An O-GlcNAc transferase pathogenic variant that affects pluripotent stem cell self-renewal

Michaela Omelková[1], Christina Dühring Fenger[2,3], Marta Murray[1], Trine Bjørg Hammer[2], Veronica M. Pravata[1], Sergio Galan Bartual[1,5], Ignacy Czajewski[1], Allan Bayat[2], Andrew T. Ferenbach[1,5], Marios P. Stavridis[1] and Daan M. F. van Aalten[1,4,5*]

[1]Division of Molecular, Cell and Developmental Biology, School of Life Sciences, University of Dundee, Dundee, UK

[2]Department of Epilepsy Genetics, Filadelfia Danish Epilepsy Centre, Dianalund, Denmark

[3]Amplexa Genetics A/S, Odense, Denmark

[4]Institute of Molecular Precision Medicine, Xiangya Hospital, Central South University, Changsha, China

[5]Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark

*Correspondence to: daan@mbg.au.dk

Key words: OGT, O-GlcNAc, CDG, intellectual disability, stem cells, self-renewal

Summary statement

We show that the C921Y O-GlcNAc transferase variant found in patients with intellectual disability leads to a defect in pluripotent stem cell self-renewal and decreased levels of stem cell markers.

Abstract

O-linked β-N-acetylglucosamine (O-GlcNAc) transferase (OGT) is an essential enzyme that modifies proteins with O-GlcNAc. Inborn *OGT* genetic variants were recently shown to mediate a novel type of Congenital Disorder of Glycosylation (OGT-CDG) which is characterized by X-linked intellectual disability (XLID) and developmental delay. Here, we report an OGT^{C921Y} variant which co-segregates with XLID and epileptic seizures, and results in loss of catalytic activity. Colonies formed by mouse embryonic stem cells carrying OGT^{C921Y} show decreased levels of protein O-GlcNAcylation accompanied by decreased levels of Oct4, Sox2 and extracellular alkaline phosphatase (ALP), implying reduced self-renewal capacity. These data establish a link between OGT-CDG and embryonic stem cell self-renewal, providing a foundation for examining the developmental aetiology of this syndrome.

Introduction

Intellectual disability (ID) is a cognitive handicap that affects up to 1 % of the general population (Maulik et al., 2011; McKenzie et al., 2016). ID is characterized by reduced cognitive function (IQ < 70) and adaptive behaviour diagnosed before the 18th year of life (Tassé et al., 2016). It is estimated that 40 % of ID cases can be attributed to genetic causes (Posada De La Paz et al., 2017). To date, > 2000 genes have been implicated in ID aetiology (Firth and Wright, 2011; PanelApp Genomics England, 2020) and 141 of these are located on the X chromosome (Tejada and Ibarluzea, 2020), which encodes a disproportionate number of genes involved in cognitive processes (Lubs et al., 2012; Skuse, 2005; Zechner et al., 2001). Indeed, due to hemizygosity for X-linked genes, more males than females suffer from ID (Posada De La Paz et al., 2017). Based on aetiology and clinical presentation, ID comprises a heterogenous group of neurodevelopmental conditions that can either be syndromic (i.e., a part of a clinically defined set of symptoms forming a syndrome) or non-syndromic.

Congenital Disorders of Glycosylation (CDG) is an umbrella term for inborn defects in glycosylation enzymes that lead to ID in 60 – 70 % of patients (Wolfe and Krasnewich, 2013). Recently, a novel type of CDG (OGT-CDG) was diagnosed in patients carrying pathogenic variants in O-linked β-*N*-acetylglucosamine (O-GlcNAc) transferase (*OGT*) (Pravata et al., 2020b). To date, 17 pathogenic OGT-CDG variants have been identified through whole genome sequencing, and several of these have been characterized clinically and biochemically (Pravata et al., 2019; Pravata et al., 2020a; Selvan et al., 2018; Vaidyanathan et al., 2017; Willems et al., 2017). In addition to ID and delayed achievement of developmental milestones, patients with OGT-CDG commonly present with musculoskeletal problems, brain malformations, eye abnormalities, dysmorphic features and language problems (Pravata et al., 2020b). Furthermore, heart anomalies, immune system defects, genital abnormalities and digestive problems have also been reported in individual patients. OGT-CDG thus appears to affect the development and function of multiple organ systems. The mechanisms through which OGT variants mediate ID remain to be established, although several hypotheses have been put forward (Pravata et al., 2020b).

OGT is a 117 kDa enzyme encoded by the *OGT* gene located on the X chromosome and expressed ubiquitously in human tissues (Kreppel et al., 1997; Lubas et al., 1997). OGT catalyses O-GlcNAcylation, an essential, dynamic modification of nucleocytoplasmic proteins with a single *O*-linked GlcNAc sugar moiety on serine and threonine residues (Kreppel et al., 1997; Torres and Hart, 1984). As opposed to other post-translational modifications that are generally mediated by protein families, O-GlcNAcylation is controlled by only two enzymes with opposing function, OGT and O-GlcNAc hydrolase (OGA), that attach and remove the GlcNAc moiety, respectively (Gao et al., 2001; Wells et al., 2002). The sugar nucleotide donor for O-GlcNAcylation, UDP-GlcNAc, is synthesized from glucose by the hexosamine biosynthetic pathway and thus O-GlcNAcylation is thought to serve as a nutrient sensor (Hart, 2015; Medford et al., 2013). Since 1984, when O-GlcNAcylation was first discovered in lymphocytes (Torres and Hart, 1984), modification of O-GlcNAc proteins has been shown to be ubiquitous across all human tissues (Wulff-Fuentes et al., 2021). O-GlcNAcylation has been implicated in a range of biological mechanisms such as cellular metabolism (Slawson et al., 2010), gene expression and regulation (Krause et al., 2018; Sakabe et al., 2010; Streubel et al., 2017; Tan et al., 2021), cell division (Drougat et al., 2012; Lefebvre et al., 2004; Leturcq and Lefebvre,

2017; Sakabe and Hart, 2010) and stem cell differentiation (Elena-Herrmann et al., 2020; Hao et al., 2019; Zhang et al., 2019). Moreover, O-GlcNAcylation is considered essential for survival as embryonic ablation of *Ogt* in mice (Shafi et al., 2000) and zebrafish (Webster et al., 2009) is lethal, while loss of Oga activity leads to perinatal death in mice (Keembiyehetty et al., 2015; Muha et al., 2021; Yang et al., 2012).

OGT is composed of 13.5 N-terminal tetratricopeptide repeats (TPRs) that mediate substrate recognition and protein - protein interactions (Iyer and Hart, 2003; Jínek et al., 2004; Rafie et al., 2017) and a globular catalytic domain that catalyses O-GlcNAcylation. In addition to acting as an essential glycosylation enzyme, OGT also proteolytically activates transcriptional co-regulator Host Cell Factor 1 (HCF1) (Capotosti et al., 2011). HCF1 is heavily glycosylated and cleaved by OGT in the same active site, however, glycosylation and proteolysis occur through separate mechanisms (Kapuria et al., 2018; Lazarus et al., 2013). Data suggest that HCF1 may bind a large proportion of promoters in the human genome, acting as an important cell cycle progression and mitochondrial biogenesis regulator (Michaud et al., 2013). Interestingly, *HCFC1* is itself a well-known XLID gene, mutations of which are thought to alter neural stem cell maintenance and differentiation (Castro et al., 2020; Huang et al., 2012; Jolly et al., 2015).

There is growing evidence to suggest that O-GlcNAcylation is essential for embryonic stem cell (ESC) maintenance (Shafi et al., 2000) and embryogenesis (Jang et al., 2012; Shafi et al., 2000; Webster et al., 2009; Yang et al., 2012; Zhang et al., 2019; Zhu et al., 2020). Cellular O-GlcNAcylation decreases during neuronal differentiation (Liu et al., 2012), suggesting a link between dynamic O-GlcNAc cycling and tissue development. Furthermore, OGT O-GlcNAcylates transcription factors that play an important role in pluripotency and stem cell function, including Octamer binding transcription factor 4 (Oct4) (Constable et al., 2017), SRY (sex determining region Y)-box 2 (Sox2) and Signal transducer and activator of transcription 3 (STAT3) (Li et al., 2017). STAT3 is a member of the JAK/STAT signalling pathway that is activated by pluripotency stimulus Leukemia Inhibiting Factor (LIF) in mouse ESCs (mESCs). Here, we report three brothers affected by OGT-CDG that carry a novel variant, OGT^{C921Y}, in the catalytic core of OGT that is absent in their healthy brother. The variant has not been previously reported, neither in the healthy population (gnomAD database) nor in patients (HGMD database). We show that the OGT^{C921Y} variant possesses decreased glycosyltransferase activity *in vitro* and in an mESC model. Strikingly, a knock-in of OGT^{C921Y} results in abrogated self-renewal in mESCs as shown by decreased expression of alkaline phosphatase (ALP), Oct4 and Sox2 during a clonogenic assay. These results suggest that the role of OGT in ESC self-renewal and pluripotency may contribute to clinical signs seen in OGT-CDG patients.

Results and Discussion

Three brothers with intellectual disability carry an inherited catalytic OGT variant absent in their healthy brother

Four male siblings were born to a healthy non-consanguineous couple shown as family members II.5 and II.6 on the family pedigree (Fig. 1A). The first child was born in 1960 and is healthy. The second child, currently a 58-year-old male shown as proband III:2 on the family pedigree, was born at 38 weeks of gestation after a normal pregnancy. Birth weight was 2950 gram (2^{nd} - 9^{th} centile) and birth length was 52 cm (50^{th} - 75^{th} centile). Apgar scores at birth are unknown. There were no concerns following birth and the patient was discharged. Following the first two years of life, he showed delay in reaching developmental milestones, especially in areas of speech and language development. He learned his first words at the age of 14 months, and he is now able to speak in short sentences with five to seven words and understand simple instructions. At the age of twelve months, he could walk and reach out with palmar grasp, transfer objects, and put them into his mouth. Later in life, the patient presented with autistic features. He attended a special needs school, and he currently lives in a sheltered home. He underwent an operation for an inguinal hernia at the age of 48 years. Brain MRI was never performed. He has experienced at least two generalised tonic clonic seizures. The first seizure occurred around the age of 40 years. At seizure onset, his EEG showed background slowing, but without any clear interictal epileptiform abnormalities. After the second seizure, a daily treatment with oxcarbazepine was prescribed and he has since had a good seizure control. Dysmorphic features include an oval face, narrow, long and pear-shaped nose with a high nasal bridge, thin upper lip, high arched palate, large ears, sparse eyebrows, thin and short fingers with distal squaring (Fig. 1B, Fig. S1). At the age of 58, the patient has osteoporosis and scoliosis. He has been diagnosed with both a short stature (166 cm, - 1.48 SD) and a head circumference of 55.3 cm (+ 0.12 SD). Cranial nerve examination was normal. Limb examination showed neither rigidity nor tremors. The power in the limbs was five out of five (MRC Scale for Muscle Strength) and the deep tendon reflexes were normal. He presented with a slow, shuffling gait.

The mother later gave birth to monozygotic male twins, here referred to as twin 1 and twin 2 (III:3 and III:4 in pedigree) at 38 weeks of gestation following a normal pregnancy. Conception was unassisted. Twin 1 (III:3) weighed 2150 gram at birth (< 0.4^{th} centile) and had a birth length of 48cm (2^{nd} - 9^{th} centile), while twin 2 (III:4) weighed 1650 gram (< 0.4^{th} centile) and measured 46 cm (0.4^{th} – 2^{nd} centile). Apgar scores are unavailable. While twin 1 was breastfed and discharged within a week following the birth, twin 2 was tube fed for two weeks. Similar to proband III:2, the twins also showed delay in reaching developmental milestones, especially in areas of speech and language development. Twin 2 never learned to speak and twin 1 is able to pronounce a few words. The twins have autism and communicate using simple signs and by pointing. They are able to understand very simple instructions. Both make high pitched sounds, have repetitive mannerisms and clap their hands together if excited. In addition, twin 2 has self-injurious behaviour. Their motor development was also affected; although both twins learned to walk by the age of 18 months, they were diagnosed with significant gross motor difficulties and clumsiness. They have attended a special needs nursery and school and they currently live in a sheltered home. Neither of the twins underwent a brain MRI. While twin 1 has not had an epileptic seizure, twin 2 has experienced his first generalized tonic clonic seizure

at around 40 years of age. Over the subsequent eleven years, he had a total of four identical epileptic seizures.

At the time of the first seizure, the EEG in twin 2 showed background slowing, but without any clear interictal epileptiform abnormalities. After the second seizure, a daily treatment with carbamazepine was prescribed. Data on dysmorphic features from childhood were not available, but both twins currently present with an oval face, up-slanting palpebral fissures, narrow, long and pear-shaped nose with a high nasal bridge, thin upper lip, hypotrichosis and thin fingers (Fig. 1B, Fig. S1). Twin 2 has also been diagnosed with scoliosis, short stature (165 cm, - 2.1 SD) and osteoporosis. In addition, twin 2 is without any rigidity or tremors of the extremities but has a slow shuffling gait. Limited clinical data is available for twin 1.

The three brothers III:2, III:3 and III:4 thus appeared to suffer from an inherited developmental defect which was not present in their eldest healthy brother (III:1). Family history revealed that two maternal uncles (II.3 and II.4) also suffered from intellectual disability and were wheelchair bound at older age, suggesting an X-linked recessive inheritance pattern. Whole genome sequencing of patient III:2 and subsequent Sanger analysis of the *OGT* gene in samples from all three remaining brothers and their mother were performed. These analyses revealed that the affected siblings III:2, III:3 and III:4 carried a guanine to adenine substitution at 2762nd nucleotide (c.2762G>A) in the *OGT* gene that translates as cysteine to tyrosine substitution at position 921 in OGT (c.2762G>A p.(C921Y), NM_181672.2, OGT^{921Tyr}) inherited from their heterozygous mother (OGT$^{921Cys/Tyr}$). This OGT variant has not been previously reported and it is classified as likely pathogenic according to the ACMG criteria (criteria PM1 PM2 PP1 PP2 PP3 PP4) (Richards et al., 2015). The OGT^{C921Y} variant was absent in the healthy brother (II:1, OGT^{921Cys}), suggesting that this single point mutation in OGT causes intellectual disability in patients II:2, II:3 and II:4. Thus, we have identified three brothers with intellectual disability who carry an inherited catalytic OGT variant absent in healthy brother. The clinical observations are consistent with previous clinical descriptions of OGT-CDG patients and add considerable knowledge to the currently limited information on adult OGT-CDG patients (Pravata et al., 2020b). For example, despite an established association between decreased O-GlcNAcylation and Alzheimer's disease (Liu et al., 2004; Liu et al., 2009; Park et al., 2021), these patients do not present with signs of age-related neurodegenerative disorders.

OGT^{C921Y} is defective in glycosyltransferase activity toward protein substrates in vitro

Disrupted OGT stability and folding caused by pathogenic missense variants could contribute to OGT-CDG pathophysiology (Pravata et al., 2020b). Therefore, the effect of C921Y substitution on the OGT protein structure was investigated. Although we were able to produce this variant in recombinant form from an *E. coli* expression system, we were unable to grow crystals for structural analysis, unlike several of the previously reported variants (Gundogdu et al., 2018; Pravata et al., 2019; Pravata et al., 2020a; Selvan et al., 2017; Vaidyanathan et al., 2017). Nevertheless, analysis of the wild type OGT structure reveals that C921 resides in the catalytic domain of OGT, proximal to the UDP-GlcNAc binding site (~13 Å, Fig. 2A). The mutation of C921 to a bulky tyrosine residue would disrupt the C845-C921 disulfide bridge found in wild type OGT. Furthermore, C921Y could perturb the position of H911, which in the wild type enzyme establishes π–π stacking interactions with the UDP uracil ring (Fig. S2 A&B). The mutation could also affect N935, induce the

loss of the Y851-UDP interaction and promote the generation of a new interaction between the Y931 and UDP (Fig. S2 A&B). Thus, the C921Y substitution could affect protein stability, interfere with catalysis and/or alter binding of UDP-GlcNAc to OGT.

We next investigated whether the C921Y mutation disrupts stability of recombinant OGT (323 – 1044aa) *in vitro* using a thermal denaturation assay. No change in melting temperature between truncated OGT^{WT} (T_m = 45 ± 1 ºC) and OGT^{C921Y} (T_m = 45 ± 1 ºC) was observed (n = 3 replicates, each consisting of three technical repeats, Fig. S3). Next, we examined the impact of the C921Y mutation on the glycosyltransferase activity of full-length OGT. We first used a steady-state kinetics assay with varying concentrations of sugar donor UDP-GlcNAc against an established acceptor peptide (Ac-KENSPAVTPVSTA-NH_2,(Pathak et al., 2015)) (Fig. 2B). The maximal reaction velocity (V_{max}) and the Michaelis Menten constant (K_m) were reduced for OGT^{C921Y} (V_{max} = 0.70 x 10^{-3} ± 0.01 µmol/s, K_m = 14 ± 4 µM) compared to OGT^{WT} (V_{max} = 2.8 x 10^{-3} ± 0.2 µmol/s, K_m = 78 ± 15 µM). However, the catalytic efficiency (k_{cat} / K_m) of the mutant and wild type enzyme were similar (OGT^{WT} k_{cat} / K_m = 0.07, OGT^{C921Y} k_{cat} / K_m = 0.10, n = 3 replicates, each consisting of three technical repeats). Thus, the catalytic activity of OGT^{C921Y} towards a peptide substrate is intact compared to OGT^{WT}.

We next evaluated effects of the mutation on OGT activity towards TAK1-binding protein 1 (TAB1), a well characterised protein substrate (Pathak et al., 2012; Rafie et al., 2017), revealing a loss of OGT^{C921Y} catalytic activity compared to OGT^{WT} (Fig. 2C). Western blot analysis of TAB1 O-GlcNAcylation showed that the time dependent increase in TAB1 O-GlcNAc signal produced by OGT^{WT} was absent in the OGT^{C921Y} catalyzed reaction (Fig. 2D). The mean level of TAB1 O-GlcNAcylation catalyzed by OGT^{C921Y} after 6 h was 28-fold lower than that catalyzed by OGT^{WT} (Two-way Anova with Sidak's multiple comparison testing, p < 0.0001).

These data reveal that while OGT^{C921Y} catalytic activity towards protein substrates is abrogated, it appears intact towards peptides, implying that disruption of substrate interactions with the catalytic domain of OGT caused by the C921Y mutation prevents accommodation of large acceptor substrates. Interestingly, this suggests a previously unappreciated role of catalytic domain residues in recognition of protein substrate features beyond the site of O-GlcNAc modification. Together, these data show that OGT^{C921Y} is defective in glycosyltransferase activity toward protein substrates *in vitro*.

OGT^{C921Y} disrupts O-GlcNAc homeostasis in undifferentiated mESCs

To investigate the impact of the OGT^{C921Y} variant on O-GlcNAc cycling, we engineered male OGT^{C921Y} mESC lines using CRISPR/Cas9. The murine and human OGT protein sequences are identical in the mutated region (Fig. S4A). Three independent CRISPR/Cas9 lines carrying OGT^{C921Y} were generated through clonal expansion from three distinct founder cells (Fig. S4B). OGT^{C921Y} mESC lines were morphologically identical to wild type cells and their cell cycle was not affected (Fig. S5). mRNA and protein levels of key pluripotency factors Oct4 and Sox2 remained the same as in OGT^{WT} mESC (Fig. S6). All the following experiments were performed using three different cell clones per genotype and repeated over multiple passages, unless stated otherwise.

