Generation of an Unbiased Interactome for the Tetratricopeptide Repeat Domain of O-GlcNAc Transferase Indicates a Role for the Enzyme in Intellectual Disability

Hannah M. Stephen, Jeremy L. Praissman, and Lance Wells*

Department of Biochemistry and Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, United States of America

Running Title: The OGT-TPR Interactome

Keywords: OGT, TPR, O-GlcNAc, Proximity Proteomics, Mass Spectrometry, Biotin ligase, X-Linked Intellectual Disability

Abbreviations:

BirA* - promiscuous Bifunctional ligase/repressor BirA

CID – collision-induced dissociation

IF – immunofluorescence

IP – immunoprecipitation

NSAF – normalized spectral abundance factor

OGT – O-GlcNAc Transferase

OMIM - Online Mendelian Inheritance in Man

TPR – Tetratricopeptide Repeat

WB – Western Blot

XLID – X-Linked Intellectual Disability

* Correspondence to: lwells@ccrc.uga.edu

Abstract

The O-GlcNAc transferase (OGT) is localized to the nucleus and cytoplasm where it regulates nucleocytoplasmic proteins by modifying serine and threonine residues with a non-extended monosaccharide, β -N-Acetyl-Glucosamine (O-GlcNAc). With thousands of known O-GlcNAc modified proteins but only one OGT encoded in the mammalian genome, a prevailing question is how OGT selects its substrates. Prior work has indicated that the N-terminal tetratricopeptide repeat (TPR) domain of OGT, rather than its C-terminal catalytic domain, is responsible for subcellular targeting and substrate selection. An additional impetus for exploring the OGT TPR domain interactome is the fact that missense mutations in *OGT* associated with X-linked intellectual disability (XLID) are primarily localized to the TPR domain without substantial impact on activity or stability of the enzyme. Therefore, we adapted the BioID labeling method to identify interactors of a TPR-BirA* fusion protein in HeLa cells. We identified 115 high confidence interactors representing both known and novel O-GlcNAc modified proteins and OGT interactors. The TPR interactors are highly enriched in processes in which OGT has a known role (e.g. chromatin remodeling, cellular survival of heat stress, circadian rhythm), as well as processes in which OGT has yet to be implicated (e.g. pre-mRNA processing). Importantly, the identified TPR interactors are involved in several disease states but most notably are highly enriched in pathologies featuring intellectual disability. These proteins represent candidate interactors that may underlie the mechanism by which mutations in *OGT* lead to XLID. Furthermore, the identified interactors provide additional evidence of the importance of the TPR domain for OGT targeting and/or substrate selection. Thus, this defined interactome for the TPR domain of OGT serves as a jumping off point for future research exploring the role of OGT, the TPR domain, and its protein interactors in multiple cellular processes and disease mechanisms, including intellectual disability.

Introduction

The O-GlcNAc transferase (OGT) is a nucleocytoplasmic glycosyltransferase that modifies substrate proteins with a β -N-acetylglucosamine (O-GlcNAc) on serine and threonine residues. OGT is a unique mammalian glycosyltransferase in that it modifies intracellular proteins outside of the secretory pathway, and the O-GlcNAc modification it creates is non-extended, dynamic, and inducible¹. The O-GlcNAc modification is often compared to phosphorylation, given their similar characteristics and the fact that both occur on thousands of nuclear and cytosolic proteins². In fact, OGT and Ser/Thr kinases often compete for the same sites on certain protein substrates² and can regulate each other by post-translational modification^{3,4}. However, unlike protein phosphorylation which is mediated by hundreds of kinases, there is only one gene encoding intracellular O-GlcNAc Transferase in mammals. Thus, not surprisingly, OGT is essential for life⁵ and is involved in many intracellular processes including nutrient sensing, transcription, and cellular stress¹. OGT also has been implicated in many diseases including cancer, Alzheimer's disease, diabetes, and more recently, in X-Linked Intellectual Disability (XLID) as identified originally by our team in collaboration with clinical partners⁶ and further confirmed and expanded on by our group and others⁷⁻⁹.

Given the wide diversity of OGT substrates and functions, and the existence of only one mammalian OGT, a prevailing question in the O-GlcNAc field is how OGT selects its substrates. Previous research suggests that the N-terminal tetratricopeptide repeat (TPR) domain of OGT (consisting of 13.5 repeats in the full-length protein), rather than its C-terminal catalytic domain, is responsible for OGT substrate selectivity¹⁰⁻¹². However, the hypothesis of the TPR domain mediating protein-protein interactions has only been directly tested for a few select proteins¹³⁻¹⁶. A few early attempts to define the full-length OGT-interactome using co-immunoprecipitation have also been made^{17,18}. An unbiased approach to identifying proteins that interact specifically with the TPR domain would lend further support to the hypothesis of the TPR domain mediating OGT substrate selectivity, and allow for the identification of new potential substrates and "partner

proteins", which interact with the TPR domain of OGT to target it to specific subcellular regions and/or protein complexes.

An additional impetus for TPR interaction studies is the observation that the majority of reported missense mutations in *OGT* causal for XLID are localized to the TPR domain and do not grossly affect catalytic activity or stability, suggesting a potential protein-protein interaction-based mechanism^{6,7}. Therefore, to demonstrate that the TPR domain of OGT is capable of substrate selection, and to capture endogenous OGT TPR interactors including transient interactors, we took advantage of the BioID method, utilizing a fusion protein consisting of the full-length OGT TPR domain with a modified biotin ligase in place of the catalytic domain of OGT.

BioID is a well-established proximity proteomic labeling method that utilizes a promiscuous biotin ligase (BirA*) to label nearby proteins with biotin, which allows them to easily be extracted and identified using mass-spectrometry based proteomics¹⁹. Using a TPR-BirA* fusion protein in a HeLa cell system (and a eGFP-BirA* fusion protein as a negative control), we have identified over 100 high-confidence OGT TPR interactors, including both known and novel OGT substrates and interactors. This work strongly suggests that the TPR domain, through protein-protein interactions, plays a major role in OGT substrate selectivity. Exploiting these interactions may allow for fine-tuning of the modification of specific O-GlcNAc modified substrates which has been explored using other techniques^{20,21}. This protein set also further confirms OGT's role in many cellular processes and reveals potential novel pathways in which O-GlcNAc may play an intricate role. Finally, the interactome is highly enriched in proteins involved in neurological disorders that present with intellectual disability. These proteins represent a set of candidate interactors to explore for future mechanistic studies of the functional role of OGT and the O-GlcNAc modification in XLID.

Experimental Procedures

Plasmid Constructs

99 Plasmids for proximity proteomics were constructed on a CMV promoter with a C-terminal BirA*.
 100 Fusion gene construction is as follows. For TPR BirA*: TPR - 3X GGGGS linker - BirA* - 2X FLAG.
 101 For eGFP-BirA*: eGFP - 3X GGGGS Linker - BirA* - 2XFLAG. The BirA* sequence was obtained
 102 from the original paper describing BioID¹⁹. For TPR-BirA*, residues 1-473 of OGT (consisting of
 103 the 13.5 TPR repeats of full-length OGT, Uniprot Accession O15294) were used. Full fusion
 104 protein sequences are in Supplementary Table 1.

105
 106 *Cell culture and BioID*
 107 HeLa cells were grown in DMEM with 10% FBS on 14.5cm plates, passaged for maintenance
 108 every 4-6 days (1:4-1:10). For expression of fusion proteins, cells were transfected at ~70%
 109 confluency using lipofectamine 2000 (Invitrogen) according to manufacturer recommended ratios
 110 with 60µg plasmid DNA for TPR-BirA* and 6µg plasmid DNA for eGFP-BirA* (eGFP-BirA*
 111 expresses at a much higher level than TPR-BirA* - see **Fig. 1 C/D**). After 24 hours, cell media
 112 was replaced with media containing 50uM biotin to induce labeling for 24 hours. Cells were then
 113 collected and subjected to nucleocytoplasmic lysis to collect protein. Briefly, cells were lysed in
 114 hypotonic buffer A (10mM Tris-HCl pH 5.5, 500uM DTT, 500uM EDTA, protease and phosphatase
 115 inhibitor cocktails (Sigma)) for 10 minutes, vortexed twice for 30 seconds, then lysed in hypertonic
 116 buffer B (10mM Tris-HCl pH 7.5, 800mM NaCl, protease and phosphatase inhibitors, and
 117 universal nuclease (ThermoFisher Scientific)) for 5 minutes, followed by 30 seconds of high-
 118 speed vortexing and an additional 5 minute incubation. Membranes were pelleted by
 119 centrifugation, and the supernatant containing cytosolic and nuclear proteins used in further
 120 analysis.

