group's median log2 signal was subtracted from all individual signals, followed by adding the median for the control group to all values regardless of treatment. These median-corrected log2 values were then converted back to non-

logarithmic values for further analysis. Note that these median-normalized values were only used for comparisons between OGT-treated and untreated arrays- any direct comparisons between wild-type and 5N5A OGT used data that had not been median normalized as the degree of background glycosylation was expected to be corrected for similarly based upon the ER-based activity correction above.

*Statistical comparison of different array treatments for multiple replicates*

Comparisons between different groups of arrays (not treated with OGT, treated with wild-type OGT, or treated with 5N5A mutant) were carried out as described by Merbl and Kirschner, with some modifications.<sup>21</sup> Briefly, the base 2 logarithm of all data was taken for statistical comparison. For control-wild-type and control-5N5A comparisons, moderated t-statistics<sup>23</sup> were generated for each protein based on accession number; for tests between control arrays and either 5N5A or wild-type-treated arrays, this was carried out on activity-corrected and median normalized data, while for wild-type OGT-5N5A comparison this was carried out on activity-corrected data that had not been median normalized. Those proteins whose moderated t values suggested no difference between the tested groups were used to generate a null distribution of t-statistics as described by Yang and Churchill.<sup>24</sup> For the comparisons between control and OGT-treated arrays, a one-tailed t-test for OGT-treated arrays having greater signal than controls selected which accession numbers were used to generate the null distribution, while for the direct comparison between wild-type and 5N5A OGT, a two-tailed t-test was used. The null distributions generated this way were used in to generate a p-value for all proteins, and the distribution of p-values was used to determine false-discovery rate and generate a q-value as described by Storey.<sup>22</sup>

Determination of which proteins were “hits” for differences in glycosylation between two different groups were carried out using both false-discovery rate and effect-size cutoffs. For false-discovery based cutoffs, proteins (on a per-accession number basis) were considered significantly different between two conditions (ie “hits”) if they had a q-value less than 0.05, representing a 5% false discovery rate. In addition, they needed to meet an effect size cutoff. For the effect-size cutoffs, median values for each protein within a group of arrays (no OGT control, wild-type OGT, or 5N5A mutant-treated) were calculated, giving a median for each protein in each group. For comparisons between OGT-treated groups and control this was done using median-corrected data, while activity-corrected data without median correction was used for direct comparison between wild-type OGT and 5N5A (see Figure S9).

The ratios between groups ( $\frac{\text{wild-type treated arrays (median corrected)}}{\text{control arrays}}$ ,  $\frac{\text{5N5A treated arrays (median corrected)}}{\text{control array}}$ , or  $\frac{\text{wild-type treated array (activity corrected)}}{\text{5N5A array (activity corrected)}}$ ) were taken based upon the median values of all similarly treated arrays for each protein. Proteins needed to have a minimum two-fold difference in activity based upon this ratio to be considered a hit. For the direct comparison between wild-type and 5N5A OGT, only those proteins that had scored as hits for glycosylation by wild-type or 5N5A OGT were considered as possible hits, as we considered differences in degree of nonspecific glycosylation by the two variants as unlikely to be enlightening regarding OGT’s biochemical behavior, and no proteins had scored as hits for the 5N5A mutant (Figure S10C–D).



### Preparation of HeLa cell extract

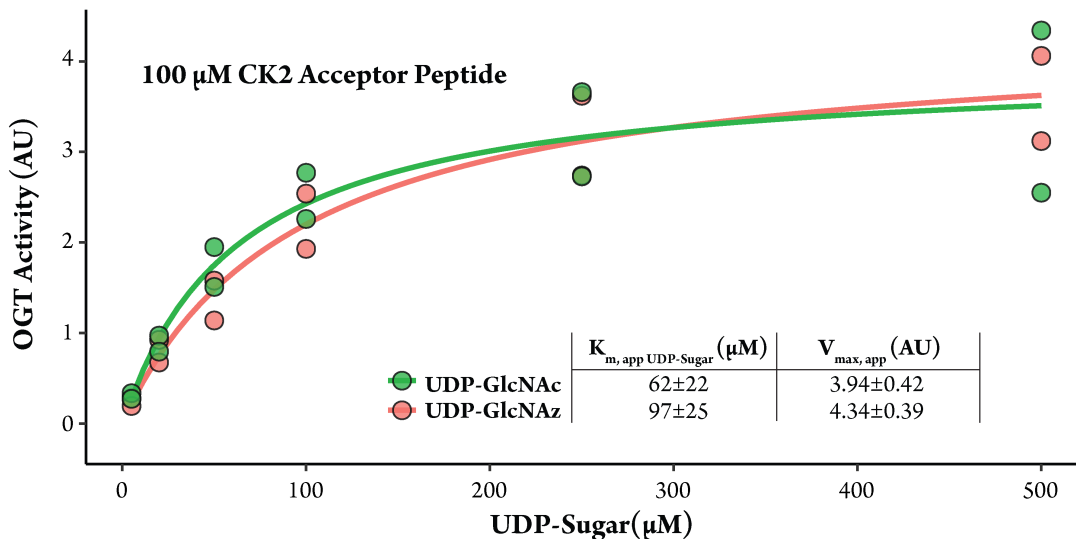
HeLa cell extract was prepared as previously described<sup>21,25</sup> with some modifications. HeLa S3 cells were not synchronized, but instead grown to  $2 \times 10^6$  cells/mL in a 1 L spinner flask then pelleted and flash frozen for later lysis. The cell pellet was partially thawed on ice at which point it was resuspended in swelling buffer (pH 7.5, 40 mM HEPES, 5 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1 mM DTT, 1 EDTA-free protease inhibitor cocktail tablet (Roche) per 25 mL) of approximately 80% the volume of the pelleted cell mass. Energy mix (7.5 mM creatine phosphate, 1 mM ATP, 0.1 mM EGTA, pH 7.7, 1 mM MgCl<sub>2</sub>) was added, and the resuspended cell mass was stirred at 4 °C under 1000 psi pressure nitrogen gas for 30 minutes, after which the pressure was release and the cellular lysate collected, followed by a second round of pressurization and pressure release to ensure complete lysis. Precipitates were pelleted at 15000 rcf for 30 minutes and cell extracts were aliquoted and flash frozen.

### Glycosylation of HeLa cell extract by OGT

HeLa cell extract was thawed on ice and any remaining precipitate was removed by centrifuging at 4 °C for 20 minutes at 45000 rcf. Extract was aliquoted and UDP-GlcNAc was added to a final concentration of 1 mM followed by the addition of recombinant OGT of the listed isoform to 2.5 μM final concentration. An equal amount of buffer (50 mM Tris, 150 mM NaCl, pH 7.4) was added to the no-added-OGT control to correct for extract dilution due to OGT addition. These aliquots were incubated at 37 °C for the listed time after OGT addition, followed by quenching via dilution in 2x Laemmli buffer and incubation at 95 °C for 5 minutes.

Reactions were analyzed by immunoblotting after SDS-PAGE and transfer to nitrocellulose. Results are shown in Figures 3B and S11.