Western blot analyses from pluripotent OGT^{C921Y} mESCs revealed a significant decrease in global O-GlcNAc levels compared to wild type cells (Figs. 3A,B, OGT^{WT} n = 15 biological replicates, OGT^{C921Y} n = 13

biological replicates, Unpaired t test, p = 0.0072), suggesting an alteration in O-GlcNAc homeostasis. We detected a significant upregulation of Ogt protein levels in OGT^{C921Y} cells compared to the wild type (Fig. 3A & C, OGT^{WT} n = 15 biological replicates, OGT^{C921Y} n = 13 biological replicates, Unpaired t test, p < 0.0001). However, given the decreased global O-GlcNAcylation levels in the mutant mESC lines, these data suggest that, despite being present at higher levels, OGT^{C921Y} glycosyltransferase activity is not sufficient to maintain O-GlcNAc homeostasis in mESCs. Oga protein levels were unchanged in the mutant cell line (Figs. 3A,D, OGT^{WT} n = 15 biological replicates, OGT^{C921Y} n = 13 biological replicates, Unpaired t test, p = 0.14), whereas previously characterized OGT-CDG variants showed a reduction in OGA protein levels when modelled in mESCs or patient-derived lymphoblastoids whilst OGT levels remained unchanged compared to wild type (Pravata et al., 2019; Pravata et al., 2020a; Vaidyanathan et al., 2017).

We next evaluated *Ogt* and *Oga* expression with RT-PCR analysis of three biological replicates with up to three OGT^{C921Y} and OGT^{WT} lines. These experiments showed that *Ogt* mRNA expression is unchanged in OGT^{C921Y} lines compared to wild type (Fig. 3E, n = 3 RT-PCR runs, each consisting of two to three biological replicates, Two-way Anova, p = 0.82), implying that Ogt may be stabilised at the protein level in OGT^{C921Y} mESCs, possibly through decreased protein degradation in cultured cells. Qian and colleagues (2018) revealed a bidirectional feedback mechanism between Ogt and Oga at the transcriptional level in primary mouse hepatocytes using an overexpression system. Furthermore, OGT has been shown to direct the transcriptional repressor Sin3A-HDAC1 complex to the *OGA* promoter in OGT-CDG patient-derived lymphoblastoids, thus modulating *OGA* expression (Vaidyanathan et al., 2017). Indeed, *Oga* mRNA levels were significantly decreased in OGT^{C921Y} mESCs (Fig. 3F, n = 3 RT-PCR runs, each consisting of two to three biological replicates, unpaired t test, p = 0.0007) despite protein levels being unchanged. Taken together, these data reveal that OGT^{C921Y} disrupts O-GlcNAc homeostasis in pluripotent mESCs despite compensatory changes in the levels of Ogt protein and *Oga* expression.

HCF1 processing is unchanged in pluripotent OGT^{C921Y} mESCs

In addition to catalyzing O-GlcNAcylation, OGT also proteolytically processes and activates the transcriptional coregulator HCF1 (Capotosti et al., 2011). HCF1 is encoded by *HCFC1*, which itself has been reported to be an intellectual disability gene (Huang et al., 2012; Wongkittichote et al., 2021). Biochemical analysis of two previously reported OGT–CDG variants showed decreased HCF1 proteolysis (Pravata et al., 2019; Willems et al., 2017). HCF1 proteolysis occurs within the same active site of OGT as O-GlcNAcylation (Kapuria et al., 2016; Lazarus et al., 2013). Given the proximity of the C921Y substitution to the OGT active site (Figs. 2A, S1) and the impact of the C921Y variant on OGT glycosyltransferase activity, we investigated whether this variant leads to deficient HCF1 O-GlcNAcylation and processing. First, we performed an *in vitro* HCF1 repeat 1 (HCF1rep1) cleavage assay using full length recombinant hOGT^{WT} or hOGT^{C921Y} (Fig. S7). The uncleavable mutant HCF1rep1^{E10D} was used in this assay as a negative control to allow us to distinguish between unspecific HCF1 degradation and *bona fide* proteolytic products (Fig. S7). In agreement with the loss of O-GlcNAcylation activity on TAB1, we observed decreased levels of O-GlcNAcylation of HCF1rep1^{WT} and HCF1rep1^{E10D} by OGT^{C921Y} (Fig. 4A). However, both the wild type and mutant OGT were able to catalyse the formation of HCF1 proteolytic products (Fig. 4A). To corroborate this further, we investigated HCF1 cleavage in the mESC model (Fig. 4B). Since HCF1 translocates to the nucleus, the abundance of HCF1

proteolytic fragments was inspected in both the cytoplasmic and nuclear fractions of undifferentiated mESCs (Fig. 4B). We observed no difference in HCF1 signal between wild type OGT and OGT^{C921Y} cells in either of the cellular compartments (Fig. 4C, n = 6 biological replicates, One-way Anova with Tukey comparison test, cytoplasmic fraction $p = 0.99$, nuclear fraction $p = 0.54$). RT-PCR of HCF1 revealed that HCF1 mRNA levels remain stable in OGT^{C921Y} mESCs (Fig. 4D, n = 3 RT-PCR runs, each consisting of two to three biological replicates, unpaired *t* test, p value = 0.172). Taken together, these data show that HCF1 processing is unchanged in undifferentiated OGT^{C921Y} mESCs.

OGT^{C921Y} abrogates mESC self-renewal capacity

Analogous to previously reported OGT-CDG mutations (Pravata et al., 2020b), patients carrying the OGT^{C921Y} variant present with a broad array of phenotypes including developmental delay, brain abnormalities and musculoskeletal defects. Embryonic development and patterning are crucially dependent on the ability of the inner cell mass of the early embryo to respond to the presence of differentiation stimuli or the absence of self-renewal stimuli. The ability of the inner cell mass as well as cultured ESCs to develop into the three primary germ cell layers is called pluripotency. Pluripotent cells maintain their identity through self-renewal. In the context of embryonic development, self-renewal and differentiation are meticulously orchestrated by various signalling molecules. In cell culture, mESCs are commonly propagated through LIF supplementation in the growth media (Smith et al., 1988; Williams et al., 1988). Upon LIF withdrawal, mESCs lose expression of pluripotency markers such as Oct4 and Nanog, colonies acquire a flattened morphology, and cells begin to differentiate (Chen et al., 2015; Cherepkova et al., 2016; He et al., 2017). O-GlcNAcylation has been previously shown to be important for core and auxiliary pluripotency factor function (Constable et al., 2017; Hao et al., 2019; Jang et al., 2012; Kim et al., 2021; Myers et al., 2016). Therefore, we hypothesized that the OGT^{C921Y} variant may impact the ability of stem cells to maintain an undifferentiated state. To test this hypothesis, we assayed colony formation and maintenance of stemness upon LIF withdrawal.

OGT^{C921Y} and wild type mESCs were subjected to clonogenic conditions in presence of LIF (positive control) or followed by LIF withdrawal for 24-96 h before fixing and alkaline phosphatase (ALP) staining (Fig. 5A). High expression of ALP is considered to be a marker of pluripotency, with ALP staining visualising pluripotent colonies (Štefková et al., 2015). The ALP-stained colonies were examined under a light microscope and classified into three categories (undifferentiated, mixed and differentiated) based on the intensity of ALP stain and colony morphology. Examples of colony scoring are shown in Fig. S8. In the wild type cell lines, the percentage of differentiated colonies increased after LIF withdrawal (Figs. 5B,C, n = 9 biological replicates, one-way Anova with Dunnett's multiple comparison test, $p < 0.0001$) and the percentage of undifferentiated colonies significantly decreased with time of culture in absence of LIF (Figs. 5B,D, n = 9 biological replicates, one-way Anova with Dunnett's multiple comparison test, $p = 0.012$). However, the percentages of undifferentiated and differentiated mutant OGT^{C921Y} colonies did not change significantly in relation to the length of LIF deprivation (Figs. 5B-D, n = 9 biological replicates, one-way Anova with Dunnett's multiple comparison test, $p = 0.10$ and $p = 0.29$ respectively). We noted that OGT^{C921Y} mESCs produced significantly higher percentage of differentiated colonies than wild type mESCs at every LIF withdrawal time point apart from 96 h (Fig. 5B,C, n = 9 biological replicates, two-way Anova, 24 h, $p < 0.0001$, 48 h $p = 0.02$, 72 h $p =$

0.004, 96 h p = 0.13). Remarkably, OGT^{C921Y} colonies showed a significantly increased number of differentiated colonies even in presence of LIF (n = 9 biological replicates, two-way Anova, p = 0.0001) compared to wild type. The number of mixed colonies remained constant over the course of the assay in both wild type and OGT^{C921Y} mESCs. Furthermore, the overall number of colonies formed by OGT^{C921Y} mESCs was significantly lower than for wild type mESCs (Fig. S9, n = 45 wells scored, unpaired t-test, p < 0.0001). This observation can be explained by a decrease in clonogenic potential of OGT^{C921Y} compared to wild type, and/or increased cell death following cell plating at limiting density. Lastly, OGT^{C921Y} mESCs formed colonies with significantly larger surface area than wild type cells (Fig. S9, OGT^{WT} n = 2270 colonies measured, OGT^{C921Y} n = 1533 colonies measured, unpaired t test, p < 0.0001; observations were pooled together from all three independent experiments performed using three clones per genotype). Given the loss of ALP staining in OGT^{C921Y}, the observed increase in surface area may reflect flattening and concomitant expansion of OGT^{C921Y} colonies due to differentiation.

To further corroborate the O-GlcNAcylation status and enzymatic activity of OGT^{C921Y} mESCs under clonogenic conditions, we sought to investigate the levels of global protein O-GlcNAcylation and protein levels of key pluripotency transcription factors (Sox2 and Oct4). OGT^{C921Y} and OGT^{WT} mESCs were cultured for 7 days under clonogenic conditions in the presence of LIF (Fig. 6A). As opposed to the decreased glycosyltransferase activity of the OGT^{C921Y} variant *in vitro* and in confluent mESCs, the decrease in global O-GlcNAcylation in OGT^{C921Y} colonies compared to OGT^{WT} colonies was not significant (Fig. 6B,D, OGT^{WT} n = 8 biological replicates, OGT^{C921Y} n = 9 biological replicates, unpaired t test, p < 0.07) despite increased levels of the Ogt protein (Fig. 6B,C, n = 9 biological replicates, unpaired t test, p = 0.0004). Previous reports indicated that global O-GlcNAcylation is highest in pluripotent stem cells and decreases as pluripotency is lost (Liu et al., 2012; Sheikh et al., 2020). Our observation may therefore reflect the shift from predominantly pluripotent stem cell population present in confluent mESCs propagated in LIF to a mixed population of cells present at day six of the clonogenic assay (Fig. 5C-E). Furthermore, Western blot analysis of lysates derived from OGT^{C921Y} colonies revealed a significant decrease of Oct4 (Fig. 6B,G, n = 9 biological replicates, unpaired t test, p = 0.004) and Sox2 (Fig. 6B,F, n = 9 biological replicates, unpaired t test, p = 0.002) levels compared to OGT^{WT}. This finding was corroborated by immunofluorescence staining of OGT^{WT} and OGT^{C921Y} colonies (Fig. S10). It is worth noting that we observed a variability in Sox2 and Oct4 protein levels among the three lines of OGT^{C921Y} mESCs, possibly stemming from the fact that each cell line is derived from a separate CRISPR/Cas9 event, therefore representing three cell populations that arose from three different founder cells. To discern whether differentiating OGT^{C921Y} mESCs assume specific germ layer preferentially, colonies were immunolabelled for markers specific for the mesoderm (Brachyury), endoderm (sox17) and ectoderm (pax6). There was no difference in germ layer marker protein expression in the mutant and wild type colonies (Fig. S11, S12, S13 and S14). However, this result is limited by the short timeframe of the assay. Taken together, these data show that independent of LIF signalling, OGT^{C921Y} mESCs show reduced ALP staining under clonogenic conditions compared to OGT^{WT}, suggesting defects in the stem cell renewal. Furthermore, OGT^{C921Y} colonies grown in the presence of LIF show reduced levels of Sox2 and Oct4 transcription factors, implying that OGT^{C921Y} affects self-renewal of mESC under these conditions.

Concluding remarks

Pathogenic missense variants in the *OGT* gene co-segregate with intellectual disability in at least 17 known affected families, seven of which have been described in detail (Bouazzi et al., 2015; Pravata et al., 2019; Pravata et al., 2020a; Selvan et al., 2018; Vaidyanathan et al., 2014; Willems et al., 2017). Through biochemical, cellular, and animal model assays, recent studies revealed that these single *OGT* point mutations (hemizygous or heterozygous) underpin pathogenesis, *via* unknown biological mechanisms. This condition was recently classified as a distinct type of CDG, termed OGT-CDG (Pravata et al., 2020b). Here, we describe a family affected by a C921Y variant in the catalytic domain of OGT. Three male siblings suffering from ID harbour the OGT^{C921Y} variant, while their brother who possesses a wild type copy of *OGT* is healthy. All three affected brothers present with dysmorphic features such as long and pear-shaped nose, as well as behavioural and language problems. These features are in line with the recently reported general clinical phenotype description of OGT-CDG patients (Pravata et al., 2020b). The severity of the OGT^{C921Y} patients' ID varied. Interestingly, the level of ID manifestations was different even between the two twins affected by this novel OGT variant. Similar differences in ID severity were reported by Pravata and colleagues in female monozygotic twins suffering from OGT-CDG (Pravata et al., 2019).

The C921Y mutation is predicted to cause structural defects within the OGT catalytic core, albeit without measurable effects on stability *in vitro*. Interestingly, only two of the OGT-CDG variants biochemically characterized to date (OGT^{N567K} and OGT^{N648Y}) did not lead to changes in OGT stability (Pravata et al., 2019; Pravata et al., 2020a). OGT^{N567K}, OGT^{N648Y} and OGT^{C921Y} are the only three described variants directly affecting the catalytic domain, whereas other OGT-CDG mutations reside in the TPR domain and result in destabilization of the protein. Collectively, these data suggest that either loss of functional OGT protein or disruption of OGT catalytic activity lead to the same clinical phenotype. Biochemical characterization of further variants affecting catalysis, solution of their crystal structures and examination of OGT localization in OGT-CDG cells will contribute to testing this hypothesis.

Previous research has revealed a role of the OGT TPR domain in protein - protein interactions and substrate recognition (Iyer and Hart, 2003; Jínek et al., 2004; Lazarus et al., 2013; Levine et al., 2018; Pathak et al., 2015; Rafie et al., 2017), demonstrating the role of an asparagine ladder in substrate binding. Our data imply a potential role of catalytic domain residues in substrate recognition, substantiated by predicted structural changes in the proximity of the active site and a loss of glycosyltransferase activity towards a protein substrate, without loss of activity towards a peptide substrate. We also revealed that Ogt activity in cultured mESCs was affected by the C921Y variant, leading to global hypo O-GlcNAcylation. Decreased O-GlcNAcylation in cultured cells as a result of an OGT-CDG mutation has previously only been observed in the OGT^{N648Y} variant (Pravata et al., 2020a). O-GlcNAc homeostasis is tightly regulated by feedback loops between OGT, OGA and O-GlcNAc levels (Decourcelle et al., 2020; Lin et al., 2021; Muthusamy et al., 2015; Tan et al., 2021; Zhang et al., 2014). This has been shown to occur through both translational and post-translational mechanism, for example, through regulation of intron detention, which alters the abundance of productive transcripts of *OGT* and *OGA* (Tan et al., 2021). Unlike in Ogt^{C921Y} cells, this generally manifests as downregulation of OGA at the protein level, particularly upon OGT inhibition or in OGT-CDG catalytic domain mutations (Ortiz-Meoz et al., 2015; Pravata et al., 2019; Pravata et al., 2020a). However, reducing

O-GlcNAcylation through inhibiting the hexosamine biosynthetic pathway has been shown to increase OGT protein levels (Lin et al., 2021). Importantly, OGA is also implicated in neurodevelopment (Olivier-Van Stichelen et al., 2017) and cognitive functioning (Muha et al., 2020), both in model animals and potentially in humans (as determined by a genome wide association study on intelligence) (Savage et al., 2018). These data together suggested the hypothesis that reduction of OGA levels in itself could be a potential mechanism that underpins OGT-CDG (Pravata et al., 2020b). However, unlike other characterized OGT-CDG mESC cell lines, the OGT^{C921Y} variant did not induce OGA downregulation in mESCs. Furthermore, based on Western blot analysis, HCF1 proteolytic processing and localization was not affected in cultured mESCs in our study, even though O-GlcNAcylation and HCF1 proteolytic processing occur within the same active site of OGT. A previously reported OGT-CDG mutation affecting the TPR domain results in a similar phenotype: unaffected HCF1 processing and reduced glycosyltransferase activity (Gundogdu et al., 2018; Vaidyanathan et al., 2017). These data point to loss of O-GlcNAcylation on specific proteins as a likely link between OGT-CDG missense mutations and patient phenotypes.

To determine whether OGT-CDG may arise due to decreased O-GlcNAcylation in stem cells, we assayed self-renewal and differentiation in mESCs harbouring the C921Y mutation in *Ogt*. In these cells, the pluripotency marker ALP and the key pluripotency transcription factors Oct4 and Sox2 are significantly decreased in the presence of LIF, compared to wild type colonies, implying a defect in the maintenance of an undifferentiated state. The colonies formed by OGT^{C921Y} mESCs were flat and spread-out, resembling differentiating colonies in their morphology. In addition, global O-GlcNAcylation in OGT^{C921Y} mESC colonies is significantly reduced compared to OGT^{WT} mESC colonies. This is in line with a previous study indicating that preventing O-GlcNAcylation blocks mESC self-renewal (Jang et al., 2012). O-GlcNAcylation has also been shown to decrease during development (Liu et al., 2012), further supporting the hypothesis that decreased O-GlcNAcylation may be pathogenic through reducing stem cell pluripotency and self-renewal. However, the OGT^{C921Y} variant does not result in preferential differentiation towards a specific germ layer, despite previous research indicating that O-GlcNAcylation of SOX2 can promote differentiation towards an ectodermal lineage (Kim *et al.,* 2021). These findings suggest that a potential contributor to OGT-CDG clinical manifestation is a misregulation of exit from a pluripotent state in stem cells. This may occur at several stages of development, with recent evidence suggesting normal O-GlcNAcylation is required for maintenance not only of ESCs (Jang *et al.,* 2012), but also neural stem cells (NSCs) (White *et al.,* 2020, Shen *et al.,* 2021).

Materials & Methods

Sequencing of patient genetic material

The inclusion of patients in research studies has been approved by the local ethical committee and consent for publication of data and photographs was given by the family. Whole exome sequencing was performed on patient III:2. Subsequent screening for the OGT-variant was done by Sanger sequencing in the mother, the healthy brother (III:1) and the affected twins (III:3 and III:4). All genetic analysis was done at an accredited clinical laboratory in Amplexa Genetics A/S, Sverigesgade 24, Odense C, Denmark.