121
 122 *Quantification of FLAG-tagged proteins*
 123 FLAG-tagged fusion proteins were quantified from HeLa cell lysate using a FLAG ELISA (Cayman
 124 Chemical) according to the manufacturer's protocol.

125

126 *Identification of biotinylated proteins*

127 Biotinylated proteins were purified from cellular lysate using an anti-biotin immunoprecipitation
 128 protocol as previously described²². 25% of eluate was reserved for anti-biotin western blot, then
 129 remaining biotinylated proteins were then run out on a 10% SDS-PAGE gel using the Bio-Rad
 130 Mini-PROTEAN gel system. The gels were not stained. Each lane was cut into four equal
 131 fractions based on molecular weight markers, then extracted, reduced, alkylated, and digested
 132 with trypsin as previously described²². Dried peptides were resuspended in 40μL sample buffer
 133 (10% Buffer B (80% acetonitrile, 0.1% formic acid), 90% Buffer A (0.1% formic acid), and 15uL of
 134 this was injected for each instrument run. LC-MS/MS was performed on an Orbitrap Fusion Tribrid
 135 mass spectrometer (ThermoFisher Scientific) equipped with an Ultimate 3000 RSLCnano HPLC
 136 system (Thermofisher Scientific). Peptides were separated on an Acclaim™ PepMap™ RSLC
 137 C18 column (75 μm ID × 15 cm; 2 μm particle size) at a flow rate of 0.200μL/min over a 150min
 138 linear gradient of 1-99% Buffer B with a total run time of 180min. Precursor scans were collected
 139 using the Orbitrap mass analyzer with a scan range of 300-2000m/z and mass resolution of
 140 60,000. Most intense ions were fragmented using 38% CID collision energy and detected in the
 141 Ion Trap with 1 microscan and dynamic exclusion for 15 seconds after one occurrence. Samples
 142 were run in randomized pairs within a replicate, with each pair consisting of a corresponding gel
 143 fraction in eGFP-BirA*, run first, then TPR-BirA*, run second, with one 90 min wash in between
 144 each sample and two 90 minute washes in between pairs (20uL 10% Buffer B injection for
 145 washes). The raw data for all 24 LC-MS/MS analyses (12 control and 12 experimental) has been
 146 deposited to the MassIVE database (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>,
 147 Dataset ID: MSV000085626).

148 *Experimental design and statistical rationale*

149 Three biological replicates were performed, each consisting of one TPR-BirA* and one eGFP-
 150 BirA* sample, with TPR-BirA* representing the experimental condition and eGFP-BirA*

representing the negative control. An n of 3 was selected to allow us to perform statistical analyses with sufficient statistical power. Spectral counts were converted to normalized spectral abundance factors (NSAFs) for analysis²³. As described below, the normal logarithm (ln) of NSAF values was computed to create a more Gaussian/normal distribution, and the Student's T-Test was utilized to determine the significance of the difference in abundance between experimental and negative control conditions for each potential interacting protein.

Data analysis

Raw files were searched with no prior peaklist selection by both ProteinMetrics Inc. Byonic (v3.8.13) and ThermoFisher Scientific Proteome Discoverer - SequestHT (2.2.0.338). The proteomic database consisted of all SwissProt annotated human protein sequences (obtained 09-2019 – 20,434 sequences), plus a list of common contaminants (trypsin, keratins, and serum albumins – 179 sequences, available in Supplementary Table 1) and the sequences for eGFP and BirA*. A concatenated database was generated for FDR calculations by including reversed protein sequences for all proteins in the database, creating a final database of 41230 sequences, all of which were searched unbiasedly. Tryptic cleavage was fully specific at Arg and Lys, with two missed cleavages allowed. For both search engines, precursor mass tolerance was 20ppm and fragment mass tolerance was 0.5Da. Carbamidomethylation on Cys was set as a fixed modification, and variable modifications were allowed: Oxidation of Met, HexNAc on Ser/Thr, and Biotin on Lys. Full peptide lists were exported from PMI-Byonic and SequestHT with no score cuts.

Peptide lists from each sample (four fractions each) were then combined using ProValt²⁴. Peptide cuts were made based on peptide FDR, set at 2% for the eGFP-BirA* and 1% for TPR-BirA* - with the lower stringency of filtering for eGFP-BirA* samples selected to help decrease the incidence of potential false positives in the final protein set. Protein lists were generated by removing all peptides not identified by both Byonic and Proteome Discoverer (peptides required

177 to have a positive Byonic Score and Sequest Xcorr), and by removing any proteins that were
 178 identified by fewer than two peptides. Only proteins identified as the top protein from among
 179 isoforms were kept in the final protein list. For initial quantification and putative TPR interacting
 180 protein list generation, ln(NSAF) values were calculated for each identified protein and compared
 181 between experimental condition and negative control using the Microsoft Excel t.test function.
 182 Protein IDs with a pvalue of 0.05 or less and with at least a 3-fold higher NSAF were kept in the
 183 final TPR-BirA* interactome.

184 Further quantification was carried out by reconstructed ion chromatogram (RIC) analysis of
 185 peptide intensity performed in Xcalibur Qual Browser (v2.0.3.2). Peptides for analysis were
 186 selected based on their appearance in all three replicates in both TPR-BirA* and eGFP-BirA*.
 187 Peptide intensity was examined for all gel fractions in which that peptide occurred (For HCF1:
 188 fractions 1 and 2; for OGT: fraction 1 [rep 3] and fraction 2 [reps 1,2]; for KNL1: fraction 1). Time
 189 ranges for intensity analysis were selected to be the same in TPR-BirA* and eGFP-BirA*. The
 190 time range was selected based on the area of overlap between the corresponding peaks, or, in
 191 the case of peaks with a slight time offset leading to no overlap, the time range was selected so
 192 that the time was evenly split between the two peaks. Peptide intensity was determined by the
 193 normalization level (NL) of the monoisotopic peak. Peptide identity was validated by recorded
 194 retention time in ProteomeDiscoverer 2.2.

195

196 *Western Blots and Antibodies*

197 SDS-PAGE gels (4-15%) were run using the BioRad Mini-PROTEAN gel system. Gels were
 198 transferred onto Immobilon-FL PVDF membranes (Sigma) using the BioRad Trans-Blot SD Semi-
 199 Dry Transfer Cell. Membranes were blocked in 1% cold water fish skin gelatin (Sigma), then
 200 incubated with primary antibody at these ratios: anti-FLAG (1:2500, Sigma F3165), anti-biotin
 201 (1:1000, Jackson 200-002-211), anti-OGT (1:1000, Santa Cruz sc-74546), anti- α tubulin (WB
 202 1:5000, Cell Signaling 3873) Histone H3 (1:1000, Cell Signaling 14269), anti-GAPDH (WB, Cell

203 Signaling 2118). Secondary antibodies were Li-Cor IRDye: 680RD donkey-anti mouse 680
204 (1:10000), 800CW Goat anti-rabbit (1:20000). Three washes in TBST (0.1% tween) were
205 performed after each antibody incubation. All Western blots were imaged on a LiCor Odyssey Clx
206 system. Densitometric measurements were made using Image Studio Lite v5.2

207

208 *Localization studies (Nucleocytoplasmic Fractionation)*

209 Protein localization was determined using UniProt²⁵. Nuclear and cytoplasmic fractions from HeLa
210 cells were obtained using subcellular fractionation as previously described²⁶ and analyzed via
211 western blot as above.

212

213 *Pathway Analysis*

214 Gene ontology analysis was performed using The Gene Ontology Resource
215 (geneontology.org)^{27,28}. All GO lists were filtered at pvalue less than 0.01, FDR of less than 0.01,
216 and a minimum of 5-fold enrichment over expected number of proteins found in that category in
217 a random protein dataset. Biological process and molecular function analysis were performed
218 using the GO Ontology Database Released 2019-12-09. ReViGo²⁹ was used to generate
219 condensed lists of GO terms and CirGo³⁰ to generate plots from the condensed data. PANTHER
220 pathway analysis was performed using PANTHER version 15 released 2020-02-14. Reactome
221 data was also obtained from The Gene Ontology Resource, using Reactome version 65 released
222 2019-12-22, and parsed at FDR less than 5E-9. Condensed GO term lists and the full reactome
223 pathway list are available in Supplementary Table 5.

