Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability

Veronica M. Pravataa,1, Villo Muhab,1, Mehmet Gundogdua,1, Andrew T. Ferenbachb, Poonam S. Kakadeb, Vasudha Vandad, Ariane C. Wilmesb, Vladimir S. Borodkinb, Shelagh Jossb, Marios P. Stavridissb, and Daan M. F. van Aaltena,b,c

Edited by Carolyn R. Bertozzi, Stanford University, Stanford, CA, and approved June 17, 2019 (received for review January 29, 2019)

O-GlcNAc transferase (OGT) is an X-linked gene product that is essential for normal development of the vertebrate embryo. It catalyses the O-GlcNAc posttranslational modification of nucleocytoplasmic proteins and proteolytic maturation of the transcriptional coregulator Host cell factor 1 (HCF1). Recent studies have suggested that conservative missense mutations distal to the OGT catalytic domain lead to X-linked intellectual disability in boys, but it is not clear if this is through changes in the O-GlcNAc proteome, loss of protein–protein interactions, or misprocessing of HCF1. Here, we report an OGT catalytic domain missense mutation in monozygotic female twins (c. X:70779215 T > A, p. N567K) with intellectual disability that allows disruption of these effects. The patients show limited IQ with developmental delay and skewed X-inactivation. Molecular analyses revealed decreased OGT stability and disruption of the substrate binding site, resulting in loss of catalytic activity. Editing this mutation into the Drosophila genome results in global changes in the O-GlcNAc proteome, while in mouse embryonic stem cells it leads to loss of O-GlcNAcase and delayed differentiation down the neuronal lineage. These data imply that catalytic deficiency of OGT could contribute to X-linked intellectual disability.

intellectual disability | O-GlcNAc | neurodevelopment

The O-linked β-N-acetyl-α-glucosamine (O-GlcNAc) transferase (OGT) and the hydrolyase O-GlcNAcCase (OGA) together orchestrate protein O-GlcNAcylation, a dynamic co/posttranslational modification cycling on thousands of nucleocytoplasmic proteins (1, 2). OGT catalyses the covalent attachment of the monosaccharide O-GlcNAc to serine or threonine. This modification has been suggested to affect transcription (3–6), translation (1), protein stability (7, 8), and subcellular localization (9, 10). O-GlcNAcylation plays a key role in regulating stress response (11–13), differentiation (14, 15), nutrient sensing (16, 17), and autophagy (18).

Genetic studies have highlighted the importance of OGT and protein O-GlcNAcylation in development. OGT is essential for embryonic stem cell viability (19, 20) and mouse embryonic development (21). Zebrafish lacking OGT function exhibit shortened body axis and smaller brains (22). The Drosophila melanogaster OGT encoded by the Polycomb group gene super sex combss (sex) plays a critical role in establishing Drosophila segment identity (23, 24).

The O-GlcNAc modification is particularly abundant in the brain (25, 26), where it controls memory formation (27–29), circadian rhythm (30, 31), and appetite (32, 33). Numerous neuron-specific proteins are O-GlcNAcylated, such as the Microtubule-associated proteins tau (34) and CRMP2 (35), synaptic vesicle proteins (36), the transcriptional factor cyclic-AMP response element binding protein (CREB) (28), and the AMPA receptor GluA2 subunit (27). Growing evidence suggests that O-GlcNAcylation is essential for normal development and function of the mammalian nervous system (21, 33, 37–39).

Although O-GlcNAcylation has been long implicated in chronic metabolic diseases, such as type II diabetes mellitus, neurodegeneration, and cancer, its role in neurodevelopmental disorders has only recently become apparent. In the past year, a small number of missense mutations within the human OGT gene have been discovered in patients with X-linked intellectual disability (XLID) (40–43). Intellectual disability (ID) refers to a broad range of developmental disorders characterized by limited intellectual capacity, IQ below 70, and poor adaptive behavior with onset before the age of 18, affecting 1 to 3% of the population worldwide (44). Genetic factors are the major cause of this developmental condition involving over 700 ID genes (45). These genes encode for proteins that are required for neuronal development and activity, contributing to neuronal structure and function through several signaling pathways (46). Mutations in genes located on the X-chromosome account for ~5 to 10% of all ID causative genes affecting predominantly male individuals (44, 47).