Protein expression and purification

Human OGT^{WT} and OGT^{C921Y} constructs (full length and a 323 – 1044aa shortened construct with truncated TPR domain fused with N-terminal His tag (full length) and GST tag (323 – 1044 OGT)) were expressed in *E. coli* BL21 cells in plasmids carrying an ampicillin resistance cassette. Transformed colonies that incorporated the plasmids were selected on ampicillin agar plates, expanded overnight at 37 ºC in a shaking incubator in 5 x ampicillin LB broth as a starter culture, and then grown in desired quantity (6 – 12 liters) at 37 ºC until reaching 0.5 – 0.6 OD_{600}. Subsequently, the temperature of the shaking incubator was decreased to 18 ºC, cultures were induced with 100 µM IPTG and further cultured for 16 h. Following overnight culture, bacteria were pelleted for 45 min at 4 ºC, resuspended in base buffer (0.1 M Tris-HCl, pH 7.5, 0.15 M NaCl, 0.5 mM TCEP) and lysed using a French press in the presence of 0.1 mg/ml DNase I, 0.5 mg/ml lysozyme and protease inhibitor cocktail (1 mM benzamidine, 0.2 mM PMSF, 5 mM leupeptin). Resulting lysates were then pelleted at 33 000 rcf at 4 ºC for 45 min. Supernatant was filtered with a 0.2 µM filter and the pellet was discarded. Clarified supernatant was exposed either to glutathione Sepharose 4B beads for GST tagged protein or NiNTA resin for His tagged protein for affinity purification. Bound protein was either eluted using 50 mM glutathione (for GST tagged protein) or 300 mM imidazole (for His tagged protein) in base buffer, buffer exchanged into base buffer only and the tag was cleaved off using 80 units of PreScission protease per prep. Protein was further purified using size exclusion chromatography and stored in solution with 25 % glycerol at – 80 ºC.

Enzyme assays

Michaelis Menten kinetics of full length recombinant OGT^{WT} and OGT^{C921Y} against acceptor peptide (Ac-KENSPAVTPVSTA-NH_2) were tested in a fluorometric *in vitro* assay as described by Pathak and colleagues (Rafie et al., 2017). The reaction time was 4 h at room temperature. Experiments were repeated three times on different days, each repeat consisting of three technical replicates. Glycosyltransferase activity of recombinant OGT^{WT} and OGT^{C921Y} (323 – 1044) against protein substrate was tested *in vitro* using TAB1 (7 – 420) as described previously (Rafie et al., 2017). Experiments were repeated on five times on different days.

Differential scanning fluorimetry (DSF) assay

To prepare assay mixtures, OGT^{WT} and OGT^{C921Y} (323 – 1044) were diluted to 1.2 µM in base buffer containing 50 mM HEPES/NaOH pH 7.5, 150 mM NaCl, 0.5 mM TCEP and mixed with 1:5000 SYPRO Orange Protein Gel Stain (Sigma). 50 µL of assay mixture was dispensed per well in a white bottom qPCR plate and each condition was performed in technical triplicate. The experiment was repeated three times on different days. The prepared assay plate was exposed to temperature increases from 25 ºC to 95 ºC with 1 ºC increments for 5 s each using a CFX Connect Real-Time PCR Detection System (BioRad). SYBR fluorescence was detected after every temperature increase. Data were truncated using Excel and analysed in Prism (GraphPad) as described previously (Huynh and Partch, 2016).

Cloning and CRISPR/Cas9 generation

Generation of OGT^{C921Y} mESCs was performed as described previously by (Pravata et al., 2019) with mutation specific parameters. A repair template was generated to create the C921Y mutation. This was cloned as a *Bam*HI-*Not*I fragment into a plasmid based on pGEX6P1. Paired gRNA sequences were selected and cloned as annealing oligos into pBABED-U6 and pX335-U6 plasmids. Silent mutations to remove the gRNA recognition sequences, in addition to a change at codon 935, were introduced by PCR of a gene block followed by restrictionless cloning into the cloned genomic region. The change at position 935 was introduced by the use of the C921Y wobble primer along with the reverse screening primer to generate a mutagenic PCR product and introduced by restrictionless cloning. The repair template was confirmed by DNA sequencing. Sequences of the listed custom-made reagents are listed in Supplementary Table 1.

Clones were screened using paired screening primers. These generated a 600bp PCR product which was then digested with Fastdigest *Xce*I. Wild type clones show two fragments – 277bp and 323bp. Successful incorporation of the repair template results in loss of restriction of the PCR product. Intact PCR products were confirmed by DNA sequencing. Following DNA confirmation of several clones, RNA was extracted, and one-step RT-PCR was carried out using Takara Primescript High fidelity RT-PCR kit with mOGT_solid_ex11_fwd and mOGT_solid_end_rev primers. The resulting PCR product was sequenced to confirm that the region outwith the repair template had not been damaged during repair and mRNA expression included the change.

Tissue culture

AW2 mESC culture was performed as described previously in (Pravata et al., 2019). AW2 line is derived from E14-TG2a.IV (129/Ola) ES cells kindly donated from the MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, University of Edinburgh (Zhou et al., 2013). Cells were routinely tested for mycoplasma and kept in strictly sterile conditions with no visible bacterial or fungal infection.

Western blotting

Protein was extracted from confluent mESCs. Plates or flasks with attached cells were washed twice with pre-warmed 1 x PBS and then covered with 20 µl per cm^2 with ice cold lysis buffer (50 mM Tris pH 7.4, 1 mM EDTA-NaOH, 1 mM EGTA-NaOH, 1% Triton X-100, 1 mM Na_3VO_4, 50 mM NaF, 0.27 M Sucrose, 5 mM protease inhibitor cocktail). Then, prepared plates were frozen at – 20 ºC overnight. Cells were then scraped off the plates, vortexed for 10 s and centrifuged for 45 min at 15 000 rcf at 4 ºC. Protein concentration in the

clarified lysate was determined using Pierce™ 660 nm Protein Assay Reagent (Thermo). For sample resolution and protein detection, 20 µg of protein was mixed with LDS buffer, boiled, loaded onto a 4 – 12 % Bis – Tris gel (Invitrogen) and then transferred onto 0.2 µm nitrocellulose membrane (GE Healthcare). Following membrane blocking with 5 % BSA in 1 x TBS, primary antibodies were applied: Anti-O-GlcNAc Transferase (DM-17, Sigma-Aldrich, catalogue number O6264, 1:5000), RL2 (Novus, catalogue number NB300, 1:1000), O-GlcNAcase (1:500, Sigma, catalogue number SAB420026), Oct3/4 (C-10, Santa Cruz Biotechnologies, catalogue number sc-5279, 1:500), Sox2 (Santa Cruz Biotechnologies, catalogue number sc-365823, 1:500), Histone 3 (Cell Signaling Technologies, catalogue number 9715, 1:2000) and α-Tubulin (Proteintech, catalogue number 11224-1-AP, 1:5000). After incubation with the corresponding LI-COR secondary antibodies (1:10 000), signal was detected using a LI-COR Odyssey scanning system and quantified using Image Studio Lite (LI-COR). Data were normalized and analyzed in Prism (GraphPad). Number of repeats per experiment is detailed in the figure legends.

For HCF1 cleavage detection, nuclear fractionation was performed. Briefly, confluent cells were detached from plates using accutase, spun down and washed. Pelleted cells were then covered with ice cold buffer A (10 mM HEPES pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.05% NP40 and protease inhibitor cocktail), vortexed and centrifuged at 3000 rpm for 10 min at 4 °C. The resulting supernatant was used as the cytoplasmic lysate fraction. Pelleted material was treated with ice cold buffer B (5 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT and 26% glycerol (v/v)), homogenized on ice and centrifuged at full speed for 10 min at 4°C. The resulting supernatant was used as the nuclear lysate fraction. Due to low protein concentration, protein was precipitated from samples as described previously by (Clark et al., 2013). 20 µg of protein was then loaded onto 4 – 12 % gel as described above. Primary antibodies used were HCF1 (Bethyl, 1:1000), Laminin B1 (ZL-5, Abcam, 1:5000) and β-Actin (Proteintech, 1: 10 000). After incubation with corresponding LI-COR secondary antibody (1:10 000), signal was detected using a LI-COR Odyssey scanning system and quantified using Image Studio Lite (LI-COR). Data were normalized and analyzed in Prism (GraphPad).

RT - PCR

RNA was extracted from confluent cells using the RNeasy kit (Qiagen). RNA purity was determined using Nanodrop 1000 (Thermo) and RNA concentration was measured using the Qubit™ RNA Broad Range kit (Thermo). RNA was transcribed into cDNA using the qScript cDNA Synthesis Kit (Quantabio). qPCR reactions were set up in 384 well plates in a total volume of 10 µl using PerfeCTa® SYBR® Green FastMix® for iQ™ (Quanta) with 5 ng of cDNA and 250 nM forward + reverse primer mix. Three housekeeping genes (GAPDH, 18 S and Actin β) were used as reference genes and each condition was run in technical triplicate. No template as well as no reverse transcription controls were used. Reactions were run in a BioRad CFX 384 real time detection system (BioRad). Primer sequences are summarized in Supplementary Table 2. Results were analysed using BioRad CFX Manager and Prism (GraphPad). Number of repeats per experiment is detailed in the figure legends.

Alkaline phosphatase assay

For alkaline phosphatase (ALP) assays, cells were well resuspended, passed through a cell strainer to remove clumps and doublets and counted using a hemocytometer. Then, 100 wild type or OGT^{C921Y} mESCs were plated into five wells of a 6-well plate pre-treated with 0.1 % (w/v) porcine gelatin in complete GMEM as described above. Cells were grown in complete media for two days to start forming colonies. From day three until day seven, media was changed in dedicated wells into no-LIF GMEM (GMEM (Gibco) supplemented with 5 % FBS fraction V (Gibco), 1 x NEAA (Gibco), 1 x sodium pyruvate (Gibco), 0.1 mM BME (Thermo)). One well per plate remained without media change. On day seven, media was aspirated, colonies were washed twice with 1 x PBS and fixed with 1 % paraformaldehyde for 2 min. Fixed colonies were stained with the Alkaline Phosphatase kit (Merck) according to manufacturer's instructions. ALP-treated plates were scanned on a flat-bed scanner, and colonies counted and scored (undifferentiated, mixed, differentiated) under a microscope. Colonies were subsequently re-stained with Coomassie to visualize ALP-negative colonies and plates re-scanned. Experiment has been repeated three times using three cell lines per genotype. Colony diameters were determined using ImageJ based on Coomassie scans using a macro attached in the supplementary material. Briefly, colonies were segmented by creating a binary image by setting an automatic threshold using the "MaxEntropy" algorithm and measured using the "Analyze Particles" function. Macro is provided in supplementary materials. Data were analysed using Prism (GraphPad).

Cell cycle analysis

Cell cycle analysis was performed using near-confluent OGT^{C921Y} (two lines) and OGT^{WT} (one line) mESCs collected over three subsequent passages. DNA content was assessed using Abcam Propidium Iodide kit (ab139418) according to manufacturer's instructions. Flow cytometry analysis was performed on BD FACSCanto (BD). Data were analysed using FlowJo and Prism (GraphPad) software.

Immunofluorescence analysis of colonies

Colonies were plated on coverslips coated with 0.1% (w/v) porcine gelatin in complete GMEM as described above. 6 days after plating, colonies were fixed for 15 min using 4% PFA. PFA was then quenched with 1 x TBS for 10 min. Fixed colonies were then washed with 1 x PBS twice for 5 min each. Samples were subsequently permeabilized and precipitated using ice cold 90% (w/w) methanol at -20 ℃ for 10 min. Samples were then washed twice with 1 x PBS and blocked blocking buffer (1 x PBS, 1% Tween20 and 5% BSA) for at least 15 min. Primary antibodies (Oct3/4 (D6C8T, Cell Signaling Technology, catalogue number 83932, 1:1000), Sox2 (D9B8N, Cell Signaling Technology, catalogue number 23064, 1:1000), Pax6 (H-295, Santa Cruz, catalogue number sc-11357, 1:1000), Sox17 (Santa Cruz, catalogue number sc-130259, 1:1000) and Brachyury (D10, Santa Cruz, catalogue number sc-166962, 1:1000)) were added in blocking buffer overnight. Following primary antibody stain, cells were washed three times for 15 min with 1 x PBS and secondary antibodies (1:2000) were added in blocking buffer. Colonies were washed with 1 x PBS and DAPI, mounted and imaged on a confocal microscope. Images were analyzed using an ImageJ macro, attached in the supplementary material. Briefly, nuclear regions were determined by performing a series of global and local thresholding steps. Non-nuclear regions were determined based on subtraction of nuclear regions from a binary image created by applying an automatic threshold to individual channels using the "MinError" algorithm.

Acknowledgements

This work was funded by a Wellcome Trust Investigator Award (WT110061) to D.M.F.v.A. M.O. was funded by a Wellcome Trust 4-year PhD studentship (WT102132). I.C. was funded by a studentship from the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs, T001682). We would like to thank the patients and their family who consented to participate in this study. We acknowledge Dr. Lindsay Davidson for advice on experiment design.

Author contributions

M.O., T.B.H., Ch.D.F. and D.M.F.v.A conceived the study; A.B. and T.B.H. performed clinical assessment, Ch.D.F. performed the exome analysis, M.O. performed experiments; M.M. and M.O. performed microscopy and cell cycle experiments, I.C. wrote code for alkaline phosphatase and microscopy images analysis, S.G.B. performed molecular docking simulation, V.M.P. generated the CRISPR cell line; A.T.F. performed molecular biology; M.O., M.P.S., A.B. and D.M.F.v.A. interpreted the data and wrote the manuscript with input from all authors.

Conflict of interest

Authors declare no conflict of interest.

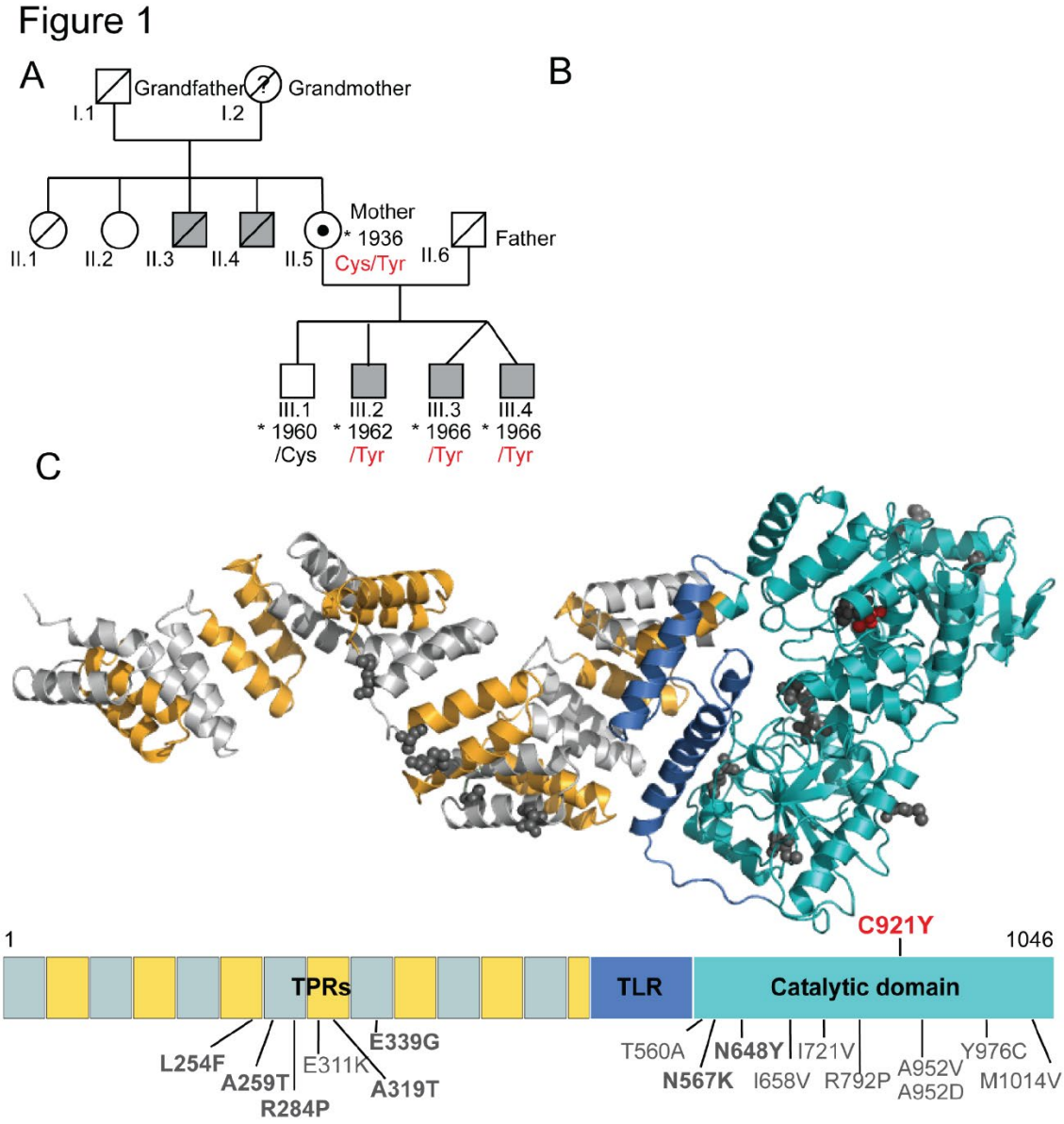


Figure 1: Brothers with intellectual disability carry an inherited variant in the catalytic core of OGT that is absent in their healthy brother

(a) Pedigree of the affected patients carrying the OGT^{C921Y} variant. Grey shapes denote individuals affected by intellectual disability. Where available, the OGT variant at position 921 is provided. Black dot indicates a female carrier of the OGT^{C921Y} variant.

(b) ***Figure panel censored for BioRxiv submission.*** Facial photographs of two brothers carrying OGT^{C921Y} variant with corresponding family tree annotation. Patient III.2 is depicted at age of 58 and patients III.3 and III.4 are depicted at age of 54.

(c) 3D model of full-length human OGT and the corresponding linear domain structure. TPRs are depicted in alternating gray and yellow, the TLR in dark blue and the catalytic core in cyan. Previously known OGT variants leading to ID and their corresponding residues are highlighted in dark gray. OGT^{C921Y} is highlighted in red. UDP and acceptor peptide molecules are shown yellow and magenta

sticks respectively. The 3D model was produced from OGT catalytic core Protein Data Bank code 5C1D, TPR domain Protein Data Bank code 1W3B.