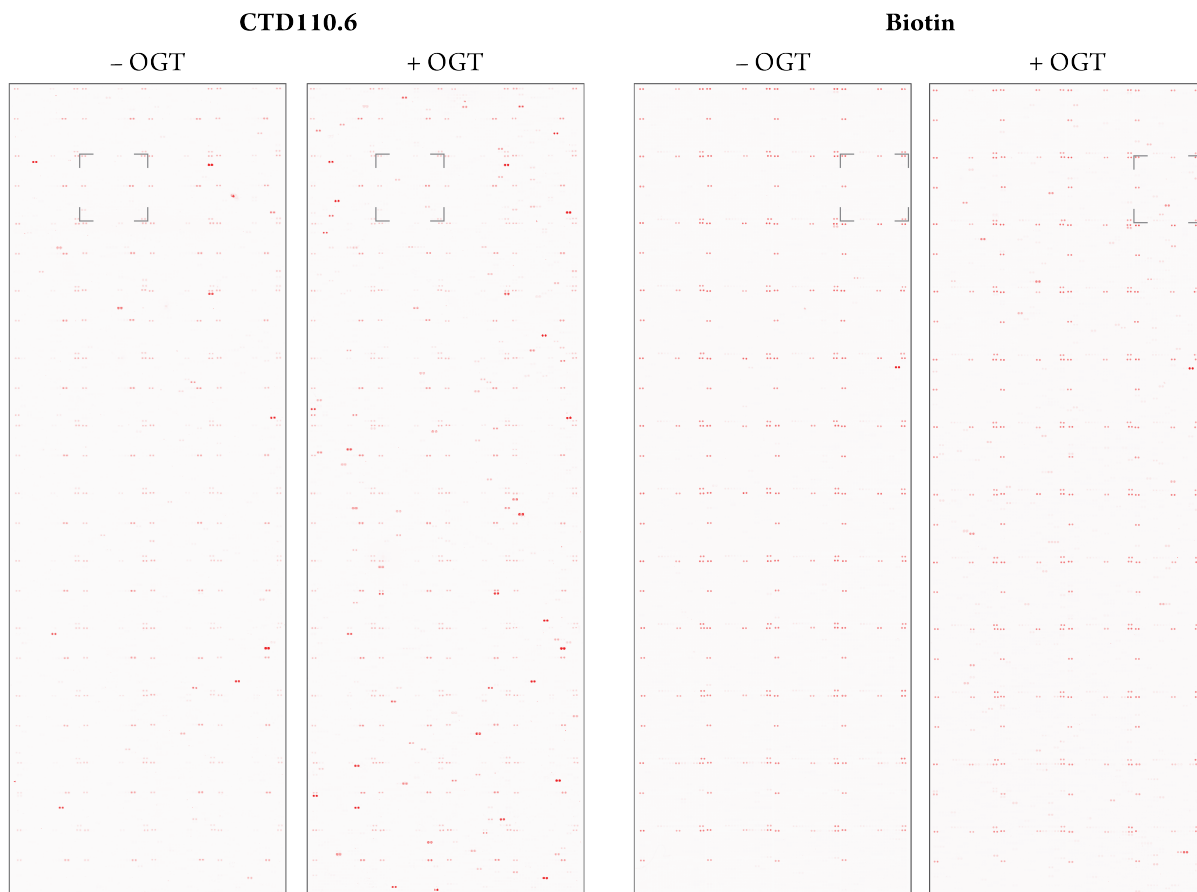
### Supplementary Figures and Tables



**Figure S1: OGT effectively uses UDP-GlcNAz *in vitro* with similar parameters to UDP-GlcNAc**

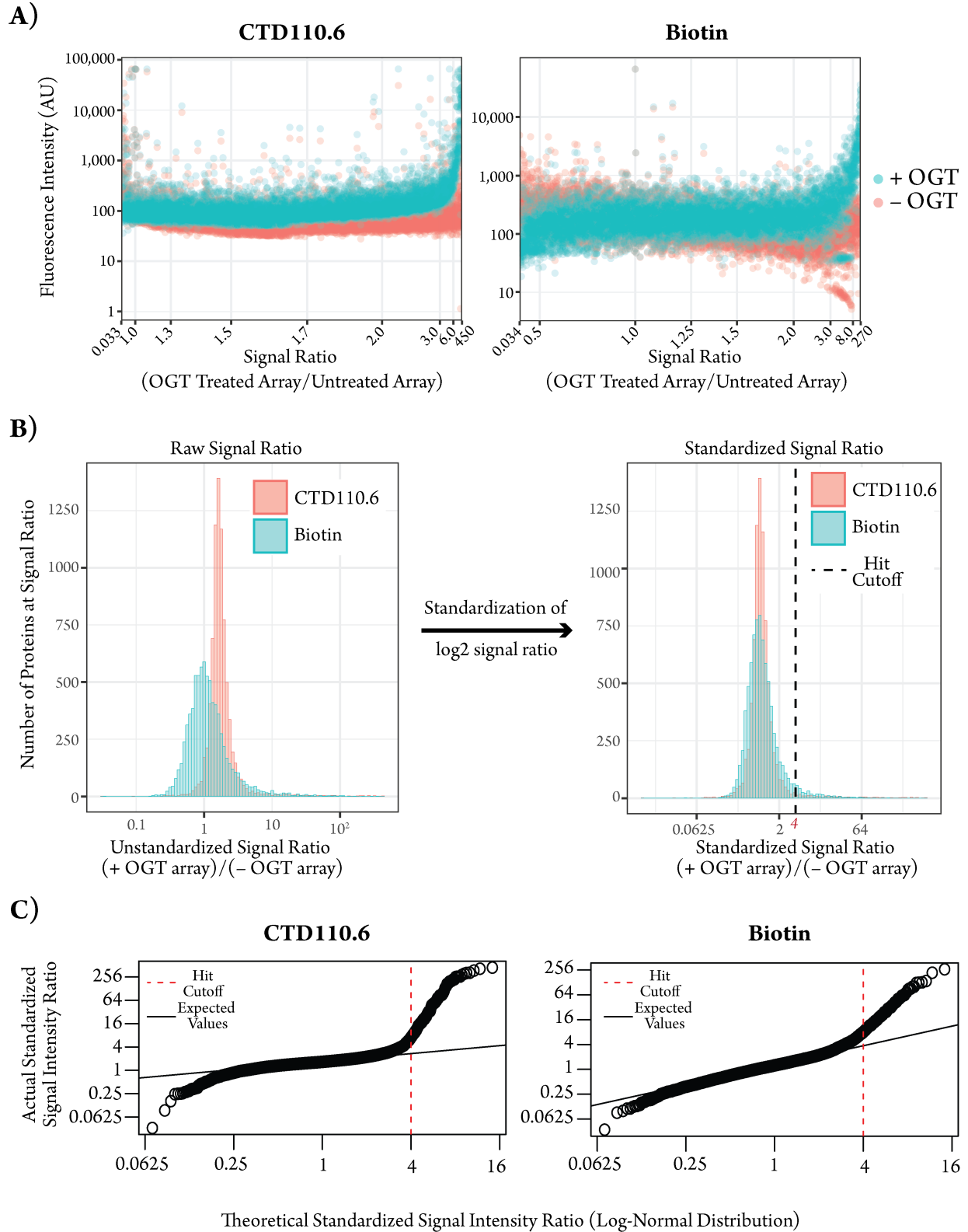
OGT was used to glycosylate CK2 acceptor peptide for 6 minutes using a variety of UDP-GlcNAc or UDP-GlcNAz concentrations. Activity was measured based upon UDP production using the UDP-Glo

detection assay<sup>®</sup> (Promega) and is reported in arbitrary units (AU) with n=2 data points per concentration. Reactions were carried out using 100  $\mu$ M CK2 acceptor peptide. Apparent Michaelis-Menten parameters are reported  $\pm$  standard error.



**Figure S2: Chemoenzymatic biotin labeling reduces background signal compared to antibodies (full arrays)**

Full-array images in contrast-adjusted color. Gray dashed boxes highlight regions of arrays that correspond to images used in Figure 1B. Arrays were treated with UDP-GlcNAc (or UDP-GlcNAz for the biotin-labeling strategy) with or without OGT as labeled. Arrays are paired to show both arrays analyzed for Figure 1 and Figure S2 for both the CTD110.6 antibody-based detection (left) and chemoenzymatic biotin detection methods (right).



**Figure S3: Signal intensity ratio allows comparison of detection methods and determination of OGT activity on single-replicate microarrays**

A) Averaged intensity values for each protein for both CTD110.6-based detection and biotin-based detection. The x-axis is an even spacing of every protein on the array sorted by ratio of signal between OGT treated arrays and untreated controls, while the y-axis is average signal for a given protein on the microarray indicated by the color of the point. The proteins have been sorted for the given detection method; this leads to different ordering of proteins between the graphs for the two detection methods. Selected signal ratios are shown below the x-axis to aid comparison to (B) and (C).

B) Standardization of log-transformed data allows direct comparison between staining methods. On the left, non-standardized data is plotted as a histogram of the ratio in fluorescent signal between OGT treated and untreated arrays for both detection methods, with the Y-axis indicating the number of proteins whose average signal falls in any given bin of signal intensity ratios. The signal intensity ratios for biotin are in blue, while the ratios for CTD-treated arrays are in red. The data were standardized as the log<sub>2</sub> data, leading to the same center between the two detection methods and similar positions at which the data tails off, allowing a similar cutoff to be drawn to declare proteins as hits for glycosylation. This cutoff, at a standardized signal ratio of 4, is indicated by the vertical dashed line.