To date, XLID-associated mutations in the OGT gene have only been reported for male patients, causing developmental delay and severe cognitive disability. Accompanying clinical phenotypes were dysmorphic features, such as clinodactyly, eye abnormalities, and microcephaly. A common molecular trait of the missense XLID OGT variants is that they appear to retain substantial OGT catalytic activity. This is in agreement with these mutations all being localized to the C-terminal noncatalytic tetra peptide repeat (TPR) domain of OGT that is responsible for substrate recognition (39, 40). The catalytic and noncatalytic domains of OGT interact in a manner that depends on the protein context, and disruption of these interactions may underpin pathogenesis (47).

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 6IBO).


Significance
Protein O-GlcNAcylation is a posttranslational modification essential for development. Recently, mutations in the O-GlcNAc transferase (OGT) substrate binding domain have been described that lead to intellectual disability, but the mechanisms underpinning pathogenesis remain to be explored. This work describes the first point mutation in the OGT catalytic domain leading to effects on O-GlcNAcylation and Host cell factor 1 processing in vitro and in a stem cell/Drosophila model, resulting in delayed neuronal differentiation. This establishes a potential link between OGT activity and intellectual disability.

www.pnas.org/cgi/doi/10.1073/pnas.1900065116

PNAS Latest Articles | 1 of 10
specificity and scaffolding (40–43). In cultured cells, global O-GlcNAc levels on proteins were unaltered (41–43), and a decrease in OGA protein expression to compensate for the (presumed) moderate loss of OGT activity was only apparent in patient-derived fibroblast cells (43). In addition to its glycosyltransferase activity, mammalian OGT is involved in the unusual proteolytic maturation of Host cell factor 1 (HCF1) (48–50), a transcriptional cofactor implicated in cell cycle control (51), also identified as an ID gene (52–54). HCF1 is O-GlcNAc–modified and subsequently cleaved by OGT using the same active site for both enzymatic activities (49). Among XLID variant OGT models, a moderate effect on HCF1 processing was only detected for one of the mutations (41, 43).

From the currently available OGT XLID mutations, it remains unclear if the patient phenotypes observed are linked to changes in the O-GlcNAc proteome, loss of protein–protein interactions,

**Fig. 1.** Clinical images from patients with the N567K mutation in the catalytic domain of OGT. (A) Picture of twin 1 showing dysmorphic features at age 4 y (Left) and 10 y (Right). (B) Picture of twin 2 showing dysmorphic features at age 4 y (Left) and 10 y (Right). (C) Schematic diagram of OGT highlighting the TPRs, TPR-like repeat, N-Cat, and C-Cat, and the site of N567K, as well as the previously identified XLID-associated mutations in OGT. The inset shows the sequence alignment of the region encompassing the N567K mutation across the commonly used model organisms. N-Cat and C-Cat: N- and C-terminal lobes of OGT catalytic domain; TPR: tetratricopeptide repeat domain; TLR: tetratricopeptide repeat-like domain.
or misprocessing of HCF1. Here, we characterize an OGT missense mutation in the catalytic domain as observed in female twins with ID and developmental delay. In vitro biochemical analysis of recombinant N567K OGT missense variant protein revealed that this mutation abrogates OGT and HCF1 proteolytic processing activity. Protein crystallography data explain the molecular basis of the observed loss of substrate binding. Introduction of this mutation into D. melanogaster by genome editing reveals global effects on the O-GlcNAc proteome. In mouse embryonic stem cells (mESCs) carrying this mutation, loss of OGA protein expression was observed, as well as defective HCF1 processing. The mutation caused defects in neurite outgrowth during neuronal differentiation. Taken together, our data provide evidence that loss of OGT activity could contribute to the XLID phenotype observed in the patients.

