Identification and characterization of a missense mutation in the *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) transferase gene that segregates with X-linked intellectual disability

Received for publication, December 5, 2016, and in revised form, March 8, 2017 Published, Papers in Press, March 16, 2017, DOI 10.1074/jbc.M116.771030

Krithika Vaidyanathan[‡], Tejasvi Niranjan[§], Nithya Selvan[‡], Chin Fen Teo[‡], Melanie May[¶], Sneha Patel[‡], Brent Weatherly[‡], Cindy Skinner[¶], John Opitz[‖], John Carey[‖], David Viskochil[‖], Jozef Gecz[**], Marie Shaw[**], Yunhui Peng[‡‡], Emil Alexov[‡‡], Tao Wang[§], Charles Schwartz[¶1], and Lance Wells[‡2]

*From the [‡]Department of Biochemistry and Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, the [§]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21287, the [¶]Greenwood Genetic Center, Greenwood, South Carolina 29646, [‖]Pediatrics (Medical Genetics), Pediatric Pathology, Human Genetics, Obstetrics, and Gynecology, University of Utah School of Medicine, Salt Lake City, Utah 84132, the **Department of Paediatrics and Robinson Research Institute, University of Adelaide, Adelaide, South Australia 5006, Australia, and the [‡‡]Department of Computational Biophysics and Bioinformatics, Clemson University, Clemson, South Carolina 29634*

Edited by Eric R. Fearon

O*-GlcNAc is a regulatory post-translational modification of nucleocytoplasmic proteins that has been implicated in multiple biological processes, including transcription. In humans, single genes encode enzymes for its attachment (*O*-GlcNAc transferase (OGT)) and removal (*O*-GlcNAcase (OGA)). An X-chromosome exome screen identified a missense mutation, which encodes an amino acid in the tetratricopeptide repeat, in *OGT* (759G>T (p.L254F)) that segregates with X-linked intellectual disability (XLID) in an affected family. A decrease in steady-state OGT protein levels was observed in isolated lymphoblastoid cell lines from affected individuals, consistent with molecular modeling experiments. Recombinant expression of L254F-OGT demonstrated that the enzyme is active as both a glycosyltransferase and an HCF-1 protease. Despite the reduction in OGT levels seen in the L254F-OGT individual cells, we observed that steady-state global *O*-GlcNAc levels remained grossly unaltered. Surprisingly, lymphoblastoids from affected individuals displayed a marked decrease in steady-state OGA protein and mRNA levels. We observed an enrichment of the OGT-containing transcriptional repressor complex mSin3A-HDAC1 at the proximal promoter region of *OGA* and correspondingly decreased *OGA* promoter activity in affected cells. Global transcriptome analysis of L254F-OGT lymphoblastoids compared with controls revealed a small subset of genes that are differentially expressed. Thus, we have begun to unravel the molecular consequences of the 759G>T (p.L254F) mutation in *OGT* that uncovered a compensation mechanism, albeit imperfect, given the phenotype of affected individuals, to maintain steady-state *O*-GlcNAc levels. Thus, a single amino acid substitution in the regulatory domain (the tetratricopeptide repeat domain) of OGT, which catalyzes the *O*-GlcNAc post-translational modification of nuclear and cytosolic proteins, appears causal for XLID.

1–3% of the world population is affected by intellectual disability (ID)[3] (1–4). ID is a leading problem of socio-economic health care in Western countries as reported by the Centers for Disease Control and Prevention (5, 6), due to the lifetime support required by the affected individuals. ID is characterized by an intelligence quotient of 70 or lower. In addition to ID, affected individuals exhibit two or more behavioral deficits in terms of social, conceptual, or practical adaptation (3, 7). A variety of factors cause ID, and they can range from malnutrition during pregnancy to chromosomal abnormalities (8). Recently, deletions, duplications and missense mutations have been the focus in the field, given the advent of the human genome (9) as well as the subsequent explosion in sequencing technologies. Monogenic causes have been mainly attributed to genes found on the X-chromosome. 5–10% of ID in males is inherited in an X-linked pattern (10), and epidemiological studies consistently show a 30–50% excess of males over females diagnosed with ID. To date, mutations in at least 102 genes result in 81 of the known 160 X-linked intellectual disability (XLID) syndromes (10). However, the cause of ID in nearly 50% of males with XLID remains unknown (10).

Nuclear and cytoplasmic proteins can be modified with *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) on the hydroxyl

This work was supported in part by a grant from the W.M. Keck Foundation (to L. W.); NIGMS, National Institutes of Health (NIH) Grant P41GM103490 (to L. W.); NINDS, NIH, Grant RO1NS073854 (to C. E. S. and T. W.); and a grant from the South Carolina Department of Disabilities and Special Needs (to C. E. S.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article contains a supplemental R program and supplemental Table 1.

The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE74263.

[1] To whom correspondence may be addressed. E-mail: ceschwartz@ggc.org.
[2] To whom correspondence may be addressed. E-mail: lwells@ccrc.uga.edu.

[3] The abbreviations used are: ID, intellectual disability; CADD, combined annotation-dependent depletion; PUGNAc, *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene) amino *N*-phenyl carbamate; GlcNAc, β-*N*-acetylglucosamine; OGT, *O*-GlcNAc transferase; OGA, *O*-GlcNAcase; TPR, tetratricopeptide repeat; XLID, X-linked intellectual disability; PDB, Protein Data Bank; qPCR, quantitative PCR.

groups of serines or threonines to influence various intracellular processes (11–16). The *O*-GlcNAc modification is a dynamic and inducible post-translational modification that cycles on and off intracellular proteins in response to the cellular environment similar to protein phosphorylation (17–19). In mammals, individual genes encode for the enzymes responsible for the addition and removal of *O*-GlcNAc, *O*-GlcNAc transferase (OGT), and *O*-GlcNAcase (OGA), respectively (20–22). The sugar nucleotide UDP-GlcNAc, which is the end product of the hexosamine biosynthetic pathway and whose cellular concentrations are responsive to nutritional status, is the donor substrate for OGT (23–25). *O*-GlcNAc modification of transcription factors, chromatin remodeling enzymes, and histones has established this modification as a regulator of gene expression (26). The *O*-GlcNAc modification has also been associated with several human diseases that involve changes in metabolism, including cancer, diabetes, and Alzheimer's disease (25, 27–30).

The *O*-GlcNAc modification of proteins is regulated by the opposing actions of OGT and OGA (11–16). OGT is encoded by a single gene (*OGT*) and maps to chromosome Xq13.1 (31). *OGT* was cloned and partially characterized in 1997 (20, 32, 33) and contains a C-terminal catalytic domain (20, 32) and an N-terminal tetratricopeptide repeat (TPR) domain that varies in length (33, 34). This TPR domain is thought to regulate protein-protein interactions and to bring the catalytic domain into proximity to its protein substrates (35, 36). Murine *ogt* knockouts are embryonic lethal, demonstrating its requirement for cell survival (31). OGA is encoded by a single gene (*MGEA5*) located on chromosome 10q24.1–24.3 (37) and was cloned (38) and partially characterized in 2002 (39). OGA has a catalytic N-terminal *O*-GlcNAcase domain and a C-terminal domain with low sequence identity to histone acetyltransferase that lacks its enzymatic activity (40). *Oga* homozygous null mice are perinatal lethal (41).

In this study, we have characterized a missense mutation in *OGT* (759G>T (p.L254F)) that segregates with XLID in a family. This mutation results in an amino acid substitution in the TPR domain of OGT (p.L254F). We demonstrate that the protein is active but unstable. Surprisingly, *O*-GlcNAc levels remain unaltered in the L254F-OGT-expressing cells. Reduced OGA protein levels compensate for lowered OGT protein levels in the cells derived from XLID individuals. OGA protein levels are decreased by means of transcriptional regulation, because both *OGA* mRNA and promoter reporter expression are decreased in XLID lymphoblastoid cells. OGT, in addition to *O*-GlcNAc-modified proteins, mSin3A and HDAC1, are enriched at the *OGA* promoter in L254F-OGT individual cells. These results suggest that L254F-OGT regulates *OGA* gene expression in XLID lymphoblastoids in a co-repressor fashion. In parallel, RNA deep sequencing has revealed a small subset of genes that are regulated in a disorder-specific context, including several chromatin components. Taken together, these data begin to delineate the molecular impact of the L254F substitution on OGT function and uncover a cellular compensation mechanism, albeit imperfect, for maintaining global *O*-GlcNAc levels.

Results

Mutations in OGT segregate with XLID

Exome sequencing of multiple family members (family K9427), revealed a mutation (759G>T (p.L254F)) encoding an amino acid in the seventh TPR of OGT perfectly segregating with XLID (Fig. 1*A*). A total of 26 single nucleotide substitutions in 22 annotated genes were identified after filtering by affected kindred/cross-cohort analysis. Among the 26 variants, 14 were located in deep introns far away from conserved splicing consensus sequence, and seven were located within 3′-UTR regions. Among the five variants located within the coding regions, two resulted in no change in amino acids, and three resulted in missense variants. Among the three missense variants, one, GABRE-p.R87C, did not segregate with ID phenotype in the proband's family. A second missense change, p.R175C, in PRICKLE3 did segregate and had a CADD score of 15.9, which could be indicative of pathogenesis. However, the gene does not appear to be expressed in the brain based on data in UNIGENE and the lack of amplification using a multiple tissue cDNA panel (data not shown). However, the OGT variant, p.L254F, did segregate with the ID phenotype in family K9427, and bioinformatic analysis (CADD score of 26.5) as well as the presence of expression in the brain made this variant the only likely candidate for the phenotype in K9427, and therefore, it was further pursued. Three family members displayed ID (Fig. 1*A* and Table 1). The affected males (II-4, III-4, and IV-1; Fig. 1) spanned three generations and also presented with a small head circumference, 5th finger clinodactyly, and genital abnormalities. Other minor findings were also noted (Table 1). Lymphoblastoids from affected individuals and controls (II-1, -3, -7, and -8; Fig. 1) were isolated for further studies. Corroborating the prediction of causality was the fact that all examined carrier females exhibited highly skewed X-inactivation (98–100% in patients II-5, -6, and -8 and III-3; Fig. 1), as has been previously seen in many families with XLID (42).

L254F-OGT protein is unstable

We first modeled the impact of the L254F substitution on the known structure of the human OGT TPR repeats. The known structure of the OGT TPR repeats (PDB code 1W3B) was used to calculate $\Delta\Delta G$ for the L254F variant. The $\Delta\Delta G$ predictions, conducted using seven different web services, for the variant predicted an average loss in stability of 0.72 kcal/mol, with six of the seven programs predicting that the L254F variant was destabilizing (Table 2). The average predicted $\Delta\Delta G$ upon substitution is relatively small, indicating that the single amino acid replacement probably does not grossly affect overall protein stability. Consistent with this finding, we visualized the structures of wild-type and L254F variant OGT TPR domain with UCSF Chimera, and the substitution sites are marked in *red* (Fig. 1*B*). The side chain conformations of the residues within 3 Å of the wild-type position or variant position are shown with a *stick representation*. The amino acid replacement is located in a helix region, but it is also close to a tight turn connecting two helices. The wild-type Leu residue is almost totally buried with low accessible surface area. Thus, the substitution with Phe, which is a bulkier residue, would be difficult to accom-