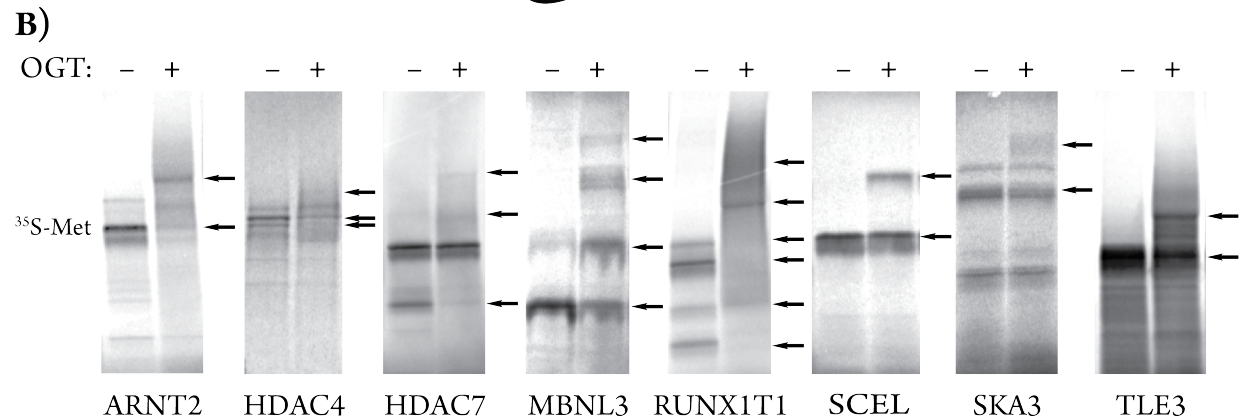
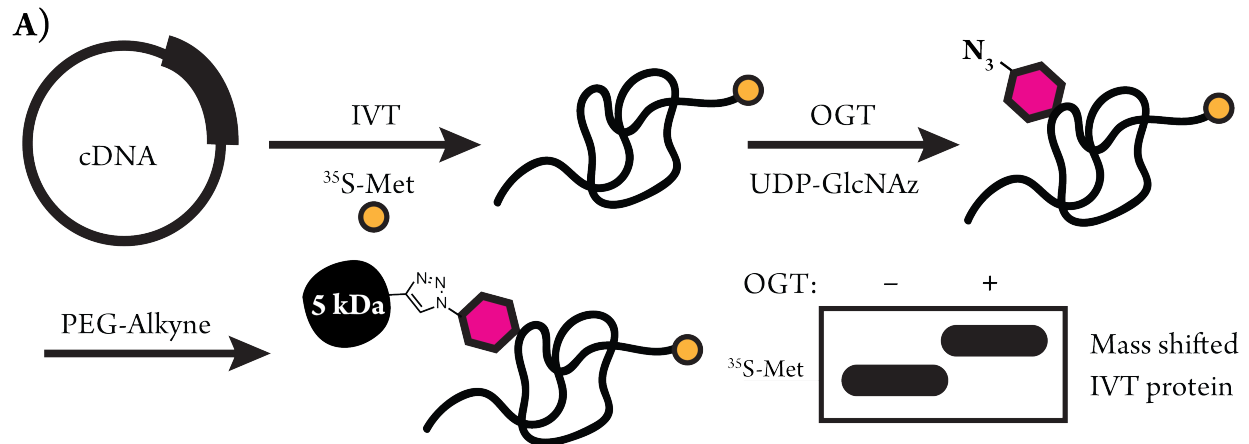
C) Comparison between a theoretical log-normal distribution and the actual standardized signal ratio data suggests a cutoff of 4 as where the signal ratios appears to cease corresponding to the expected distribution. All proteins are plotted in order, with their position on the x-axis giving the value that would be expected if the data followed a log-normal distribution, while the Y-axis position gives the actual signal ratio observed. The solid black line indicates where all data would fall if the data was log-normally distributed. There is a clear inflection in the data for both detection methods at a value of roughly 4, which is where the cutoff was set (indicated by the vertical dashed red line)

**Table S1 (Excel File, separate): Comparative studies of detection methods for the study of OGT signal detection on protein microarrays**

This table contains all proteins that were analyzed on the microarrays with accession numbers, HGNC symbols, and descriptions. Control array signal, treated array signal, and standardized signal ratio is provided for each detection method (CTD110.6 antibody and chemoenzymatic biotin tagging), as is whether the protein is known to be O-GlcNAc modified from either the PhosphoSitePlus database<sup>12</sup> or proteomic studies by Wang et al.<sup>16</sup>

<b>Protein on Array (Former)</b>	<b>Hit by CTD110.6 Detection in this study</b>	<b>Hit by Biotin Detection</b>
BAIAP2	Yes	Yes
DMTN (EPB49)	Yes	Yes
E2F8	Yes	Yes
HGS	Yes	Yes
MEF2D	Yes	Yes
MEF2A	Yes	Yes
NR3C1	No	No
NOL4L (C20orf112)	Yes	Yes
SSBP2	Yes	Yes
SSBP3	Yes	Yes

“Hit” means protein exceeded standardized signal ratio of 4 (as reported in Table S1). “Former” refers to gene symbols reported in Ortiz-Meoz et al.<sup>18</sup>

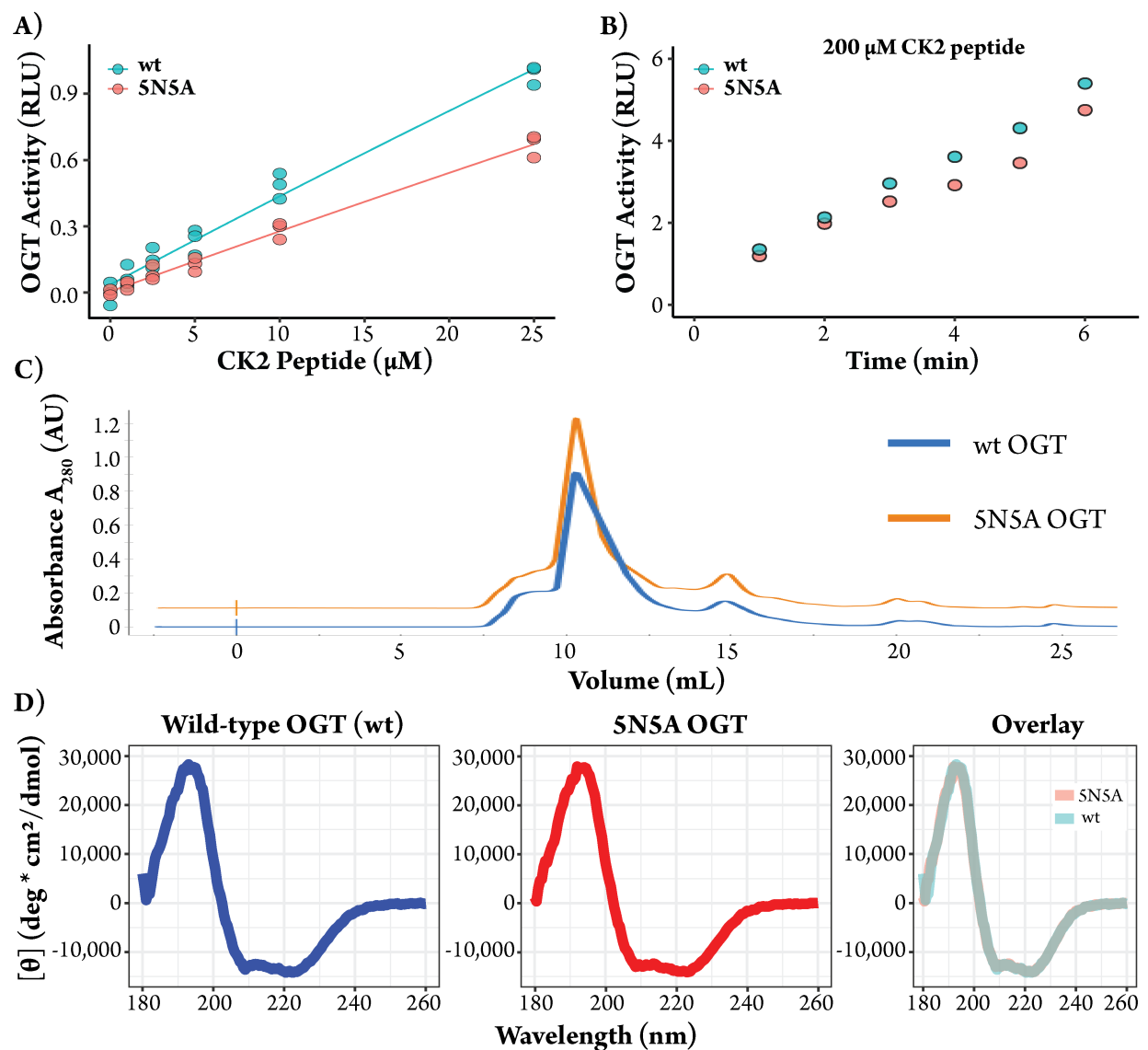


**Figure S4: *In vitro* mass shift shows activity against new biochemical substrates of OGT**

**A)** Schematic of mass-tagging method. A cDNA of a protein is translated by *in vitro* transcription/translation (IVT) in the presence of  $^{35}\text{S}$  Methionine. The resulting radiolabeled protein is glycosylated by OGT using the nucleotide sugar UDP-GlcNAz to label sites of *de novo* glycosylation. The azide is used to add a 5 kDa polyethylene glycol (PEG) to glycosylated proteins via strain-promoted azide-alkyne reaction, leading to a mass shift on SDS-PAGE if they are substrates for OGT.