Results and Discussion

Monozygotic Twins with ID and Developmental Delay Possess a De Novo OGT Mutation. Twin girls (twin 1 and twin 2) were born at 33-wk gestation by semielective caesarean section for maternal preeclampsia. Conception was unassisted. Twin 1 weighed 1,670 g at birth and required continuous positive airway pressure (CPAP) to aid breathing for 3 wk and supplemental oxygen for 58 d. Twin 2 weighed 2,150 g and needed CPAP for 2 d and supplemental oxygen for 20 d. Both babies were tube-fed initially but bottle feeding was established from 3 to 4 wk of age. The twins showed delay in reaching developmental milestones, especially in areas of speech and language development. By 4 y of age, twin 1 only used 3 to 4 single words; at 10 y she was able to put 2 to 3 words together. Language development of twin 2 was more advanced than her sister; at 4 y she was using short phrases and at 10 y she was speaking in sentences and able to recognize letters of the alphabet. Their movement and physical development were also affected: twin 1 by age 4 y and twin 2 by the age of 2 y were able to walk independently. However, twin 2 had significant gross motor difficulties and clumsiness at the age of 8 y. Both twins were diagnosed with cerebral visual impairment, twin 1 more severe than twin 2. They have attended an additional needs nursery and school.

The twins (Fig. 1 A and B) appeared to be affected by a similar developmental syndrome, including small nose, high arched palate, and fifth finger clinodactyly. Twin 1 has ataxic gait with knees bent, walking on toes with inverted feet. An MRI brain scan showed prominence of the lateral ventricles and an enlarged cistern magna in twin 1 at age 2 y, suggesting possible hypoplasia of the inferior cerebellar vermis. Although both siblings have kept generally good health, hypotonia was apparent in infancy and they required iron supplementation intermittently. Twin 1 had an adenoidectomy and treatment for gastro-esophageal reflux and constipation.

Sanger sequencing was performed as part of the Deciphering Developmental Disorders initiative, which currently lists genomic and clinical data from over 14,000 child patients with severe undiagnosed developmental disorders (55). This revealed a single pathogenic missense mutation in the OGT gene (X:70779215 T > A, N567K). The mutation was found absent in control population of 141,456 samples, comprising 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals reported in the genome aggregation database, gnomAD (56). The mutation affected both monozygotic twins, whereas both of their parents were noncarriers. A 98:2 skew in X-inactivation was detected in both children using polymorphic markers that are differentially methylated on the active and inactive chromosome (SI Appendix, Fig. S1). However, being a de novo mutation, which 1 of the 2 chromosomes is predominantly inactivated remains unclear. Taking these data together, we have identified monozygotic twins with ID and developmental delay that possess a de novo OGT mutation.

The N567K Mutation Abrogates OGT Activity In Vitro. To understand the potential effects of the OGT N567K (OGT_N567K) mutation, we next investigated changes in catalytic activity compared with full-length wild-type OGT (OGT_WT) protein. Steady-state kinetics of OGT glycosyltransferase activity was measured against 2 acceptor peptide substrates, P1 (KKVPVSRA) and P2 (KKVAVSRA), that only varied at the −2 position adjacent to the O-GlcNAcylation site (57). OGT_N567K activity was 12-fold reduced against P1 (Fig. 2A and Table 1) and not detectable against P2 (Fig. 2B). Next, OGT activity was measured against an intact protein substrate, TAK-1 binding protein (TABI), where the reaction was followed with a TAB1 O-GlcNAc site-specific antibody (58). OGT_N567K showed negligible glycosyltransferase activity (Fig. 2B). OGT possesses a second catalytic activity in the form of promoting autocatalytic cleavage and activation of HCF1, which itself is an ID-associated gene (52). We next investigated the effects of the N567K mutation on OGT-mediated HCF1 processing. OGT proteolytic activity was measured using a GST-fusion construct of a minimal HCF1 fragment (HCF1-repl1) containing one of the PRO repeats that are known to be the targets of OGT-mediated proteolysis (49). Wild-type OGT not only promotes HCF1 proteolysis but was also able to hyperglycosylate this protein fragment (Fig. 2C). HCF1_ex01090, which lacks the key glutamate required for proteolytic processing, was not cleaved by wild-type OGT. Strikingly, even after 8 h of reaction time, OGT_N567K was able to neither glycosylate nor cleave HCF1-repl1 (Fig. 2C). Taken together, the data show that the N567K mutation abrogates OGT activity.