224 Disease association for proteins was identified using the OMIM catalog³¹. Disorders were
225 categorized manually, where “Intellectual Disability” refers to any disorder with the symptom
226 intellectual disability (or several other related terms), “Immunodeficiency” refers to disorders
227 causing immunodeficiency, “Malignancy” refers to any of several cancers, “Congenital, other”
228 refers to congenital disorders not featuring intellectual disability, “Neurological, other” refers to

229 non-congenital neurological disorders, and “Hormone” refers to disorders of the endocrine
230 system.

231

232 **Results**

233 *Defining the OGT TPR Interactome in HeLa Cells*

234 To identify OGT TPR interactors, we utilized a fusion protein strategy using promiscuous biotin
235 ligase BirA*. We generated a fusion protein TPR-BirA*, essentially replacing the catalytic domain
236 of OGT with BirA*, and also created a eGFP-BirA* fusion protein to serve as a negative control
237 for nonspecific protein interactions or promiscuous labeling (Supplementary Table 1). Each fusion
238 protein was transiently expressed in HeLa cells and induced with biotin for 24 hours for labeling
239 of proximal proteins (**Fig. 1A**). Note that for transfections, 10x more TPR-BirA* plasmid was used
240 than eGFP-BirA*, due to eGFP-BirA* expressing at a much higher level than TPR-BirA* (**Fig.**
241 **1C/D**). Following labeling, we isolated biotinylated proteins with a biotin immunopurification (**Fig.**
242 **1E**). It is noteworthy that even though there is significant biotin labeling in both TPR-BirA* and
243 eGFP-BirA*, the band pattern differs significantly indicating a change in the specificity of
244 biotinylation between eGFP-BirA* and TPR-BirA*. A sectioned SDS-PAGE gel was subjected to
245 in-gel digestion and the resulting peptides separated by nanoflow reverse-phase liquid
246 chromatography in-line to a tribrid mass spectrometer for protein identification (see methods).
247 This entire procedure (transfection to LC-MS/MS analyses) was carried out in 3 independent
248 biological replicates for both TPR-BirA* and eGFP-BirA*.

249 For analysis of the mass spectrometry data, we opted for a multi-algorithm search to increase the
250 stringency of our protein IDs (**Fig. 1B**). Raw mass spectrometry data was searched using both
251 PMI-Byonic and Sequest HT (through ProteomeDiscoverer 2.2) against the human database
252 (Swissprot 09/2019), and only peptides identified by both algorithms were used to generate the
253 final protein set. The negative control (eGFP-BirA*) protein set was searched at a looser peptide
254 FDR (2%) than the TPR protein set (1%) to further enhance the stringency of the final protein IDs.

255 To generate proteins for a final TPR interactors list, we combined uniquely identified proteins in
 256 the TPR-BirA* analyses with those that were enriched in the TPR-BirA* protein lists as compared
 257 to the eGFP-BirA* (negative control) protein lists. This enrichment was required to be significant
 258 according to the Student's t-test with the cut-off for significance being a p-value of 0.05, and we
 259 also required the average NSAF²³ to be at least 3 times higher in the TPR-BirA* condition
 260 compared to the eGFP-BirA* condition. These proteins represent a stringent list of 115 high
 261 confidence OGT TPR interactors (**Table 1**, Supplementary Table 4).

262 46 of the proteins (indicated by an *italicized* protein name in Table 1) had already been identified
 263 as O-GlcNAc modified by previous -omic datasets^{32,33}. This result suggests that the TPR domain
 264 alone can select OGT substrate proteins without the presence of the catalytic domain. In addition,
 265 8 (indicated by an ***italicized and bold*** protein name in Table 1) of the identified proteins had
 266 previously been demonstrated to specifically interact with OGT, where "interaction" here is defined
 267 as a one or two directional co-immunoprecipitation^{14–16,34–37}. Together, these factors lend
 268 confidence to the novel protein IDs in the dataset.

269 It is noteworthy that OGT itself is identified in the screen. Although many peptides attributed to
 270 OGT are due to the overexpression of the TPR-BirA* fusion protein, several high-confidence
 271 peptides in the catalytic domain of OGT were also identified. This indicates that endogenous OGT,
 272 which normally exists as a dimer¹¹, did complex with the fusion TPR protein. Several well-studied
 273 OGT interactors were also identified, including HCF1¹⁶, mSin3a¹⁴, and Tet2¹⁵. We also identify
 274 the O-GlcNAc hydrolase OGA, which OGT is known to regulate both pre- and post-
 275 translationally^{6,38}.

276 Many of the interactors identified here are members of protein complexes which may imply that
 277 OGT does not directly interact with all of the proteins in the complex. One such complex is the
 278 HIRA protein complex, previously shown to interact with OGT³⁴. We identified all three members
 279 of this complex (HIRA, UBN1, and CABIN). We also identified a novel TPR interaction with the
 280 SWI/SNF complex. Six members of the SWI/SNF complex were identified (SMRC2, SMRC1,

281 ACL6A, SMCA4, ARI1A, ARI1B). SWI/SNF proteins, like OGT³⁹, function in chromatin
282 remodeling⁴⁰, but OGT has, to our knowledge, never been shown to interact with these proteins.
283 In addition to many protein interactors involved in known OGT functions, we also identified
284 proteins with roles in RNA processing, an area of cellular biology for which limited research exists
285 on the role of OGT and the O-GlcNAc modification. These interactors include proteins with known
286 and putative roles in pre-mRNA splicing (SF3A1, PCF11, PRC2A, PR40A), polyadenylation
287 (CPSF1), and RNA binding (ZN106, TNF6B, RBM33, RBM25, RBM26).

288

289 *Validation of Proteins Identified in both TPR-BirA* and eGFP-BirA**

290 72 of the 115 OGT TPR interactors were only observed in the TPR interactome. Several protein
291 IDs (43 – indicated by a **bold** Uniprot accession in Table 1) were identified in both the TPR-BirA*
292 and eGFP-BirA* samples, but were significantly enriched in TPR-BirA* at the level of average
293 NSAF (Student's t-test p value < 0.05, fold enrichment of average NSAF >3) (**Fig. 2A**). To further
294 confirm the validity of the inclusion of these proteins in the final dataset, we examined MS1
295 reconstructed ion chromatograms for peptides identified in both TPR-BirA* and eGFP-BirA*. OGT
296 itself was identified in both, although it is highly enriched in the TPR-BirA*, likely in part due to the
297 overexpression of the TPR-BirA* fusion protein. To confirm that endogenous OGT labeling is
298 enriched in the TPR-BirA* sample, we compared the intensity of a catalytic domain peptide
299 between the TPR-BirA* samples and the eGFP-BirA* samples. The average intensity of this
300 peptide in the TPR-BirA* samples is 9.1 times higher than in the eGFP-BirA* samples, supporting
301 specific interaction of TPR-BirA* with full-length endogenous OGT (**Fig. 3A**).

302 To further confirm the enrichment of relevant OGT interactors in TPR-BirA*, we next examined
303 the intensity of peptides from HCF-1, a well-studied OGT interactor with roles in intellectual
304 disability^{16,41}. Average intensity for an HCF-1 peptide are 36.4-fold higher in TPR-BirA* than
305 eGFP-BirA* (**Fig. 3B**). Finally, we confirmed a fold increase in peptide intensity for the protein
306 closest to our fold enrichment cutoff, KNL1 (**Fig. 3C**). The peptide shown from KNL1 has an

average intensity in TPR-BirA* that is 3.93 higher than in eGFP-BirA*. Taken together, these results indicate that although some proteins were identified in both the experimental and negative control conditions, their inclusion in the final protein interactor list due to higher enrichment is supported by the raw mass spectrometry data.

Subcellular Localization of TPR-BirA Interactors*

OGT localizes primarily to the nucleus, but also to the cytoplasm, in the mammalian cell⁴². To confirm that the TPR-BirA* fusion protein also localized to both compartments, we examined the subcellular localization of OGT and our fusion proteins. By subcellular fractionation (**Fig. 4A/B**), endogenous OGT localizes primarily to the nucleus with some expression in the cytoplasm, as expected⁴². In contrast, the TPR-BirA* fusion protein localizes more highly to the cytoplasm, although it is also present in the nucleus. This result is unsurprising as previous research has shown that overexpressed OGT localizes more highly to the cytoplasm than native OGT⁴². eGFP-BirA* also localizes to both the nucleus and cytoplasm. The subcellular localization profiles of TPR-BirA* and eGFP-BirA* are very similar, supporting the use of eGFP-BirA* as a sufficient negative control for nonspecific labeling by BirA* in both the nuclear and cytosolic compartments. We expected most TPR-BirA* interactors to be primarily nuclear, as most recorded OGT interactors are as well¹. Analysis of the subcellular localization of identified TPR-BirA* interactors supports this (Visualized in an UpsetR plot **Fig. 4C**, and in a pie chart in **Fig. 4D**). 67 of the 115 identified interactors are exclusively nuclear, with an additional 23 occurring in both the nucleus and the cytoplasm. Several interactors localize specifically to the nuclear pore. Only four proteins exclusively localize to the cytoplasm, all of which are novel OGT interactors (RGPD5, WNK1, TNR6B, and ANKH1). Taken together, this supports the physiologic relevance of the identified TPR interactors and is consistent with prior studies suggestion that OGT interacts primarily with nuclear proteins¹.