**B)** Proteins that are substrates for OGT in IVT assay are listed below the corresponding radioactive gels. Arrows to the right of each gel highlight bands that shift in an OGT-dependent manner; multiple bands indicate the likely presence of multiple sites of glycosylation with differing levels of occupancy, leading to different numbers of attached PEG per molecule which is resolved as different bands on the gel.

Table S3: Detection of newly validated substrates by different detection methods		
Protein	Hit by CTD110.6 Detection	Hit by Biotin Detection
ARNT2	Yes	No
HDAC4	Yes	Yes
HDAC7	Yes	Yes
MBNL3	Yes	No
RUNX1T1	Yes	Yes
SCEL	No	Yes
SKA3	No	Yes
TLE3	Yes	Yes



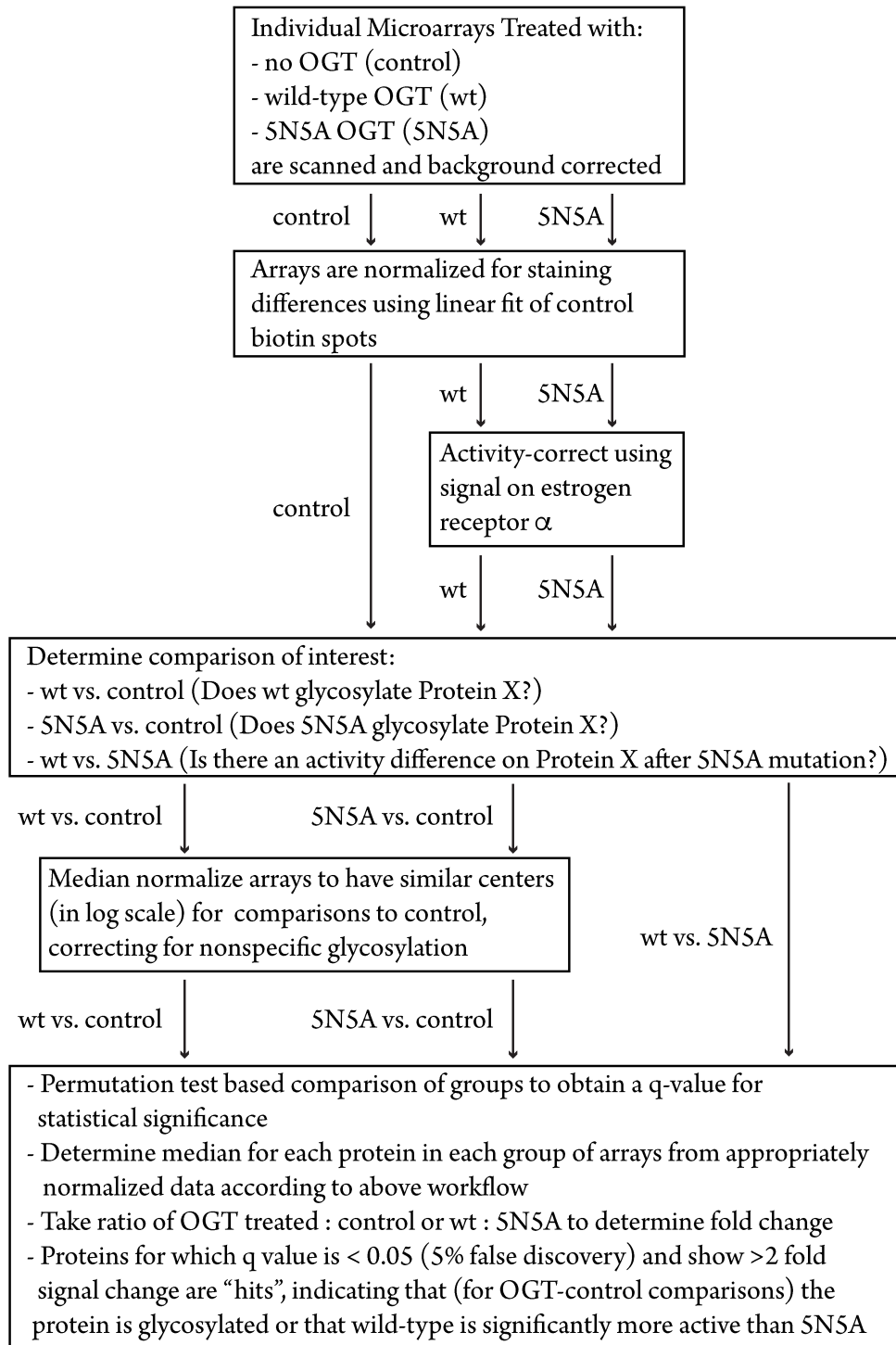
**Figure S5: Characterization of 5N5A mutant shows it is active and properly folded**

**A)** Activity of wild-type or 5N5A mutant OGT against short CK2 acceptor peptide at 6 minutes. Data is a zoom in of lower concentrations from Figure 2B.  $n = 3$  replicates per peptide concentration. Activity is measured in relative luciferase units (RLU), equivalent to AU in Figure 2B (different name is used to distinguish from absorbance units, AU, used in panel C of this figure). Data was generated using UDP-Glo<sup>®</sup> (Promega). Lines indicate Michaelis-Menten fit as reported in Figure 2B

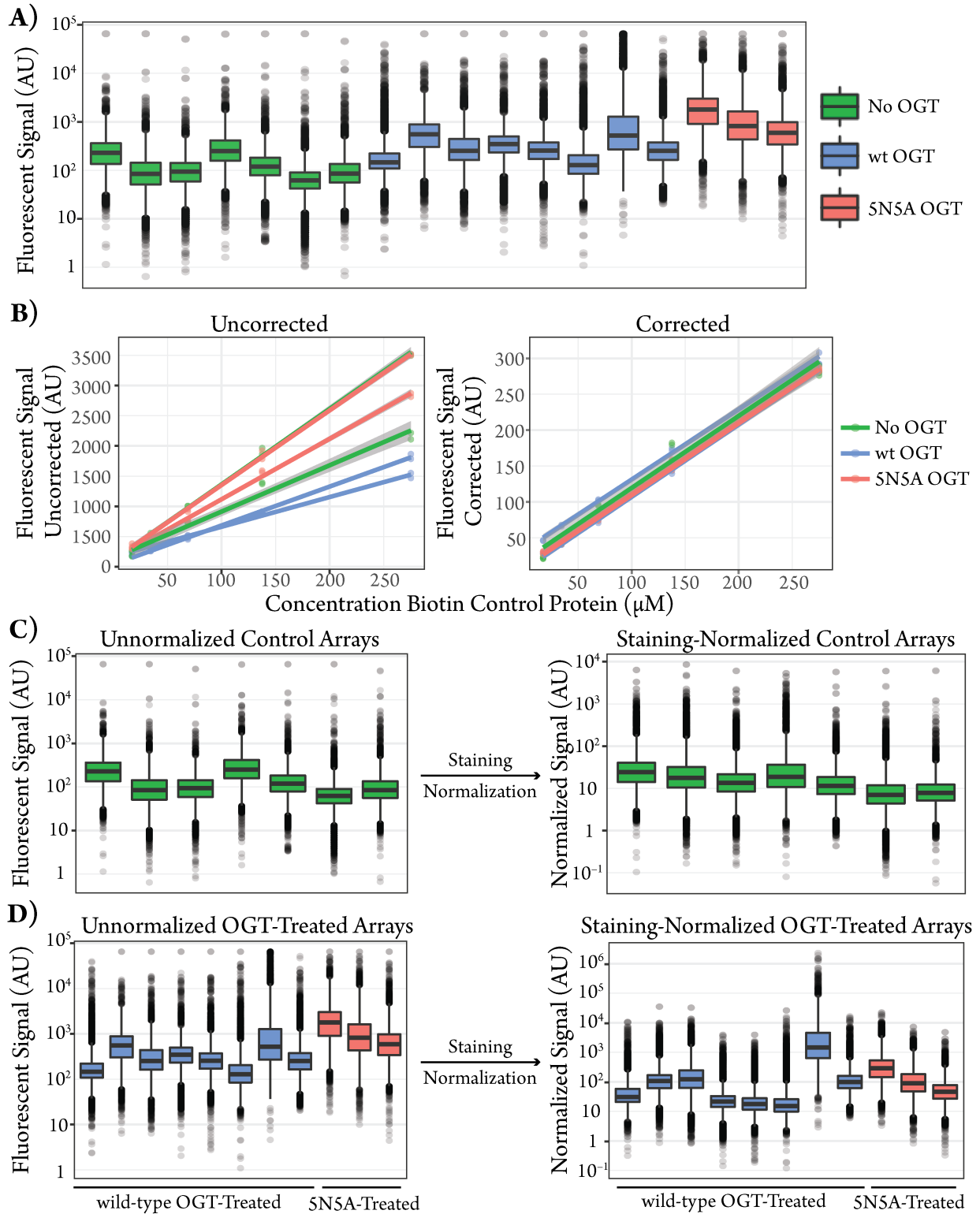
**B)** Activity over time against CK2 peptide at 200  $\mu\text{M}$  peptide. Both wild-type and 5N5A mutant OGT were used to glycosylate this peptide for varying lengths of time and RLU from UDP-Glo assay is reported.  $n = 1$  per time point per mutant.