The N567K Mutation Disrupts the OGT Acceptor Binding Site. To understand the molecular basis of the observed loss of OGT catalytic activity, we next initiated structural characterization of OGT_N567K. Inspection of the wild-type OGT structure reveals that Asn567 maps to a loop in the N-terminal lobe of the catalytic domain that is completely conserved through evolution from Caenorhabditis elegans to man (Fig. 1C). The Asn567 side chain borders the −2 subsite as part of the OGT acceptor substrate binding cleft (Fig. 2D). Therefore, we hypothesized that the N567K mutation would affect OGT acceptor substrate binding, without affecting the overall OGT fold. Recombinant OGT_N567K was obtained from Escherichia coli using construct boundaries previously employed to crystallize OGT_WT-substrate complexes (57). OGT_N567K was then co-crystallized with a donor analog, UDP-55-GlcNAc, and 2 OGT acceptor peptide substrates, P1 (KKVPVSRA) and P2 (KKVAVSRA), previously used for activity measurements (Fig. 2D). Only the condition containing P2, which contains an alanine at the −2 subsite, yielded crystals that diffracted to 2.3 Å and allowed structure solution by molecular replacement and refinement (Table 2). The OGT_N567K structure superposed onto that of OGT_WT with an rmsd of 0.4 Å across 675 Cα atoms, suggesting the mutation does not induce large conformational changes in the catalytic domain. Initial unbiased difference electron density maps revealed partial density corresponding to the acceptor peptide (Fig. 2D). Surprisingly, this only covered the P2 backbone in the +2 through to −2 subsites, suggesting that the N567K mutation may induce increased flexibility at the N-terminal tail of the acceptor peptide (Fig. 2D and SI Appendix, Fig. S2). In agreement with this, a comparative B-factor analysis suggested that the acceptor peptide bound to OGT_N567K is substantially more flexible than that bound to OGT_WT (SI Appendix, Fig. S2). Close inspection of the OGT_N567K acceptor substrate binding cleft revealed that the protrusion caused by the N567K mutation partially occupies the position of the −2 proline frequently observed in OGT acceptor peptides (57, 59, 60). In summary, the N567K mutation appears to disrupt the OGT acceptor binding site.
Fig. 2. The N567K mutation abrogates OGT activity due to disrupted acceptor binding site. (A) Michaelis–Menten kinetics of OGT glycosyltransferase activity against P1 and P2 peptides. (B) Immunoblots showing OGT glycosyltransferase activity against TAB1. (C) Immunoblots showing OGT glycosyltransferase and proteolytic activities against HCF1-rep1, HCF1-rep1: GST-tagged host cell factor-1 fragment containing the first PRO repeat. W: wild type. E: E101Q. (D) Crystal structures of OGT ternary complexes of OGT<sub>WT</sub> and OGT<sub>N567K</sub>, active site in complex with UDP-5S-GlcNAc (U5G, turquoise, orange atoms) and acceptor peptide. OGT<sub>WT</sub> is shown in complex with RB2 (Retinoblastoma-like protein 2 peptide, light gray carbon atoms) [PDB ID code: 5C1D (57)]; OGT<sub>N567K</sub> is shown with P2 peptide (pink carbon atoms) (PDB ID code 6IBO). K567 is shown in red. An unbiased F<sub>o</sub>–F<sub>c</sub> difference map before inclusion of the peptide is shown as a mesh.
The sxcN595K Mutation Leads to a Reduction of Protein O-GlcNAcylation in D. melanogaster. Lack of functional OGT causes lethality and developmental arrest in most metazoan model organisms, including mice, frogs, zebrafish, and fruit flies (19, 22, 24, 61). To probe if the N567K mutation directly affects OGT catalytic activity in vivo, we exploited the genetically tractable system D. melanogaster, where phenotypes arisen due to null and hypomorphic alleles of sxc (the fly Ogt ortholog) have been described (62, 63). Drosophila sxc null mutants (sxc− and sxc−/−) die at the pharate adult stage with homeotic transformation defects (62). A hypomorph allele, sxcH537A, having substantially reduced OGT catalytic activity (64), shows significantly decreased protein O-GlcNAcylation levels (63) yet produce fertile offspring. Crucially, the fruit fly provides a platform to dissect the different effects of the OGT N567K mutation. In the fly, feedback between proteome O-GlcNAcylation levels and OGA/OGT protein/mRNA levels has not been observed. Furthermore, although Drosophila Hcf1 is heavily O-GlcNAc modified, proteolytic processing is catalyzed by the Taspase I protease instead of OGT (65).