Figure 2

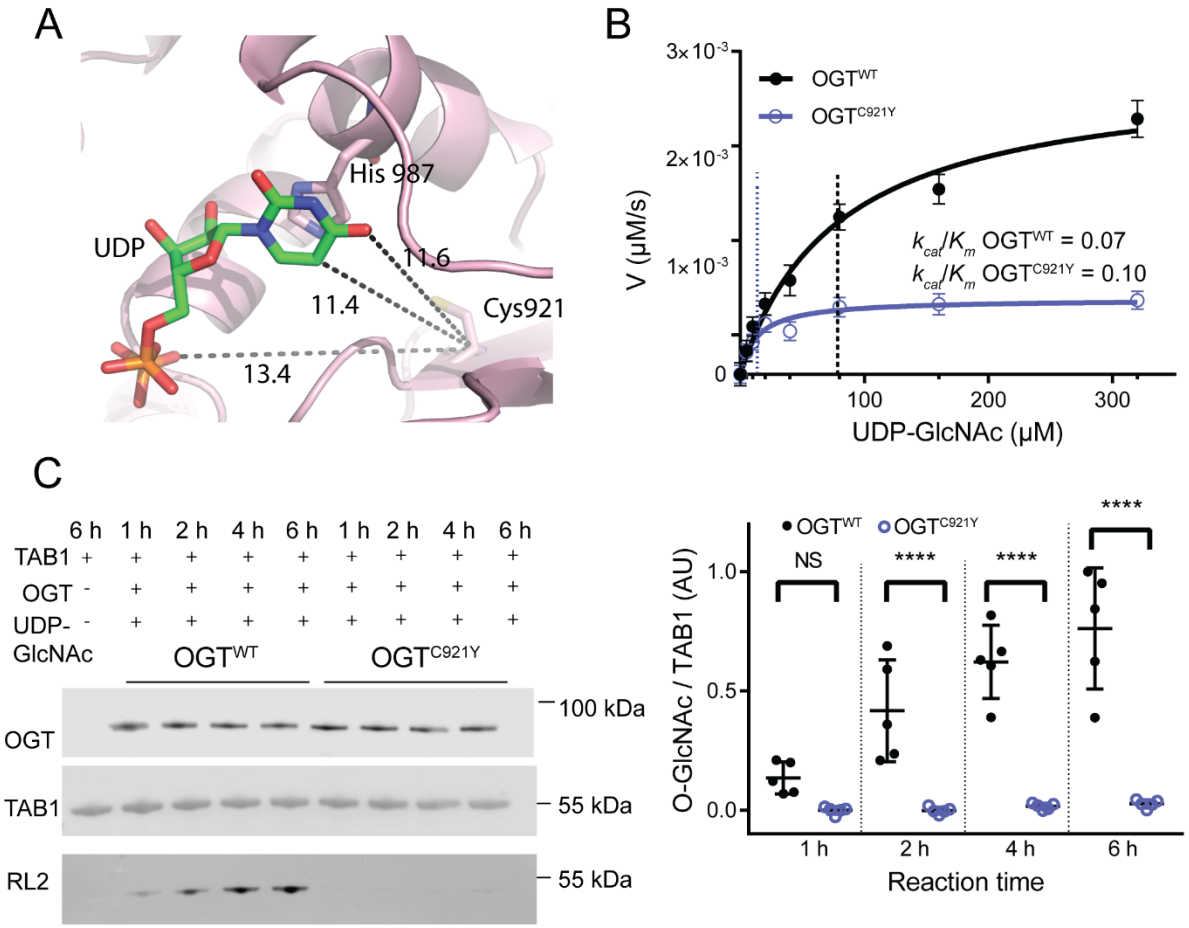


Figure 2: OGT^{C921Y} is defective in glycosyltransferase activity towards protein substrates *in vitro*

(a) Zoomed view into catalytic domain of OGT. UDP is shown as sticks. Distances between C921 residue and UDP are highlighted by dotted lines (13.3 Å from C921 to the uracil ring in green, 14.3 Å from C921 to the pyrophosphate linker in orange).

(b) Michaelis-Menten kinetics of recombinant full length OGT^{WT} and OGT^{C921Y} incubated with acceptor peptide (Ac-KENSPAVTPVSTA-NH2) and varying concentrations of UDP-GlcNAc, n = 3 repeats, each consisting of three technical replicates.

(c) Immunoblot detection of TAB1 O-GlcNAcylation by recombinant OGT^{WT} and OGT^{C921Y} (323-1041) over the course of 1 h, 2 h, 4 h and 6 h *in vitro*. Corresponding immunoblot signal quantification is shown on the right. n = 5 reactions using the same batch of recombinant protein. Repeated Measures two-way ANOVA with Sidak's multiple comparison test, 0 h $p = 0.14$, 2 h $p < 0.0001$, 4 h $p < 0.0001$, 2 h $p < 0.0001$. Error bars represent standard deviation.

Figure 3

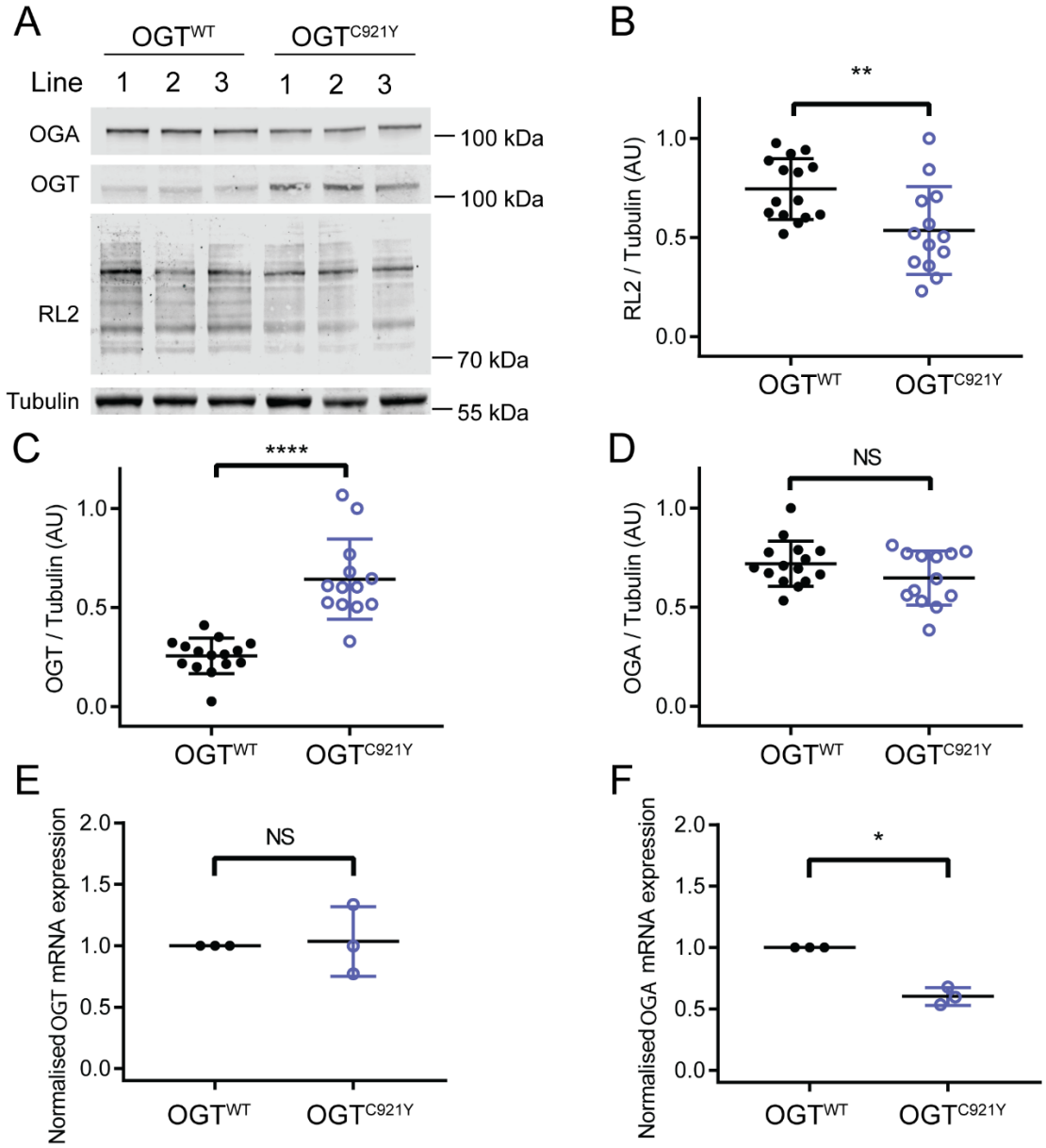


Figure 3: The OGT^{C921Y} variants disrupt O-GlcNAcylation machinery in undifferentiated mESCs.

(a) Immunoblot showing detection of O-GlcNAcylation (RL2), OGT and OGA in mESCs carrying wild type OGT or the OGT^{921Y} variant. The quantification shown in panels B – D is based on results obtained from three different cell clones per genotype and repeated over four to five passages per clone. Error bars represent standard deviation.

(b) Quantification of RL2 immunoblot signal. RL2 signal was normalized to Tubulin. OGT^{WT} n = 15, OGT^{C921Y} n = 13, Unpaired t test, $p = 0.007$.

(c) Quantification of OGT immunoblot signal. OGT signal was normalized to Tubulin. OGT^{WT} n = 15, OGT^{C921Y} n = 13, Unpaired t test, $p < 0.0001$.

(d) Quantification of OGA immunoblot signal. OGA signal was normalized to Tubulin. OGT^{WT} n = 15, OGT^{C921Y} n = 13, Unpaired t test, $p = 0.14$.

(e) RT – PCR analysis of OGT mRNA normalized to GAPDH, 18 S and Actin β. Data points representing the mean expression calculated from three separate RT-PCR runs are shown. Each RT - PCR run was set up using several OGT^{WT} and OGT^{C921Y} as biological replicates. Unpaired *t* test, $p = 0.84$.

(f) RT – PCR analysis of OGT mRNA normalized to GAPDH, 18 S and β-Actin. Data points representing the mean expression calculated from three separate RT-PCR runs are shown. Each RT-PCR run was set up using several OGT^{WT} and OGT^{C921Y} lines as biological replicates. Unpaired *t* test, p = 0.0007.

Figure 4

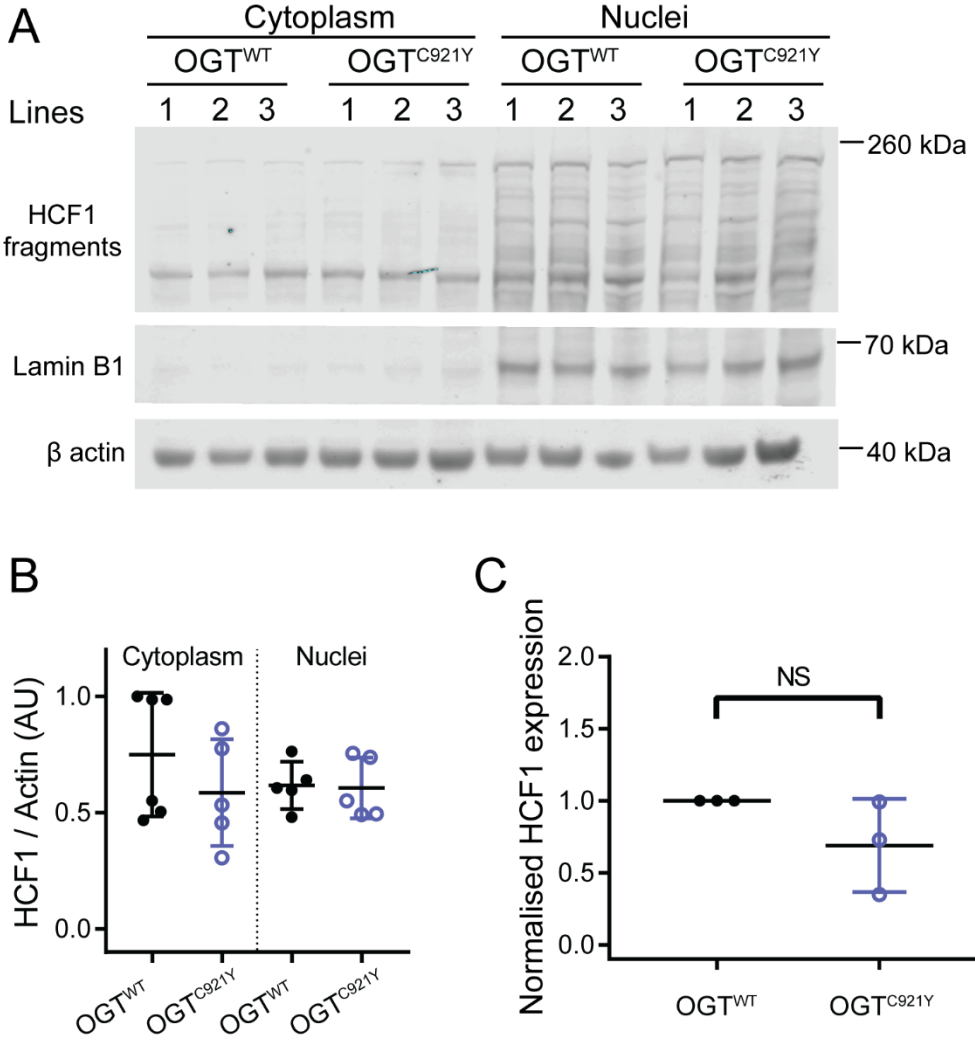


Figure 4: HCF1 processing by undifferentiated OGT^{WT} and OGT^{C921Y} in mESCs.

(a) Immunoblot of HCF1 proteolytic fragments in the cytoplasm and nucleus of mESCs harboring either wild type or OGT^{C921Y} variant. Lamin B1 was used as marker for nuclear fraction. This experiment was performed using three cell clones per genotype and repeated over two passages.

(b) Quantification of HCF1 signal. HCF1 signal was normalized to Actin β. n = 6 biological replicates, One-way Anova with Tukey comparison test, cytoplasmic fraction p = 1, nuclear fraction p = 0.54. Error bars represent standard deviation.

(c) RT – PCR analysis of HCF1 mRNA expression normalized to GAPDH, 18 S and Actin β. Data points representing the mean expression calculated from three separate RT-PCR runs are shown. Each RT-PCR run was set up using several OGT^{WT} and OGT^{C921Y} lines as biological replicates. Unpaired *t* test, p = 0.172.

Figure 5

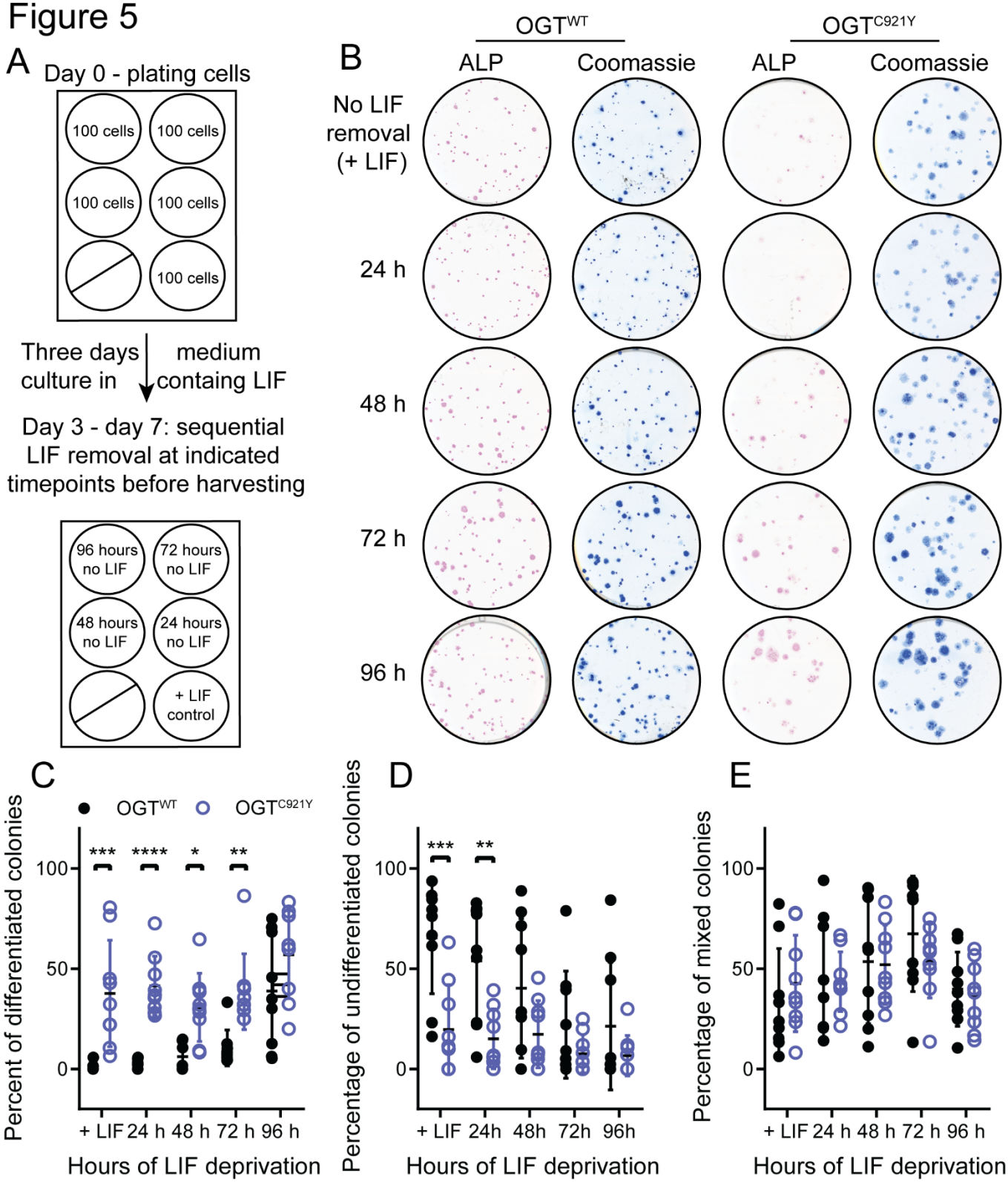


Figure 5: The OGT^{C921Y} variant leads to increased differentiation in response to LIF withdrawal in mESCs.

(a) Schematic representation of experimental set up. This experiment has been performed using three cell clones per genotype and repeated over three passages.

(b) Scans of alkaline phosphatase (ALP, pink) stained colonies of wild type or OGT^{C921Y} mESCs at different time points after LIF withdrawal. Each well was re-stained by Coomassie (blue) to reveal ALP negative colonies.

(c) Plot of percentages of differentiated colonies in wild type and OGT^{C921Y} mESCs. n = 9 biological replicates, Two-way Anova, +LIF p = 0.0001 24 h p = 0.005, 48 h p = 0.21, 72 h p = 0.68, 96 h p = 0.67.

(d) Plot of percentages of undifferentiated colonies in wild type and OGT^{C921Y} mESCs. n = 9 biological replicates, Two-way Anova, +LIF p = 0.0006, 24 h p < 0.0001, 48 h p = 0.017, 72 h p = 0.004, 96 h p = 0.13.

(e) Plot of percentages of mixed colonies in wild type and OGT^{C921Y} mESC. n = 9 biological replicates, Two-way Anova, all p values are non-significant.