**C)** Size exclusion chromatography on a Superdex 200 Increase 10/300 GL column shows similar elution profiles for wild-type and 5N5A OGT during purification. Note the trace for 5N5A is offset from wild-type (wt) by  $\sim 0.11$  AU for visibility.

**D)** Circular dichroism spectra (CD) show near identical spectra for wildtype OGT (left) and the 5N5A mutant (middle, with overlay of both on right). Spectra is reported in mean residue ellipticity versus wavelength, and is consistent with proper folding of the alpha-helical TPR domain.<sup>26</sup>



**Figure S6: Workflow diagram of how data was analyzed for comparison replicates of SN5A, wild-type, and no OGT control arrays**



**Figure S7: Normalization for staining reduces variation between arrays not treated with OGT**

**A)** Boxplots show fluorescent signal levels *before* correction for staining for all non-control spots on each microarray. Each boxplot is one array, and the replicates are colored by the treatment used. Dark bar in



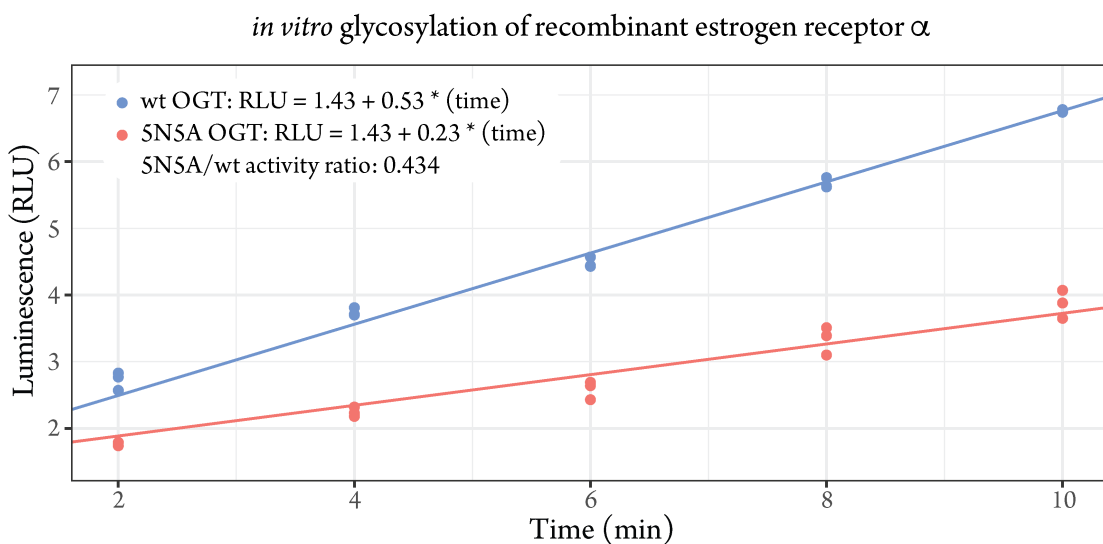
middle of box indicates median fluorescent signal, while black dots indicate outliers (more than 1.5 times the interquartile range outside of the 1<sup>st</sup> or 3<sup>rd</sup> quartile, in log scale)

**B)** Selected biotin control protein spots from 6 blocks (two per array, with three microarrays represented) are shown. The line of best fit based upon robust linear regression is shown for each individual block. On the left is the fit of the raw data, while on the right is the resulting data after correcting by the slope.

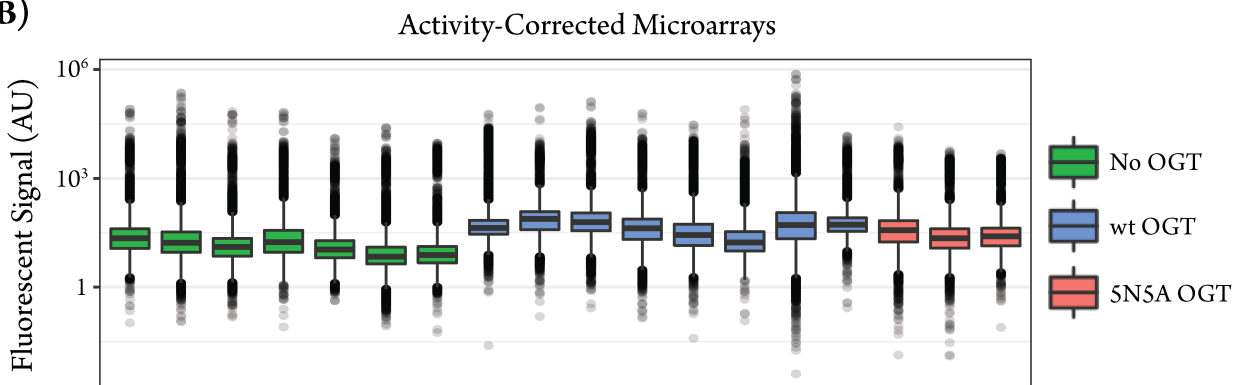
**C)** Boxplots show fluorescent signal levels for control microarrays before and after correction for staining. The staining normalization led to improved overlap between control arrays.

**D)** Boxplots show fluorescent signal levels for OGT-treated microarrays (both wild-type and 5N5A) before and after correction for staining. Unlike with the control arrays, the staining normalization exposed a high degree of variability we attribute to variation in activity between array replicates.