We introduced the OGT N567K equivalent mutation (N595K) into D. melanogaster using a CRISPR/Cas9 gene-editing toolkit. Vasa::Cas9 embryos were injected in-house with a mixture of plasmids codings for the guide RNA and repair template DNA. In total, 100 potential candidate flies were screened exploiting restriction fragment-length polymorphism arising from the loss of a 250 base pairs upstream and downstream from the desired restriction site (SI Appendix, Fig. S3). We recovered 2 independent knockout lines (10.3 and 19.1) carrying the mutation, sxc\textsuperscript{N595K}. Presence of the mutation was confirmed by sequencing the region ~250 base pairs upstream and downstream from the mutation. Additionally, the lines were further validated by sequencing the full-length sxc mRNA. Both lines developed to adulthood without apparent homeotic defects. Transheterozygous sxc\textsuperscript{N595K}/sxc\textsuperscript{H537A}, sxc\textsuperscript{N595K}/sxc\textsuperscript{N595K-19.1}, and sxc\textsuperscript{N595K}/sxc\textsuperscript{N595K-19.1} flies with presumably further reduced OGT activity were viable (SI Appendix, Fig. S4).

To probe the effect of the N595K mutation on OGT enzymatic activity and stability in vivo, we subjected adult fly head and embryo lysates to Western blotting with an O-GlcNAc antibody, RL2 (Fig. 3A, B, and C). Drosophila sxc mutants displayed significant reduction of O-GlcNAcylation (40). 1 N567K- and N595K-representative flies with substantially reduced OGT catalytic activity (64), shows significantly decreased protein O-GlcNAcylation levels (63) yet produce fertile offspring. Crucially, the fruit fly provides a platform to dissect the different effects of the OGT N567K mutation. In the fly, feedback between proteome O-GlcNAcylation levels and OGA/OGT protein/mRNA levels has not been observed. Furthermore, although Drosophila Hcf1 is heavily O-GlcNAc modified, proteolytic processing is catalyzed by the Taspase I protease instead of OGT (65).

The N567K Mutation Leads to Misprocessing of HCF1. We next attempted to uncover the molecular mechanisms underpinning the observed differentiation phenotypes. We tested the levels of protein O-GlcNAcylation using Western blotting (Fig. 4A). Surprisingly, levels of protein O-GlcNAcylation in pluripotent mESCs appeared comparable in 3HA-OGT\textsuperscript{N567K} and 3HA-OGT\textsuperscript{WT} cells. Previous studies have shown that (an as yet to be discovered) feedback mechanism regulates OGA and OGT expression levels in response to protein O-GlcNAcylation levels (42, 43, 66, 67). Indeed, while OGT levels appeared to be unaltered, OGA levels were abrogated in 3HA-OGT\textsuperscript{N567K} pluripotent mESCs (Fig. 4B), presumably to compensate for reduced OGT activity, but appeared normal during differentiation (Fig. 3C and D).

Next, we investigated the effects of the N567K mutation on the posttranslational modification and cellular localization of HCF1 in pluripotent mESCs. Western blot analysis on cytoplasmic and nuclear fractions revealed an elevated ratio of full-length versus total HCF1 in the nuclear fractions of 3HA-OGT\textsuperscript{N567K} compared with 3HA-OGT\textsuperscript{WT} mESCs, while the levels of HCF1 proteolytic products were reduced (Fig. 4C and D). This is in agreement with previous published male patients carrying OGT mutations in the TPR domain, all suffered from neurodevelopmental delay. We next explored the early events of neuronal development by studying effects on pluriptopy and differentiation in mESCs, a cellular system amenable to genetic modification. We generated male mESC lines expressing C-terminal triple hemagglutinin (3HA) -tagged version of wild-type OGT (3HA-OGT\textsuperscript{WT}) and carrying the N567K mutation (3HA-OGT\textsuperscript{N567K}) using 2 rounds of CRISPR/ Cas9-mediated gene editing at the endogenous locus (SI Appendix, Figs. S6 and S7). At each step, at least 2 clones with the correct genotype were isolated and selected for further analysis.