Ontology Analyses of the OGT TPR-Interactome

To further understand the enrichment of various processes in our OGT TPR interactome, we performed several different Gene Ontology (GO) analyses using the Gene Ontology Resource (GeneOntology.org) (**Fig. 5**). Performance of a PANTHER Overrepresentation Test for biological processes in our interactome further confirms OGT interactors having frequent roles in transcriptional and chromatin regulation (**Fig. 5A**). General chromatin organization is a strongly enriched category, along with the related peptidyl-lysine modification (indicative of histone modification). Of note is the enrichment of OGT interactors specifically involved in gene silencing. OGT is a Polycomb Group Protein (sxc in *Drosophila melanogaster*), which is responsible for the silencing of Hox genes during developmental patterning^{43,44}. The interactors identified here further confirm OGT's tendency toward roles in gene silencing and may reveal further avenues by which OGT regulates gene silencing. Of additional note is the enrichment of interactors involved in the regulation of cellular response to heat, since previous work has demonstrated a role for OGT in coping with cellular heat shock⁴⁵; however, limited work has been published exploring the specific OGT interactions that help it to perform this function. Finally, the enrichment of proteins involved in rhythmic process and circadian rhythm aligns with previous research demonstrating that OGT is involved in circadian rhythm regulation⁴⁶. To confirm these enriched processes, we also performed a PANTHER Overrepresentation Test for molecular function pathways in our OGT TPR interactome (**Fig. 5B**). Many molecular functions corroborate our findings of enriched biological processes, including consistent high enrichment in chromatin and transcriptional regulation. It is interesting to note that RNA Pol II transcription factor binding in particular is an enriched molecular function, since OGT is known to interact with and regulate RNA Pol II-mediated transcription^{13,47,48}. Also enriched are processes specifically relating to histone modification, further confirming the enrichment of peptidyl-lysine modification of histones as identified in biological process enrichment and consistent with the O-GlcNAc modification being part of the histone code⁴⁹. Finally, enrichment of nuclear pore components and nuclear

localization sequence binding confirms the long-standing role for OGT in nuclear pore structure and/or regulation⁵⁰.

We further examined enriched Reactome pathways (**Table 2**) among TPR interactors, specifically the mostly highly enriched pathways at an FDR of less than 5E-9. This pathway analysis validates our previous GO analyses that reveal roles for OGT in chromatin regulation, transcriptional regulation, and nuclear pore processes. Reactome pathway enrichment also uniquely reveals several roles for OGT TPR interactors in viral infection, nuclear import, and processing. OGT has been demonstrated to play a role in a limited number of specific viral infections^{38,51} but these enriched pathways point to a potentially broader role for OGT and its interactors more generally in viral infection. Interactors are also enriched in the reactome pathway “regulation of glucokinase by glucokinase regulatory protein”. OGT has already been shown to regulate glucokinase⁵² as well as other proteins involved in glucose metabolism including phosphofructokinase¹⁵³. Panther pathway enrichment analysis (**Fig. 5C**) reveals OGT interactor involvement in basic leucine-zipped transcription factor mediated transcriptional regulation, the Wnt signaling pathway, and Huntington disease related processes. The interplay with basic leucine-zipped transcription factors points to another potential avenue for OGT’s regulation of transcription. Furthermore, OGT has already been shown to interface with the Wnt pathway by modulating β -Catenin stability⁵⁴. The interactors identified here involved in this pathway may point to other mechanisms by which OGT modulates Wnt signaling.

Pathophysiology Analyses of the OGT TPR-Interactome and Orthogonal validation of XLID-related Interactors

The identification of Huntington’s disease (**Fig. 5C**) as an enriched disease process among the TPR interactors prompted us to examine whether identified TPR interactors are involved in other disease processes. Unsurprisingly, as determined using the OMIM catalogue, many TPR interactors are involved in disease processes with which OGT is already associated, including malignancy⁵⁵ and neurological⁵⁶ disorders (**Fig. 6A/B**).

Strikingly, of the 44 TPR interactors involved in an OMIM-classified phenotype, 24 are linked to disorders which feature intellectual disability. Three interactors are linked to two different intellectual disability-related disorders each, bringing the total count of intellectual disability disorders associated to OGT interactors to 27 (**Fig 6A/B**). Note that one interactor is also linked to two different malignancy disorders, bringing the total number of disorders associated to TPR interactors to 48. Considering that OGT has recently been linked to X-Linked Intellectual Disability⁶⁻⁹, and that the majority of confirmed XLID-causing mutations occur in the TPR domain, these interactors are of high interest in assisting with the elucidation of the mechanism behind variants of OGT being causal for XLID. Furthermore, several of the XLID-associated OGT variants have been demonstrated to be catalytically normal, leading to the hypothesis that the XLID mutations may interrupt protein-protein interactions^{6,7,57}. Therefore, these 24 protein interactors are of significant interest in the search for a mechanism underlying the OGT XLID phenotype.

Discussion

One of the prevailing mysteries in the O-GlcNAc field is how the O-GlcNAc Transferase (OGT) enzyme is able to select from among thousands of possible substrates given that there is only one gene encoding the protein in the mammalian cell. A prevailing hypothesis in the field is that post-translational modification of OGT and protein-protein interactors are responsible for OGT substrate selection. An existing model is that, unlike phosphorylation specificity that evolved by gene duplication followed by divergence driven by evolutionary selective pressures leading to the hundreds of protein Ser/Thr kinases in the mammalian proteome⁵⁸, the O-GlcNAc modification specificity arose from protein-protein associations, perhaps primarily through the TPR domain of the enzyme, evolving to bring substrates into proximity. This mechanism may be similar to RNA PolIII that transcribes all protein-coding genes but is exquisitely regulated by protein-protein associations including the basal transcriptional machinery and transcription factors⁵⁹. The role of the TPR domain of OGT in promoting highly specific substrate selection has been demonstrated in part in structural biology studies^{11,12,60,61}, but has only been explored for a limited number of

specific substrates. Here, we demonstrated that the TPR domain of OGT is capable of interacting with substrate proteins even without the presence of the catalytic domain, and using the BioID technique we have identified 115 high-confidence TPR interactors, representing both known and novel OGT interactors.

OGT is known to localize to the nucleus and the cytoplasm, but usually primarily resides in the nucleus⁴². The fact that most interactors found here are exclusively localized to the nucleus despite the TPR-BirA* fusion protein being localized more dominantly to the cytoplasm is an intriguing observation, suggesting that OGT more strongly and/or frequently interacts with nuclear proteins regardless of localization. Future work is necessary to biochemically confirm each interaction, as well as to determine the degree to which given interactions are transient or stable, under what conditions they occur, and what functional roles they play in the cell.

It is important to note that the identified interactors fall into several different classes. Many of the identified proteins are O-GlcNAc modified, and these may be the effector substrates by which OGT modulates cellular status. Another group, not mutually exclusive to the first, may represent partner proteins; that is, proteins that interact with the TPR domain of OGT to target it to specific substrates or intracellular regions, thus affecting the substrates OGT can access. These proteins may or may not themselves be functionally O-GlcNAc modified. As an example, Tet2 binds to OGT and directs it to histones, but the O-GlcNAc modification on Tet2 has no observed effect on its function¹⁵. Finally, it is likely that some of the identified interactors do not directly interact with OGT but rather are members of a complex, a subset of which interact directly with OGT. The Swi/Snf complex is an example of this – we have identified 6 of at least 20 possible subunits in our TPR-BirA* interaction list. It is likely that the TPR domain does not directly interact with all 6 identified components and instead interacts with a subset, but additional complex members are labeled due to the diffusion of the reactive biotin intermediate released by the BirA* protein. This would also explain why we fail to identify the full complex; additional members of the complex may be too distant in space to be biotin labeled. The interactors identified here also help to narrow

the pool of possible proteins that are directly interacting with OGT, as opposed to a co-immunoprecipitation which would likely pull down the entire stable protein complex. Further work is required to identify direct versus indirect interactors.