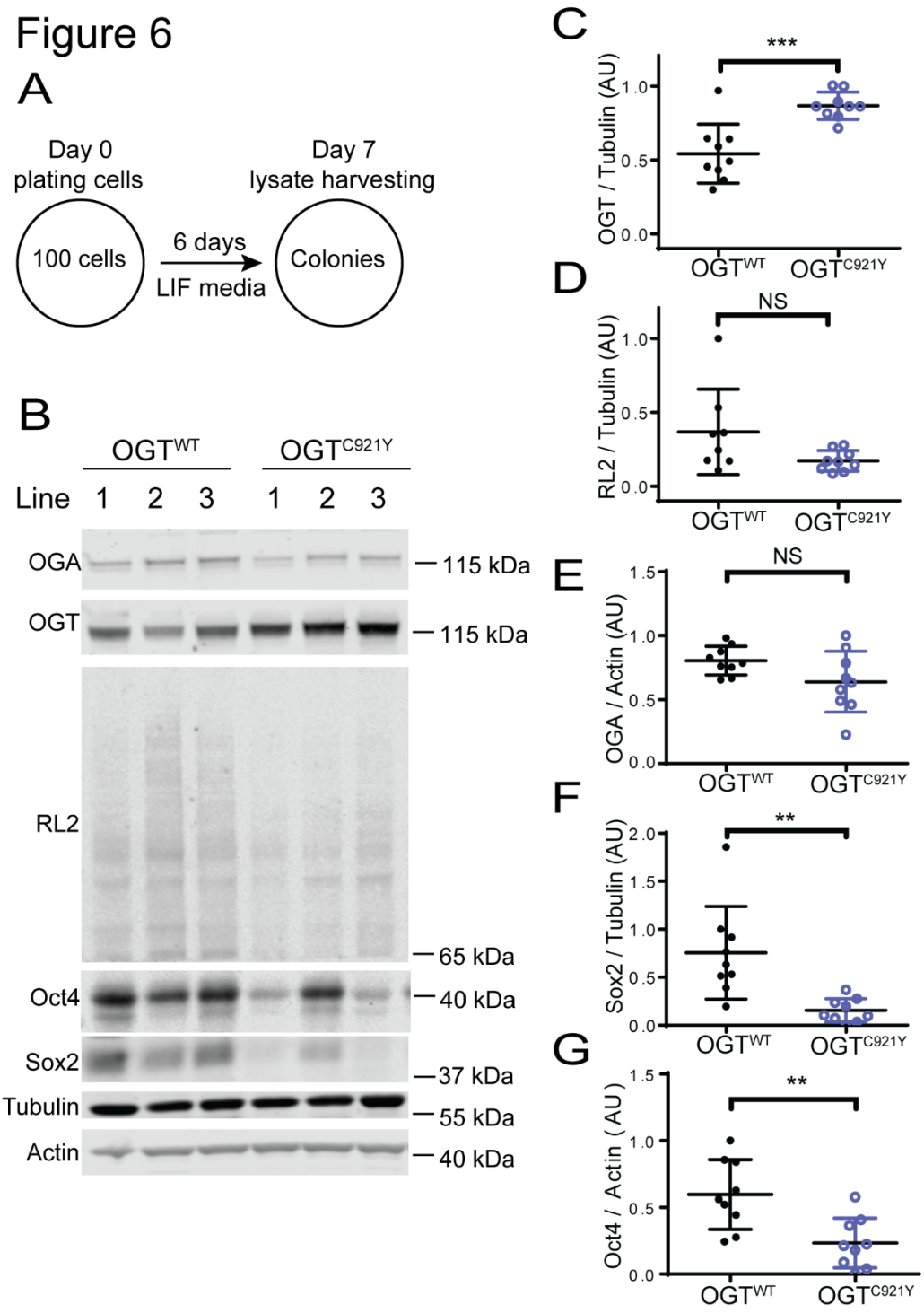


Figure 6: mESCs harboring the OGT^{C921Y} variant downregulate Sox2 and Oct4 in response to clonogenic conditions.

(a) Schematic representation of experimental set up. The following experiments were performed using three cell clones per genotype and repeated over three passages.

(b) Immunoblot showing detection of O-GlcNAcylation (RL2), OGT, OGA, Oct4 and Sox2 in mESC lines carrying wild type or OGT^{921Y} variant. Lysates were derived from cells grown under clonogenic conditions in presence of LIF as shown in (a).

(c) Quantification of OGT immunoblot signal. RL2 signal was normalized to Tubulin. OGT^{WT} n = 9 biological replicates, OGT^{C921Y} n = 9 biological replicates, Unpaired t test, p = 0.0004. Error bars represent standard deviation.

(d) Quantification of RL2 immunoblot signal. RL2 signal was normalized to Tubulin. OGT^{WT} n = 8 biological replicates, OGT^{C921Y} n = 9 biological replicates, Unpaired t test, p = 0.067. Error bars represent standard deviation.

(e) Quantification of OGA immunoblot signal. OGA signal was normalized to Actin. OGT^{WT} n = 9 biological replicates, OGT^{C921Y} n = 9 biological replicates, Unpaired t test, p = 0.077. Error bars represent standard deviation.

(f) Quantification of Sox2 immunoblot signal. Sox2 signal was normalized to Tubulin. OGT^{WT} n = 9 biological replicates, OGT^{C921Y} n = 9 biological replicates, Unpaired t test, p = 0.0023. Error bars represent standard deviation.

(g) Quantification of Oct4 immunoblot signal. Oct4 signal was normalized to Actin. OGT^{WT} n = 9 biological replicates, OGT^{C921Y} n = 9 biological replicates, Unpaired t test, p = 0.0038. Error bars represent standard deviation.

References

Bouazzi, H., Lesca, G., Trujillo, C., Alwasiyah, M. K. and Munnich, A. (2015). Nonsyndromic X-linked intellectual deficiency in three brothers with a novel MED12 missense mutation [c.5922G>T (p.Glu1974His)]. *Clin. case reports* **3**, 604–609.

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.

Castro, V. L., Reyes, J. F., Reyes-Nava, N. G., Paz, D. and Quintana, A. M. (2020). Hcfc1a regulates neural precursor proliferation and asxl1 expression in the developing brain. *BMC Neurosci.* **21**, 1–17.

Chen, C. Y., Lee, D. S., Yan, Y. T., Shen, C. N., Hwang, S. M., Lee, S. T. and Hsieh, P. C. H. (2015). Bcl3 Bridges LIF-STAT3 to Oct4 Signaling in the Maintenance of Naïve Pluripotency. *Stem Cells* **33**, 3468–3480.

Cherepkova, M. Y., Sineva, G. S. and Pospelov, V. A. (2016). Leukemia inhibitory factor (LIF) withdrawal activates mTOR signaling pathway in mouse embryonic stem cells through the MEK/ERK/TSC2 pathway. *Cell Death Dis.* **7**, e2050.

Clark, P. M., Rexach, J. E. and Hsieh-Wilson, L. C. (2013). Visualization of O-GlcNAc glycosylation stoichiometry and dynamics using resolvable poly(ethylene glycol) mass tags. *Curr. Protoc. Chem. Biol.* **5**, 281–302.

Constable, S., Lim, J. M., Vaidyanathan, K. and Wells, L. (2017). O-GlcNAc transferase regulates transcriptional activity of human Oct4. *Glycobiology* **27**, 927–937.

Decourcelle, A., Loison, I., Baldini, S., Leprince, D. and Dehennaut, V. (2020). Evidence of a compensatory regulation of colonic O-GlcNAc transferase and O-GlcNAcase expression in response to disruption of O-GlcNAc homeostasis. *Biochem. Biophys. Res. Commun.* **521**, 125–130.

Drougat, L., Olivier-Van Stichelen, S., Mortuaire, M., Foulquier, F., Lacoste, A. S., Michalski, J. C., Lefebvre, T. and Vercoutter-Edouart, A. S. (2012). Characterization of O-GlcNAc cycling and proteomic identification of differentially O-GlcNAcylated proteins during G1/S transition. *Biochim. Biophys. Acta - Gen. Subj.* **1820**, 1839–1848.

Elena-Herrmann, B., Montellier, E., Fages, A., Bruck-Haimson, R. and Moussaieff, A. (2020). Multi-platform NMR Study of Pluripotent Stem Cells Unveils Complementary Metabolic Signatures towards Differentiation. *Sci. Rep.* **10**, 1–11.

Firth, H. V. and Wright, C. F. (2011). The Deciphering Developmental Disorders (DDD) study. *Dev. Med. Child Neurol.* **53**, 702–703.

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 beta-N-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276**, 9838–9845.

Gundogdu, M., Llabrés, S., Gorelik, A., Ferenbach, A. T., Zachariae, U. and van Aalten, D. M. F. (2018). The O -GlcNAc Transferase Intellectual Disability Mutation L254F Distorts the TPR Helix. *Cell Chem. Biol.* **25**, 513--518.e4.

Hao, Y., Fan, X., Shi, Y., Zhang, C., Sun, D. en, Qin, K., Qin, W., Zhou, W. and Chen, X. (2019). Next-generation unnatural monosaccharides reveal that ESRRB O-GlcNAcylation regulates pluripotency of

mouse embryonic stem cells. *Nat. Commun.* **10**, 1–13.

Hart, G. W. (2015). Nutrient regulation of transcription and signalling by O-GlcNAcylation. *Perspect. Sci.* **6**, 49–57.

He, R., Xhabija, B., Al-Qanber, B. and Kidder, B. L. (2017). OCT4 supports extended LIF-independent self-renewal and maintenance of transcriptional and epigenetic networks in embryonic stem cells. *Sci. Rep.* **7**,.

Huang, L., Jolly, L. A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubridge, C., Wieczorek, D., Tzschach, A., Cohen, M., et al. (2012). A Noncoding, Regulatory Mutation Implicates HCFC1 in Nonsyndromic Intellectual Disability. *Am. J. Hum. Genet.* **91**, 694–702.

Huynh, K. and Partch, C. L. (2016). Current Protocols in Protein Science: Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* 1–19.

Iyer, S. P. N. and Hart, G. W. (2003). Roles of the Tetratricopeptide Repeat Domain in O-GlcNAc Transferase Targeting and Protein Substrate Specificity. *J. Biol. Chem.* **278**, 24608–24616.

Jang, H., Kim, T. W., Yoon, S., Choi, S.-Y., Kang, T.-W., Kim, S.-Y., Kwon, Y.-W., Cho, E.-J. and Youn, H.-D. (2012). O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network. *Cell Stem Cell* **11**, 62–74.

Jínek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A. and Conti, E. (2004). The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nat. Struct. Mol. Biol.* **11**, 1001–1007.

Jolly, L. A., Nguyen, L. S., Domingo, D., Sun, Y., Barry, S., Hancarova, M., Plevova, P., Vlckova, M., Havlovicova, M., Kalscheuer, V. M., et al. (2015). HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain. *Hum. Mol. Genet.* **24**, 3335–3347.

Kapuria, V., Röhrig, U. F., Bhuiyan, T., Borodkin, V. S., van Aalten, D. M. F., Zoete, V. and Herr, W. (2016). Proteolysis of HCF-1 by Ser/Thr glycosylation-incompetent O-GlcNAc transferase:UDP-GlcNAc complexes. *Genes Dev.* **30**, 960–972.

Kapuria, V., Röhrig, U. F., Waridel, P., Lammers, F., Borodkin, V. S., Van Aalten, D. M. F., Zoete, V. and Herr, W. (2018). The conserved threonine-rich region of the HCF-1 PRO repeat activates promiscuous OGT:UDP-GlcNAc glycosylation and proteolysis activities. *J. Biol. Chem.* **293**, 17754–17768.

Keembiyehetty, C., Love, D. C., Harwood, K. R., Gavrilova, O., Comly, M. E. and Hanover, J. A. (2015). Conditional knock-out reveals a requirement for O-Linked N-Acetylglucosaminase (O-GlcNAcase) in metabolic homeostasis. *J. Biol. Chem.* **290**, 7097–7113.

Kim, D. K., Lee, J. S., Lee, E. Y., Jang, H., Han, S., Kim, H. Y., Hwang, I. Y., Choi, J. W., Shin, H. M., You, H. J., et al. (2021). O-GlcNAcylation of Sox2 at threonine 258 regulates the self-renewal and early cell fate of embryonic stem cells. *Exp. Mol. Med.* **53**, 1759–1768.

Krause, M., Love, D. C., Ghosh, S., Wang, P., Yun, S., Fukushige, T. and Hanover, J. A. (2018). Nutrient-Driven O-GlcNAcylation at Promoters Impacts Genome-wide RNA Pol II Distribution. *Front. Endocrinol. (Lausanne).* **9**, 521.

Kreppel, L. K., Blomberg, M. a and Hart, G. W. (1997). Dynamic Glycosylation of Nuclear and Cytosolic Proteins. *J. Biol. Chem.* **272**, 9308–9315.

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 (80-.).* **342**, 1235–1239.

Lefebvre, T., Baert, F., Bodart, J. F., Flament, S., Michalski, J. C. and Vilain, J. P. (2004). Modulation of O-GlcNAc glycosylation during xenopus oocyte maturation. *J. Cell. Biochem.* **93**, 999–1010.

Leturcq, M. and Lefebvre, T. (2017). O-GlcNAcylation and chromatin remodeling in mammals : an up-to-date overview. *Biochem. Soc. Trans.* **45**, 323–338.

Levine, Z. G., Fan, C., Melicher, M. S., Orman, M., Benjamin, T. and Walker, S. (2018). O -GlcNAc Transferase Recognizes Protein Substrates Using an Asparagine Ladder in the Tetratricopeptide Repeat (TPR) Superhelix. *J. Am. Chem. Soc.* **140**, 3510–3513.

Li, X., Zhang, Z., Li, L., Gong, W., Lazenby, A. J., Swanson, B. J., Herring, L. E., Asara, J. M., Singer, J. D. and Wen, H. (2017). Myeloid-derived cullin 3 promotes STAT3 phosphorylation by inhibiting OGT expression and protects against intestinal inflammation. *J. Exp. Med.* **214**, 1093–1109.

Lin, C. H., Liao, C. C., Chen, M. Y. and Chou, T. Y. (2021). Feedback Regulation of O-GlcNAc Transferase through Translation Control to Maintain Intracellular O-GlcNAc Homeostasis. *Int. J. Mol. Sci.* **22**,.

Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. and Gong, C.-X. (2004). O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci.* **101**, 10804–10809.

Liu, F., Shi, J., Tanimukai, H., Gu, J., Gu, J., Grundke-Iqbal, I., Iqbal, K. and Gong, C. X. (2009). Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease. *Brain* **132**, 1820–1832.

Liu, Y., Li, X., Yu, Y., Shi, J., Liang, Z., Run, X., Li, Y., Dai, C., Grundke-Iqbal, I., Iqbal, K., et al. (2012). Developmental Regulation of Protein O-GlcNAcylation, O-GlcNAc Transferase, and O-GlcNAcase in Mammalian Brain. *PLoS One* **7**, e43724.

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.

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.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T. and Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Res. Dev. Disabil.* **32**, 419–436.

McKenzie, K., Milton, M., Smith, G. and Ouellette-Kuntz, H. (2016). Systematic Review of the Prevalence and Incidence of Intellectual Disabilities: Current Trends and Issues. *Curr. Dev. Disord. Reports* **3**, 104–115.

Medford, H. M., Chatham, J. C. and Marsh, S. A. (2013). 2012 Chronic ingestion of a Western diet increases O-linked-β-N.pdf. *Life Sci.* **90**, 883–888.

Michaud, J., Praz, V., Faresse, N. J., JnBaptiste, C. K., Tyagi, S., Schütz, F. and Herr, W. (2013). HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. *Genome Res.* **23**, 907–916.

Muha, V., Fenckova, M., Ferenbach, A. T., Catinozzi, M., Eidhof, I., Storkebaum, E., Schenck, A. and van Aalten, D. M. F. (2020). O-GlcNAcase contributes to cognitive function in Drosophila. *J. Biol. Chem.* **295**, 8636–8646.

Muha, V., Authier, F., Szoke-Kovacs, Z., Johnson, S., Gallagher, J., McNeilly, A., McCrimmon, R. J., Teboul, L. and van Aalten, D. M. F. (2021). Loss of O-GlcNAcase catalytic activity leads to defects in mouse embryogenesis. *J. Biol. Chem.* **296**, 100439.

Muthusamy, S., Hong, K. U., Dassanayaka, S., Hamid, T. and Jones, S. P. (2015). E2F1 Transcription Factor Regulates O-linked N-acetylglucosamine (O-GlcNAc) Transferase and O-GlcNAcase Expression. *J. Biol. Chem.* **290**, 31013–31024.

Myers, S. A., Peddada, S., Chatterjee, N., Friedrich, T., Tomoda, K., Krings, G., Thomas, S., Maynard, J., Broeker, M., Thomson, M., et al. (2016). SOX2 O-GlcNAcylation alters its protein-protein interactions and genomic occupancy to modulate gene expression in pluripotent cells. *Elife* **5**, e10647.

Olivier-Van Stichelen, S., Wang, P., Comly, M., Love, D. C. and Hanover, J. A. (2017). Nutrient-driven O-linked N-acetylglucosamine (O-GlcNAc) cycling impacts neurodevelopmental timing and metabolism. *J. Biol. Chem.* **292**, 6076.

Ortiz-Meoz, R. F., Jiang, J., Lazarus, M. B., Orman, M., Janetzko, J., Fan, C., Duveau, D. Y., Tan, Z. W., Thomas, C. J. and Walker, S. (2015). A small molecule that inhibits OGT activity in cells. *ACS Chem. Biol.* **10**, 1392.

PanelApp Genomics England (2020). Intellectual Disability panel, (Version 3.651).

Park, J., Ha, H. J., Chung, E. S., Baek, S. H., Cho, Y., Kim, H. K., Han, J., Sul, J. H., Lee, J., Kim, E., et al. (2021). O-GlcNAcylation ameliorates the pathological manifestations of Alzheimer's disease by inhibiting necroptosis. *Sci. Adv.* **7**,.

Pathak, S., Borodkin, V. S., Albarbarawi, O., Campbell, D. G., Ibrahim, A. and Van Aalten, D. M. F. (2012). O-GlcNAcylation of TAB1 modulates TAK1-mediated cytokine release. *EMBO J.* **31**, 1394–1404.

Pathak, S., Alonso, J., Schimpl, M., Rafie, K., Blair, D. E., Borodkin, V. S., Schüttelkopf, A. W., Albarbarawi, O. and Van Aalten, D. M. F. (2015). The active site of O-GlcNAc transferase imposes constraints on substrate sequence. *Nat. Struct. Mol. Biol.* **22**, 744–749.

Posada De La Paz, M., Taruscio, D. and Groft, S. C. (2017). Rare Diseases Epidemiology: Update and Overview. *Adv. Exp. Med. Biol.* **1031**,.

Pravata, V. M., Muha, V., Gundogdu, M., Ferenbach, A. T., Kakade, P. S., Vandadi, V., Wilmes, A. C., Borodkin, V. S., Joss, S., Stavridis, M. P., et al. (2019). Catalytic deficiency of O-GlcNAc transferase leads to X-linked intellectual disability. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14961–14970.

Pravata, V. M., Gundogdu, M., Bartual, S. G., Ferenbach, A. T., Stavridis, M., Õunap, K., Pajusalu, S., Žordania, R., Wojcik, M. H. and van Aalten, D. M. F. (2020a). A missense mutation in the catalytic domain of O-GlcNAc transferase links perturbations in protein O-GlcNAcylation to X-linked intellectual disability. *FEBS Lett.* **594**, 717–727.