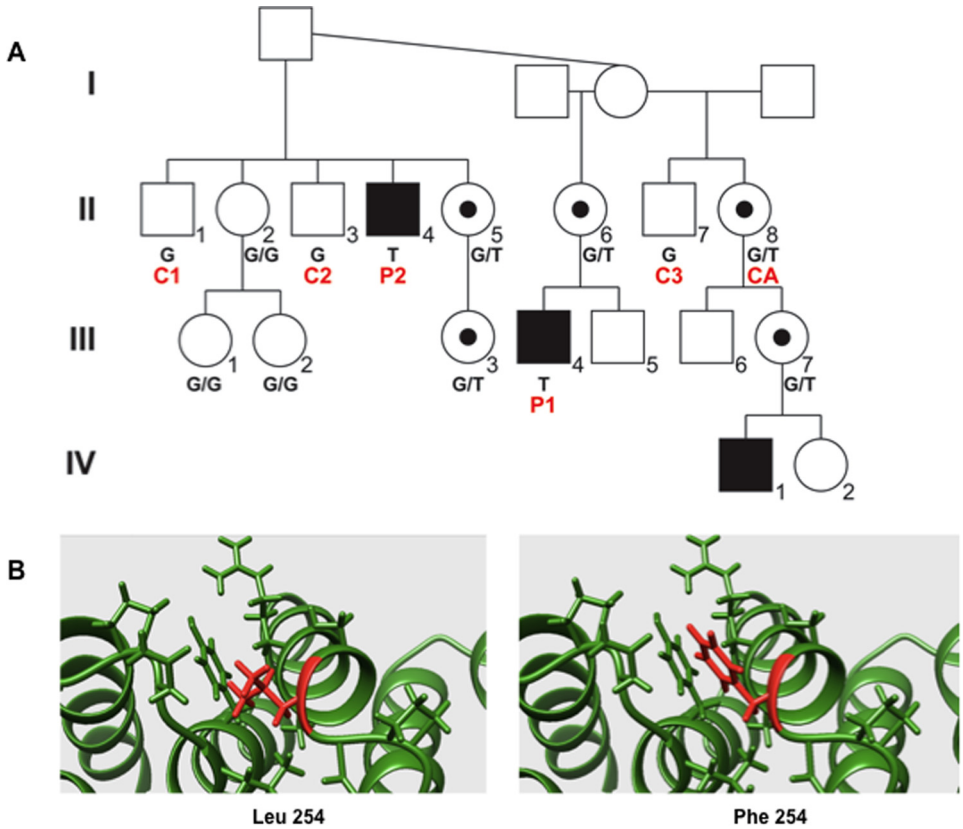


Figure 1. A mutation, predicted to destabilize the protein, in *OGT* segregates with XLID. *A*, partial pedigree of L254F family (K9427). Segregation of the mutation, c.759G>T, in OGT in family K9427. Genotype is given *below* the pedigree symbol. *Red letters below symbols* indicate control cell line (C1–C3), carrier female cell line (CA), or affected cell line (P1 and P2). *Black squares* indicate affected males, whereas a *black dot inside* the *circle* indicates a confirmed carrier female with skewed X-inactivation. *B*, modeling of the side chain conformation of the L254F variant in the OGT protein predicts loss in stability.

Table 1
Clinical features of the L254F family (K9427)

Clinical features	Patient		
	II-4 (P2)	III-4 (P1)	IV-1
Age (years)	56.5	14.7	6.7
Height (cm)	159 (< 3rd percentile)	154 (10th–15th percentile)	114 (10th–25th percentile)
Weight (pounds)	153.4 (<25th percentile)	74.4 (<3rd percentile)	45.5 (50th percentile)
Head circumference	55.2 (15th percentile)	52.4 (3rd percentile)	51.3 (25th percentile)
Scalp	Male pattern balding	Bilateral frontal upsweeps	2 hair whorls, cowlicks in front
Facial	Thin upper lip, full philtrum	Mild synophrys, open mouth	Increased size of lips, open mouth
Hands	Minimal 5th finger clinodactyly	5th finger clinodactyly	Short 5th finger clinodactyly
Genital	Small testis	Mild first-degree hypospadias	Repaired hypospadias
Intelligence quotient	49	61	58

Table 2
Predicted protein folding free energy changes upon missense mutation in OGT

$\Delta\Delta G < 0$ indicates mutation destabilizing the protein, whereas $\Delta\Delta G > 0$ indicates stabilization. Six of the seven servers as well as the average $\Delta\Delta G$ suggest that L254F is destabilizing.

Web server	L254F[a] $\Delta\Delta G$
	kcal/mol
I-MUTANT	−1.55
PopMuSiC	−0.21
DUET	−1.12
mCSM	−1.01
SDM	−0.29
Eris	−2.11
SAFFEC	1.25
AVERAGE	−0.72

[a] PDB ID 1W3B.

modate by the protein. This is consistent with the folding free energy predictions that L254F is a destabilizing variant of OGT.

Because modeling predicted that the L254F-OGT protein is partially unstable, we examined steady-state levels of OGT in both affected and control lymphoblastoids. Immunoblotting with multiple antibodies to OGT showed that affected males P1 (III-4) and P2 (II-4) exhibited significantly lower OGT steady-state levels (Fig. 2*A*), compared with a heterozygous female carrier CA (II-8) as well as unaffected control males C1 (II-1), C2 (II-3), and C3 (II-7) (Fig. 2*B*). We used two different loading controls, β-actin and α-tubulin, to account for any possible changes in housekeeping proteins. We next investigated the half-life of L254F-OGT to explain the decrease in steady-state levels observed in affected samples. Following blocking of translation, the L254F-OGT in affected P1 turned over faster compared with control C1 (Fig. 2*C*). The half-life of WT OGT in normal control C1 was 13 h, compared with the L254F-OGT in affected male P1, whose half-life was 5.5 h (Fig. 2*C*). β-Actin

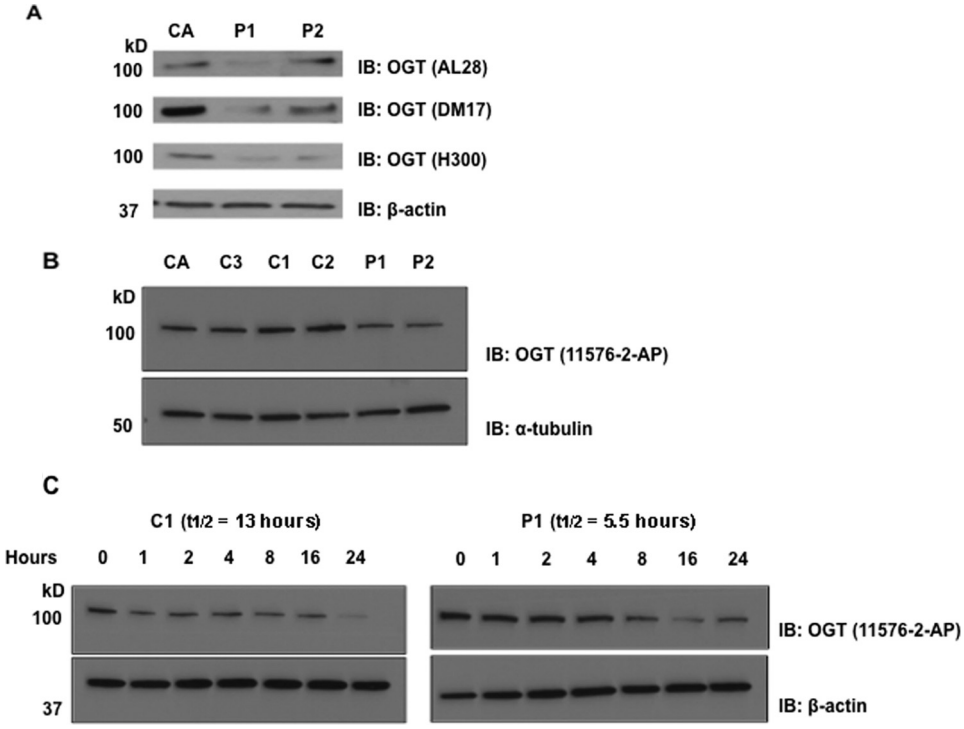


Figure 2. L254F-OGT has reduced protein levels and half-life. *A*, immunoblotting (*IB*) of equal amounts of crude lysates displays a decrease in steady-state OGT levels in XLID lymphoblastoids (*P1* and *P2*) when compared with a control female carrier (*CA*) using three independent antibodies to OGT. β-Actin immunoblotting was performed to confirm equal loading. *B*, immunoblotting of equal amounts of crude lysates from XLID lymphoblastoids (P1 and P2) displays less OGT protein when compared with control male relatives (C1–C3) or the control female carrier (CA) using a fourth α-OGT (11576-2-AP) antibody. α-Tubulin immunoblotting was performed to confirm equal loading. *C*, OGT has a reduced half-life in an XLID lymphoblastoid (P1) compared with control (C1) as measured by immunoblotting over time following blocking of translation with cycloheximide. For β-actin, which has a half-life in excess of 48 h, immunoblotting was performed to confirm equal loading. Blots shown are representative of results from three independent replicates.

was used as the loading control due to its extended half-life of 2–3 days (43). Thus, L254F-OGT appears to be a slightly unstable protein.

L254F-OGT protein is an active enzyme

We next wanted to explore whether the OGT variant protein had enzymatic activity. When recombinant L254F-OGT is exogenously introduced into HEK293F cells via transient transfection, it is expressed and capable of elevating *O*-GlcNAc levels on nucleocytoplasmic proteins comparable with WT OGT overexpression (Fig. 3*A*). Immunopurified recombinant wild-type and L254F variant proteins from HEK293F cells were also catalytically competent using a CK2α-derived peptide as an acceptor (data not shown). We also evaluated purified, recombinantly expressed WT and L254F variant proteins from an *E. coli* expression system, as well as a catalytically inactive variant (K852M) (44), for activity toward recombinantly expressed CK2α protein and saw no difference in the ability of wild type and L254F to glycosylate CK2α protein *in vitro* (Fig. 3B). Because OGT has also been demonstrated to have catalytic protease and glycosylation activity toward HCF-1 (45–47), we examined these activities for the wild type and L254F variant but found no differences between the WT and variant enzyme (Fig. 3*C*). Thus, L254F-OGT appears to be an active enzyme.

Steady-state O-GlcNAc levels remain unaltered in XLID lymphoblastoids

Considering that steady-state levels of OGT are decreased in L254F-OGT-expressing lymphoblastoids, we expected to observe a global decrease in *O*-GlcNAc levels. Surprisingly, global *O*-GlcNAc levels, as detected by immunoblotting with anti-*O*-GlcNAc antibodies, remained unaltered in affected males (P1 and P2) relative to both female carrier (CA) and unaffected male relatives (C1, C2, and C3) (Fig. 4). We probed for *O*-GlcNAc-modified proteins using multiple anti *O*-GlcNAc antibodies (110.6 and mAb14 antibodies (shown in Fig. 4) and mAb10, mAb3, and RL-2 (data not shown)). Thus, steady-state *O*-GlcNAc levels remain unaltered in the L254F-OGT-containing lymphoblastoids compared with controls.

Steady-state OGA protein and mRNA levels and OGA promoter expression are decreased in L254F-OGT XLID lymphoblastoids

As a result of global *O*-GlcNAc persisting at normal levels, we examined the other cycling enzyme of this dynamic modification in the L254F-OGT individual cells. Interestingly, OGA steady-state protein levels were severely diminished in affected individuals (P1 and P2) when compared with a female carrier (CA) and unaffected male relatives (Fig. 5, *A* and *B*). Thus, we have apparently uncovered a compensation mechanism that is in play in affected lymphoblastoids that accounts for the lack of change in *O*-GlcNAc levels in the XLID cell lines with reduced OGT levels.