**A)**



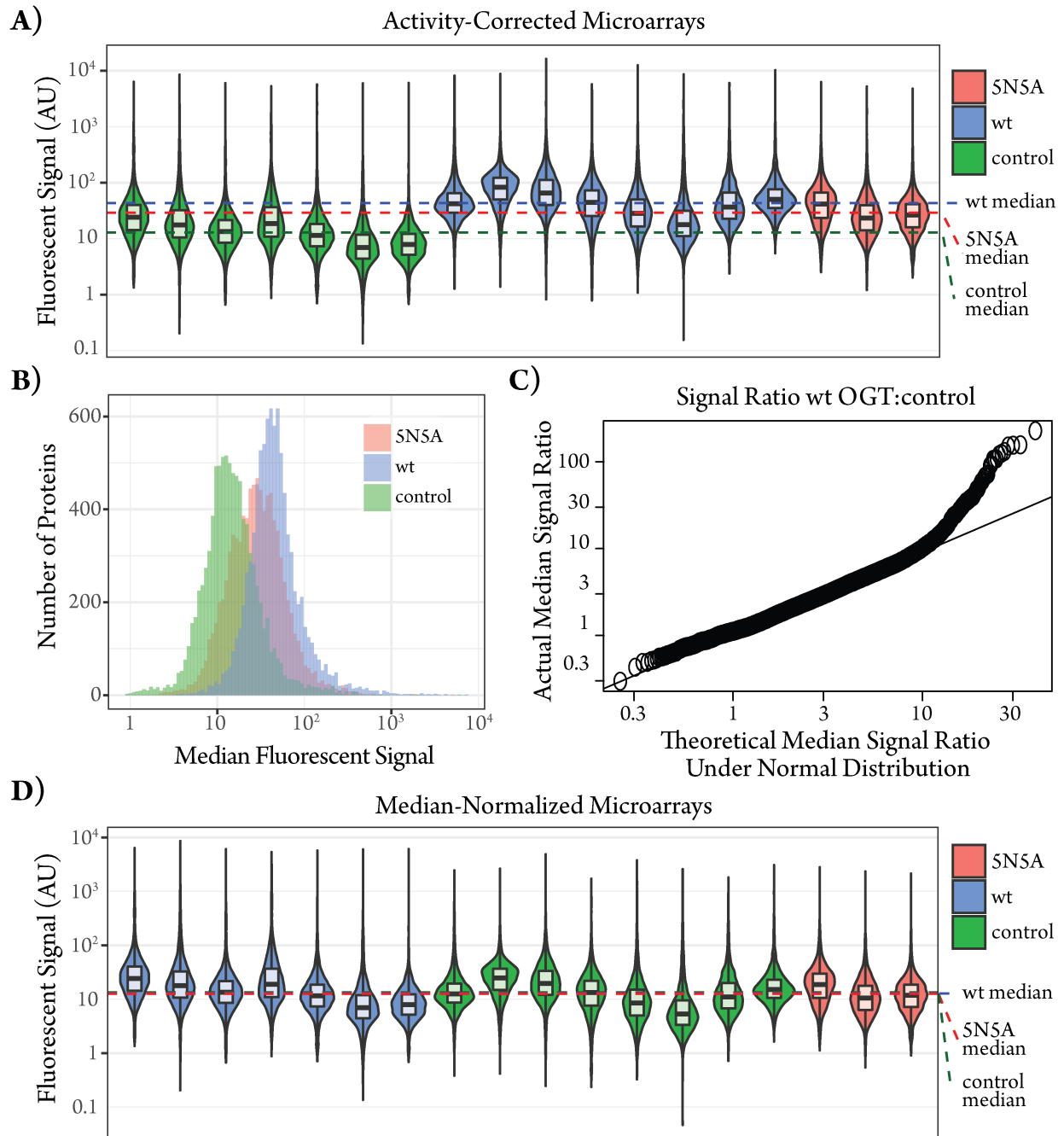
**B)**



**Figure S8: Use of estrogen receptor  $\alpha$  allows normalization of differences in OGT activity**

**A)** *In vitro* glycosylation of recombinant estrogen receptor  $\alpha$  (ER) shows more efficient glycosylation by wild-type OGT than 5N5A, with an activity ratio of 0.434.

**B)** Boxplots of individual arrays after activity correction (on the basis of the ER $\alpha$  control signal on each block) show a greater degree of similarity between distributions of OGT-treated arrays. The boxplot is configured as described for Figure S6A.



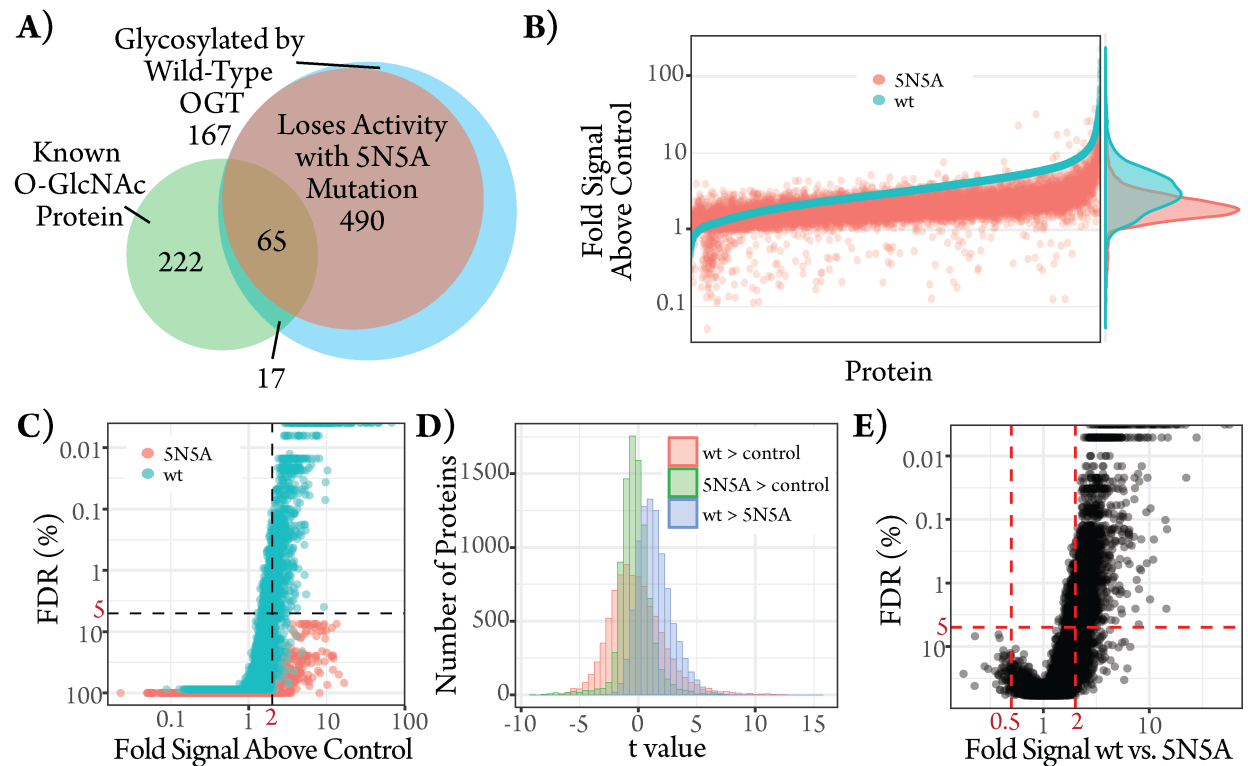
**Figure S9: Median normalization is required to distinguish non-specific and specific OGT activity**

**A)** Violin plots of activity-corrected microarrays. Each microarray is a separate shape, with number of proteins at a given level of signal represented by the width of the shape in the x-axis and the signal level along the y-axis. Arrays are colored by treatment, and the boxplot inside each shape indicates the median, first, and third quartiles of each microarray signal, as represented in Figure S7B, but without marking outlier values. The overall median of each group of arrays is indicated by the dashed colored lines, which are labeled on the right; the wild-type and SNSA arrays have higher median signal than the control arrays.

**B)** Histogram of protein signal by group provide an alternate illustration of the degree to which OGT-treated arrays have higher signal for almost every protein. For every protein on the microarray the median value across all microarrays was taken (on a per-accession number basis), then the histogram of signals is plotted together on the same log scale axis.

**C)** Comparing the ratio of median wild-type signals to median control signal (as done for single arrays in Figure S2) shows a distribution that is similar to what would be expected for a normal distribution. The y-axis is the actual distribution of signal ratio of median wild-type signal divided by median control signal for each protein; the x-axis is the value expected if the data followed a normal distribution, with the solid diagonal line showing the expected relationship between actual and theoretical data. Similar to seen on single arrays, there are a number of proteins with higher signal than predicted at the higher end of signals, suggests a set of proteins that are specifically glycosylated beyond the degree of nonspecific glycosylation seen across most proteins

**D)** Each group was median-normalized relative to one another, and re-plotted as shown in (A)



**Figure S10: SNSA has lower specific glycosylation and limited signal above background**

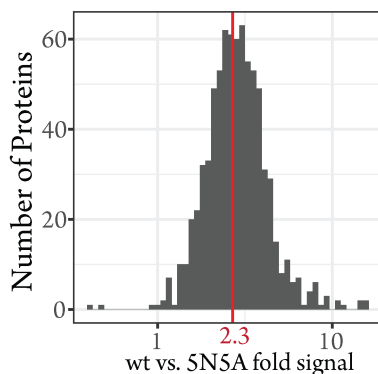
**A)** Venn diagram shows how many proteins score as hits at q-value (false-discovery rate) and fold-change cutoffs of 5% and 2-fold respectively. The blue circle marks proteins glycosylated by wild-type (i.e. proteins that were hits in a comparison between control arrays and wild-type), the red circle is proteins that are hits for loss of activity upon SNSA mutation (only proteins scoring as glycosylated by wild-type were considered, hence the complete overlap), and the green circle is proteins known to be glycosylated in cells from proteomics.<sup>12,16</sup> Note that the set of microarrays used in these analyses had higher overlap than those used for comparing detection methods, leading to a larger number of proteins analyzed (7427 in total by HGNC symbol).