In general, the identified interactors confirm OGT's role as a high-level regulator of cellular function. OGT has previously been characterized as a "rheostat" rather than a switch¹, and the data here supports this notion. Most of the TPR interactors we identified are "modulators" themselves, e.g. they are not enzymes with a direct effect on a given substrate, but rather affect cellular physiology at a global level by modulating transcription, protein stability, or transport. This gives a perspective of OGT as a modulator of the modulators; that is, OGT regulates cellular function by making many subtle changes in global regulators, adding up to a more significant functional outcome. One such global regulation avenue is chromatin remodeling, which is a previously known function of OGT and a function in which many of our TPR interactors are involved. While our data does not determine how OGT's interaction with these chromatin remodelers affects their function, many of our identified interactors are involved in lysine modification of histones, pointing to a potential avenue for OGT's regulation of chromatin remodeling. Indeed, OGT has already been noted to interact with histone modifying enzymes including HDACs¹⁴. We are unsure why we did not identify any HDACs in our screen – it is possible that they interact with OGT as a part of a protein complex but at a distance outside of the BirA* biotin labeling radius.

Interestingly, we have also identified proteins involved in survival during cellular heat stress. OGT has already been implicated in cellular survival of heat shock⁴⁵. Many of the interactors that fall into this ontology category are nuclear pore proteins. This may indicate that OGT's role in heat shock is mediated by its modification of nuclear pore proteins, as previously suggested⁶².

We have also identified interactors involved in biological processes in which OGT has yet to be implicated, most noteworthy in RNA processing and transport. Interestingly, OGA, which removes the O-GlcNAc modification, has previously been shown to localize to the nucleolus⁶³, indicating

the presence and possible role of O-GlcNAc modified proteins in this subcellular structure involved in RNA processing. Future work will be necessary to determine the specific role OGT plays in these processes.

Finally, we have identified many OGT interactors that are involved in disease. OGT and the O-GlcNAc modification are already known to be involved in many disease states including diabetes, cancer, and neurological disorders^{1,2}, but this is often only a correlative connection. The TPR interactors we present here may represent avenues for future research uncovering mechanistic proteins underlying OGT's role in various disease states. Of high interest in the field right now is the mechanism underlying OGT mutations leading to X-Linked Intellectual Disability (XLID). One prominent hypothesis that we have previously suggested^{6,7,57} is that mutations in the TPR domain disrupt OGT protein interactions, leading to downstream developmental effects that lead to the XLID phenotype. Here, we have identified 24 OGT TPR interactors directly involved in disorders with intellectual disability. While it is possible that a novel interactor or set of interactors underlies the OGT-XLID mechanism, these interactors represent a strong set of candidate interactors that may contribute to the phenotype. The high number of interactors with connections to intellectual disability may also indicate that there may be a more global interruption in protein-protein interactions caused by XLID variants in OGT. Rather than one specific interactor failing to interact with OGT and leading to XLID, there may be a more subtle reduction in interaction with many proteins, leading to global neurodevelopmental abnormalities.

Our lab is currently undertaking BioID and immunoprecipitation studies to identify any perturbations in the OGT interactome in neural lines harboring XLID-linked OGT variants. The BioID method described here will be a valuable tool to identify potentially tissue-/cell type-specific TPR interactors that fail to interact with XLID-associated OGT variants. In a more directed approach, intellectual disability-related interactors identified here are being screened for protein interaction with XLID-linked OGT variants to determine if they may represent protein interactors underlying the XLID-OGT phenotype. Thus, the work presented here lays a groundwork for

additional studies to understand OGT substrate selectivity and the role of OGT and the O-GlcNAc modification in a plethora of biological processes and human pathophysiology including XLID.

Acknowledgments

We would like to dedicate this paper to the memory of our beloved colleague Dr. Brent Weatherly, who developed the proteomic analysis workflows used in this manuscript. We thank Dr. Kelley Moremen for technical advice and plasmid constructs. This work was supported by a grant from the W.M. Keck foundation (L.W. Co-PI); an NICHD National Institutes of Health (NIH) grant R21HD097652 (L.W); and an NICHD Grant F30 HD098828 (H.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The final TPR interactors list along with statistical analyses is available in supplementary table 2. All peptide matches (supplementary table 3) and protein IDs from each replicate (supplementary table 4) along with GO term lists (supplementary table 5) are attached as supplementary data. Raw mass spectrometry data (24 .raw files) is deposited in the MassIVE database (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>, Dataset ID: MSV000085626).

Author Contributions

H.S. and L.W. conceived and coordinated the study and wrote the manuscript. H.S. performed all experiments and data searches. J.P. performed calculations and statistical analyses on final protein lists and assisted in revision of the manuscript.

References

1. Hart, G. W. Nutrient regulation of signaling and transcription. (2019). *J. Biol. Chem.* 294, 2211–2231
2. Hart, G., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-glcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. (2011). *Annu. Rev. Biochem.* 80, 825–858
3. Li, Z. *et al.* Checkpoint kinase 1-induced phosphorylation of O-linked β -N-acetylglucosamine transferase regulates the intermediate filament network during cytokinesis. (2017). *J. Biol. Chem.* 292, jbc.M117.811646
4. Tarrant, M. K. *et al.* Regulation of CK2 by Phosphorylation and O-GlcNAcylation Revealed by Semisynthesis. (2012). *Nat. Chem. Biol.* 8, 262–269
5. Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. (2000). *Proc. Natl. Acad. Sci.* 97, 5735–5739
6. Vaidyanathan, K. *et al.* Identification and Characterization of a Missense Mutation in the O-GlcNAc Transferase Gene that Segregates with X-Linked Intellectual Disability. (2017). *J. Biol. Chem.* 292, jbc.M116.771030
7. Selvan, N. *et al.* O-GlcNAc transferase missense mutations linked to X-linked intellectual disability deregulate genes involved in cell fate determination and signaling. (2018). *J. Biol. Chem.* 293, 10810–10824
8. Pravata, V. M., Gundogdu, M., Bartual, S. G., Ferenbach, A. T. & Wojcik, M. H. A missense mutation in the catalytic domain of O -GlcNAc transferase links perturbations in protein O -GlcNAcylation to X-linked intellectual disability. (2019). 1–11 doi:10.1002/1873-3468.13640
9. Pravata, V. M. *et al.* Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability. (2019). 116, 14961–14970
10. Iyer, S. P. N. & Hart, G. W. Roles of the Tetratricopeptide Repeat Domain in O-GlcNAc Transferase Targeting and Protein Substrate Specificity. (2003). *J. Biol. Chem.* 278, 24608–24616
11. Jinek, M. *et al.* The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. (2004). *Nat Struct Mol Biol* 11, 1001–1007
12. Levine, Z. *et al.* O-GlcNAc transferase recognizes protein substrates using an asparagine ladder in the TPR superhelix. (2018). *J. Am. Chem. Soc.* jacs.7b13546 doi:10.1021/jacs.7b13546
13. Comer, F. I. & Hart, G. W. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. (2001). *Biochemistry* 40, 7845–7852
14. Yang, X., Zhang, F. & Kudlow, J. E. Recruitment of O -GlcNAc Transferase to Promoters by Corepressor mSin3A : Coupling Protein O -GlcNAcylation to Transcriptional Repression. (2002). *Cell* 110, 69–80
15. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Xiaochun, Y. Tet2 promotes histone O-GlcNAcylation during gene transcription. (2013). *Nature* 493, 561–564
16. Capotosti, F. *et al.* O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. (2011). *Cell* 144, 376–388
17. Deng, R. P. *et al.* Global identification of O-GlcNAc transferase (OGT) interactors by a human proteome microarray and the construction of an OGT interactome. (2014). *Proteomics* 14, 1020–1030
18. Gao, J. *et al.* Proteomic analysis of the OGT interactome: novel links to epithelial-mesenchymal transition and metastasis of cervical cancer. (2018). *Carcinogenesis* 1–13 doi:10.1093/carcin/bgy097
19. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. (2012). *J. Cell Biol.* 196, 801–810
20. Ramirez, D. H. *et al.* Engineering a proximity-directed O-GlcNAc transferase for selective protein O-GlcNAcylation in cells. (2020). *ACS Chem. Biol.* 15, 1059–1066
21. Zhu, Y. & Hart, G. W. Targeting the O-GlcNAc Transferase to Specific Proteins Using RNA Aptamers. (2020). *FASEB J.* 31, 1–1
22. Yu, S. H. *et al.* Selective exo-enzymatic labeling detects increased cell surface sialoglycoprotein expression upon megakaryocytic differentiation. (2016). *J. Biol. Chem.* 291, 3982–3989
23. Zybaïlov, B. *et al.* Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. (2006). *J. Proteome Res.* 5, 2339–2347
24. Weatherly, D. B. *et al.* A Heuristic Method for Assigning a False-discovery Rate for Protein