To investigate this reduction in OGA, we investigated whether the decrease in OGA steady-state levels was due to a decrease in transcription. We performed quantitative RT-PCR to assess mRNA levels of *OGA*. XLID lymphoblastoids exhibited a significant decrease in *OGA* mRNA (Fig. 5*C*). We also performed cycloheximide half-life assessment and ruled out the

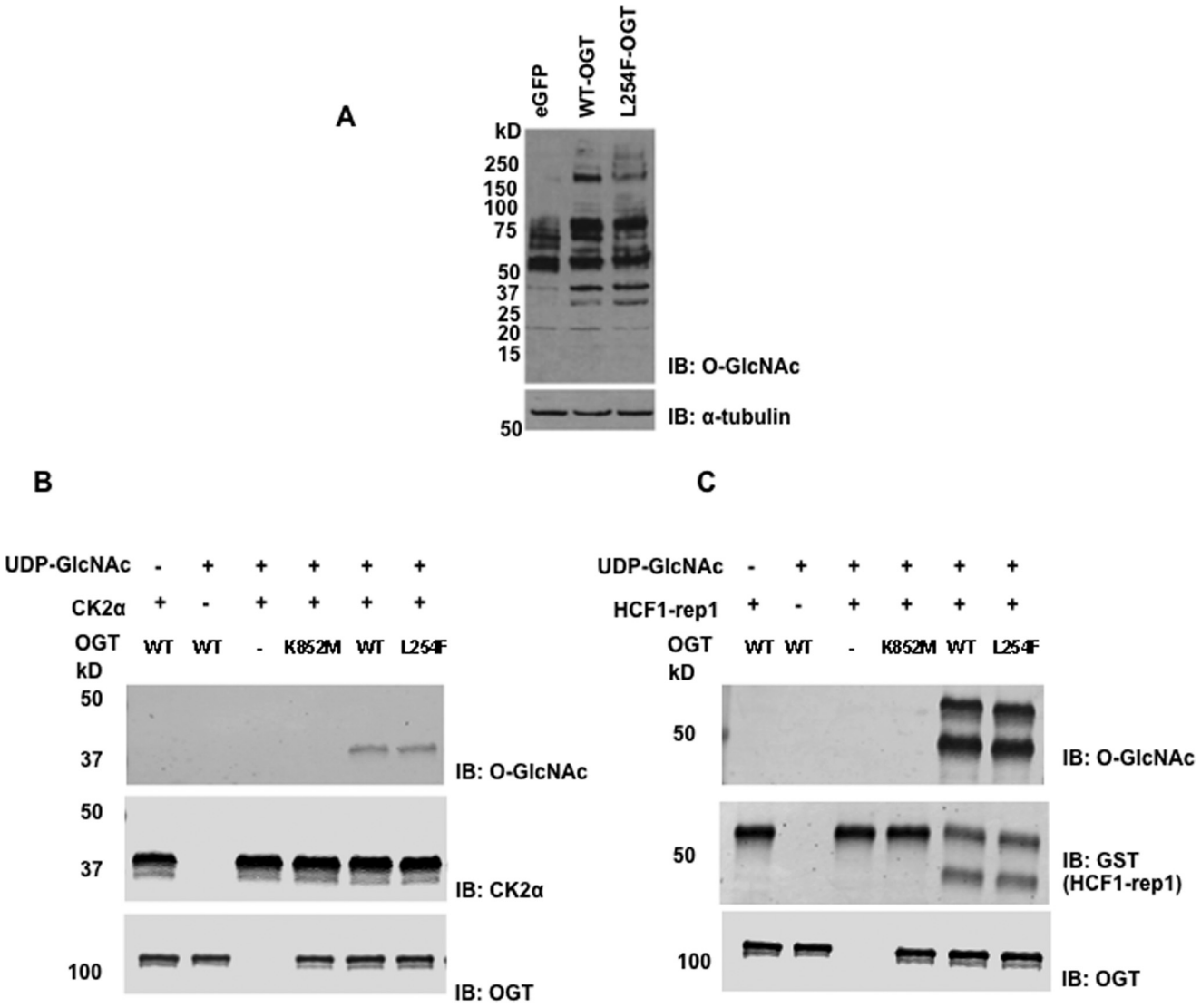


Figure 3. L254-OGT is an active glycosyltransferase and HCF1-protease. *A*, *O*-GlcNAc levels are elevated when WT or L254F-OGT is transiently overexpressed, compared with control (enhanced GFP (*eGFP*)), in HEK293F cells as measured by immunoblotting (*IB*) with a pan-*O*-GlcNAc antibody (110.6). *B*, purified WT and L254F-OGT are capable of glycoslyating purified CK2α. *C*, purified WT and L254F-OGT enzymes can glycosylate and cleave HCF1-rep1, whereas the catalytically compromised K852M-OGT can catalyze neither reaction. Blots shown are representative of results from at least three independent replicates.

possibility that lowered OGA protein levels might be due to protein turnover (data not shown), because there was no change in protein stability during the 24-h treatment. To further strengthen our hypothesis that *OGA* is transcriptionally down-regulated, we transfected a 2-kb proximal promoter region of *OGA* tagged to a luciferase reporter into the XLID and control lymphoblastoids. After 48 h of transfection, we observed that there was significantly lower expression from the *OGA* promoter in the affected lymphoblastoids compared with the control (Fig. 5*D*). These results suggest that *OGA* is transcriptionally regulated by OGT.

OGT, O-GlcNAc-modified proteins, and components of a chromatin silencing complex are enriched at the OGA promoter in XLID cells

The OGA down-regulation in a transcription-dependent manner led us to further probe the *OGA* promoter region. ChIP was performed using OGT- and *O*-GlcNAc-specific antibodies and qPCR with primers designed to amplify the proximal promoter region of *OGA*. OGT was significantly enriched, whereas *O*-GlcNAc-modified proteins displayed a similar trend of enrichment at the XLID OGA promoter (Fig. 6, *A* and *B*). These results suggest that whereas there is less OGT protein in the whole cell, there is more OGT at the *OGA* promoter in the XLID lymphoblastoids. It has been established previously that OGT exists in co-repressor complexes that down-regulate gene expression (48). This led us to hypothesize that OGT might exist in a similar co-repressor complex at the *OGA* promoter. Thus, we tested for mSin3A and HDAC1 enrichment at the *OGA* promoter by performing ChIP using the same proximal promoter primers to *OGA* as described above. We observed an increase in both HDAC1 and mSin3A at the *OGA* promoter in affected cells compared with controls (Fig. 6, *C* and *D*). It has previously been demonstrated that both mSin3A and HDAC1 are substrates of OGT and that OGT activity is required for maximal gene silencing (48). These results suggest that the

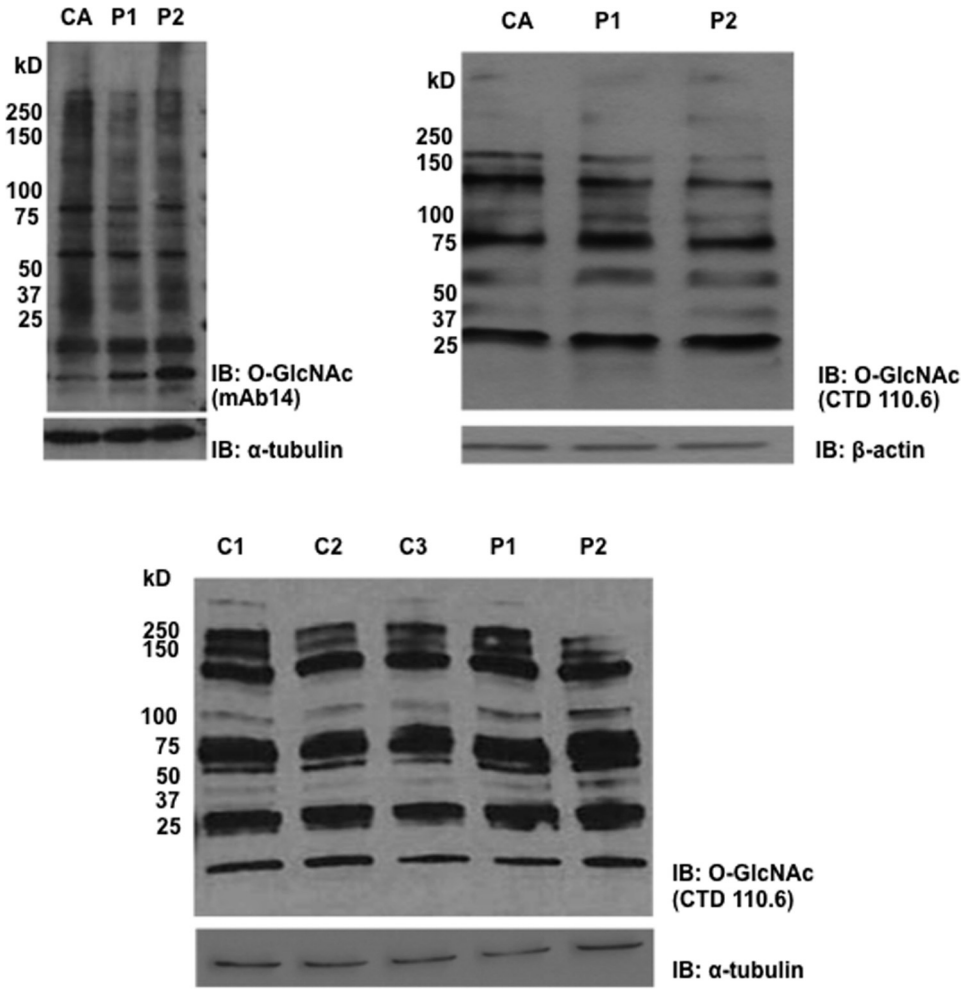


Figure 4. Global *O*-GlcNAc levels remain unaltered in XLID lymphoblastoids. Immunoblotting (*IB*) of equal amounts of crude lysates with pan-*O*-GlcNAc antibodies (mAb14 and 110.6) from carrier (CA)-affected L254F-OGT (P1 and P2) and unaffected (C1–C3) males detects no significant changes. Blots shown are representative of results from at least three independent replicates.

L254F-OGT as part of the mSin3A-HDAC1 complex compared with wild-type OGT acts preferentially at the *OGA* promoter to down-regulate *OGA* transcription.

Subsets of differentially expressed genes segregate by XLID as revealed by global transcriptome analysis

Due to the previously fortified role of OGT in transcription and our combined knowledge of the transcriptional down-regulation of OGA in the L254F-OGT-expressing cells, we elected to perform Illumina Hiseq 2000 RNA sequencing (Illumina, Inc., San Diego, CA) to see whether genes beyond *OGA* were being differentially regulated by the L254F-OGT. We chose to compare two unaffected males (C1 and C2) with the two affected males (P1 and P2) in the L254F-OGT XLID family. Three of the four males (C1, C2, and P2) are brothers, whereas the other affected male (P1) is a nephew. Using R software (for the algorithm, see the supplemental material) to perform a Spearman correlation (49), we compared each set of data with all of the other sets to determine which data sets showed the highest degree of similarity. We demonstrated that transcript expression segregates tighter with disease than generation (Fig. 7*A*). A plethora of genes involved in various cellular processes exhibited a differential expression pattern in the affected cells compared with the controls. Following the stringency applied to our data set described under "Methods," we were able to quantify 8800 genes (see supplemental Table 1). The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus (50). When we applied -fold filters, there were 349 genes that were differentially expressed 2-fold, 89 genes that changed 3-fold, and 38 genes that changed 4-fold (Fig. 7*B*). Given that our data set is from only four males, we generated two mock data sets by taking averages of groupings of our samples not based on disease (see "Experimental procedures" and supplemental Table 1). When we compared the mock set with the set sorted by XLID diagnosis, we observed a 1.3-fold enrichment of disease over natural variation in the human subjects for all 8800 genes. Of note, there was a 3.9-fold enrichment of disease over natural variation in the list of genes changing 3-fold in expression (Fig. 7, *B* and *D*, and Tables 3 and 4). In total, we saw ~1% of all quantified genes being differentially regulated in the affected *versus* the controls in the 3-fold group (Fig. 7*C*). Of the 89 genes in the 3-fold group, 67% of the genes were up-regulated in disease, whereas 33% were down-regulated (Fig. 7*C* and Tables 3 and 4). We orthogonally validated *HIST1H4B* (histone H4) to be up-regulated and *HIST1H3A* (histone H3.1) to be down-regulated in the XLID

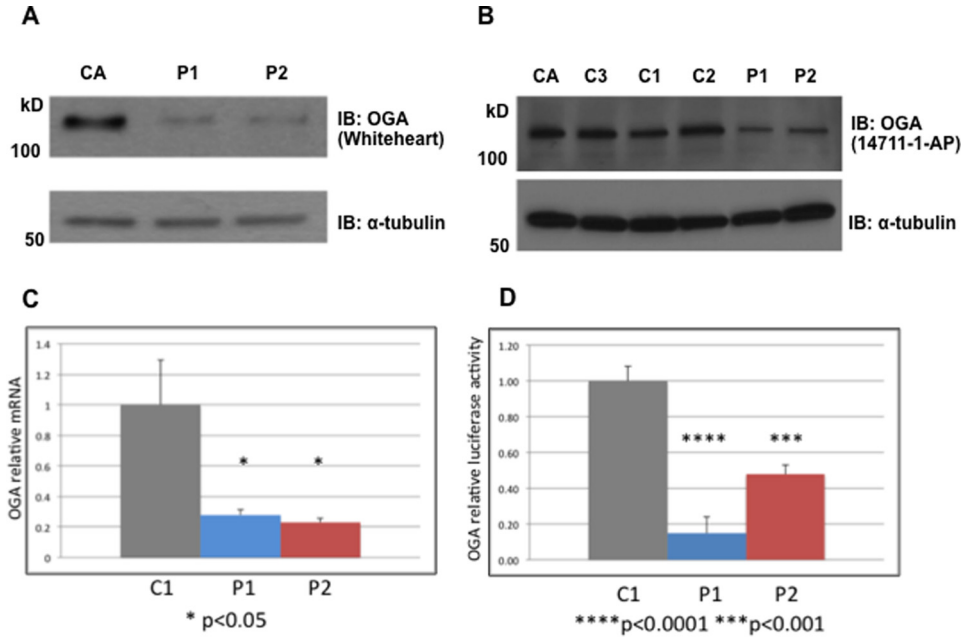


Figure 5. OGA protein, mRNA, and promoter activity are reduced in XLID lymphoblastoids. *A*, immunoblotting (*IB*) of equal amounts of crude lysates displays a decrease in steady-state OGA levels in XLID lymphoblastoids (P1 and P2) when compared with a control female carrier (CA) using an α-OGA antibody (Whiteheart Ab). *B*, immunoblotting of equal amounts of crude lysates from XLID lymphoblastoids (P1 and P2) displays less OGA protein when compared with control related males (C1–C3) using an α-OGA antibody (14711-1-AP). Blots shown in *A* and *B* are representative of results from at least three independent replicates. *C*, steady-state *OGA* mRNA levels assayed by quantitative RT-PCR show decreased transcript levels in XLID lymphoblastoids (P1 and P2) compared with a control male, C1 ($n = 3$, $p < 0.05$). *D*, XLID lymphoblastoids (P1 and P2) displayed less luciferase activity compared with a control male (C1) when transfected with a plasmid containing 2 kb of the proximal promoter region of *OGA* driving luciferase expression ($n = 3$, $p \leq 0.001$).