**B)** Proteins are plotted along the x-axis arranged by fold signal above control arrays in a waterfall plot, and the corresponding signal above control arrays is plotted on the y-axis for both arrays treated with wild-type OGT (blue) and with SN5A (red). On the right, the overall probability density of proteins is shown, allowing a comparison of number of proteins from each category at each level. Fold signal values are the ratio of median activity-corrected signal for a protein on arrays treated with the appropriate form of OGT to the same protein's median signal on control arrays. Plotted data is not median-normalized. Overall, proteins with high signal for wild-type (on the right) show less signal for SN5A, though there are similar general trends in median activity.

**C)** A volcano plot comparing median fold signal above control (as described in B) to false discovery rate (FDR) percentage as quantified by a Storey's q value<sup>22</sup> for both wild-type (blue) and SN5A mutant (red) OGT. The black dashed lines show the cutoffs for being considered a hit as glycosylated in fold-change (vertical, 2-fold) and FDR (horizontal, <5% FDR). Note that while some proteins exceed the median fold change cutoff none meet the FDR criteria for glycosylation by SN5A, suggesting a failure for this mutant to consistently glycosylate these proteins in a manner that is clearly distinguishable from background.

**D)** Histograms of moderated t-statistics for each protein in each comparison (wild-type vs. control: red; SN5A vs. control: green; wild-type vs. SN5A, blue). These t-values were compared to values generated by random permutation as described by Yang and Churchill<sup>24</sup> to determine statistical significance. Note that wild-type vs. control has a large positive tail, while the overall distribution of SN5A vs. wild-type is right shifted from 0, suggesting a global phenotype of reduced glycosylation by SN5A rather than a specific subset of proteins being less recognized. The SN5A vs. control comparison lacks any clear positive tail, explaining the lack of significance of proteins at 5% FDR.

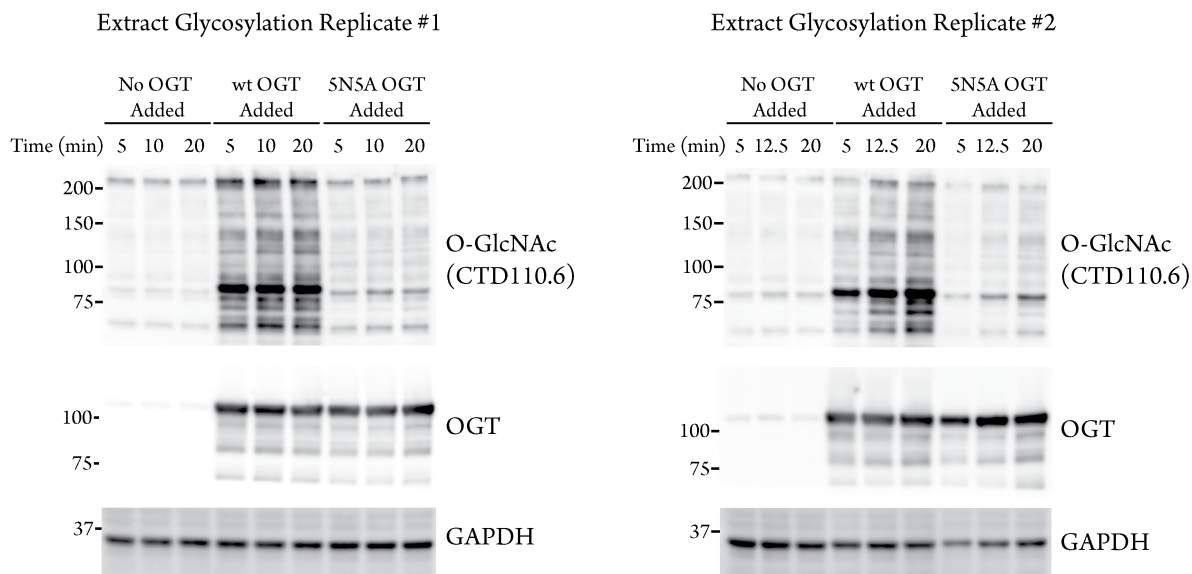
**D)** Volcano plot of a direct comparison between wild-type and SN5A mutant OGT shows that a large number of proteins are more glycosylated by wild-type OGT than the SN5A mutant. The x-axis is the ratio of median activity corrected signal for wild-type OGT to the SN5A mutant on a per-protein basis. The y-axis is a negative log scale of q-value, representing FDR. The dotted red line are the 5% FDR and 2-fold change cutoffs used to call hits- note there are no hits for SN5A being more active (top left) but a number of proteins that are hits for wild-type glycosylating better than SN5A (top right).



**Figure S11: TAB1, a substrate against which SN5A is known to be less active, is a representative protein in terms of loss of activity upon SN5A mutation**

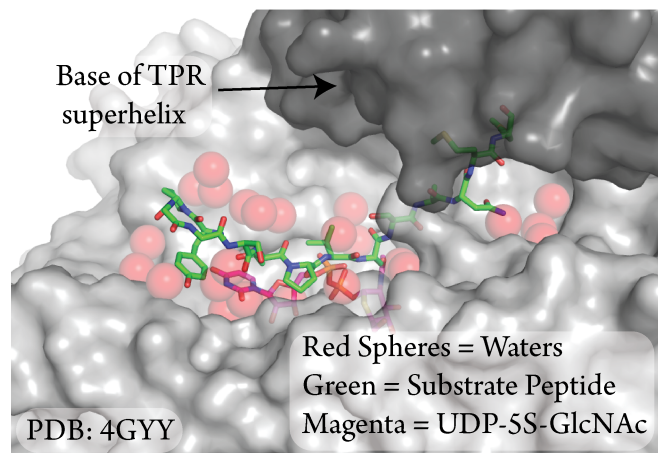
Histogram of ratios of wild-type to SN5A median activity-corrected signal on the microarrays. TAB1 (red vertical line), which is known to not to be glycosylated effectively by SN5A,<sup>27</sup> has an average change

in signal relative to other proteins. The 2.3-fold change is consistent with the 50% loss of activity upon mutation seen in the literature.<sup>27</sup>



**Figure S12: Western blot of HeLa cell extracts glycosylation with recombinant OGT**

Figure relates to Figure 3B in main text. Two replicates of extract glycosylation performed with different batches of HeLa S3 extracts and different batches of recombinant enzyme. Full molecular weight ranges showing staining are shown for each blot, with molecular weight markers indicated at left. Replicate #1 (left) is shown with reduced molecular weight ranges in Figure 3B.



**Figure S13: The OGT active site has a number of water molecules that can be displaced to accommodate diverse sequences**

This figure, from PDB 4GY Y,<sup>28</sup> highlights the OGT active site. The CK2 peptide substrate is shown in green above the UDP-GlcNAc (maroon sticks, crystallized as the thiosugar analogue UDP-5S-GlcNAc), with waters near the peptide highlighted as red spheres. Note the large number of waters near the side chains of peptide substrates, demonstrating that there is room in the shallow peptide cleft at the site of glycosylation to accommodate a wide variety of substrates in extended form.

**Table S4 (Excel File, separate): Results of multiple-replicate microarray studies comparing wild-type OGT, 5N5A mutant OGT, or no enzyme control arrays**

This table contains all proteins analyzed on the microarray as enumerated in Table S1. Median staining and activity normalized signal for each group (wild-type, control, and 5N5A) is listed (for wild-type and 5N-5A, median is given both before and after median normalization), as is the q-value measure of FDR-corrected statistical significance for each comparison. Whether proteins score as hits in comparisons between control and wild-type, control and 5N5A, or wild-type and 5N5A is listed as well.