- Identifications from Mascot Database Search Results. (2005). *Mol. Cell. Proteomics* 4, 762–772
25. UniProt Consortium, T. UniProt: A worldwide hub of protein knowledge. (2018). *Nucleic Acids Res.* 47, D506–D515
26. Schreiber, E., Matthias, P., Muller, M. M. & Schaffner, W. Rapid detection of octamer binding proteins with ‘mini-extracts’, prepared from a small number of cells. (1989). *Nucleic Acids Res.* 17, 6419
27. Gene Ontology Consortium, T. *et al.* The Gene Ontology Consortium, Michael Ashburner¹, Catherine A. Ball³, Judith A. Blake⁴, David Botstein³, Heather Butler¹, J. Michael Cherry³, Allan P. Davis⁴, Kara Dolinski³, Selina S. Dwight³, Janan T. Eppig⁴, Midori A. Harris³, David P. Hill⁴, Laurie Is. (2000). *Nat. Genet.* 25, 25–29
28. Gene Ontology Consortium, T. The Gene Ontology Resource: 20 years and still GOing strong. (2019). *Nucleic Acids Res.* 47, D330–D338
29. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. Revigo summarizes and visualizes long lists of gene ontology terms. (2011). *PLoS One* 6,
30. Kuznetsova, I., Lugmayr, A., Siira, S. J., Rackham, O. & Filipovska, A. CirGO: An alternative circular way of visualising gene ontology terms. (2019). *BMC Bioinformatics* 20, 1–7
31. Online Mendelian Inheritance in Man, OMIM. <https://omim.org/> (2020) *McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD)*
32. Alfaro, J. F. *et al.* Tandem mass spectrometry identifies many mouse brain O -GlcNAcylated proteins including EGF domain-specific O -GlcNAc transferase targets. (2012). *Pnas* 109, 7280–7285
33. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. (2015). *Nucleic Acids Res.* 43, D512–D520
34. Lee, J. S. & Zhang, Z. O-linked N-acetylglucosamine transferase (OGT) interacts with the histone chaperone HIRA complex and regulates nucleosome assembly and cellular senescence. (2016). *Proc. Natl. Acad. Sci. U. S. A.* 113, E3213–E3220
35. Slawson, C., Lakshmanan, T., Knapp, S. & Hart, G. W. A Mitotic GlcNAcylation/Phosphorylation Signaling Complex Alters the Posttranslational State of the Cytoskeletal Protein Vimentin. (2008). *Mol. Biol. Cell* 19, 4130–4140
36. Dey, A. *et al.* Loss of the tumor suppressor BAP1 causes myeloid transformation. (2012). *Science (80-)*. 337, 1541–1546
37. Kreppel, L. K. & Hart, G. W. Regulation of a Cytosolic and Nuclear O-GlcNAc Transferase. (1999). *J. Biol. Chem.* 274, 32015–32022
38. Groussaud, D. *et al.* Hijacking of the O-GlcNAcZYME complex by the HTLV-1 Tax oncoprotein facilitates viral transcription. (2017). *PLoS Pathog.* 13, 1–23
39. Hardville, S. & Hart, G. Nutrient Regulation of Signaling, Transcription, and Cell Physiology by O-GlcNAcylation. (2014). *Cell Metab.* 20, 208–213
40. Brownlee, P. M., Meisenberg, C. & Downs, J. A. The SWI/SNF chromatin remodelling complex: Its role in maintaining genome stability and preventing tumorigenesis. (2015). *DNA Repair (Amst)*. 32, 127–133
41. Huang, L. *et al.* A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. (2012). *Am. J. Hum. Genet.* 91, 694–702
42. Kreppel, L. K., Blomberg, M. a & Hart, G. W. Dynamic Glycosylation of Nuclear and Cytosolic Proteins. (1997). *J. Biol. Chem.* 272, 9308–9315
43. Gambetta, M. C., Oktaba, K. & Muller, J. Essential Role of the Glycosyltransferase Sxc/Ogt in Polycomb Repression. (2009). *Science (80-)*. 325, 93–96
44. Sinclair, D. A. R. *et al.* Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). (2009). *Proc. Natl. Acad. Sci. U. S. A.* 106, 13427–13432
45. Martinez, M. R., Dias, T. B., Natov, P. S. & Zachara, N. E. Stress-Induced O-GlcNAcylation, an Adaptive Process of Injured Cells. (2017). *Biochem. Soc. Trans.* 45, 237–249
46. Kaasik, K. *et al.* Glucose Sensor O-GlcNAcylation Coordinates with Phosphorylation to Regulate Circadian Clock. (2013). *Cell Metab.* 17, 291–302
47. Kelly, W. G., Dahmus, M. E. & Hart, G. W. RNA polymerase II is a glycoprotein. (1993). *J. Biol. Chem.* 268, 10416–10424
48. Ranuncolo, S. M., Ghosh, S., Hanover, J. A., Hart, G. W. & Lewis, B. A. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. (2012). *J. Biol.*

Chem. 287, 23549–23561

49. Sakabe, K., Wang, Z. & Hart, G. W. β -N-acetylglucosamine (O-GlcNAc) is part of the histone code. (2010). *Proc. Natl. Acad. Sci. U. S. A.* 107, 19915–19920
50. Eustice, M., Bond, M. R. & Hanover, J. A. O-GlcNAc cycling and the regulation of nucleocytoplasmic dynamics. (2017). *Biochem. Soc. Trans.* 45, 427–436
51. Wang, Q. *et al.* O-GlcNAc transferase promotes influenza A virus-induced cytokine storm by targeting interferon regulatory factor-5. (2020). *Sci. Adv.* 6, 1–13
52. Baldini, S. F. *et al.* Glucokinase expression is regulated by glucose through O-GlcNAc glycosylation. (2016). *Biochem. Biophys. Res. Commun.* 478, 942–948
53. Yi, W. *et al.* PFK1 Glycosylation Is a Key Regulator of Cancer Cell Growth and Central Metabolic Pathways Wen. (2013). *Science* (80-.). 337, 975–980
54. Olivier-Van Stichelen, S. *et al.* O-GlcNAcylation stabilizes β -catenin through direct competition with phosphorylation at threonine 41. (2014). *FASEB J.* 28, 3325–3328
55. Hanover, J. A., Chen, W. & Bond, M. R. O- GlcNAc in cancer : An Oncometabolism-fueled vicious cycle. (2018). *J. Bioenerg. Biomembr.*
56. Akan, I., Stichelen, S. O.-V., Bond, M. R. & Hanover, J. A. Nutrient-driven O-GlcNAc in proteostasis and neurodegeneration. (2017). *J. Neurochem.* 12, 3218–3221
57. Pravata, V. M. *et al.* An intellectual disability syndrome with single-nucleotide variants in O-GlcNAc transferase. (2020). *Eur. J. Hum. Genet.* 28, 706–714
58. Oruganty, K. & Kannan, N. Design principles underpinning the regulatory diversity of protein kinases. (2012). *Philos. Trans. R. Soc. B Biol. Sci.* 367, 2529–2539
59. Danino, Y. M., Even, D., Ideses, D. & Juven-Gershon, T. The core promoter: At the heart of gene expression. (2015). *Biochim. Biophys. Acta - Gene Regul. Mech.* 1849, 1116–1131
60. Joiner, C. M., Levine, Z. G., Aonbangkhen, C., Woo, C. M. & Walker, S. Aspartate residues far from the active site drive O-GlcNAc transferase substrate selection. (2019). *J. Am. Chem. Soc.* jacs.9b06061 doi:10.1021/jacs.9b06061
61. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. (2011). *Nature* 2 469, 564–567
62. Li, B. & Kohler, J. J. Glycosylation of the nuclear pore. (2014). *Traffic* 15, 347–361
63. Zeidan, Q., Wang, Z., Maio, A. De & Hart, G. W. O-GlcNAc Cycling Enzymes Associate with the Translational Machinery and Modify Core Ribosomal Proteins. (2010). *Mol. Biol. Cell* 21, 1922–1936