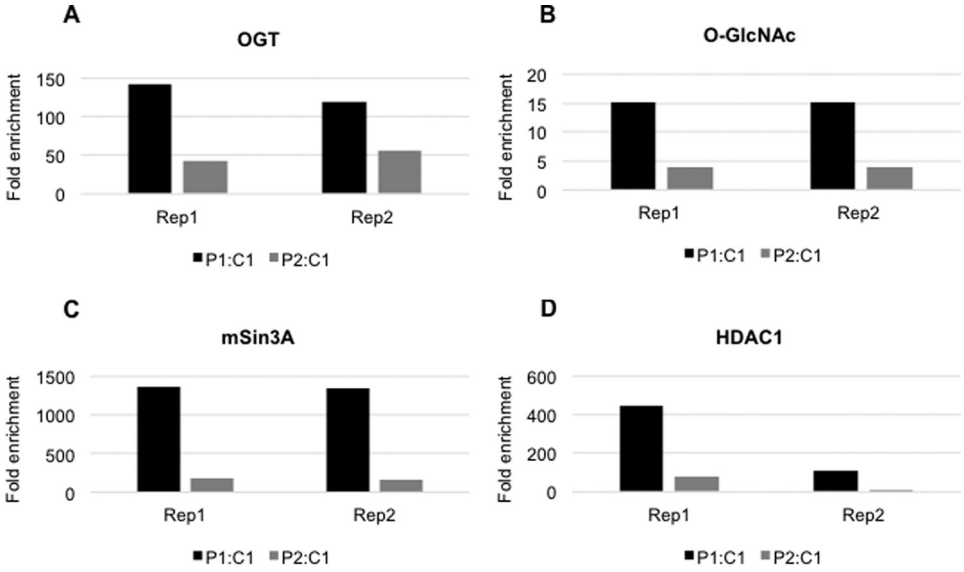
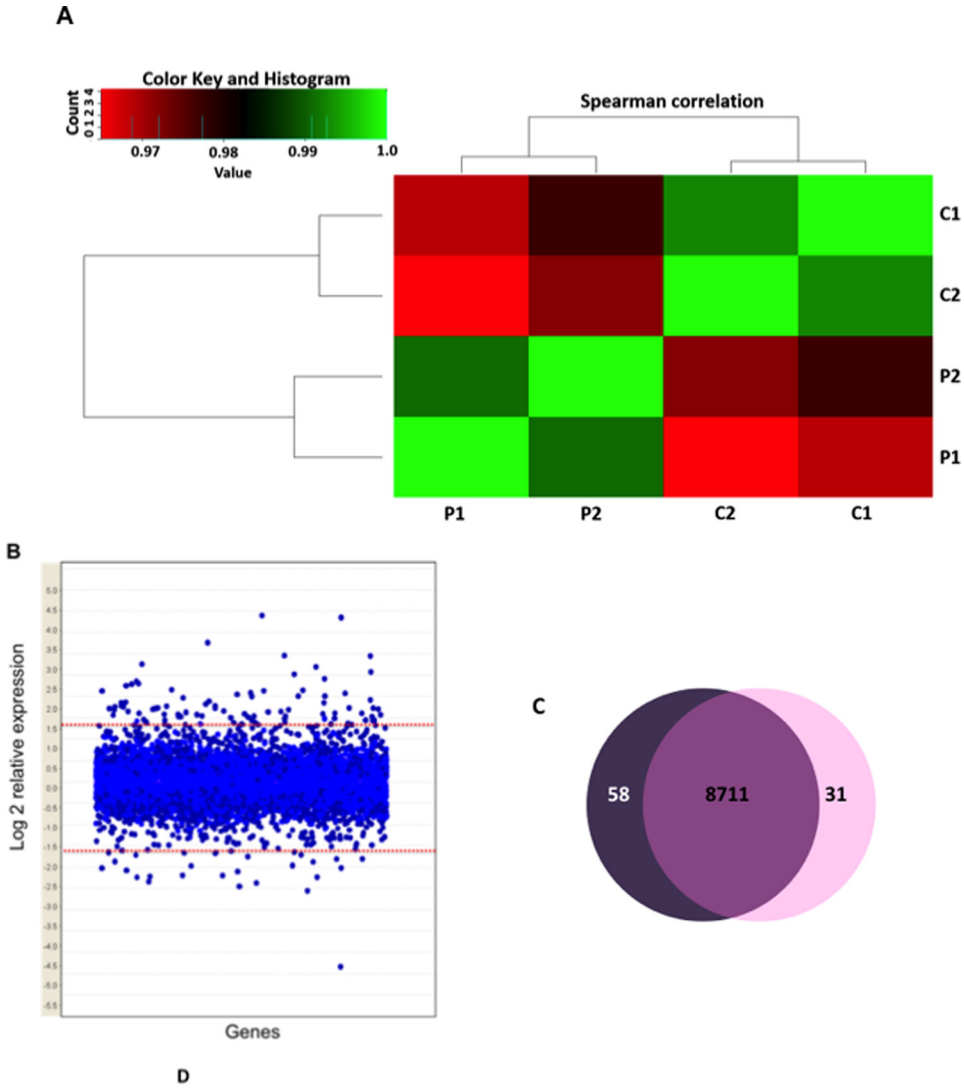


Figure 6. OGT-mSin3A-HDAC1 complex is enriched at the *OGA* promoter in XLID lymphoblastoids. ChIP analyses of the *OGA* proximal promoter following enrichment with antibodies to OGT (*A*), *O*-GlcNAc (*B*), mSin3A (*C*), and HDAC1 (*D*) show enrichment in XLID lymphoblastoids (P1 and P2) compared with a control male (C1). All values were determined using qPCR relative to the percentage of input and are presented as -fold enrichment on the *y* axis with different replicates on the *x* axis.

samples, as seen in the transcriptome data (data not shown). This analysis, in conjunction with the previously described role of OGT in transcriptional regulation under various settings, demonstrates that OGT influences the expression of a subset of genes in the L254F-OGT XLID cells.

Discussion

Although both XLID and O-GlcNAc have been studied for several decades, we have for the first time characterized a mutation in *OGT* that segregates with XLID. *O*-GlcNAc modification of almost 2000 proteins impacts several crucial cellular processes, and there is only one enzyme (OGT) for its addition (51). A novel mutation has been identified in *OGT* by a focused X-chromosome exome sequencing in XLID-affected individuals. The observation that a missense mutation in *OGT* segregates with disease is the first example of the OGT enzyme being directly linked to a disease. OGT is highly expressed in the brain (20), and a recent study has demonstrated that placental OGT is



D

Total no of genes	2-fold	3-fold	4-fold
8800	349	89	38
Disease/natural variation	4.8 fold	3.9 fold	3.3 fold

Figure 7. Gene expression is differentially regulated in XLID lymphoblastoids. *A*, Spearman correlation analysis shows segregation of RNA-seq data by disease (P1 and P2 expression data are more similar to one another than to either C, and this is also true for C1 and C2). *B*, the vast majority of transcripts change by $<$ 3-fold ($\log_2 = 1.58$). *C*, approximately 1% of quantifiable transcripts are altered 3-fold in XLID lymphoblastoids (P1 and P2) compared with control males (C1 and C2). *D*, number of quantifiable genes under differential -fold expression sets by disease *versus* natural variation.

crucial for hypothalamic gene expression in the mouse neonatal brain (52).

The X-chromosome only contains 4% of all human genes; however, 10% of all Mendelian diseases are assigned to the X-chromosome (53). X-linked inheritance pattern was observed in ID as early as the 1930s (54), and it was only in the 1980s that the first successes were seen in gene mapping and causal gene identification (55–57). Subsequent studies have consistently illustrated a 30% excess in males with XLID (10). 1:600 to 1:1000 males of the total world population are afflicted with XLID (1, 10), and only 50% of the cases have been assigned a defective gene (10). The remaining cases are of unknown etiology.

In this study, we have identified a mutation, resulting in an amino acid replacement in the TPR domain region of the encoded enzyme, in *OGT* associated with XLID in a family (Fig. 1 and Table 1). For L254F-OGT, we demonstrated that OGT steady-state protein levels are decreased in lymphoblastoids from affected individuals as a result of reduced half-life (Fig. 2). This finding validates the bioinformatic predictions of instability for this variant of OGT (Table 2). Complete loss of OGT is lethal in mammals (31). Therefore, we moved forward with the knowledge that OGT would probably be active in cells derived from affected individuals and validated that the variant is indeed catalytically competent (Fig. 3). Given that the mutation resulted in a variant in the regulatory protein-protein interac-

Table 3

Genes up-regulated 3-fold in affected cells (P1 and P2) compared with control cells (C1 and C2)