Next, we examined whether neuronal differentiation was altered in 3HA-OGT\textsuperscript{N567K} cells. Expression of pluriptopy markers Sox2 and Oct4 was unaffected during the first 6 d of differentiation (Fig. 3C and D). Neurite outgrowth was assessed after 8 d. Neurite length was significantly reduced in 3HA-OGT\textsuperscript{N567K} cells compared with 3HA-OGT\textsuperscript{WT} control cells immune-labeled against the neuron-specific β-tubulin III protein (Fig. 3E and F). Thus, the N567K mutation leads to an mESC neuronal differentiation phenotype.

Table 2. X-ray diffraction data processing and refinement statistics of the OGT\textsubscript{N567K} ternary complex

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Cell dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>F222</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>137.7, 150.6, 200.0</td>
</tr>
<tr>
<td>A, B, C (Å)</td>
<td>90.0, 90.0, 90.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>45.31–2.17 (2.17–2.23)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.07 (0.91)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (99)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.2 (5.3)</td>
</tr>
</tbody>
</table>

Refinement

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Cell dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>45.31–2.17 (2.17–2.23)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>286,887 (23,302)</td>
</tr>
<tr>
<td>Rwork/Rfree</td>
<td>0.20/0.25</td>
</tr>
</tbody>
</table>

| No. atoms       | Protein           | 5,530 |
|                 | Nucleotide sugar  | 39   |
|                 | Peptide           | 35   |
|                 | Water             | 328  |

B-factors

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Cell dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>53.2</td>
</tr>
<tr>
<td>Nucleotide sugar</td>
<td>46.6</td>
</tr>
<tr>
<td>Peptide</td>
<td>91.9</td>
</tr>
<tr>
<td>Water</td>
<td>51.3</td>
</tr>
</tbody>
</table>

Rmsd

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Cell dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond lengths (Å)</td>
<td>0.02</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Numbers in brackets show the highest resolution bin.
Fig. 3. The N567K OGT mutation leads to altered O-GlcNAc homeostasis in D. melanogaster and neurite outgrowth defect in mESCs. (A) Immunoblot on 3- to 5-d-old adult Drosophila head lysates. (B) Quantification of O-GlcNAc modification on proteins and OGT protein level, normalized to actin signal. Protein O-GlcNAc modification is decreased in sxcN595K 19.1 (***P < 0.0001) and sxcN595K 10.3 (***P < 0.0001) lines to a comparable level found in the hypomorphic sxcH537A flies (***P < 0.0001). n = 4, mean ± SD. ANOVA with Tukey test. (C) Immunoblots on full cell lysate of 3HA-OGTWT and 3HA-OGTN567K mESCs showing OGA, Oct4, Sox2, Histone 3 (H3), and protein O-GlcNAcylation (RL2) levels. mESCs were differentiated for 0, 2, 4, and 6 d in N2B27 medium. (D) Quantification of immunoblots of OGA, Oct4, Sox2, and global protein O-GlcNAcylation in differentiated 3HA-OGTWT and 3HA-OGTN567K mESCs normalized to H3 signal. n = 4, mean ± SEM; **P = 0.012. Multiple t test using the Holm–Sidak method. (E) Immunofluorescence microscopy of 3HA-OGTWT and 3HA-OGTN567K mESCs after 8 d of neuronal differentiation in N2B27 medium. Cells were stained for tubulin III (green) and DAPI (magenta). (Scale bar, 50 μm.) (F) Quantification of neurite outgrowth. Total neurite length was measured per microscopic image by tracking β-tubulin III signal and then normalized to the number of nuclei (DAPI). (3HA-OGTWT, 108 to 800 nuclei per image; 3HA-OGTN567K, 260 to 740 nuclei per image) **P = 0.0079, Student t test, mean ± SD, n = 3.
the impaired OGT-mediated proteolytic cleavage of HCF1 observed in vitro (Fig. 2C). HCF1 contributes to the regulation of transcriptional programs associated with controlling cell cycle and pluripotency (68, 69). Interestingly, the expression of some of the HCF1 interactors are also regulated by HCF1 itself (51, 66, 69). As an initial attempt to uncover possible candidate genes affected by HCF1 misprocessing, we investigated the expression of known HCF1 partners by qRT-PCR (Fig. 4E). The Ets transcription factor GA-binding protein subunit α (GABPA), a component of the nuclear respiratory factor-2 (NRF2) complex, showed a 2.3-fold increase in mRNA level in 3HA-OGT

differentiation from cobalamin deficiency and expressed a much milder form of congenital anomalies than the severe HCF1 group, suggesting that a degree of HCF1 activity is retained. Interestingly, we observed an increase in full-length HCF1 protein level and decreased level of the mature proteolytic fragment in nuclei, showing that HCF1 processing is affected in cellulo, albeit with retention of some activity.