<b>Table S5: Peptides and double-stranded DNA Fragments used in this study.</b>	
Construct	Sequence
dsDNA fragment for construction of 5N5A mutant	5' – CCGCTCTGCGTCTGTGCCCGACCCACGCTGACTCTCTGAACGCACT GGCTAACATCAAACGTGAACAGGGTAACATCGAAGAAGCTGTTTCG CTGTACCGTAAAGCTCTGGAAGTTTTCCCGGAATTCGCTGCTGCTC ACTCTGCACTGGCTTCTGTTCTGCAGCAGCAGGGTAAACTGCAGGA AGCTCTGATGCACTACAAAGAAGCTATCCGTATCTCTCCGACCTTC GCTGACGCTTACTCTGCAATGGGTAACACCCTGAAAGAAATGCAGG ACGTTTCAGGGTGCTCTGCAGTGCTACACCCGTGCTATCCAGATCAA CCCGGCTTTCGCTGACGCTCACTCTGCACTGGCTTCTATCCACAAA GACTCTGGTAACATCCCGGAAGCTATCGCTTCTTACCGTACCGCTC TGAAACTGAAACCGGACTTCCCGGACGCCTACTGCGCACTGGCTCA CTGCCTGCAGATCGTTTGGGACTGGACCGACT – 3'
HCF-E10S peptide	NH <sub>2</sub> – KKKYVRVCSNPPCSTHQGTNTATTATSNMAGQH – CONH <sub>2</sub>
CK2 peptide	NH <sub>2</sub> – KKKYPPGGSTPVSSANMM – CONH <sub>2</sub>

**Table S6: Protein microarrays used in study (data available on NIH GEO, series [GSE107911](#))**

Microarray Name	Barcode	Description	Array Batch
CTDcont1	52641	CTD-stained array, no enzyme	HA20301
CTDwt1	52630	CTD-stained array, wild-type OGT-treated	HA20301
Biotin_cont1	74577	Biotinylated array, no enzyme (replicate 1)	HA20358
Biotin_wt1	74576	Biotinylated array, wild-type OGT-treated (replicate 1)	HA20358
Biotin_cont2	74584	Biotinylated array, no enzyme (replicate 2)	HA20358
Biotin_cont3	91412	Biotinylated array, no enzyme (replicate 3)	HA20446
Biotin_cont4	74618	Biotinylated array, no enzyme (replicate 4)	HA20358
Biotin_cont5	74616	Biotinylated array, no enzyme (replicate 5)	HA20358
Biotin_cont6	74581	Biotinylated array, no enzyme (replicate 6)	HA20358
Biotin_cont7	74582	Biotinylated array, no enzyme (replicate 7)	HA20358
Biotin_wt2	74585	Biotinylated array, wild-type OGT-treated (replicate 2)	HA20358
Biotin_wt3	74586	Biotinylated array, wild-type OGT-treated (replicate 3)	HA20358
Biotin_wt4	74617	Biotinylated array, wild-type OGT-treated (replicate 4)	HA20358
Biotin_wt5	74619	Biotinylated array, wild-type OGT-treated (replicate 5)	HA20358
Biotin_wt6	74583	Biotinylated array, wild-type OGT-treated (replicate 6)	HA20358
Biotin_wt7	74579	Biotinylated array, wild-type OGT-treated (replicate 7)	HA20358
Biotin_wt8	74578	Biotinylated array, wild-type OGT-treated (replicate 8)	HA20358
Biotin_SNSA1	74600	Biotinylated array, SNSA OGT-treated (replicate 1)	HA20358
Biotin_SNSA2	91384	Biotinylated array, SNSA OGT-treated (replicate 2)	HA20446
Biotin_SNSA3	91454	Biotinylated array, SNSA OGT-treated (replicate 3)	HA20446

#### References:

- 1) Vocollo, D. J.; Hang, H. C.; Kim, E. J.; Hanover, J. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9116–9121.
- 2) Gross, B. J.; Kraybill, B. C.; Walker, S. *J Am Chem Soc* **2005**, *127*, 14588–14589.
- 3) Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.; Hutchison, C.A. 3<sup>rd</sup>; Smith, H.O. *Nat. Methods*, **2009**, *6*, 343–345.
- 4) R Core Team; *R: A language for statistical computing*, version 3.3.2. R Foundation for Statistical Computing: Vienna, Austria, 2016
- 5) RStudio Core Team; *RStudio: Integrated Development for R*, version 1.0.136. RStudio, Inc: Boston, MA, 2016
- 6) H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009
- 7) Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; and Smyth, G.K.; *Nucleic Acids Res.* **2015**, *43*, e47. (limma version 3.30.13)
- 8) *MASS package*, version 7.3-47, described in: Venables, W. N.; Ripley, B. D.; *Modern Applied Statistics with S*. Fourth Edition: Springer, New York. ISBN 0-387-95457-0
- 9) Wickham, H.; Francois, R.; Henry, L.; Müller, K.. *dplyr: A Grammar of Data Manipulation*. R package: version 0.7.1, 2017.
- 10) Wickham, H.; *J. Stat. Softw.* **2007**, *21*, 1-20.

- 11) (a) Ritchie, M.E.; Silver, J.; Oshlack, A.; Silver, J.; Holmes, M.; Diyagama, D.; Holloway, A.; Smyth, G.K.; *Bioinformatics* **2007**, *23*, 2700–2707. (b) Silver, J.; Ritchie, M.E.; Smyth, G.K.; *Biostatistics*, **2009**, *10*, 352–363.
- 12) Hornbeck, P.V.; Kornhauser, J.M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M.; *Nucleic Acids Res.* **2012**, *40*, D261–D270.
- 13) (a) Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W.; *Nat. Protoc.* **2009**, *4*, 1184–1191. (b) Durinck, S.; Moreau, Y.; Kasprzyk; Davis, A.; De Moor, B.; Brazma, A.; Huber, W.; *Bioinformatics*, **2005**, *21*, 3439–3440.
- 14) O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C.M.; Goldfarb, T.; Gupta, T.; Haft, D.; Hatcher, E.; Hlavina, W.; Joardar, V.S.; Kodali, V.K.; Li, W.; Maglott, D.; Masterson, P.; McGarvey, K.M.; Murphy, M.R.; O'Neill, K.; Pujar, S.; Rangwala, S.H.; Rausch, D.; Riddick, L.D.; Schoch, C.; Shkeda, A.; Storz, S.S.; Sun, H.; Thibaud-Nissen, F.; Tolstoy, I.; Tully, R.E.; Vatsan, A.R.; Wallin, C.; Webb, D.; Wu, W.; Landrum, M.J.; Kimchi, A.; Tatusova, T.; DiCuccio, M.; Kitts, P.; Murphy, T.D.; Pruitt, K.D.; *Nucleic Acids Res.* **2016**, *44*, D733–D745.
- 15) The Uniprot Consortium; *Nucleic Acids Res.*, **2017**, *45*, D158–D169 .
- 16) Wang, S.; Yang, F.; Petyuk, V.A.; Shukla, A.K.; Monroe, M.E.; Gritsenko, M.A.; Rodland, K.D.; Smith, R.D.; Qian, W.-J.; Gong, C.-X.; Liu, T.; *J. Pathol.* **2017**, *243*, 78–88.
- 17) Gray, K.A.; Yates, B.; Seal, R.L.; Wright, M.W.; Bruford, E.A.; *Nucleic Acids Res.* **2015**, *43*, D1079–D1085.
- 18) Ortiz-Meoz, R.F.; Merbl, Y.; Kirschner, M.W.; Walker, S.; *J. Am. Chem. Soc.*, **2014**, *136*, 4845–4848.
- 19) Greenfield, N.J.; *Nat. Protoc.*, **2006**, *1*, 2876–2890.
- 20) Ortiz-Meoz, R.F.; Jiang, J.; Lazarus, M.B.; Orman, M.; Janetzko, J.; Fan, C.; Dubeau, D.Y.; Tan, Z.-W.; Thomas, C.J.; Walker, S.; *ACS Chem. Biol.*, **2015**, *10*, 1392–1397.
- 21) Merbl, Y.; Kirschner, M.W.; *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 2543–2548.
- 22) (a) Storey, J. D.; *J. R. Stat. Soc. B*, **2002**, *64*, 479–498. (b) Storey, J.D.; Tibshirani, R.; *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9440–9445.
- 23) (a) Smyth, G.K.; *Stat. Appl. Genet. Mol. Biol.* **2004**, *3*, 3. (b) Phipson, B.; Lee, S.; Majewski, I.J.; Alexander, W.S.; and Smyth, G.K.; *Ann. Appl. Stat.* **2004**, *10*, 946–963.
- 24) Yang, H.; Churchill, G.; *Bioinformatics* **2007**, *23*, 38–43.
- 25) Rape, M.; Kirschner, M.W. *Nature*, **2004**, *432*, 588–595.
- 26) Carrigan, P.E.; Sikink, L.A.; Smith, D.F.; Ramirez-Alvarado, M.; *Protein Sci.*, **2006**, *15*, 522–532.
- 27) Rafie, K.; Raimi, O.; Ferenbach, A.T.; Borodkin, V.S.; Kapuria, V.; van Aalten, D.M.F. *Open Biol.* **2017**, *7*, 170078.
- 28) Lazarus, M.B.; Jiang, J.; Gloster, T.M.; Zandberg, W.F.; Vocadlo, D.J.; Walker, S.; *Nat. Chem. Biol.* **2012**, *8*, 966–968.