Sequence ID	HGNC[a] ID	HGNC symbol	Sequence name	FC[b]	Log_2(FC)	*p*
ENSG00000152952	HGNC:9082	*PLOD2*	Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2 (HGNC; Acc:9082)	14.93	3.9	0.0019
ENSG00000080854	HGNC:32326	*IGSF9B*	Immunoglobulin superfamily, member 9B (HGNC; Acc:32326)	14.42	3.85	0.0035
ENSG00000173110	HGNC:5239	*HSPA6*	Heat-shock 70-kDa protein 6 (HSP70B′) (HGNC; Acc:5239)	9.58	3.26	0.0045
ENSG00000163207	HGNC:6187	*IVL*	Involucrin (HGNC; Acc:6187)	7.73	2.95	0.0116
ENSG00000003096	HGNC:22931	*KLHL13*	Kelch-like 13 (*Drosophila*) (HGNC; Acc:22931)	7.73	2.95	0.0015
ENSG00000168785	HGNC:17753	*TSPAN5*	Tetraspanin 5 (HGNC; Acc:17753)	6.77	2.76	0.0014
ENSG00000070404	HGNC:3973	*FSTL3*	Follistatin-like 3 (secreted glycoprotein) (HGNC Acc:3973)	6.45	2.69	0.0046
ENSG00000137203	HGNC:11742	*TFAP2A*	Transcription factor AP-2 α (activating enhancer-binding protein 2 α) (HGNC; Acc:11742)	5.94	2.57	0.0099
ENSG00000070018	HGNC:6698	*LRP6*	Low-density lipoprotein receptor-related protein 6 (HGNC; Acc:6698)	5.74	2.52	0.0055
ENSG00000172476	HGNC:18283	*RAB40A*	RAB40A, member RAS oncogene family (HGNC; Acc:18283)	5.35	2.42	0.0001
ENSG00000134072	HGNC:1459	*CAMK1*	Calcium/calmodulin-dependent protein kinase I (HGNC; Acc:1459)	5.13	2.36	0.007
ENSG00000123700	HGNC:6263	*KCNJ2*	Potassium inwardly rectifying channel, subfamily J, member 2 (HGNC; Acc:6263)	4.92	2.3	0.0099
ENSG00000205018			Uncharacterized protein; cDNA FLJ26728 fis, clone PNC06635 (UniProtKB/TrEMBL; Acc:Q6ZP14)	4.89	2.29	0.0077
ENSG00000204991	HGNC:30623	*SPIRE2*	Spire homolog 2 (*Drosophila*) (HGNC; Acc:30623)	4.76	2.25	0.0046
ENSG00000143434	HGNC:10740	*SEMA6C*	Sema domain, transmembrane domain, and cytoplasmic domain, (semaphorin) 6C (HGNC; Acc:10740)	4.47	2.16	0.0072
ENSG00000167588	HGNC:4455	*GPD1*	Glycerol-3-phosphate dehydrogenase 1 (soluble) (HGNC; Acc:4455)	4.44	2.15	0.0102
ENSG00000143889	HGNC:25127	*HNRNPLL*	Heterogeneous nuclear ribonucleoprotein L-like (HGNC; Acc:25127)	4.41	2.14	0.012
ENSG00000116833	HGNC:7984	*NR5A2*	Nuclear receptor subfamily 5, group A, member 2 (HGNC; Acc:7984)	4.38	2.13	0.013
ENSG00000178297	HGNC:30079	*TMPRSS9*	Transmembrane protease, serine 9 (HGNC; Acc:30079)	4.26	2.09	0.019
ENSG00000229689			Unknown	4.2	2.07	0.0003
ENSG00000163069	HGNC:10806	*SGCB*	Sarcoglycan, β (43-kDa dystrophin-associated glycoprotein) (HGNC; Acc:10806)	4.14	2.05	0.013
ENSG00000101311	HGNC:15889	*FERMT1*	Fermitin family member 1 (HGNC; Acc:15889)	4.03	2.01	0.0005
ENSG00000148225	HGNC:21421	*WDR31*	WD repeat domain 31 (HGNC; Acc:21421)	4.03	2.01	0.0118
ENSG00000055813	HGNC:29400	*CCDC85A*	Coiled-coil domain-containing 85A (HGNC; Acc:29400)	4	2	0.0068
ENSG00000221886	HGNC:30804	*ZBED8*	Chromosome 5 open reading frame 54 (HGNC; Acc:30804)	3.92	1.97	0.0023
ENSG00000105854	HGNC:9205	*PON2*	Paraoxonase 2 (HGNC; Acc:9205)	3.78	1.92	0.0069
ENSG00000107281	HGNC:7899	*NPDC1*	Neural proliferation, differentiation, and control, 1 (HGNC; Acc:7899)	3.71	1.89	0.0124
ENSG00000125966	HGNC:7172	*MMP24*	Matrix metallopeptidase 24 (membrane-inserted) (HGNC; Acc:7172)	3.66	1.87	0.0017
ENSG00000171877	HGNC:28214	*FRMD5*	FERM domain-containing 5 (HGNC; Acc:28214)	3.61	1.85	0.0042
ENSG00000100583	HGNC:18631	*SAMD15*	Sterile α motif domain-containing 15 (HGNC; Acc:18631)	3.61	1.85	0.0114
ENSG00000254918			Uncharacterized protein (UniProtKB/TrEMBL; Acc:H3BTG1)	3.53	1.82	0.008
ENSG00000004838	HGNC:19412	*ZMYND10*	Zinc finger, MYND type-containing 10 (HGNC; Acc:19412)	3.43	1.78	0.0081
ENSG00000198223	HGNC:2435	*CSF2RA*	Colony-stimulating factor 2 receptor, α, low-affinity (granulocyte-macrophage) (HGNC; Acc:2435)	3.43	1.78	0.0005
ENSG00000101460	HGNC:6838	*MAP1LC3A*	Microtubule-associated protein 1 light chain 3 α (HGNC; Acc:6838)	3.41	1.77	0.0058
ENSG00000128383	HGNC:17343	*APOBEC3A*	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3A (HGNC; Acc:17343)	3.39	1.76	0.0071
ENSG00000173258	HGNC:23384	*ZNF483*	Zinc finger protein 483 (HGNC; Acc:23384)	3.39	1.76	0.0096
ENSG00000251184			Unknown	3.34	1.74	0.0247
ENSG00000136866	HGNC:12863	*ZFP37*	Zinc finger protein 37 homolog (mouse) (HGNC; Acc:12863)	3.32	1.73	0.0116
ENSG00000180938	HGNC:26758	*ZNF572*	Zinc finger protein 572 (HGNC; Acc:26758)	3.32	1.73	0.0067
ENSG00000205307	HGNC:41908	*SAP25*	Sin3A-associated protein, 25 kDa (HGNC; Acc:41908)	3.29	1.72	0.015
ENSG00000165891	HGNC:23820	*E2F7*	E2F transcription factor 7 (HGNC; Acc:23820)	3.29	1.72	0.0163
ENSG00000130005	HGNC:4136	*GAMT*	Guanidinoacetate *N*-methyltransferase (HGNC; Acc:4136)	3.29	1.72	0.0042
ENSG00000204389	HGNC:5232	*HSPA1A*	Heat-shock 70-kDa protein 1A (HGNC; Acc:5232)	3.25	1.7	0.0109
ENSG00000137821	HGNC:25965	*LRRC49*	Leucine-rich repeat-containing 49 (HGNC; Acc:25965)	3.25	1.7	0.0349
ENSG00000137831	HGNC:15947	*UACA*	Uveal autoantigen with coiled-coil domains and ankyrin repeats (HGNC; Acc:15947)	3.23	1.69	0.0125
ENSG00000213888	HGNC:26029	*LINC01521*	CDNA FLJ20464 fis, clone KAT06158;HCG1777549;uncharacterized protein (UniProtKB/TrEMBL; Acc:Q9NX35)	3.23	1.69	0.0253
ENSG00000185437	HGNC:10822	*SH3BGR*	SH3[c] domain binding glutamic acid-rich protein (HGNC; Acc:10822)	3.16	1.66	0.0238
ENSG00000267060	HGNC:43943	*PTGES3L*	Prostaglandin E synthase 3 (cytosolic)-like (HGNC; Acc:43943)	3.14	1.65	0.0347
ENSG00000121454	HGNC:21734	*LHX4*	LIM homeobox 4 (HGNC; Acc:21734)	3.14	1.65	0.0013
ENSG00000115641	HGNC:3703	*FHL2*	Four and a half LIM domains 2 (HGNC; Acc:3703)	3.12	1.64	0.0239
ENSG00000226180			Uncharacterized protein; cDNA FLJ45526 fis, clone BRTHA2027227 (UniProtKB/TrEMBL; Acc:Q6ZSH4)	3.1	1.63	0.0241
ENSG00000143850	HGNC:17053	*PLEKHA6*	Pleckstrin homology domain-containing, family A member 6 (HGNC; Acc:17053)	3.05	1.61	0.0211
ENSG00000163519	HGNC:30698	*TRAT1*	T cell receptor-associated transmembrane adaptor 1 (HGNC; Acc:30698)	3.03	1.6	0.0135
ENSG00000091622	HGNC:21043	*PITPNM3*	PITPNM family member 3 (HGNC; Acc:21043)	3.03	1.6	0.0033
ENSG00000069702	HGNC:11774	*TGFBR3*	Transforming growth factor β receptor III (HGNC; Acc:11774)	3.01	1.59	0.023
ENSG00000163534	HGNC:18509	*FCRL1*	Fc receptor-like 1 (HGNC; Acc:18509)	3.01	1.59	0.0089
ENSG00000196381	HGNC:26745	*ZNF781*	Zinc finger protein 781 (HGNC; Acc:26745)	2.99	1.58	0.0021
ENSG00000242732	HGNC:29430	*RGAG4*	Retrotransposon GAG domain-containing 4 (HGNC; Acc:29430)	2.99	1.58	0.0145

[a] HGNC, Human Gene Nomenclature Committee.
[b] FC, -fold change.
[c] SH3, Src homology 3.

Table 4
Genes down-regulated 3-fold in affected cells (P1 and P2) compared with control (C1 and C2) cells

Sequence ID	HGNC[a] ID	HGNC symbol	Sequence name	FC[b]	Log_2(FC)	*p*
ENSG00000188126			Myosin XVB pseudogene (HGNC; Acc:14083)	0.33	−1.58	0.0234
ENSG00000157303	HGNC:28391	*SUSD3*	Sushi domain-containing 3 (HGNC; Acc:28391)	0.33	−1.62	0.0021
ENSG00000265303			Uncharacterized protein (UniProtKB/TrEMBL; Acc:J3QRE1)	0.33	−1.59	0.0262
ENSG00000162069	HGNC:33584	*CCDC64B*	Coiled-coil domain-containing 64B (HGNC; Acc:33584)	0.33	−1.62	0.0245
ENSG00000172985	HGNC:24699	*SH3RF3*	SH3[c] domain-containing ring finger 3 (HGNC; Acc:24699)	0.32	−1.66	0.0325
ENSG00000130208	HGNC:607	*APOC1*	Apolipoprotein C-I (HGNC; Acc:607)	0.32	−1.65	0.0084
ENSG00000160888	HGNC:28871	*IER2*	Immediate early response 2 (HGNC; Acc:28871)	0.31	−1.71	0.0208
ENSG00000067082	HGNC:2235	*KLF6*	Kruppel-like factor 6 (HGNC; Acc:2235)	0.31	−1.67	0.0212
ENSG00000067606	HGNC:9412	*PRKCZ*	Protein kinase C, ζ (HGNC; Acc:9412)	0.31	−1.67	0.0035
ENSG00000163046	HGNC:35167	*ANKRD30BL*	Ankyrin repeat domain 30B-like (HGNC; Acc:35167)	0.3	−1.76	0.0064
ENSG00000140030	HGNC:4517	*GPR65*	G protein-coupled receptor 65 (HGNC; Acc:4517)	0.3	−1.72	0.0019
ENSG00000127561	HGNC:11501	*SYNGR3*	Synaptogyrin 3 (HGNC; Acc:11501)	0.29	−1.81	0.0028
ENSG00000211943	HGNC:5582	*IGHV3-15*	Immunoglobulin heavy variable 3–15 (HGNC; Acc:5582)	0.28	−1.82	0.0037
ENSG00000122224	HGNC:6730	*LY9*	Lymphocyte antigen 9 (HGNC; Acc:6730)	0.27	−1.87	0.0199
ENSG00000196154	HGNC:10494	*S100A4*	S100 calcium-binding protein A4 (HGNC; Acc:10494)	0.27	−1.88	0.0303
ENSG00000130635	HGNC:2209	*COL5A1*	Collagen, type V, α1 (HGNC; Acc:2209)	0.26	−1.97	0.0091
ENSG00000137193	HGNC:8986	*PIM1*	*pim-1* oncogene (HGNC; Acc:8986)	0.26	−1.96	0.0034
ENSG00000141401	HGNC:6051	*IMPA2*	Inositol(myo)-1(or 4)-monophosphatase 2 (HGNC; Acc:6051)	0.25	−2.02	0.0012
ENSG00000256642	HGNC:38595	*LINC00273*	Long intergenic non-protein coding RNA 273 (HGNC; Acc:38595)	0.24	−2.03	0.0088
ENSG00000180535	HGNC:22265	*BHLHA15*	Basic helix-loop-helix family, member a15 (HGNC; Acc:22265)	0.24	−2.08	0.0029
ENSG00000177685	HGNC:28703	*CRACR2B*	EF-hand calcium-binding domain 4A (HGNC; Acc:28703)	0.23	−2.11	0.0069
ENSG00000132749	HGNC:7446	*MTL5*	Metallothionein-like 5, testis-specific (tesmin) (HGNC; Acc:7446)	0.22	−2.2	0.0052
ENSG00000128965	HGNC:28680	*CHAC1*	ChaC, cation transport regulator homolog 1 (*E. coli*) (HGNC; Acc:28680)	0.22	−2.2	0.0045
ENSG00000175911			Uncharacterized protein (UniProtKB/TrEMBL; Acc:Q9HBN7)	0.21	−2.24	0.0066
ENSG00000149527	HGNC:29037	*PLCH2*	Phospholipase C, η2 (HGNC; Acc:29037)	0.21	−2.24	0.0018
ENSG00000120129	HGNC:3064	*DUSP1*	Dual-specificity phosphatase 1 (HGNC; Acc:3064)	0.21	−2.22	0.0048
ENSG00000185022	HGNC:6780	*MAFF*	v-*maf* musculoaponeurotic fibrosarcoma oncogene homolog F (HGNC; Acc:6780)	0.2	−2.34	0.0138
ENSG00000136859	HGNC:490	*ANGPTL2*	Angiopoietin-like 2 (HGNC; Acc:490)	0.19	−2.38	0.0061
ENSG00000131037	HGNC:21295	*EPS8L1*	EPS8-like 1 (HGNC; Acc:21295)	0.18	−2.46	0.0063
ENSG00000165312	HGNC:27346	*OTUD1*	OTU domain-containing 1 (HGNC; Acc:27346)	0.17	−2.56	0.0085
ENSG00000211942	HGNC:5581	*IGHV3-13*	Immunoglobulin heavy variable 3-13 (HGNC; Acc:5581)	0.05	−4.33	0.0014

[a] HGNC, Human Gene Nomenclature Committee.
[b] FC, -fold change.
[c] SH3, Src homology 3.

tion TPR domains, exploration of the OGT interactome of L254F-OGT compared with wild-type OGT is warranted, given that such interactions are thought to temporally and spatially regulate OGT (58) and that several OGT interactors, such as HUWE1, HCFC1, and histone deacetylases, have been associated with XLID.