Pravata, V. M., Omelková, M., Stavridis, M. P., Desbiens, C. M., Stephen, H. M., Lefeber, D. J., Gecz, J., Gundogdu, M., Õunap, K., Joss, S., et al. (2020b). An intellectual disability syndrome with single-nucleotide variants in O-GlcNAc transferase. *Eur. J. Hum. Genet.* 706–714.

Qian, K., Wang, S., Fu, M., Zhou, J., Singh, J. P., Li, M. D., Yang, Y., Zhang, K., Wu, J., Nie, Y., et al. (2018). Transcriptional regulation of O-GlcNAc homeostasis is disrupted in pancreatic cancer. *J. Biol. Chem.* **293**, 13989–14000.

Rafie, K., Raimi, O., Ferenbach, A. T., Borodkin, V. S., Kapuria, V. and van Aalten, D. M. F. (2017). Recognition of a glycosylation substrate by the O-GlcNAc transferase TPR repeats. *Open Biol.* **7**, 170078.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S. and Gastier-Foster, J. (2015). Standards and guidelines for the interpretation of sequence variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424.

Sakabe, K. and Hart, G. W. (2010). O-GlcNAc transferase regulates mitotic chromatin dynamics. *J. Biol. Chem.* **285**, 34460–34468.

Sakabe, K., Wang, Z. and Hart, G. W. (2010). β - N -acetylglucosamine (O-GlcNAc) is part of the histone code. *PNAS* **107**, 19915–19920.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919.

Selvan, N., Williamson, R., Mariappa, D., Campbell, D. G., Gourlay, R., Ferenbach, A. T., Aristotelous, T., Hopkins-Navratilova, I., Trost, M. and van Aalten, D. M. F. (2017). A mutant O-GlcNAcase enriches Drosophila developmental regulators. *Nat. Chem. Biol.* **13**, 882–887.

Selvan, N., George, S., Serajee, F. J., Shaw, M., Hobson, L., Kalscheuer, V., Prasad, N., Levy, S. E., Taylor, J., Aftimos, S., et al. (2018). *O*-GlcNAc transferase missense mutations linked to X-linked intellectual disability deregulate genes involved in cell fate determination and signaling. *J. Biol. Chem.* **293**, 10810–10824.

Shafi, R., Iyer, S. P. N., 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.* **97**, 5735–5739.

Sheikh, M. A., Emerald, B. S. and Ansari, S. A. (2020). Stem cell fate determination through protein O-GlcNAcylation. *J. Biol. Chem.* **296**, 100035.

Skuse, D. H. (2005). X-linked genes and mental functioning. *Hum. Mol. Genet.* **14**, 27–32.

Slawson, C., Copeland, R. J. and Hart, G. W. (2010). O-GlcNAc signaling: A metabolic link between diabetes and cancer? *Trends Biochem. Sci.* **35**, 547–555.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by. *Nature* **336**, 688–690.

Štefková, K., Procházková, J. and Pacherník, J. (2015). Alkaline phosphatase in stem cells. *Stem Cells Int.* **2015**,.

Streubel, G., Fitzpatrick, D. J., Oliviero, G., Scelfo, A., Moran, B., Das, S., Munawar, N., Watson, A., Wynne, K., Negri, G. L., et al. (2017). Fam60a defines a variant Sin3a-Hdac complex in embryonic stem cells required for self-renewal. *EMBO J.* **36**, 2216–2232.

Tan, Z. W., Fei, G., Paulo, J. A., Bellaousov, S., Martin, S. E. S., Duveau, D. Y., Thomas, C. J., Gygi, S.

P., Boutz, P. L. and Walker, S. (2021). O-GlcNAc regulates gene expression by controlling detained intron splicing. *Nucleic Acids Res.* **48**, 5656–5669.

Tassé, M. J., Luckasson, R. and Schalock, R. L. (2016). The relation between intellectual functioning and adaptive behavior in the diagnosis of intellectual disability. *Intellect. Dev. Disabil.* **54**, 381–390.

Tejada, M. I. and Ibarluzea, N. (2020). Non-syndromic X linked intellectual disability : Current knowledge in light of the recent advances in molecular and functional studies. *Clin. Genet.* **97**, 677–687.

Torres, C.-R. and Hart, G. W. (1984). Topography and polypeptide distribution of terminal N-Acetylglucosamine residues on the surfaces of intact lymphocytes.pdf. *J Biol Chem* **259**, 3308–3317.

Vaidyanathan, K., Durning, S., Wells, L. and Carbohydrate, C. (2014). Functional O-GlcNAc modifications: Implications in molecular regulation and pathophysiology. *Crit Rev Biochem Mol Biol* **49**, 140–163.

Vaidyanathan, K., Niranjan, T., Selvan, N., Teo, C. F., May, M., Patel, S., Weatherly, B., Skinner, C., Opitz, J., Carey, J., et al. (2017). 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. *J. Biol. Chem.* **292**, 8948–8963.

Webster, D. M., Teo, C., Sun, Y., Wloga, D., Gay, S., Klonowski, K. D., Wells, L. and Dougan, S. T. (2009). O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. *BMC Dev. Biol.* **9**, 28.

Wells, L., Gao, Y., Mahoney, J. A., Vosseller, K., Chen, C., Rosen, A. and Hart, G. W. (2002). GLYCOBIOLOGY AND EXTRACELLULAR MATRICES: Dynamic O- Glycosylation of Nuclear and Cytosolic Proteins: Further characterization of the nucleocytoplasmic B-N-Acetylglycosaminidase, O-GlcNAcase. *J. Biol. Chem.* **277**, 1755–61.

Willems, A. P., Gundogdu, M., Kempers, M. J. E., Giltay, J. C., Pfundt, R., Elferink, M., Loza, B. F., Fuijkschot, J., Ferenbach, A. T., van Gassen, K. L. I., et al. (2017). Mutations in *N* -acetylglucosamine (*O* -GlcNAc) transferase in patients with X-linked intellectual disability. *J. Biol. Chem.* **292**, 12621–12631.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687.

Wolfe, L. A. and Krasnewich, D. (2013). C d g i d. **225**, 211–225.

Wongkittichote, P., Wegner, D. J. and Shinawi, M. S. (2021). Novel exon-skipping variant disrupting the basic domain of HCFC1 causes intellectual disability without metabolic abnormalities in both male and female patients. *J. Hum. Genet.*

Wulff-Fuentes, E., Berendt, R. R., Massman, L., Danner, L., Malard, F., Vora, J., Kahsay, R. and Olivier-Van Stichelen, S. (2021). The human O-GlcNAcome database and meta-analysis. *Sci. Data* **8**, 25.

Yang, Y. R., Song, M., Lee, H., Jeon, Y., Choi, E.-J. J., Jang, H.-J. J., Moon, H. Y., Byun, H.-Y. Y., Kim, E.-K. K., Kim, D. H., et al. (2012). O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. *Aging Cell* **11**, 439–448.

Zechner, U., Wilda, M., Kehrer-Sawatzki, H., Vogel, W., Fundele, R. and Hameister, H. (2001). A high

density of X-linked genes for general cognitive ability: A run-away process shaping human evolution? *Trends Genet.* **17**, 697–701.

Zhang, Z., Tan, E. P., VandenHull, N. J., Peterson, K. R. and Slawson, C. (2014). O-GlcNAcase Expression is Sensitive to Changes in O-GlcNAc Homeostasis. *Front. Endocrinol. (Lausanne).* **5**, 206.

Zhang, Z., Parker, M. P., Graw, S., Novikova, L. V., Fedosyuk, H., Fontes, J. D., Koestler, D. C., Peterson, K. R. and Slawson, C. (2019). O-GlcNAc homeostasis contributes to cell fate decisions during hematopoiesis. *J. Biol. Chem.* **294**, 1363–1379.

Zhou, X., Smith, A. J. H., Waterhouse, A., Blin, G., Malaguti, M., Lin, C.-Y., Osorno, R., Chambers, I. and Lowell, S. (2013). Hes1 Desynchronizes Differentiation of Pluripotent Cells by Modulating STAT3 Activity. *Stem Cells* 1511–1522.

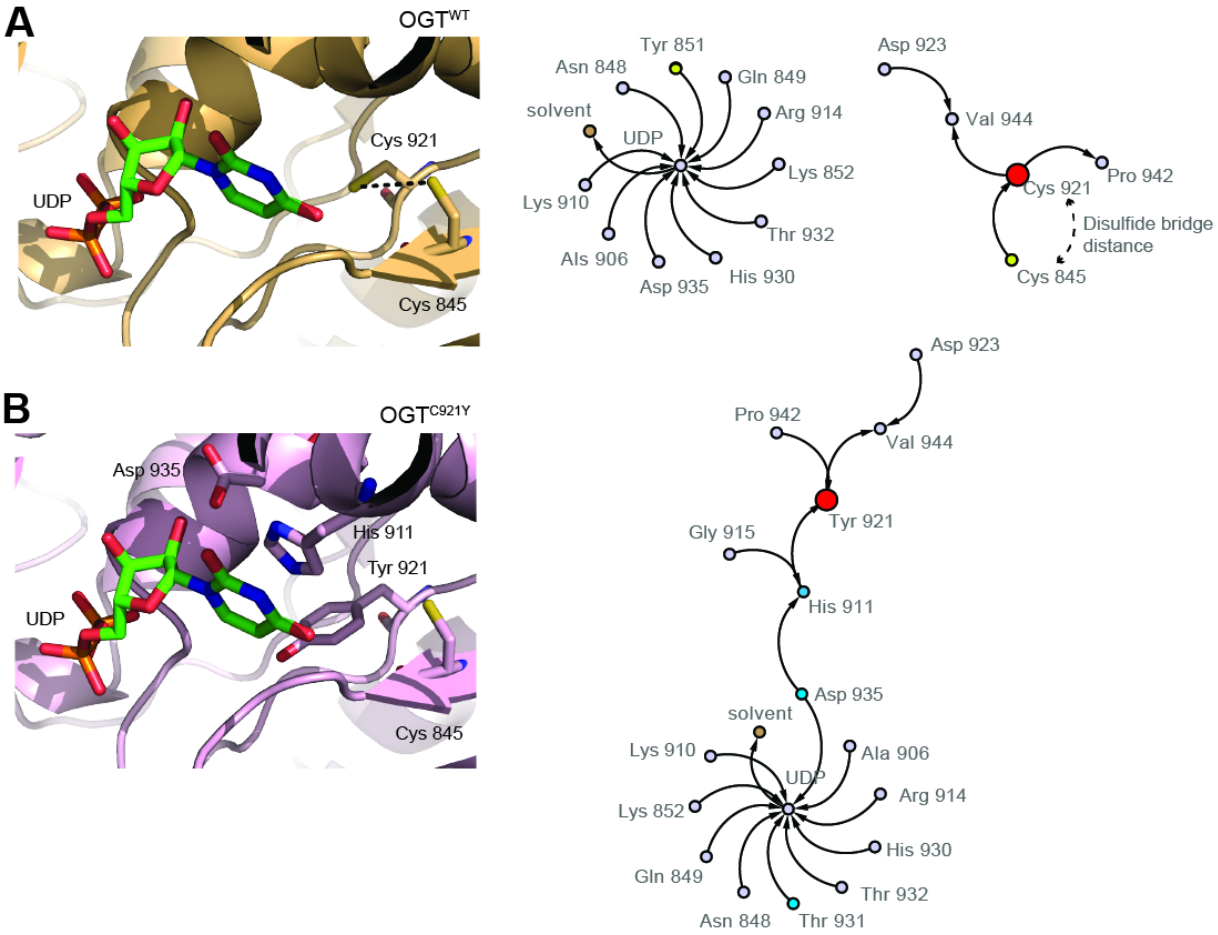
Zhu, Q., Cheng, X., Cheng, Y., Chen, J., Xu, H., Gao, Y., Duan, X., Ji, J., Li, X. and Yi, W. (2020). O-GlcNAcylation regulates the methionine cycle to promote pluripotency of stem cells. *Proc. Natl. Acad. Sci.* 201915582.

Supplementary information

Supplementary figure S1: *Figure removed for BioRxiv submission.* Profile photos and close up view of patient's eyebrows.

a) and b) Profile photographs of affected brothers carrying OGT^{C921Y} variant.

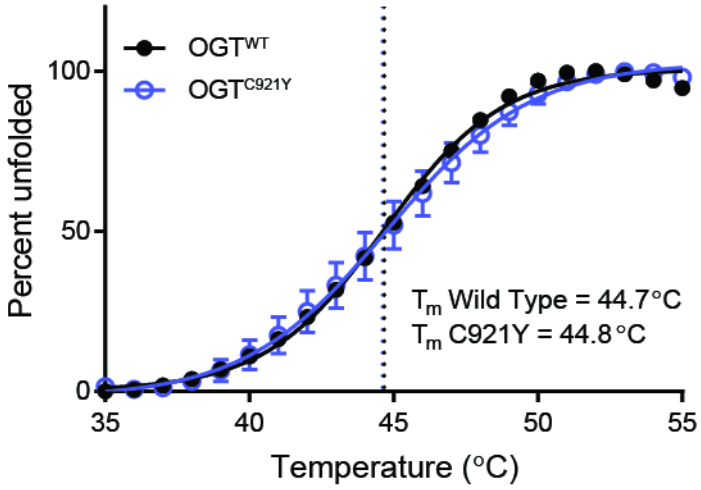
c) Proband III.2 has sparse eyebrows**.**



Supplementary figure S2: Model of the OGT C921Y substitution and changes in interactions in the catalytic core.

a) The PDB 5LWV containing a UDP molecule sitting in the active site (green sticks) and a fusion protein peptide, removed for our analysis, was used a template for the OGT^{WT} protein. Averaged MD trajectory calculated after 1000 poses is shown in the left side. For clarity, cysteine 921 is represented as a red circle.

b) The OGT^{C921Y} variant was modelled over the PDB 5LWV with the same modifications as described for OGT^{WT}. Averaged MD trajectory calculated after 1000 poses is shown in the left side. For clarity the mutant tyrosine 921 is represented with a red circle. All key residues are labelled.



Supplementary figure S3: Melting temperature determination for OGT^{WT} and OGT^{C921Y} (323 – 1044).

Thermal shift assay performed with recombinant OGT^{WT} and OGT^{C921Y} hOGT (323-1041). This experiment was performed three times, using three technical replicates each time. Error bars represent standard deviation.

A

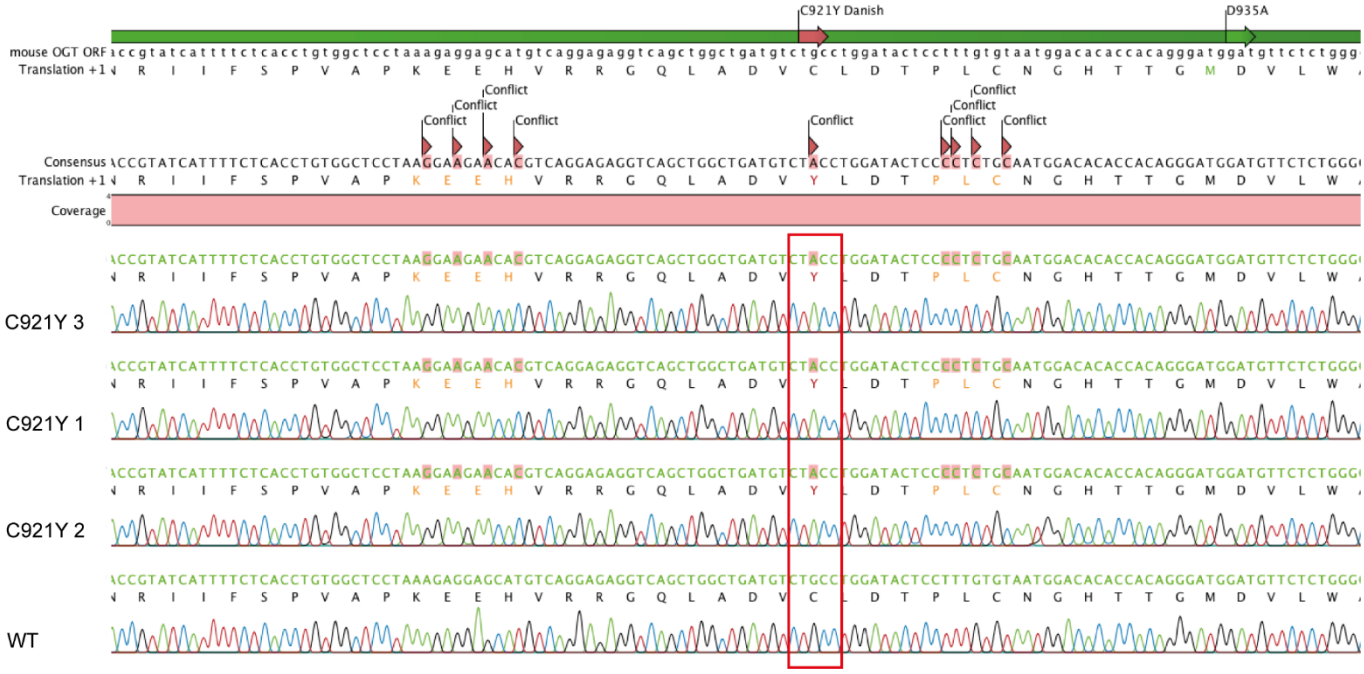
```
human  860 AIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRI      900
mouse      AIVYCNFNQLYKIDPSTLQMWANILKRVPNSVLWLLRFPAVGEPNIQQYAQNMGLPQNRI      900
           ************************************************************

human      IFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQL      960
mouse      IFSPVAPKEEHVRRGQLADVCLDTPLCNGHTTGMDVLWAGTPMVTMPGETLASRVAASQL      960
           ************************************************************

human      TCLGCLELIAKNRQEYEDIAVKLGTDLEYLKKVRGKVWKQRISSPLFNTKQYTMELERLY      1020
mouse      TCLGCLELIAKSRQEYEDIAVKLGTDLEYLKKIRGKVWKQRISSPLFNTKQYTMELERLY      1020
           **********.**********:**************************

human      LQMWEHYAAGNKPDHMIKPVEVTESA*     1046
mouse      LQMWEHYAAGNKPDHMIKPVEVTESA*     1046
           ***************************
```