Surprisingly, the expression of the transcriptional factor GABPA downstream of HCF1 was increased indicating that the N567K mutation in OGT could translate into moderate molecular changes downstream of HCF1. The opposite effect, decreased expression of GABPA, was detected upon overexpression of the HCF1 N-terminal fragment (HCF1n) (68). GABPA belongs to the E-26 family of DNA binding factors and regulates expression of several proteins required for mitochondrial DNA transcription and replication (85). Moreover, GABPA modulates the expression of genes involved in cell cycle control, apoptosis, differentiation, and hormonal regulation. Its importance in early embryonic development was demonstrated in knockout mouse studies (86). However, there is no established link between GABPA and neurodevelopment. Furthermore, expression of several other transcriptional factors targeted by HCF1—such as Creb1, E2F4, Krox-2, Sin3a, Thap2, and Zfp143—were found unaltered, suggesting that partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.

While OGT is critical for mESC maintenance of pluripotency and differentiation, we have detected no difference at the level of pluriptomy markers, Sox2 and Oct4, between 3HA-OGT

partial abrogation of HCF1 cleavage may affect just a small subset of targets, whose expression is influenced by the altered ratio of proteolytically cleaved and uncleaved HCF1.
**Fig. 4.** N567K mutation leads to defects in HCF1 processing and downstream gene expression in 3HA-OGT<sup>N567K</sup> pluripotent mESCs. (A) Immunoblots showing OGT, OGA, and protein O-GlcNAcylation (RL2) levels in 3HA-OGT<sup>WT</sup> and 3HA-OGT<sup>N567K</sup> undifferentiated mESCs. (B) Quantification of immunoblots of global protein O-GlcNAcylation, OGA, and OGT protein level in undifferentiated mESCs normalized to tubulin signal. n = 3, mean ± SD. Unpaired t-test. **P = 0.0049. (C) 3HA-OGT<sup>N567K</sup> mESCs show impaired HCF1 proteolytic cleavage. Western blot analysis for HCF1, tubulin (cytoplasmic marker), and Laminin B1 (nuclear marker) indicates decreased level of proteolytic fragments of HCF1 in 3HA-OGT<sup>N567K</sup> compared with 3HA-OGT<sup>WT</sup> mESCs. (D) Quantification of immunoblots on nuclear and cytoplasmic fractions of undifferentiated 3HA-OGT<sup>N567K</sup> mESCs compared with 3HA-OGT<sup>WT</sup> mESCs control line showing HCF1 levels. HCF1 full-length (FL) and proteolytic fragments (PF) normalized to tubulin signal for the cytosolic fraction and to Lamin B1 signal for nuclear fraction. n = 3, mean ± SD. Multiple t-test. *P = 0.0189 for FL and *P = 0.044 for PF. (E) qPCR analysis of gene expression of HCF1 targets in undifferentiated 3HA-OGT<sup>N567K</sup> and 3HA-OGT<sup>WT</sup> mESCs. Mean ± SEM, n = 3. **P = 0.002.
catalytic activity that affect neurodevelopment resulting in intellectual disability.

**Materials and Methods**

The study was approved by the University of Dundee Institutional Risk Assessment Committee, and 1 of each participants’ parents provided written, informed consent. The DDS study has UK Research Ethics Committee (REC) approval (10/H0305/83, granted by the Cambridge South Research Ethics Committee, and GEN/284/12 granted by the Republic of Ireland REC).