Table 1: Interactors identified as TPR-BirA* interactors in HeLa cells

Uniprot Accession	Sequence Name		
ZHX3_HUMAN	Zinc fingers and homeoboxes protein 3	KMT2D_HUMAN	Histone-lysine N-methyltransferase 2D
CCAR1_HUMAN	Cell division cycle and apoptosis regulator protein 1	ZHX1_HUMAN	<i>Zinc fingers and homeoboxes protein 1</i>
JHD2C_HUMAN	<i>Probable JmjC domain-containing histone demethylation protein 2C</i>	PF21A_HUMAN	PHD finger protein 21A
ZN609_HUMAN	Zinc finger protein 609	ARI1A_HUMAN	AT-rich interactive domain-containing protein 1A
QSER1_HUMAN	<i>Glutamine and serine-rich protein 1</i>	ZFH3_HUMAN	Zinc finger homeobox protein 3
RBM26_HUMAN	RNA-binding protein 26	EP300_HUMAN	Histone acetyltransferase p300
ZHX2_HUMAN	Zinc fingers and homeoboxes protein 2	SF3A1_HUMAN	Splicing factor 3A subunit 1
PRR12_HUMAN	Proline-rich protein 12	BCL9_HUMAN	B-cell CLL/lymphoma 9 protein
TRPS1_HUMAN	Zinc finger transcription factor Trps1	TRRAP_HUMAN	Transformation/transcription domain-associated protein
BCOR_HUMAN	<i>BCL-6 corepressor</i>	ACL6A_HUMAN	Actin-like protein 6A
OGA_HUMAN	Protein O-GlcNAcase ³⁵	TAF6_HUMAN	<i>Transcription initiation factor TFIID subunit 6</i>
CIC_HUMAN	Protein capicua homolog	TNR6A_HUMAN	<i>Trinucleotide repeat-containing gene 6A protein</i>
KMT2C_HUMAN	Histone-lysine N-methyltransferase 2C	POGZ_HUMAN	<i>Pogo transposable element with ZNF domain</i>
CEP85_HUMAN	Centrosomal protein of 85 kDa	AHDC1_HUMAN	AT-hook DNA-binding motif-containing protein 1
NCOA6_HUMAN	<i>Nuclear receptor coactivator 6</i>	I2BP2_HUMAN	Interferon regulatory factor 2-binding protein 2
HIRA_HUMAN	Protein HIRA ³⁴	RREB1_HUMAN	<i>Ras-responsive element-binding protein 1</i>
ARI5B_HUMAN	<i>AT-rich interactive domain-containing protein 5B</i>	NCOR2_HUMAN	<i>Nuclear receptor corepressor 2</i>
ZN318_HUMAN	Zinc finger protein 318	PHF12_HUMAN	PHD finger protein 12
RGPD8_HUMAN	<i>RANBP2-like and GRIP domain-containing protein 8</i>	OGT1_HUMAN	UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit ³⁷
RGPD5_HUMAN	RANBP2-like and GRIP domain-containing protein 5/6	ANR17_HUMAN	<i>Ankyrin repeat domain-containing protein 17</i>
ZFR_HUMAN	<i>Zinc finger RNA-binding protein</i>	RBM27_HUMAN	<i>RNA-binding protein 27</i>
RGPD3_HUMAN	<i>RanBP2-like and GRIP domain-containing protein 3</i>	NU153_HUMAN	<i>Nuclear pore complex protein Nup153</i>
ZC3HE_HUMAN	Zinc finger CCCH domain-containing protein 14	KMT2A_HUMAN	Histone-lysine N-methyltransferase 2A
BAP1_HUMAN	Ubiquitin carboxyl-terminal hydrolase BAP1 ³⁶	NUP50_HUMAN	Nuclear pore complex protein Nup50
TET2_HUMAN	Methylcytosine dioxygenase TET2 ¹⁵	TNR6B_HUMAN	Trinucleotide repeat-containing gene 6B protein
ZEP1_HUMAN	Zinc finger protein 40	HCFC1_HUMAN	Host cell factor 1 ¹⁶
RPRD2_HUMAN	<i>Regulation of nuclear pre-mRNA domain-containing protein 2</i>	ZC3H1_HUMAN	Zinc finger C3H1 domain-containing protein
PHC3_HUMAN	<i>Polyhomeotic-like protein 3</i>	GANP_HUMAN	Germinal-center associated nuclear protein
CABIN_HUMAN	Calcineurin-binding protein cabin-1	UBAP2_HUMAN	<i>Ubiquitin-associated protein 2</i>
SET1A_HUMAN	Histone-lysine N-methyltransferase SETD1A	NCOR1_HUMAN	<i>Nuclear receptor corepressor 1</i>
TBL1X_HUMAN	F-box-like/WD repeat-containing protein TBL1X	EMS1_HUMAN	<i>BRCA2-interacting transcriptional repressor EMS1</i>
MEF2D_HUMAN	<i>Myocyte-specific enhancer factor 2D</i>	KDM1A_HUMAN	Lysine-specific histone demethylase 1A
NUP62_HUMAN	<i>Nuclear pore glycoprotein p62</i>	PR40A_HUMAN	Pre-mRNA-processing factor 40 homolog A
PCF11_HUMAN	Pre-mRNA cleavage complex 2 protein Pcf11	MINT_HUMAN	<i>Msx2-interacting protein</i>
TAF4_HUMAN	<i>Transcription initiation factor TFIID subunit 4</i>	TBL1R_HUMAN	F-box-like/WD repeat-containing protein TBL1XR1
LIN54_HUMAN	Protein lin-54 homolog	KDM3B_HUMAN	Lysine-specific demethylase 3B
UBN2_HUMAN	Ubinuclein-2	EP400_HUMAN	<i>E1A-binding protein p400</i>
P121C_HUMAN	Nuclear envelope pore membrane protein POM 121C	ILF2_HUMAN	Interleukin enhancer-binding factor 2
RESF1_HUMAN	Retroelement silencing factor 1	ZN281_HUMAN	Zinc finger protein 281
TASO2_HUMAN	Protein TASOR 2	NU214_HUMAN	<i>Nuclear pore complex protein Nup214</i>
DOT1L_HUMAN	Histone-lysine N-methyltransferase, H3 lysine-79 specific	DIDO1_HUMAN	<i>Death-inducer obliterator 1</i>
NUP58_HUMAN	Nucleoporin p58/p45	SMCA4_HUMAN	Transcription activator BRG1
ASXL2_HUMAN	<i>Putative Polycomb group protein ASXL2</i>	PRC2A_HUMAN	Protein PRRC2A
RBM33_HUMAN	RNA-binding protein 33	YLPM1_HUMAN	<i>YLP motif-containing protein 1</i>
SMRC1_HUMAN	SWI/SNF complex subunit SMARCC1	NUP88_HUMAN	Nuclear pore complex protein Nup88
ICE1_HUMAN	Little elongation complex subunit 1	ANKH1_HUMAN	Ankyrin repeat and KH domain-containing protein 1
PRC2B_HUMAN	<i>Protein PRRC2B</i>	WNK1_HUMAN	<i>Serine/threonine-protein kinase WNK1</i>
NUP54_HUMAN	Nucleoporin p54	TAF5_HUMAN	Transcription initiation factor TFIID subunit 5
SIN3A_HUMAN	Paired amphipathic helix protein Sin3a ¹⁴	PRC2C_HUMAN	<i>Protein PRRC2C</i>
ARI1B_HUMAN	AT-rich interactive domain-containing protein 1B	RBM25_HUMAN	RNA-binding protein 25
CBP_HUMAN	<i>CREB-binding protein</i>	RBP2_HUMAN	<i>E3 SUMO-protein ligase RanBP2</i>
SMRC2_HUMAN	SWI/SNF complex subunit SMARCC2	BPTF_HUMAN	Nucleosome-remodeling factor subunit BPTF
LRIF1_HUMAN	Ligand-dependent nuclear receptor-interacting factor 1	CPSF1_HUMAN	Cleavage and polyadenylation specificity factor subunit 1
KDM2B_HUMAN	Lysine-specific demethylase 2B	HNRPL_HUMAN	Heterogeneous nuclear ribonucleoprotein L
NCOA2_HUMAN	<i>Nuclear receptor coactivator 2</i>	RAGP1_HUMAN	Ran GTPase-activating protein 1
UBN1_HUMAN	Ubinuclein-1 ³⁴	UBP2L_HUMAN	<i>Ubiquitin-associated protein 2-like</i>
		ZN106_HUMAN	Zinc finger protein 106
		KNL1_HUMAN	Kinetochore scaffold 1

Legend:
 Black Accession: ID'd only in TPR
 Bold Accession: ID'd in both TPR and eGFP, p value <0.05
 Italicized Protein Name: Known O-GlcNAc modified
 Bold and Italicized Protein Name: Known OGT interactors (with citations)