Notably, global *O*-GlcNAc levels remained unaltered in all affected lymphoblastoids compared with controls (Fig. 4). This was a surprising finding for L254F-OGT, given the reduction observed in steady-state enzyme levels. This led us to evaluate the other cycling enzyme of the *O*-GlcNAc modification, OGA, in these cells. Steady-state OGA protein levels were concomitantly lowered in XLID, leading to the observation that a compensation mechanism exists in the affected lymphoblastoids (Fig. 5). This result suggests that there is an active mechanism for attempting to maintain global *O*-GlcNAc levels in lymphoblastoids. Further evaluation is required to elucidate whether this compensation mechanism is cell type-specific, especially in affected tissues of the XLID individuals.

Reduction of OGA protein levels in L254F-OGT lymphoblastoids was due to a decrease in *OGA* transcription as seen by reduced transcript levels in affected cells compared with controls (Fig. 5). To further validate this finding, we performed promoter luciferase reporter assays, which further validate reduced transcription from the *OGA* promoter in affected cells (Fig. 5). To potentially explain this finding, we were able to demonstrate the enrichment of an OGT-containing corepressor complex at the *OGA* proximal promoter region in cells derived from affected individuals (Fig. 6). Interestingly, P1 appeared to be slightly more affected than P2 in terms of lack of *OGA* promoter activity (5D) as well as increased binding of the OGT-containing corepressor complex at the *OGA* promoter. To further explore the impact of the L254F-OGT variant on transcription, we examined global gene expression and found that ~1% of the 8800 genes quantified by RNA-seq were altered 3-fold or more (Fig. 7). Panther Gene list analysis software was used to assign gene ontology-based grouping (59). This assignment revealed that 16.4% of the genes up-regulated and 8.1% of the genes down-regulated in the 3-fold set are involved in modulating transcription factor activity (Tables 3 and 4).

Our initial studies presented here were conducted in the lymphoblastoid cell lines available. We have also identified a second mutation in *OGT* (1013A>G (p.E338G)) that would also lead to an amino acid replacement in the N-terminal TPR repeats of OGT in a separate XLID family but have not yet fully characterized this mutation. During the preparation of this manuscript, a meeting abstract appeared from another independent exome-sequencing study (60) that identified a third mutation, c.775G>A (p.A259T), resulting in another amino acid substitution in the seventh TPR domain, the same repeat as the L254F mutation, of OGT associated with XLID. Future studies are aimed at obtaining neural lineages derived from a single human embryonic stem cell line, which has the benefit of a homogenous genetic background, that is edited to produce the relevant OGT mutant lines using CRISPR/Cas9 technology. These studies will further enhance our understanding of XLID and the causal role of OGT. In addition to studying XLID, these studies will also serve to illuminate the regulation (and dysregulation in disease states) of OGT, OGA, and *O*-GlcNAc and their importance in maintaining normal cellular function.

Finally, to summarize our findings, we have characterized a mutation identified in *OGT* for the first time that segregates with disease in families with XLID. Thus, approximately 30 years after identification of the *O*-GlcNAc modification in the Hart laboratory (16), we now have mutations in one of the genes encoding a cycling enzyme (OGT) that segregate with a specific disease. Further, during the characterization of this mutation, a compensatory mechanism to maintain global *O*-GlcNAc levels by OGT-dependent transcriptional regulation of *OGA* in lymphoblastoids was uncovered.

Experimental procedures

Reagents

Tissue culture media, serum, and antibiotics were purchased from Gibco. *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene) amino *N*-phenyl carbamate (PUGNAc) was purchased from Toronto Research Chemicals Inc. Anti-*O*-GlcNAc (CTD 110.6) was previously generated in the laboratory of Dr. Gerald W. Hart (Johns Hopkins University) (61). Anti-*O*-GlcNAc (RL2) antibody was from Enzo Life Sciences. Protein A/G beads, normal sera, and agarose-conjugated beads were purchased from Santa Cruz Biotechnology, Inc. Anti-OGT (DM17) antibody was from Sigma-Aldrich. Anti-OGT (H300) antibody was from Santa Cruz Biotechnology. Anti-OGT (AL28) antibody was previously generated in the laboratory of Dr. Gerald W. Hart (Johns Hopkins University). Anti-OGT (11576-2-AP) and anti-OGA (14711-1-AP) were from Proteintech. mAb10 was generated at the University of Georgia (62). The Whiteheart OGA antibody was a kind gift from Dr. Sidney Whiteheart (University of Kentucky).

Study samples

Individuals with XLID and control males with normal cognitive function were recruited from Greenwood Genetic Center (Greenwood, SC). Human subject research protocols for these studies were approved by the institutional review board. Informed consent was obtained from each study member and/or their parents or legal guardians. These individuals were evaluated by clinical geneticists and underwent comprehensive laboratory evaluations for ID. All individuals were found to have a normal karyotype, a negative molecular test for fragile X syndrome, and a negative screen for common inborn errors of metabolism. For each individual, 5–10 ml of blood was collected from affected probands males with XLID to establish Epstein-Barr virus-transformed lymphoblast cell lines for preparation of genomic DNA samples.

Human X chromosome exome sequencing

Sequence libraries were prepared using a TruSeq™ genomic DNA library preparation kit (Illumina), enriched for the human X chromosome exome using a SureSelect target enrichment kit (Agilent), and sequenced using the 75-bp pair-end sequence module on HiSeq2000 (Illumina). Alignment of the fastq reads, base recalibration, and variant calling were completed using Bowtie2 and Unified Genotyper (GATK). To enrich for disease-

causing mutations, we first utilized variant filters based on dbSNP, the male-restricted portions of the 1000 Genomes Project, or the Exome Variant Server data sets for variants with a frequency of 1% or less. Later, we developed and optimized a strategy, affected kindred/cross-cohort analysis, which utilizes a cohort of affected male kindred pairs and an additional small cohort of affected unrelated males to enrich for potentially pathological variants and to remove neutral variants. This strategy achieved a substantial enrichment for mutations in known XLID genes as compared with variant filters from public reference databases (63). Evolutionary conservation of the amino acid residues involved in the identified mutations was evaluated by multiple-sequence alignment of HomoloGene. Standard bioinformatics algorithms, including SIFT and PolyPhen-2, were used to predict the functional impact of the identified mutations (63).

Mutation validation, segregation analysis, and polymorphism study

Sanger sequencing was used for validation, segregation analysis, and polymorphism studies of each mutation using the BigDye Terminator version 3.1 cycle sequencing kit on an ABI3100 automatic DNA analyzer (Applied Biosystems) following the manufacturer's instructions. Variant analysis was completed on standard sequence alignment software (CodonCode and MacVector) followed by manual investigations of the chromatograms. X-inactivation in females available in family K9427 was determined using the androgen receptor locus as described previously (64).

Bioinformatics and modeling analysis

The structure of OGT contains an N-terminal TPR domain that mediates the recognition of a broad range of target proteins and C-terminal catalytic regions. The X-ray structure of the OGT TPR repeats (PDB code 1W3B), which contains 11.5 TPR units of human OGT and covers the sequence from 26 to 410, was used to model the effects of the L254F variant of OGT. Change of folding free energy upon missense mutation ($\Delta\Delta G$) was calculated with multiple web servers. The web servers used in this study include DUET, Eris, I-Mutant 2.0, mCSM, PopMuSiC, SDM, and SAAFEC (65, 66). The structure of OGT TPR repeats (PDB code 1W3B) was used to calculate $\Delta\Delta G$ for the L254F variant. Both the wild type and L254F variant were visualized with UCSF Chimera.

Site-directed mutagenesis

To create a recombinant HA-tagged L254F-OGT, we designed primers using the QuikChange Primer Design program available through Agilent Technologies. Mutagenesis was set up using the QuikChange site-directed mutagenesis kit (catalog no. 200523) as per the manufacturer's protocol using Gateway pENTR vector. Sanger sequencing was done at the Johns Hopkins School of Medicine Synthesis and Sequencing facility to validate the mutant DNA sequences. Following validation, we performed an LR reaction using the Gateway technology to obtain the mammalian expression vector pDEST26 containing the HA-tagged WT or L254F-*OGT*.

Creation of lymphoblastoid cell line

Cell lines were obtained by immortalization of lymphocytes from blood samples using Epstein-Barr virus with standard protocols (67).

Tissue culture, transfection, and cycloheximide treatment

Lymphoblastoids were cultured in RPMI 1640 medium containing 15% fetal bovine serum and 1% antibacterial/antimycotics. Cells were passaged every week and grown in suspension to the desired density for assays at 37 °C in 10% CO_2.

Cycloheximide was added to cells at a concentration of 50 μM and harvested at 0, 1, 2, 4, 8, 16, and 24 h following treatment.

HEK293F cells were cultured in Freestyle™ 293-F (Life Technologies, Inc., catalog no. 12338026) medium at 37 °C at ~130 rpm in 5% CO_2.

Recombinant *OGT* constructs containing the mutation in the pDEST26 backbone were transfected in HEK293F cells using 293Fectin™ (Life Technologies, catalog no. 12347-019) reagent. Enhanced GFP transfection was used as a both a transfection control and a vector control. We followed the manufacturer's protocol for transfection.

Cell lysis and immunoblotting

Lymphoblastoids from both affected individuals (P1 and P2) and controls (CA, C1, C2 and C3) or HEK293F cells were lysed using 50 mM Tris (pH 8), 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, 1× protease and phosphatase inhibitors, 10 μM PUGNAc (OGA inhibitor), and 1 mM DTT, a protocol modified from Brumbaugh *et al.* (68). 10 μg of nucleocytoplasmic proteins obtained was resolved on an SDS-polyacrylamide gel by electrophoresis for Figs. 2, 4, and 5 (except in Fig. 2*C*, where 20 μg was loaded) and transferred to a nitrocellulose membrane (Bio-Rad). Membranes were blocked for 1 h at room temperature in 1× TBST, 0.2% Tween 20 supplemented with 3.7% BSA for *O*-GlcNAc antibody 110.6. For all other antibodies, 1× TBST, 0.1% Tween 20 supplemented with 3% BSA (HA tag, mAb14, and mAb10) or 5% milk (OGT, OGA) was used. Following blocking, membranes were incubated overnight at 4 °C with the different antibodies. Membranes were washed four times with 1× TBST, 0.1% Tween 20 and incubated at room temperature for 1 h with a secondary horseradish peroxidase-conjugated antibody. Then the membranes were washed four times with 1× TBST supplemented with Tween 20. Proteins were detected using SuperSignal West Pico chemiluminescent substrate (Thermo Scientific, catalog no. 34080).