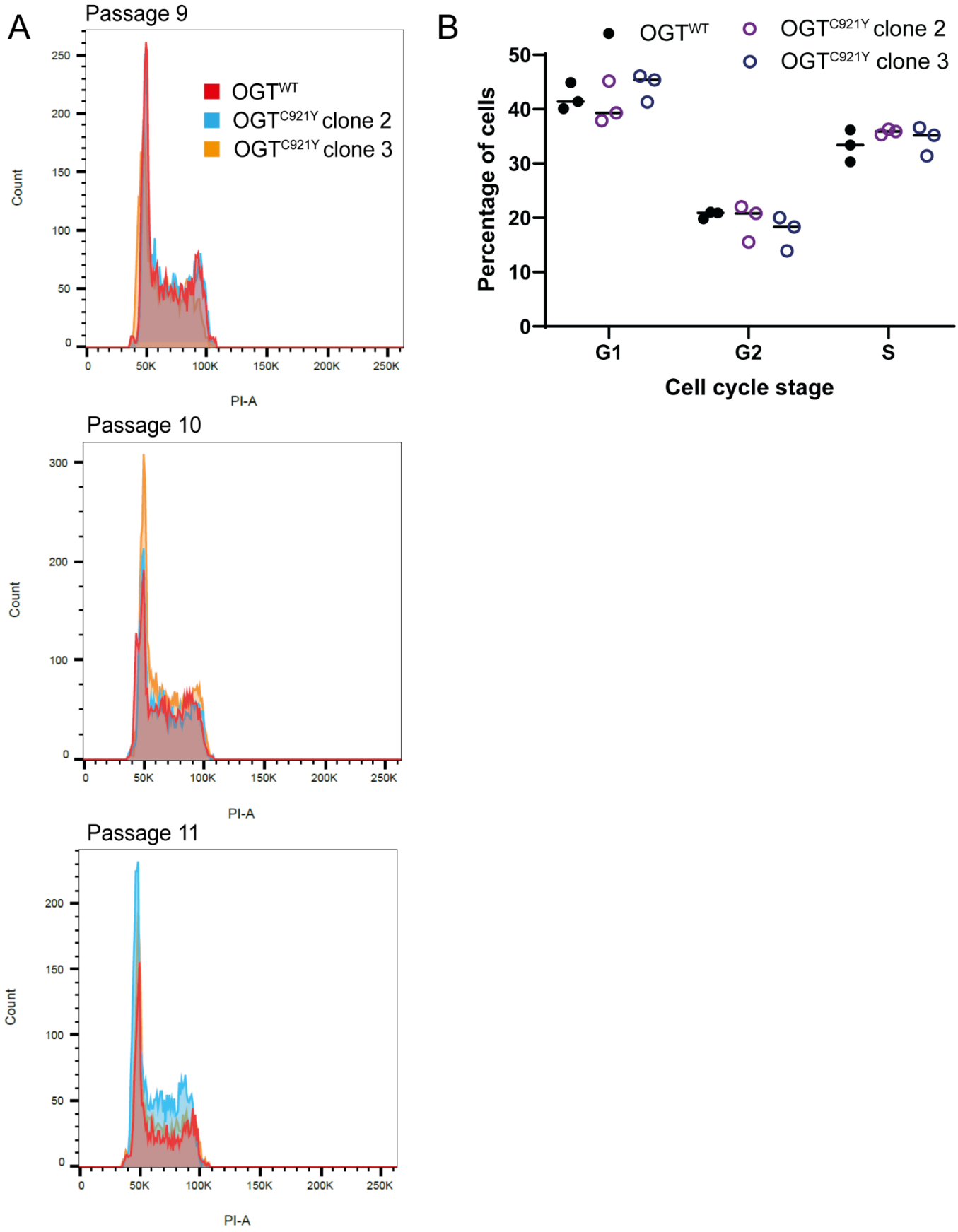
B



Supplementary figure S4: Generation of OGT^{C921Y} mESC lines.

a) Sequence alignment of human and mouse OGT (amino acids 840 – 1046). Cysteine 921 is highlighted in red.

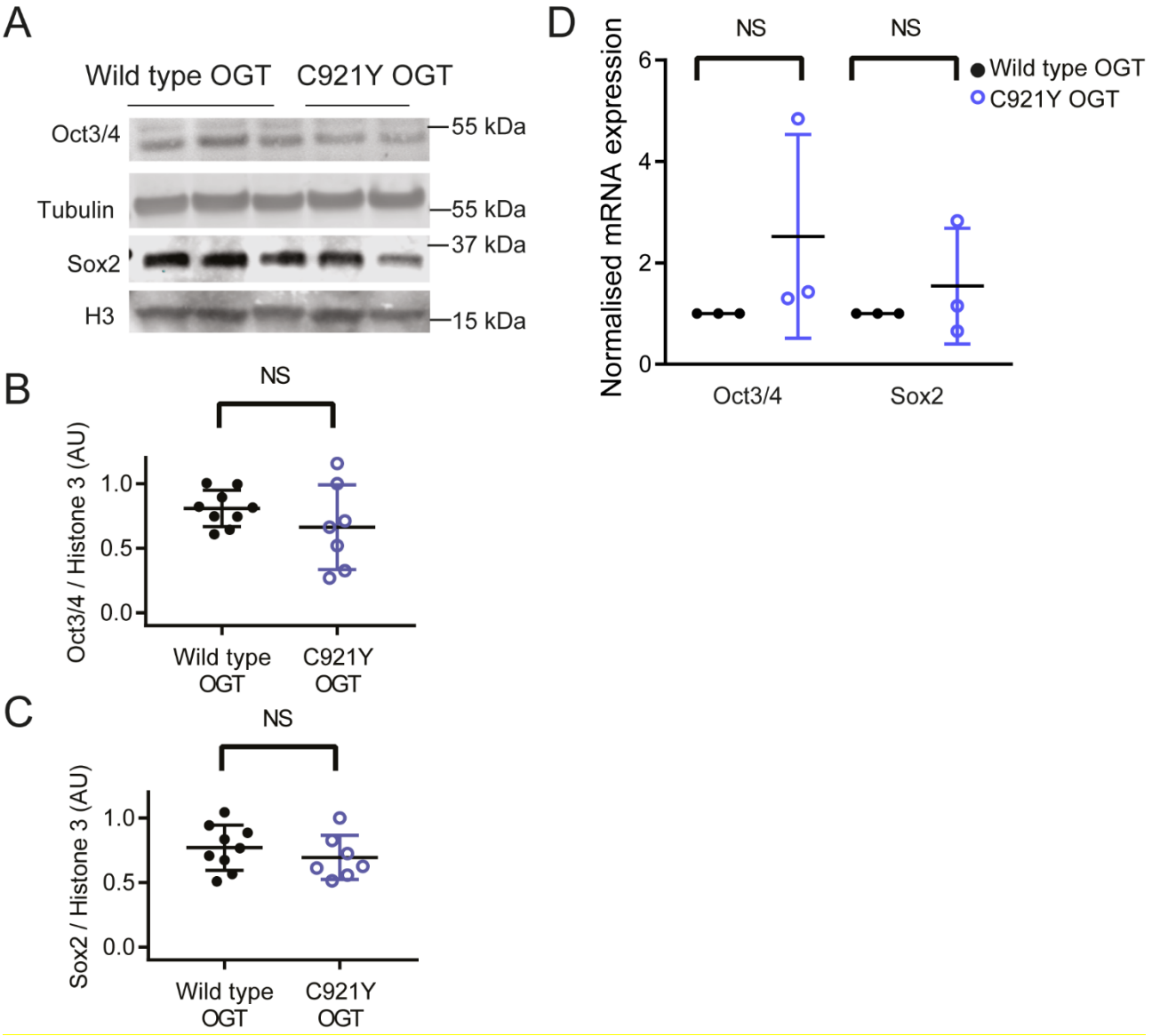
b) DNA sequencing of generated OGTC921Y mESC clones (chromas 1 – 3) and wild type mESC (Chroma 4). Positions corresponding to cysteine 921 are highlighted in red rectangle.



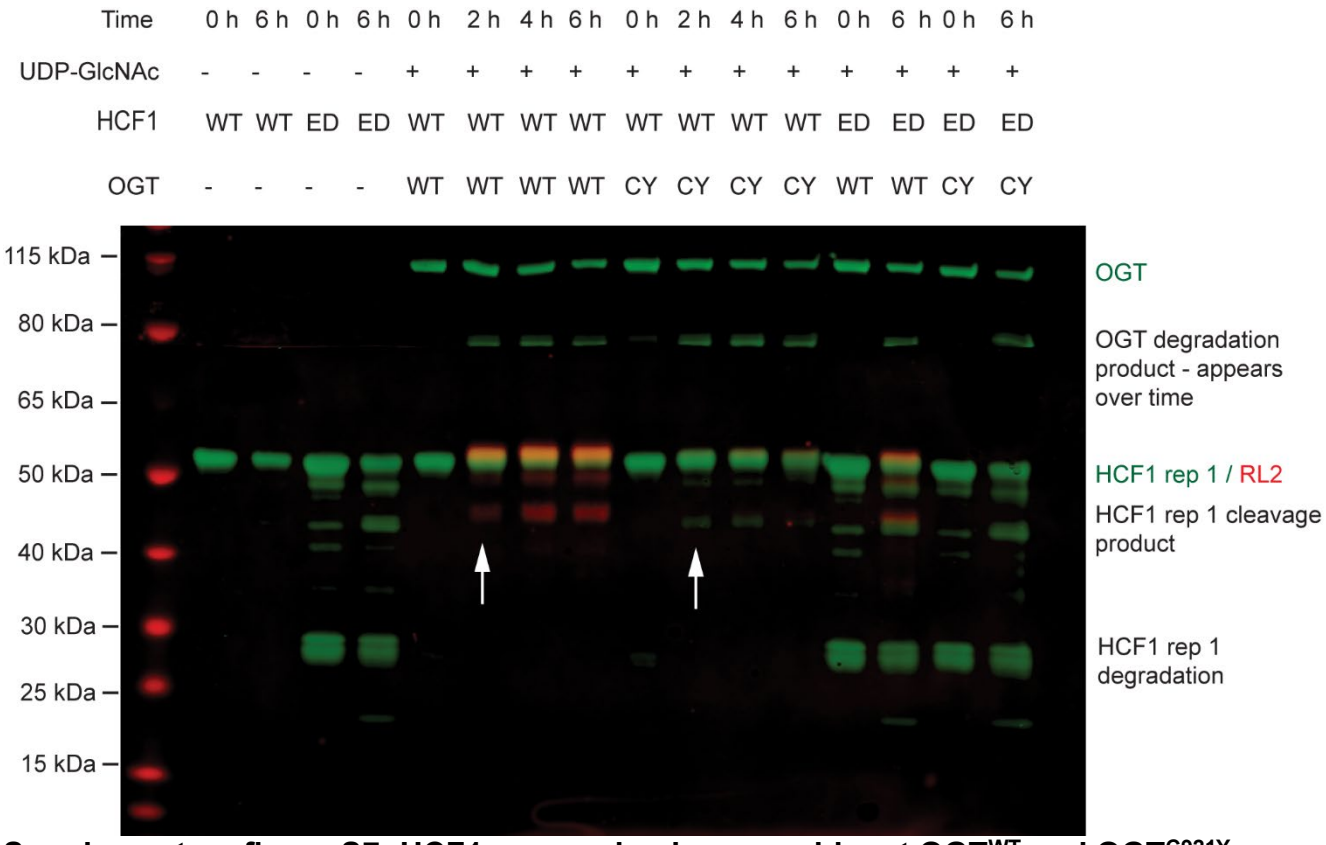
Supplementary figure S5: Cell cycle analysis of OGT^{WT} and OGT^{C921Y} mESC.

a) Cell cycle profiles of OGT^{WT} and OGT^{C921Y} across three passages using two mutant and one wild type cell line. We aimed to perform cell cycle profiling with the earliest available passages, which is the reason for not using all three cell OGT^{C921Y} cell lines in this experiment.

b) Plot of percentages of cells in each cell cycle stage.

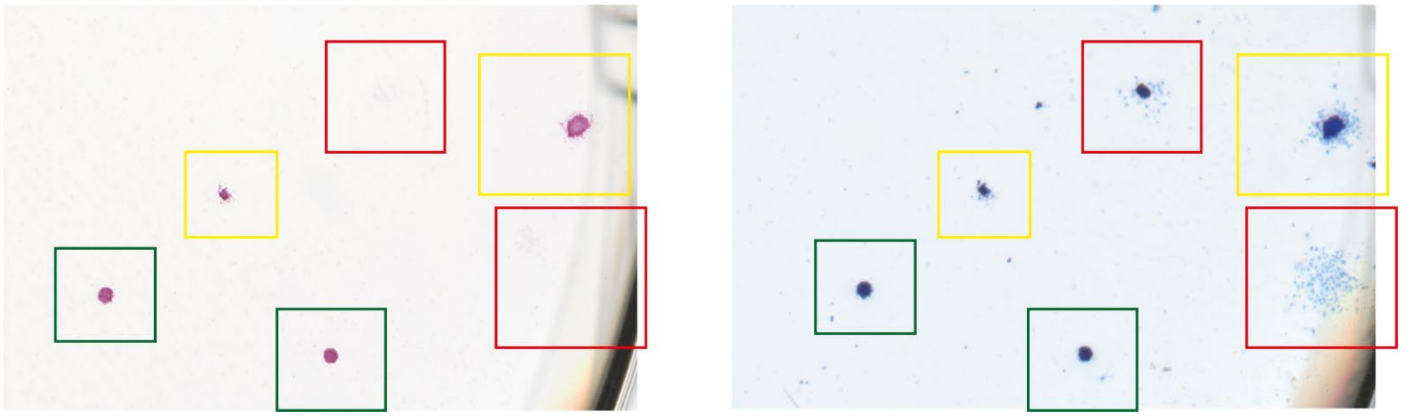


Supplementary figure S6: Oct3/4 and Sox2 expression in OGT^{WT} and OGT^{C921Y} mESC propagated in LIF. The quantification shown in panels B – D is based on results obtained from two to three different cell clones per genotype and repeated over three passages per clone.
a) Immunoblot of Oct3/4, Sox2, Tubulin and Histone3.
b) Quantification of Oct3/4 signal relative to tubulin. Unpaired *t* test, *p* value = 0.251. Error bars represent standard deviation.
c) Quantification of Sox2 signal relative to Histone 3. Unpaired *t* test, *p* value = 0.401. Error bars represent standard deviation.
d) mRNA expression of Oct3/4 and Sox2. Unpaired *t* test, Sox2 *p* value = 0.455, Oct3/4 *p* value = 0.260. Each data point represents normalised mean expression calculated from three separate RT-PCR runs. Each RT - PCR run was set up using several OGTWT and OGTC921Y as biological replicates. Error bars represent standard deviation.

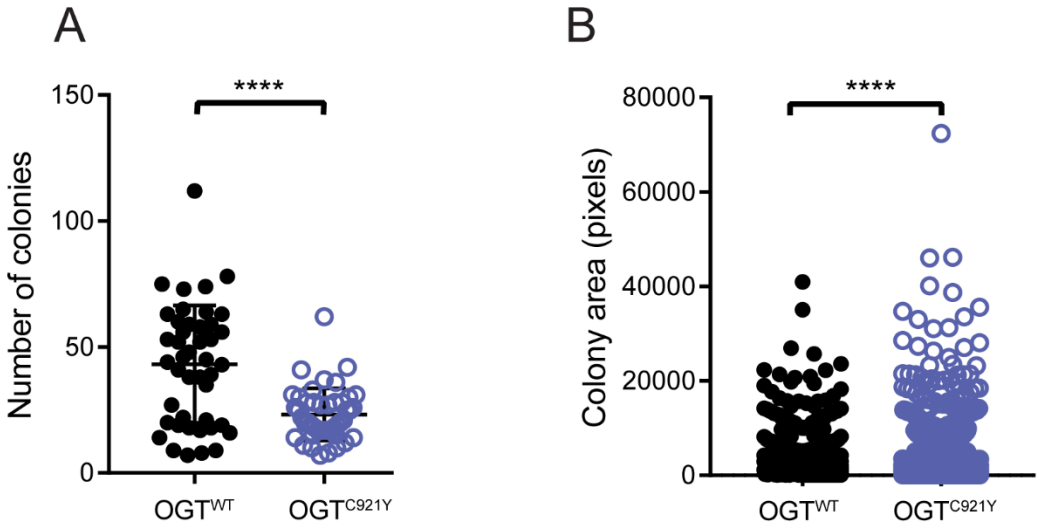


Supplementary figure S7: HCF1 processing by recombinant OGT^{WT} and OGT^{C921Y}.

Immunoblot of HCF1 proteolytic fragments produced in a time course *in vitro* reaction using the following human recombinant proteins: wild type (WT) and uncleavable (ED) HCF1 repeat 1 (HCF1rep1); OGT^{WT} and OGT^{C921Y}. HCF1 and OGT were detected in the green channel, O-GlcNAc (RL2) was detected in the red channel.



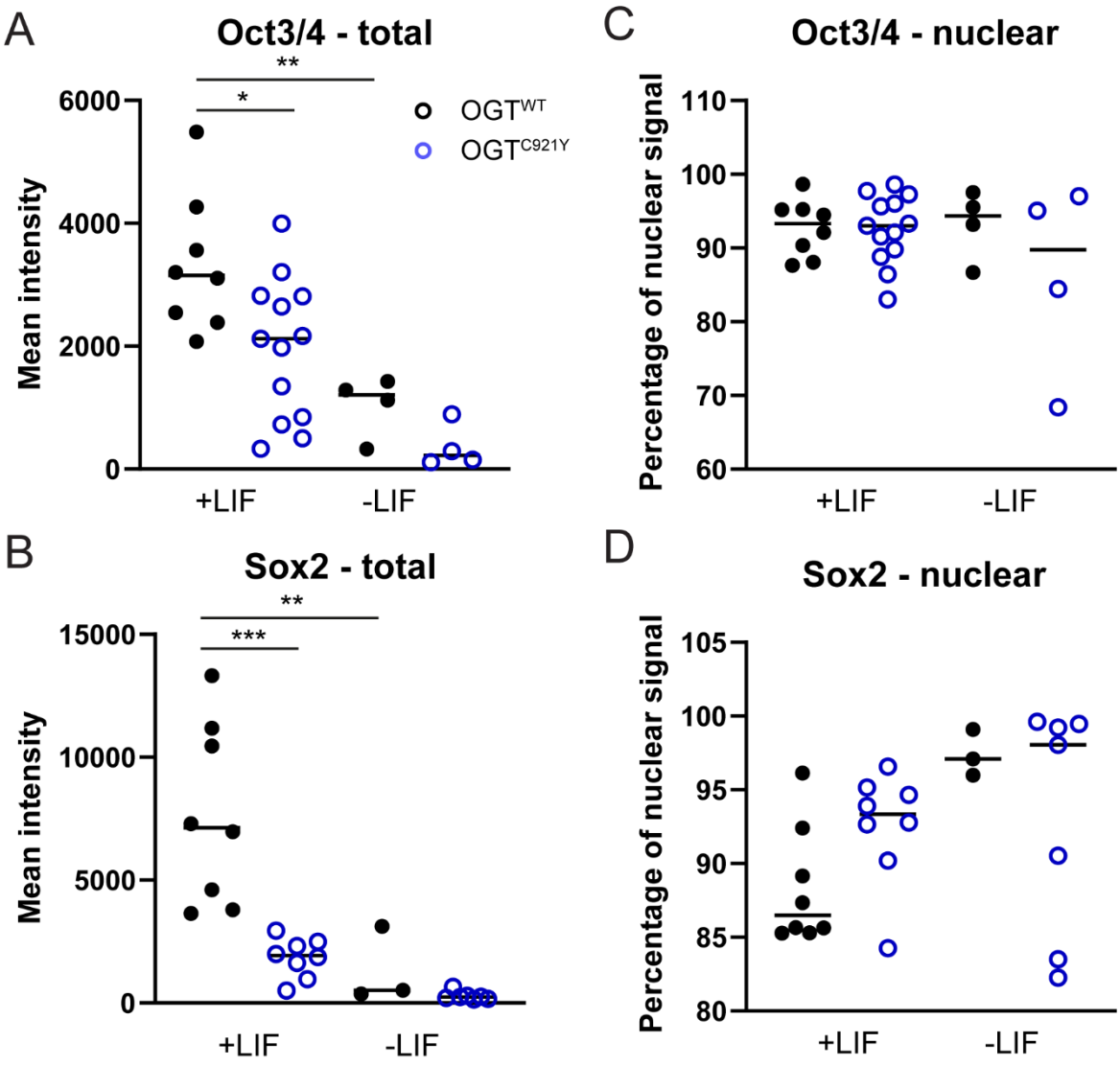
Supplementary figure S8: Colony scoring criteria. Images corresponding to the same section of a 6-well plate are shown. Image in the left represents colonies stained with ALP and the image on the right represents colonies stained with Coomassie. Compact colonies with strong ALP staining and no signs of differentiation were scored as undifferentiated (green rectangle), ALP positive colonies with signs of differentiation around the perimeter were scored as mixed (yellow rectangle) and dispersed colonies with no ALP signal were scored as differentiated (red rectangle).