Table 2: Reactome pathways enriched in TPR interactors list (FDR <5E-9)

Reactome Pathway	Identifier	Protein Number	Fold enrichment	P value	FDR
Chromatin organization	R-HSA-4839726	27	20.4	9.70E-27	1.11E-23
Chromatin modifying enzymes	R-HSA-3247509	27	20.4	9.70E-27	2.22E-23
SUMO E3 ligases SUMOylate target proteins	R-HSA-3108232	15	17.32	3.26E-14	1.86E-11
SUMOylation	R-HSA-2990846	15	16.69	5.42E-14	2.47E-11
Transcriptional regulation by RUNX1	R-HSA-8878171	16	14.22	7.48E-14	2.85E-11
Rev-mediated nuclear export of HIV RNA	R-HSA-165054	9	50.99	9.89E-13	3.23E-10
HCMV Early Events	R-HSA-9609690	12	21.76	1.15E-12	3.29E-10
Interactions of Rev with host cellular proteins	R-HSA-177243	9	47.99	1.58E-12	4.00E-10
Transport of Mature mRNA Derived from an Intronless Transcript	R-HSA-159231	9	41.84	4.58E-12	1.05E-09
Transport of Mature mRNAs Derived from Intronless Transcripts	R-HSA-159234	9	40.8	5.59E-12	1.16E-09
SUMOylation of DNA replication proteins	R-HSA-4615885	9	37.95	9.87E-12	1.61E-09
Processing of Capped Intron-Containing Pre-mRNA	R-HSA-72203	15	11.43	9.06E-12	1.72E-09
SUMOylation of RNA binding proteins	R-HSA-4570464	9	37.95	9.87E-12	1.73E-09
HCMV Infection	R-HSA-9609646	12	17.55	1.18E-11	1.80E-09
Transport of Ribonucleoproteins into the Host Nucleus	R-HSA-168271	8	50.02	2.16E-11	2.60E-09
NEP/NS2 Interacts with the Cellular Export Machinery	R-HSA-168333	8	50.02	2.16E-11	2.75E-09
Defective TPR may confer susceptibility towards thyroid papillary carcinoma (TPC)	R-HSA-5619107	8	50.02	2.16E-11	2.91E-09
Regulation of Glucokinase by Glucokinase Regulatory Protein	R-HSA-170822	8	50.02	2.16E-11	3.09E-09
Export of Viral Ribonucleoproteins from Nucleus	R-HSA-168274	8	48.35	2.73E-11	3.12E-09
Vpr-mediated nuclear import of PICs	R-HSA-180910	8	46.79	3.42E-11	3.39E-09
Late Phase of HIV Life Cycle	R-HSA-162599	12	16	3.23E-11	3.51E-09
Nuclear import of Rev protein	R-HSA-180746	8	46.79	3.42E-11	3.55E-09
SUMOylation of SUMOylation proteins	R-HSA-4085377	8	45.33	4.25E-11	3.89E-09
Transport of the SLBP independent Mature mRNA	R-HSA-159227	8	45.33	4.25E-11	4.05E-09
Nuclear Pore Complex (NPC) Disassembly	R-HSA-3301854	8	43.95	5.26E-11	4.29E-09
Transport of the SLBP Dependant Mature mRNA	R-HSA-159230	8	43.95	5.26E-11	4.45E-09
SUMOylation of chromatin organization proteins	R-HSA-4551638	9	30.79	5.20E-11	4.57E-09

Figure 1: BioID approach to define OGT TPR interactors in HeLa cells. **A**, Workflow of BioID process for identifying TPR interactors. **B**, Workflow of MS analysis and protein validation. Samples were prepped in biological triplicate. **C**, Western blot showing expression of TPR-BirA* and eGFP-BirA* fusion proteins (anti-FLAG tag), α -tubulin as loading control. 20ug/lane. **D**, ELISA quantification of FLAG-tagged fusion proteins (one representative replicate, A.U.=Arbitrary units) **E**, Representative western blot with anti-biotin antibody demonstrating enrichment of biotinylated proteins following biotin immunoprecipitation (representative blot from three replicates). Red box indicates lanes showing enrichment of biotinylated proteins by biotin IP. 10 μ g/lane. FT=flow-through. For elute and FT, 10ug/lane. For elute, 25% of total eluate loaded.

Figure 2: Proteins identified in both TPR-BirA* and eGFP-BirA* are enriched in TPR-BirA*. Fold enrichment values for all proteins identified in both TPR-BirA* and eGFP-BirA*. Fold enrichment values are average NSAF of TPR-BirA* over average NSAF of eGFP-BirA*. Note that 72 proteins were only observed in TPR-BirA*.

Figure 3: Reconstructed ion chromatograms confirm peptide-level enrichment of TPR-BirA* interactors. For **A-C**, fold increase is for the replicate shown. The average fold intensity for that peptide (averaged across all replicates and fractions in which that peptide appeared) is shown with the standard deviation. Average mass error is the absolute value of the mean across 3 replicates of both TPR-BirA* and eGFP-BirA*. NL=Normalization Level (Base Peak Intensity) **A**, Reconstructed ion chromatograms demonstrating greater enrichment of a catalytic OGT peptide in TPR-BirA* compared to eGFP-BirA*. **B**, Reconstructed ion chromatograms demonstrating greater enrichment of an HCF-1 peptide in TPR-BirA* compared to eGFP-BirA*. **C**, Reconstructed ion chromatograms demonstrating greater enrichment of a KNL1 peptide (novel interactor closest to cut-off for assignment) in TPR-BirA* compared to eGFP-BirA*.

Figure 4: TPR interactors are primarily nuclear localized. **A**, Subcellular fractionation of HeLa cells demonstrating localization of OGT (anti-OGT F12) and BirA* fusion proteins (anti-FLAG tag).

Cytoplasmic marker is GAPDH, nuclear marker is Histone H3. 10ug/lane, representative western blot of three biological replicates **B**, Ratios of nuclear to cytoplasmic expression of marker proteins (Nuclear: Histone H3, Cytoplasmic: GAPDH) and fusion proteins. Averaged across three biological replicates. **C**, UpsetR plot showing the subcellular localization of TPR interactors. **D**, Venn diagram showing the subcellular localization of TPR interactors. Numbers represent the total number of TPR interactors in that category. Localization determined using UniProt.

Figure 5: TPR interactors demonstrate enrichment in biological processes and disease states. **A**, CirGo plot showing enriched biological processes in the TPR protein interactor list **B**, CirGo plot showing enriched molecular functions in the TPR protein interactor list **C**, Venn diagram of PANTHER Pathway enrichment of TPR interactors. Numbers are the fold enrichment of the pathway process over expected enrichment.

Figure 6: OGT interacts with proteins involved in intellectual disability in HeLa cells. **A**, Venn diagram showing disease states enriched in the TPR interactors. **B**, UpsetR plot showing disease states enriched in the TPR interactors. Disease associations determined using the OMIM resource.

Figure 1

Figure 2

Figure 3

A OGT catalytic domain peptide: LVSIVADQLEK

Theoretical mass: 1214.6991 Da

Average Mass Error: Δ 2.17ppm

Fold increase: 11.7 (Average 9.1 +/- 2.3)

#53541-53786 RT 104.70-105.10 AV: 71
F: FTMS + p NSI Full ms [300.0000-2000.0000]

#51731-51978 RT 104.70-105.10 AV: 71
F: FTMS + p NSI Full ms [300.0000-2000.0000]

B HCF-1 peptide: IATGHGQQGVTVVLK

Theoretical mass: 1635.9177 Da

Average Mass Error: Δ 0.749ppm

Fold increase: 26.3 (Average 36.4 +/- 24.4)

#27673-27973 RT 59.00-59.80 AV: 173
F: FTMS + p NSI Full ms [300.0000-2000.0000]

#27508-27785 RT 59.00-59.80 AV: 179
F: FTMS + p NSI Full ms [300.0000-2000.0000]

C KNL1 peptide: AQVESCQLNNR (57.021 (C6))

Theoretical mass: 1318.6168 Da

Average Mass Error: Δ 2.12ppm

Fold increase: 3.08 (Average 3.93 +/- 1.93)

#13332-14343 RT 32.86-34.29 AV: 250
F: FTMS + p NSI Full ms [300.0000-2000.0000]

#13951-15007 RT 32.86-34.29 AV: 236
F: FTMS + p NSI Full ms [300.0000-2000.0000]

Figure 4

Figure 5

Figure 6

A

B