Escherichia coli expression, purification, and analysis of glycosyltransferase and protease activity of WT and L254F-OGT

A plasmid encoding CK2 was obtained from Gerald Hart (Johns Hopkins University and the NHLBI (National Institutes of Health) P01HL107153 Core C4). CK2α (residues 1–365) was cloned into the expression vector pGEX6P2 using the BamHI and SalI sites. The plasmids encoding ncOGT and HCF1-rep1 were obtained from Suzanne Walker (Harvard University) and

Winship Herr (University of Lausanne), respectively. All plasmids were expressed in *E. coli* and purified as described previously (44). CK2α glycosylation and HCF1-rep1 cleavage/glycosylation assays were performed as described before with minor modifications (45–47). Briefly, reaction mixtures contained 1 μM OGT, 1.5 μM CK2α or HCF1-rep1, and 1 mM UDP-GlcNAc in a final volume of 30 μl with buffer composed of 20 mM Tris, pH 8.0, 20 mM $MgCl_2$, 150 mM NaCl, and 1 mM DTT. Reactions were incubated at 37 °C for 1.5 h (CK2α) or 6 h (HCF1-rep1) and terminated by boiling in the presence of 1× Laemmli buffer. One-fourth of the final volume of the samples was resolved by SDS-PAGE and subjected to Western blotting. The following primary antibodies were used at a 1:1000 dilution: *O*-GlcNAc (Thermo, MA1072), *O*-GlcNAc transferase (Santa Cruz Biotechnology, sc-32921), CK2α (Santa Cruz Biotechnology, sc-373894), and GST (to detect GST-tagged HCF1-rep1) (Santa Cruz Biotechnology, sc-33613). Secondary antibodies were purchased from LI-COR and used at a 1:10,000 dilution. Experiments were performed in triplicate with separate preparations of recombinant WT and L254F-OGT.

RNA isolation and quantitative RT-PCR

Total RNA was extracted from affected and control lymphoblastoids using the Qiagen RNeasy Plus minikit (catalog no. 74134) by following the manufacturer's protocol. cDNA was prepared using the Bio-Rad iScript cDNA synthesis kit (catalog no. 170-8890) as per the manufacturer's instructions. The resulting cDNA was used as a template for amplification in a Bio-Rad 96-well MyIQ single-color real-time PCR detection instrument using the SYBR protocol (catalog no. 170-8880). All Quantitect assay primers for qPCR were purchased from Qiagen and used with the Bio-Rad iQ SYBR Green supermix (catalog no. 170-8880). Changes in gene expression of *OGA* (QT00085862), *HIST1H4B* (QT00207207), and *HIST1H3A* (QT00246764) were normalized using *B2M*, *RPL4*, *GAPDH*, *CYCG*, and *GUS* as the housekeeping genes. Quantification was performed using the ΔΔ*Ct* method (69).

Reporter luciferase assays

Lymphoblastoids from both affected individuals (P1 and P2) were transfected with pGL4.10 luciferase vectors using the Roche X-tremeGENE HP DNA transfection reagent (catalog no. 06366244001). pGL4.10-OGA contained a 2-kb proximal promoter region of *OGA* subcloned upstream of the coding region of the luciferase gene. The proximal promoter region of *OGA* was determined using Promoter 2.0 prediction, and the resulting plasmid was confirmed by DNA sequencing. Controls used were the *Renilla* luciferase (pGL4.74 hRluc/TK) (Promega, catalog no. E6921), SV40 promoter (pGL4.13 luc2/SV40) (Promega, catalog no. E6681), and the pGL4.10 luciferase empty vector (Promega, catalog no. E6651). After 48 h of transfection, cells were pelleted, followed by lysis and detection using the Promega Dual-Glo luciferase assay system (catalog no. E2980) as per the manufacturer's protocol on a luminometer. The resulting data were all normalized to a control male (C1 = 1).

ChIP

ChIP was performed as described previously (70). Briefly, DNA and protein were cross-linked using 2% formaldehyde and quenched with glycine. Sonicated DNA extract was precleared using protein A/G-agarose beads and mouse or rabbit IgG linked to agarose conjugate. Chromatin from 3×10^6 cells was used for each immunoprecipitation. Lysates were incubated with anti-OGT (Abcam, catalog no. 50273), anti-*O*-GlcNAc (mAb14), anti-HDAC1, or anti-mSin3A antibodies overnight at 4 °C with rotation. Protein-DNA complexes were incubated with protein-agarose A/G beads for 2 h and washed three times using buffers containing 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris, 150–500 mM NaCl, and protease inhibitors. DNA was eluted from beads using elution buffer containing 0.1% SDS and 100 mM $NaHCO_3$. Cross-linking was reversed by the addition of NaCl to a final concentration of 325 mM, and DNA was incubated overnight at 65 °C. DNA was extracted using phenol/chloroform after RNase and proteinase K treatment and analyzed by quantitative RT-PCR against the primers to the proximal *OGA* promoter (forward, 5′-aggggaaacagcggaagac-3′; reverse, 5′-tgccacctctgcgggt-3′). The primers were designed using the UCSC In-Silico PCR free tool. Results are shown as percentage of input.

RNA sequencing analysis and bioinformatics

RNA from lymphoblastoids from both affected individuals (P1 and P2) and related controls (C1 and C2) were extracted using the Qiagen RNeasy Plus minikit (catalog no. 74134) by following the manufacturer's protocol. Samples were submitted to the Genomics Services Laboratory at the Hudson Alpha Institute of Biotechnology (Huntsville, AL) for poly(A) mRNA library preparation and further sequencing and analysis. The concentration and integrity of the extracted total RNA were estimated by a Qubit® version 2.0 fluorometer (Invitrogen) and Agilent 2100 bioanalyzer (Applied Biosystems, Carlsbad, CA), respectively. RNA samples with an RNA integrity number value of ≥ 9.5 were used for further processing. From each of the four samples, 1 μg of RNA was used for poly(A) mRNA library preparation using the NEBNext poly(A) magnetic isolation module (New England Biolabs Inc., Ipswich, MA), according to the manufacturer's protocol. Samples were individually barcoded with unique in-house Genomics Services Laboratory primers and amplified through 10 cycles of PCR using KAPA HiFi HotStart Ready Mix (Kapa Biosystems, Inc., Woburn, MA). The quality of the libraries was assessed by a Qubit® version 2.0 fluorometer, and the concentration of the libraries was estimated by utilizing a DNA 1000 chip on an Agilent 2100 bioanalyzer.

Accurate quantification of the prepared mRNA libraries for downstream sequencing applications was determined using the qPCR-based KAPA Biosystems library quantification kit (Kapa Biosystems). Each library was then diluted to a final concentration of 12.5 nM and pooled equimolar before clustering. Cluster generation was carried out on a cBot version 1.4.36.0 using the TruSeq paired-end cluster kit version 3.0 (Illumina). Paired-end sequencing was performed using a 200-cycle TruSeq SBS HS version 3 kit on an Illumina HiSeq2000, running HiSeq

Control Software (HCS) version 1.5.15.1 (Illumina). Image analysis and base calling were performed using the standard Illumina Pipeline consisting of RTA (Real-Time Analysis) version 1.13. Raw reads were demultiplexed using bcl2fastq conversion software version 1.8.3 (Illumina) with default settings.

Postprocessing of the sequencing reads from RNA-seq experiments from each sample was performed as per our unique in-house pipeline. Briefly, quality control checks on raw sequence data from each sample were performed using FastQC (Babraham Bioinformatics, London, UK). Raw reads were mapped to the reference human genome hg19/GRCh37 using TopHat version 1.4 (71, 72) with two mismatches allowed and other default parameters. TopHat is a splice junction mapping tool for RNA-seq reads that utilizes an ultrafast high-throughput short read aligner Bowtie (71) in the background and then takes the mapping result and identifies the splice junctions. The alignment metrics of the mapped reads were estimated using SAMtools (73). Aligned reads were then imported onto the commercial data analysis platform, Avadis NGS (Strand Scientific Intelligence, Inc.). After quality inspection, the aligned reads were filtered on the basis of read quality metrics, where reads with a base quality score < 30, alignment score < 95, and mapping quality < 40 were removed. Remaining reads were then filtered on the basis of their read statistics, where missing mates and translocated, unaligned, and flipped reads were removed. The reads list was then filtered to remove duplicates. To reduce noise from low-signal reads, the minimum intensity was set to 8, and reads were removed if neither the average of affected individual values nor control values were >16. The intent of this filter was to ensure conservative interpretation of -fold change for low signal values. The final list was created by removing reads where the relative S.D. of the affected individuals values (P1 and P2) or control values (C1 and C2) was greater than 25%. This resulted in 8800 quantifiable transcripts. R was used to perform Spearman correlation analysis (see the supplemental material). Differential expression of genes was calculated on the basis of -fold change (using thresholds of ± 2.0, ± 3.0, and ± 4.0) observed between defined conditions. To assess data quality, two "mock" versions of the final reads list were created by grouping P1 with C1 and P2 with C2 (mock 1) and P1 with C2 and P2 with C1 (mock 2) and filtering as described above (from noise reduction on). Enrichment for differentially expressed genes (disease *versus* natural variation) was obtained by dividing the number of genes for experimental sets at preset -fold changes by the average number of genes of the two mock sets at the same -fold change.

Statistical analysis

Data are expressed as mean ± S.E. The differences between means and the effects of treatments were analyzed using Excel or GraphPad to determine statistically significant values.

Author contributions—K. V., C. S., and L. W. conceived and coordinated the study and wrote the paper. K. V., T. N., N. S., C. F. T., M. M., B. W., S. P., C. S., J. O., J. C., D. V., J. G., M. S., Y. P., E. A., T. W., C. S., and L. W. designed, performed, and/or analyzed the experiments presented. All authors reviewed and approved the final submitted version of the manuscript.

Acknowledgments—We thank the families for participation in the X-linked intellectual disability study at the Greenwood Genetic Center. We also thank Dr. Salim Aftimos (retired) (Genetic Health Services New Zealand Northern Hub) for referring his family for the study. We also thank Dr. Gerald Hart (Johns Hopkins School of Medicine), Dr. Sidney Whiteheart (University of Kentucky), Dr. Suzanne Walker (Harvard University), and Dr. Winship Herr (University of Lausanne) for providing antibodies, plasmids, and/or technical advice.

References

1. Roeleveld, N., Zielhuis, G. A., and Gabreëls, F. (1997) The prevalence of mental retardation: a critical review of recent literature. *Dev. Med. Child. Neurol.* **39,** 125–132
2. Leonard, H., and Wen, X. (2002) The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment. Retard. Dev. Disabil. Res. Rev.* **8,** 117–134
3. Raymond, F. L. (2006) X linked mental retardation: a clinical guide. *J. Med. Genet.* **43,** 193–200
4. Otte, C., and Rauch, A. (2013) [Intellectual disability: a frequent reason for referral to medical genetics]. *Praxis* **102,** 1467–1473
5. Salvador-Carulla, L., and Bertelli, M. (2008) "Mental retardation" or "intellectual disability": time for a conceptual change. *Psychopathology* **41,** 10–16
6. Ropers, H. H. (2010) Genetics of early onset cognitive impairment. *Annu. Rev. Genomics Hum. Genet.* **11,** 161–187
7. Kaufman, L., Ayub, M., and Vincent, J. B. (2010) The genetic basis of non-syndromic intellectual disability: a review. *J. Neurodev. Disord.* **2,** 182–209
8. Chelly, J., Khelfaoui, M., Francis, F., Chérif, B., and Bienvenu, T. (2006) Genetics and pathophysiology of mental retardation. *Eur. J. Hum. Genet.* **14,** 701–713
9. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., *et al.* (2001) The sequence of the human genome. *Science* **291,** 1304–1351
10. Lubs, H. A., Stevenson, R. E., and Schwartz, C. E. (2012) Fragile X and X-linked intellectual disability: four decades of discovery. *Am. J. Hum. Genet.* **90,** 579–590
11. Comer, F. I., and Hart, G. W. (1999) *O*-GlcNAc and the control of gene expression. *Biochim. Biophys. Acta* **1473,** 161–171
12. Wells, L., Vosseller, K., and Hart, G. W. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and *O*-GlcNAc. *Science* **291,** 2376–2378
13. Zachara, N. E., and Hart, G. W. (2002) The emerging significance of *O*-GlcNAc in cellular regulation. *Chem. Rev.* **102,** 431–438
14. Zachara, N. E., and Hart, G. W. (2006) Cell signaling, the essential role of *O*-GlcNAc!. *Biochim. Biophys. Acta* **1761,** 599–617
15. Ozcan, S., Andrali, S. S., and Cantrell, J. E. (2010) Modulation of transcription factor function by *O*-GlcNAc modification. *Biochim. Biophys. Acta* **1799,** 353–364
16. Torres, C. R., and Hart, G. W. (1984) Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for *O*-linked GlcNAc. *J. Biol. Chem.* **259,** 3308–3317
17. Zeidan, Q., and Hart, G. W. The intersections between *O*-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *J. Cell Sci.* **123,** 13–22
18. Haltiwanger, R. S., Kelly, W. G., Roquemore, E. P., Blomberg, M. A., Dong, L. Y., Kreppel, L., Chou, T. Y., and Hart, G. W. (1992) Glycosylation of nuclear and cytoplasmic proteins is ubiquitous and dynamic. *Biochem. Soc. Trans.* **20,** 264–269
19. Zachara, N. E., and Hart, G. W. (2004) *O*-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular func-