**X-Inactivation Analysis.** To determine if X chromosome inactivation occurred randomly or nonrandomly, the ratio of methylated and unmethylated maternal and paternal X chromosome was determined. (SI Appendix, Fig. S1). Sites are methylated on the inactive X chromosome and unmethylated on the active X chromosome. The method utilizes the androgen receptor sequence that is encoded on the X chromosome (Xq11.2-q12) and contains a highly polymorphic polyglutamine (CAG) trinucleotide repeat in exon 1. There are 20 alleles of androgen receptor, with 11 to 31 repeats reported in the general population and 85 to 90% of patients are heterozygous for this locus, thus suitable for X-inactivation analysis. For experimental procedure, see SI Appendix, SI Materials and Methods.

**Structure Solution.** Crystallization of truncated OGTWTR386 (residues 323–1044) was performed as described previously (57) (for protein expression, purification, and crystallization, see SI Appendix, SI Materials and Methods). The crystal space group was R32 of tricrural unit cell was measured by structure refinement for OGTWT [pdb ID code 5C17 (57)] as the search model. The resulting model was manually refined using Coot (95) and REFMAC (96), respectively. The editing and refinement of the model was iterated until it was in complete agreement with the data. Scaling and model building statistics can be seen in Table 2.

**Enzyme Activity Assays.** Michaelis-Menten kinetics of OGT were measured using a fluorometric assay, as described previously (97), with the exception of reduced reaction volume of 25 μL and usage of 384-well plate. As acceptor substrate, P1 (KVKVPSRA) and P2 (KVKVAVSRA) were used. Additional O-GlcNAcylation assays were performed on TAB1 protein (residues 7–420). Proteolytic assays were performed as described previously (43), except a noncleavable HCF1-rep1 fragment was used instead of OGA treatment as the negative control. For further details, see SI Appendix, SI Materials and Methods.

**mESC Culture.** mESC AW2 line was acquired from the MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, University of Edinburgh (98). mESCs were cultured in an undifferentiated state in 1% (v/v) gelatin (wt/vol)-coated plates in GMEM BHK-21 supplemented with 10% FBS (vol/vol), 0.1 mM MEM nonessential amino acids, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, and 100 units/ml LIF at 37 °C in 5% CO2. For differentiation assays, cells were cultured following a previously published protocol (99). Briefly, cells were cultured in DMEM and F12 media mixed at 1:1 ratio (vol/vol), supplemented with modified N2 (25 μg/mL insulin, 100 μg/mL apotransferrin, 6 μg/mL progesterone, 16 μg/mL putrescine, 30 nM sodium selenite, and 50 μg/mL BSA fraction) and neurobasal supplement B27. Medium was renewed every 2 d. For generation of CRISPR/Cas9 lines, Immuncytometry, Western blotting, and qPCR analyses details are described in SI Appendix, SI Materials and Methods.

**Drosophila Husbandry, Microinjection, and Genetics.** Fly stocks were maintained on Nutri-Fly Bloomingdon Formulation fly food at 25 °C. Vasa::Cas9 (#51323), sxc1, b, sp/ism5 (SM#058) and Df(2R)BSC630/CyO (#2705) stocks were obtained from the Bloomington Drosophila Stock Centre. The sxc hypomorphic allele, sxc5714, was described in ref. 63. The sxc9550 Drosophila lines were generated by CRISPR/Cas9 mutagenesis, as described in SI Appendix, SI Materials and Methods.

**Western Blotting from Drosophila Samples.** Sample preparation of embryos and fly heads and Western blotting were performed as described in SI Appendix, SI Materials and Methods.

**ACKNOWLEDGMENTS.** This work was funded by a Wellcome Trust Investigator Award (110061) (to D.M.F.v.A.). V.M.P. was supported by the SPRINT/MND-MSK Scholarship (University of Edinburgh). The X-ray diffraction experiments were performed on beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Local Contact at the ESRF for providing assistance for using beamline ID23-2. We acknowledge Dr. Louise A. Young for the aid in the X-inactivation analysis and Dr. Andrew Smith for the AW2 mouse embryonic stem cell line. We acknowledge the Deppeachinger Developmental Disorders study, presenting independent research commissioned by the Health Innovation Challenge Fund (grant no. HICF-1009-003), a parallel funding partnership between Wellcome and the Department of Health of the Wellcome/National Institute for Health Research, through the Comprehensive Clinical Research Network. The authors would also like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource.
55. E. Julien, W. Herr, Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. EMBO J. 22, 2360–2369 (2003).