tion in response to nutrition and stress. *Biochim. Biophys. Acta* **1673,** 13–28
20. Kreppel, L. K., Blomberg, M. A., and Hart, G. W. (1997) Dynamic glycosylation of nuclear and cytosolic proteins: cloning and characterization of a unique *O*-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272,** 9308–9315
21. Dong, D. L., and Hart, G. W. (1994) Purification and characterization of an *O*-GlcNAc-selective *N*-acetyl-β-D-glucosaminidase from rat spleen cytosol. *J. Biol. Chem.* **269,** 19321–19330
22. Hart, G. W., Housley, M. P., and Slawson, C. (2007) Cycling of *O*-linked β-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446,** 1017–1022
23. Hanover, J. A., Lai, Z., Lee, G., Lubas, W. A., and Sato, S. M. (1999) Elevated *O*-linked *N*-acetylglucosamine metabolism in pancreatic β-cells. *Arch. Biochem. Biophys.* **362,** 38–45
24. Hart, G. W., Haltiwanger, R. S., Holt, G. D., and Kelly, W. G. (1989) Nucleoplasmic and cytoplasmic glycoproteins. *Ciba Found. Symp.* **145,** 102–112. discussion 112–108
25. Wells, L., Vosseller, K., and Hart, G. W. (2003) A role for *N*-acetylglucosamine as a nutrient sensor and mediator of insulin resistance. *Cell Mol. Life Sci.* **60,** 222–228
26. Vaidyanathan, K., Durning, S., and Wells, L. (2014) Functional *O*-GlcNAc modifications: implications in molecular regulation and pathophysiology. *Crit. Rev. Biochem. Mol. Biol.* **49,** 140–163
27. Ma, Z., Vocadlo, D. J., and Vosseller, K. (2013) Hyper-*O*-GlcNAcylation is anti-apoptotic and maintains constitutive NF-κB activity in pancreatic cancer cells. *J. Biol. Chem.* **288,** 15121–15130
28. Yang, X., Ongusaha, P. P., Miles, P. D., Havstad, J. C., Zhang, F., So, W. V., Kudlow, J. E., Michell, R. H., Olefsky, J. M., Field, S. J., and Evans, R. M. (2008) Phosphoinositide signalling links *O*-GlcNAc transferase to insulin resistance. *Nature* **451,** 964–969
29. Shi, Y., Tomic, J., Wen, F., Shaha, S., Bahlo, A., Harrison, R., Dennis, J. W., Williams, R., Gross, B. J., Walker, S., Zuccolo, J., Deans, J. P., Hart, G. W., and Spaner, D. E. (2010) Aberrant *O*-GlcNAcylation characterizes chronic lymphocytic leukemia. *Leukemia* **24,** 1588–1598
30. Yuzwa, S. A., Shan, X., Macauley, M. S., Clark, T., Skorobogatko, Y., Vosseller, K., and Vocadlo, D. J. (2012) Increasing *O*-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **8,** 393–399
31. Shafi, R., Iyer, S. P., Ellies, L. G., O'Donnell, N., Marek, K. W., Chui, D., Hart, G. W., and Marth, J. D. (2000) The *O*-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 5735–5739
32. Kreppel, L. K., and Hart, G. W. (1999) Regulation of a cytosolic and nuclear *O*-GlcNAc transferase: role of the tetratricopeptide repeats. *J. Biol. Chem.* **274,** 32015–32022
33. Lubas, W. A., Frank, D. W., Krause, M., and Hanover, J. A. (1997) *O*-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J. Biol. Chem.* **272,** 9316–9324
34. Allan, R. K., and Ratajczak, T. (2011) Versatile TPR domains accommodate different modes of target protein recognition and function. *Cell Stress Chaperones* **16,** 353–367
35. Iyer, S. P., Akimoto, Y., and Hart, G. W. (2003) Identification and cloning of a novel family of coiled-coil domain proteins that interact with *O*-GlcNAc transferase. *J. Biol. Chem.* **278,** 5399–5409
36. Cheung, W. D., Sakabe, K., Housley, M. P., Dias, W. B., and Hart, G. W. (2008) *O*-linked β-*N*-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J. Biol. Chem.* **283,** 33935–33941
37. Comtesse, N., Maldener, E., and Meese, E. (2001) Identification of a nuclear variant of MGEA5, a cytoplasmic hyaluronidase and a β-*N*-acetylglucosaminidase. *Biochem. Biophys. Res. Commun.* **283,** 634–640
38. Gao, Y., Wells, L., Comer, F. I., Parker, G. J., and Hart, G. W. (2001) Dynamic *O*-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic β-*N*-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276,** 9838–9845
39. Wells, L., Gao, Y., Mahoney, J. A., Vosseller, K., Chen, C., Rosen, A., and Hart, G. W. (2002) Dynamic *O*-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic β-*N*-acetylglucosaminidase, *O*-GlcNAcase. *J. Biol. Chem.* **277,** 1755–1761
40. Rao, F. V., Schüttelkopf, A. W., Dorfmueller, H. C., Ferenbach, A. T., Navratilova, I., and van Aalten, D. M. (2013) Structure of a bacterial putative acetyltransferase defines the fold of the human *O*-GlcNAcase C-terminal domain. *Open Biol.* **3,** 130021
41. Yang, Y. R., Song, M., Lee, H., Jeon, Y., Choi, E. J., Jang, H. J., Moon, H. Y., Byun, H. Y., Kim, E. K., Kim, D. H., Lee, M. N., Koh, A., Ghim, J., Choi, J. H., Lee-Kwon, W., *et al.* (2012) *O*-GlcNAcase is essential for embryonic development and maintenance of genomic stability. *Aging Cell* **11,** 439–448
42. Plenge, R. M., Stevenson, R. A., Lubs, H. A., Schwartz, C. E., and Willard, H. F. (2002) Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. *Am. J. Hum. Genet.* **71,** 168–173
43. Antecol, M. H., Darveau, A., Sonenberg, N., and Mukherjee, B. B. (1986) Altered biochemical properties of actin in normal skin fibroblasts from individuals predisposed to dominantly inherited cancers. *Cancer Res.* **46,** 1867–1873
44. Selvan, N., Mariappa, D., van den Toorn, H. W., Heck, A. J., Ferenbach, A. T., and van Aalten, D. M. (2015) The early metazoan *Trichoplax adhaerens* possesses a functional *O*-GlcNAc system. *J. Biol. Chem.* **290,** 11969–11982
45. Capotosti, F., Guernier, S., Lammers, F., Waridel, P., Cai, Y., Jin, J., Conaway, J. W., Conaway, R. C., and Herr, W. (2011) *O*-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* **144,** 376–388
46. Janetzko, J., Trauger, S. A., Lazarus, M. B., and Walker, S. (2016) How the glycosyltransferase OGT catalyzes amide bond cleavage. *Nat. Chem. Biol.* **12,** 899–901
47. Lazarus, M. B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W. F., Vocadlo, D. J., Herr, W., and Walker, S. (2013) HCF-1 is cleaved in the active site of *O*-GlcNAc transferase. *Science* **342,** 1235–1239
48. Yang, X., Zhang, F., and Kudlow, J. E. (2002) Recruitment of *O*-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein *O*-GlcNAcylation to transcriptional repression. *Cell* **110,** 69–80
49. R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria
50. Edgar, R., Domrachev, M., and Lash, A. E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210
51. Hardivillé, S., and Hart, G. W. (2014) Nutrient regulation of signaling, transcription, and cell physiology by *O*-GlcNAcylation. *Cell Metab.* **20,** 208–213
52. Howerton, C. L., Morgan, C. P., Fischer, D. B., and Bale, T. L. (2013) *O*-GlcNAc transferase (OGT) as a placental biomarker of maternal stress and reprogramming of CNS gene transcription in development. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 5169–5174
53. Ross, M. T., Grafham, D. V., Coffey, A. J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G. R., Burrows, C., Bird, C. P., Frankish, A., Lovell, F. L., Howe, K. L., Ashurst, J. L., Fulton, R. S., *et al.* (2005) The DNA sequence of the human X chromosome. *Nature* **434,** 325–337
54. Lisik, M. Z., and Sieron, A. L. (2008) X-linked mental retardation. *Med. Sci. Monit.* **14,** RA221–RA229
55. Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., and Sakaguchi, A. Y. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306,** 234–238
56. Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M., and Kunkel, L. M. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323,** 646–650
57. Warren, S. T., Zhang, F., Licameli, G. R., and Peters, J. F. (1987) The fragile X site in somatic cell hybrids: an approach for molecular cloning of fragile sites. *Science* **237,** 420–423
58. Nagel, A. K., and Ball, L. E. (2014) *O*-GlcNAc transferase and *O*-GlcNAcase: achieving target substrate specificity. *Amino Acids* **46,** 2305–2316
59. Mi, H., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. (2013) Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8,** 1551–1566

60. Huq, A. (2015) A new disorder of *O*-glycosylation due to a mutation in the X-linked OGT gene is associated with microcephaly, epilepsy, abnormal sleep architecture, intellectual disability and hypothyroidism. *American Society of Human Genetics Annual Meeting 2015, Baltimore, October 6–10, 2015*, p. 2297T, American Society of Human Genetics, Bethesda, MD
61. Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A., and Hart, G. W. (2001) Characterization of a mouse monoclonal antibody specific for *O*-linked *N*-acetylglucosamine. *Anal. Biochem.* **293,** 169–177
62. Teo, C. F., Ingale, S., Wolfert, M. A., Elsayed, G. A., Nöt, L. G., Chatham, J. C., Wells, L., and Boons, G. J. (2010) Glycopeptide-specific monoclonal antibodies suggest new roles for *O*-GlcNAc. *Nat. Chem. Biol.* **6,** 338–343
63. Niranjan, T. S., Skinner, C., May, M., Turner, T., Rose, R., Stevenson, R., Schwartz, C. E., and Wang, T. (2015) Affected kindred analysis of human X chromosome exomes to identify novel X-linked intellectual disability genes. *PLoS One* **10,** e0116454
64. Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M., and Belmont, J. W. (1992) Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* **51,** 1229–1239
65. Capriotti, E., Fariselli, P., and Casadio, R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33,** W306–W310
66. Yin, S., Ding, F., and Dokholyan, N. V. (2007) Eris: an automated estimator of protein stability. *Nat. Methods* **4,** 466–467
67. Sie, L., Loong, S., and Tan, E. K. (2009) Utility of lymphoblastoid cell lines. *J. Neurosci. Res.* **87,** 1953–1959
68. Brumbaugh, J., Hou, Z., Russell, J. D., Howden, S. E., Yu, P., Ledvina, A. R., Coon, J. J., and Thomson, J. A. (2012) Phosphorylation regulates human OCT4. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 7162–7168
69. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C(T))$ method. *Methods* **25,** 402–408
70. Ekanayake, D., and Sabatini, R. (2011) Epigenetic regulation of polymerase II transcription initiation in *Trypanosoma cruzi*: modulation of nucleosome abundance, histone modification, and polymerase occupancy by *O*-linked thymine DNA glucosylation. *Eukaryot. Cell* **10,** 1465–1472
71. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25
72. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111
73. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079