Supplementary figure S9: Number and size of OGT^{WT} and OGT^{C921Y} colonies.

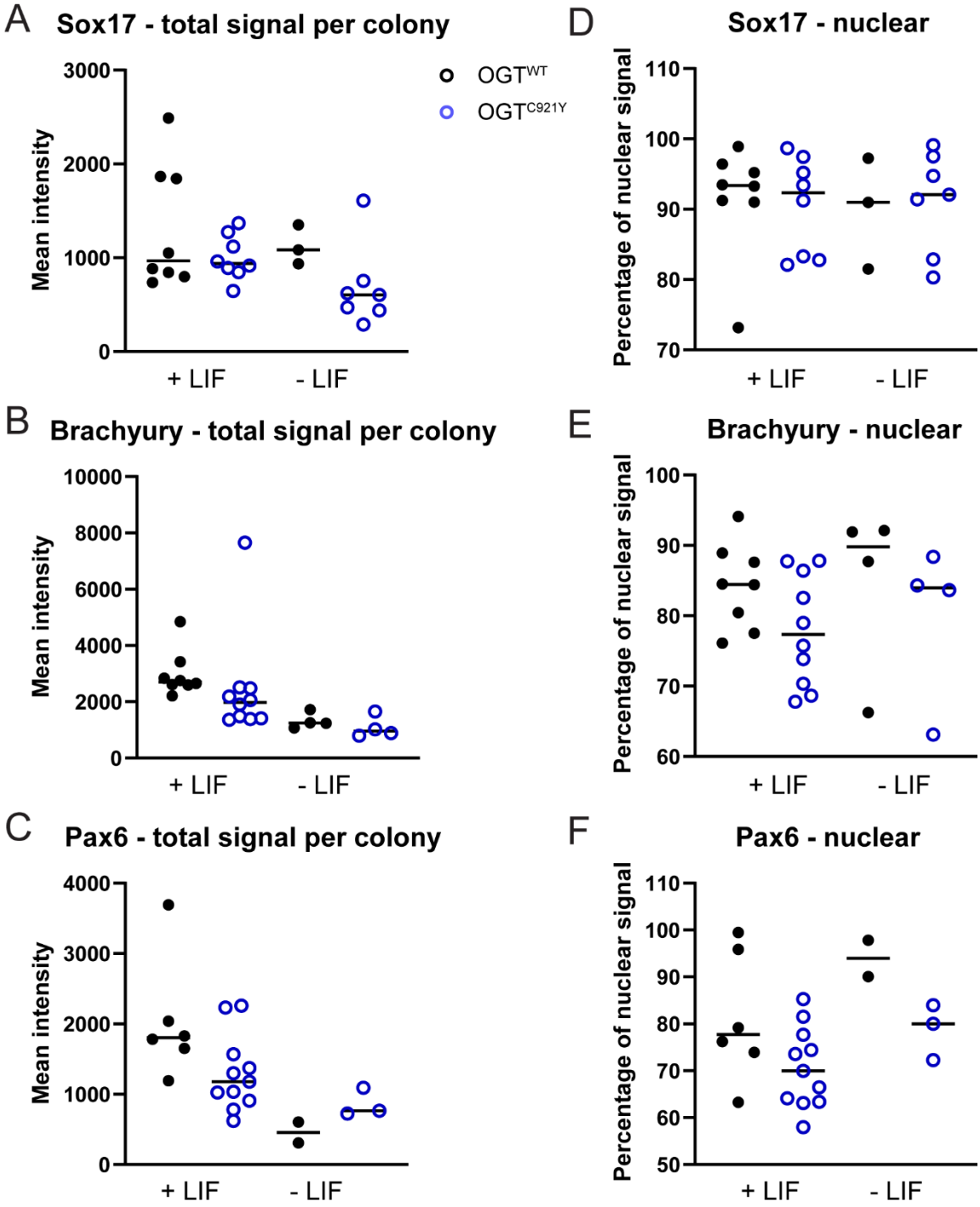
(f) Total number of colonies produced by wild type or OGT^{C921Y} mESCs in any condition. Unpaired t test, n = 45 wells scored, $p < 0.0001$. Error bars represent standard deviation.

(g) Mean colony area with SEM measured by automated ImageJ macro based on scans of colonies stained with Coomassie. Unpaired t test, wild type n = 2270 measured colonies, OGT^{C921Y} n = 1533 measured colonies, $p < 0.0001$.



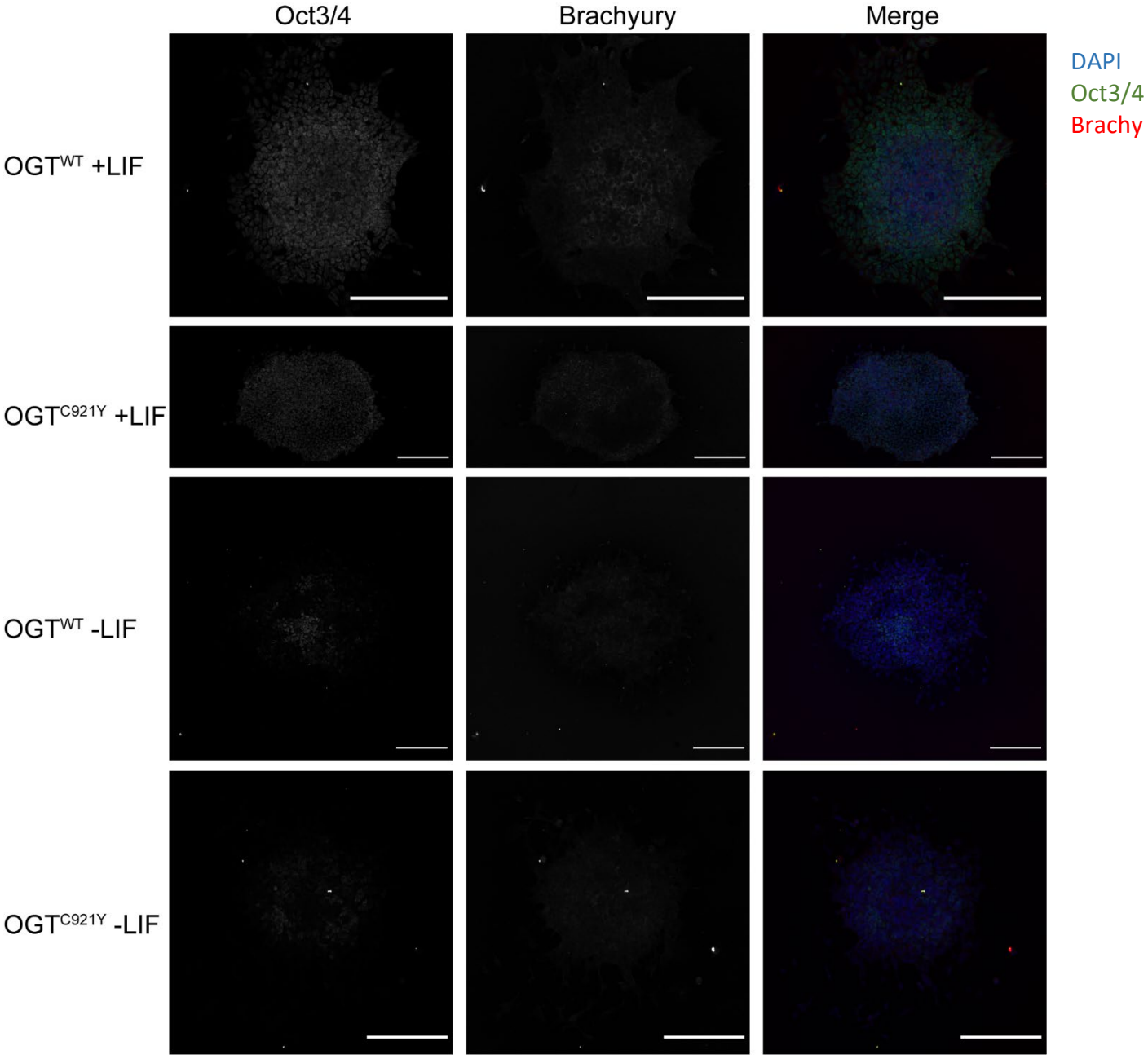
Supplementary figure S10: Oct3/4 and Sox2 levels in OGT^{WT} and OGT^{C921Y} colonies.

(A, B) Each data point represents the mean intensity per colony calculated as sum signal intensity divided by colony area. Oct3/4 +LIF OGT^{WT} versus OGT^{C921Y}, *p* value = 0.035 (One-way ANOVA followed by Sidak's multiple comparison test). Oct3/4 +LIF OGT^{WT} versus -LIF OGT^{C921Y}, *p* value = 0.006 (One-way ANOVA followed by Sidak's multiple comparison test). Sox2 +LIF OGT^{WT} versus OGT^{C921Y}, *p* value = 0.0001 (One-way ANOVA followed by Sidak's multiple comparison test). Sox2 +LIF OGT^{WT} versus -LIF OGT^{WT}, *p* value = 0.002 (One-way ANOVA followed by Sidak's multiple comparison test). (D – F) Each data point represents the percentage of nuclear signal calculated as sum nuclear signal divided by sum signal intensity. Two lines per genotype over two passages were used for this experiment. All the colonies present on a coverslip were imaged. One-way ANOVA showed no significant differences in nuclear signal between tested groups.

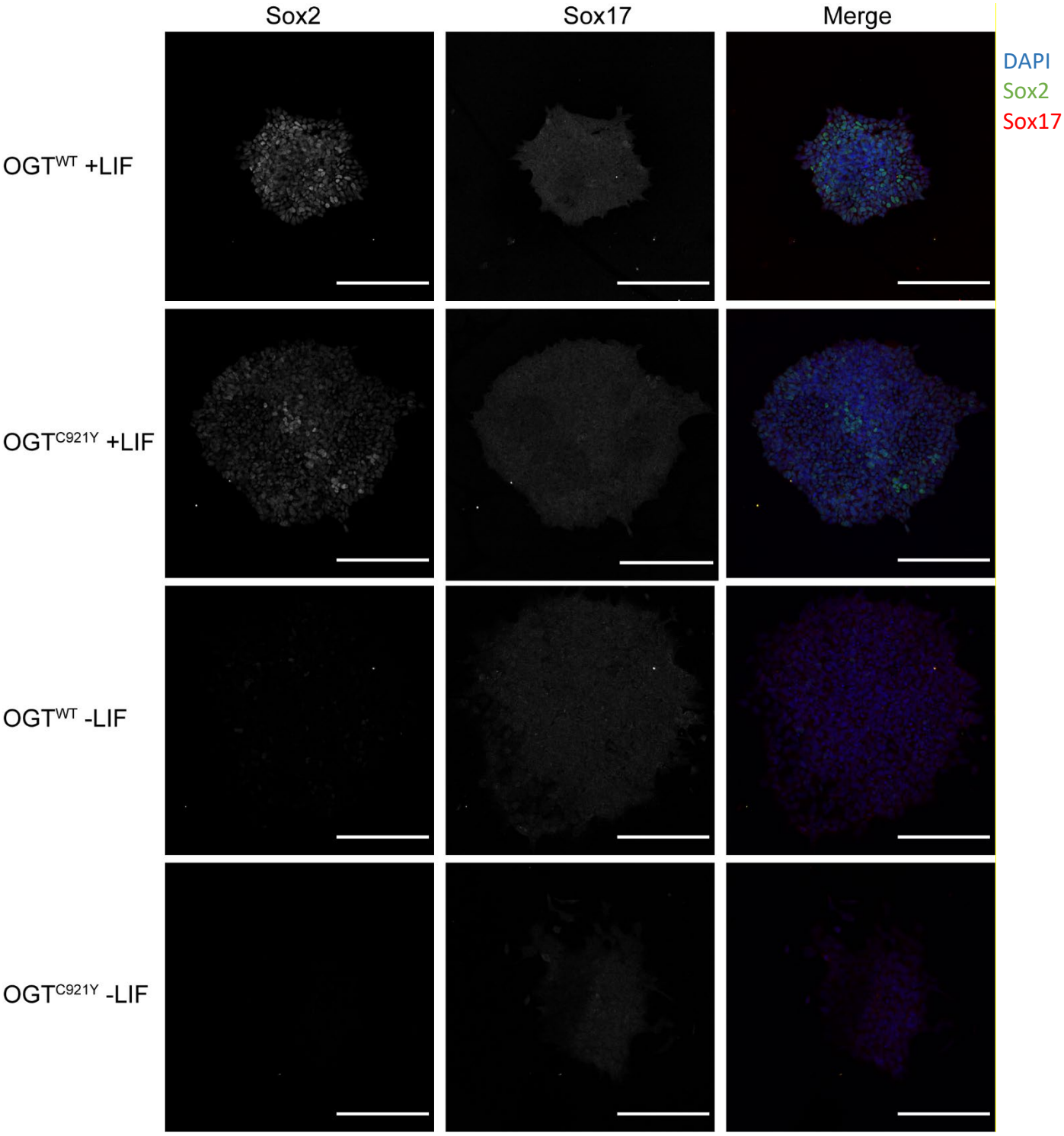


Supplementary figure S11: Sox17, Brachyury and Pax6 levels in OGT^{WT} and OGT^{C921Y} colonies.

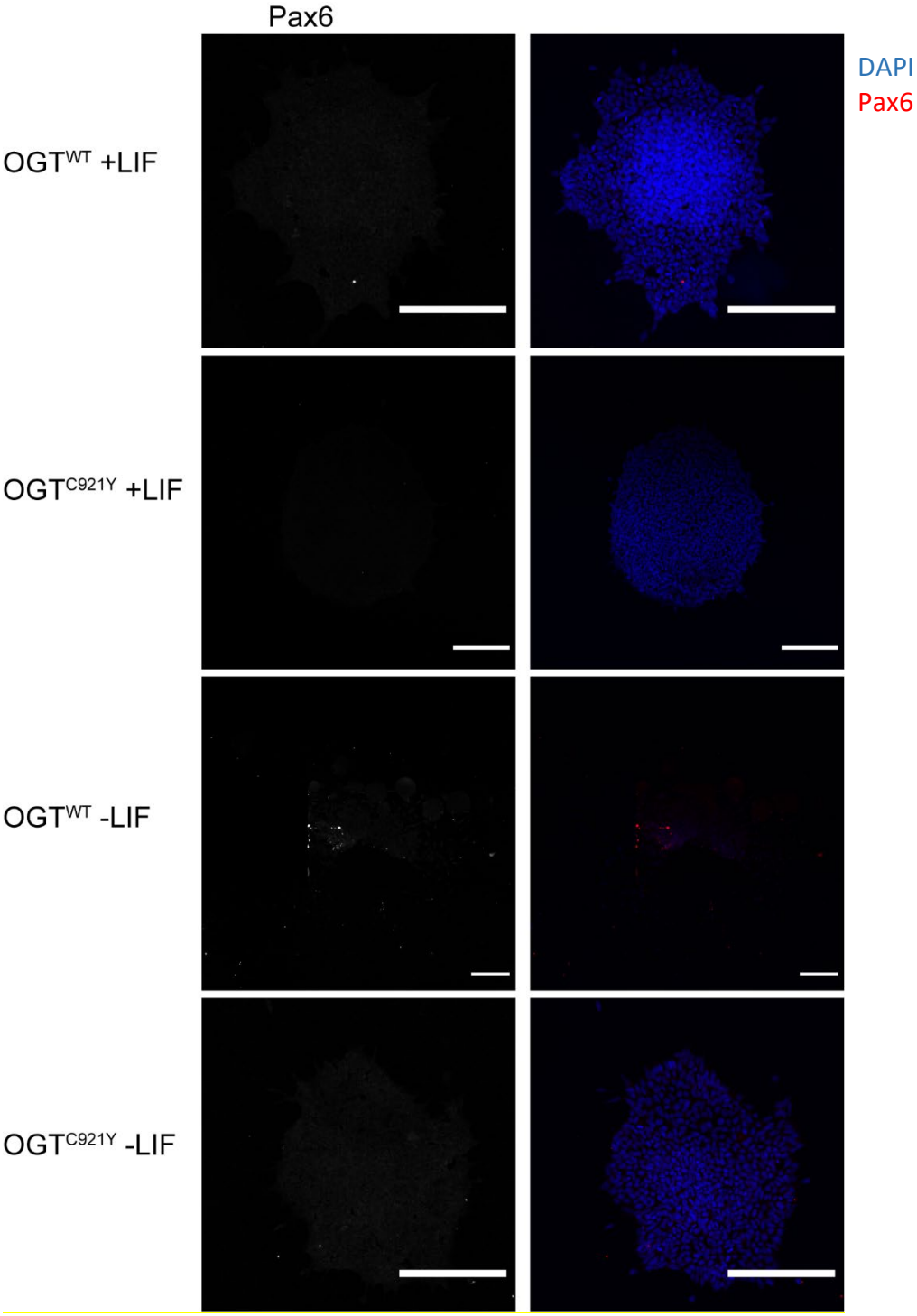
(A – C) Each data point represents the mean intensity per colony calculated as sum signal intensity divided by colony area. (D – F) Each data point represents the percentage of nuclear signal calculated as sum nuclear signal divided by sum signal intensity. Two lines per genotype over two passages were used for this experiment. All the colonies present on a coverslip were imaged. One-way ANOVA showed no significant differences between tested groups.



Supplementary figure S12: representative images of Oct3/4 and Brachyury staining in OGT^{WT} and OGT^{C921Y} colonies. Scale bar = 200 µM.



Supplementary figure S13: representative images of Sox2 and Sox17 staining in OGT^{WT} and OGT^{C921Y} colonies. Scale bar = 200 µM.



Supplementary figure S14: representative images of Pax6 staining in OGT^{WT} and OGT^{C921Y} colonies. Scale bar = 200 µM.

Supplementary table 1: Sequences of custom-made reagents used for generation of OGT^{C921Y} mESC lines.

Guide RNA	M935_left_F	CACC G ctcctgacatgctcctcttt
Guide RNA	M935_left_R	AAAC aaagaggagcatgtcaggagC
Guide RNA	M935_right_F	CACC G ctggatactcctttgtgtaa
Guide RNA	M935_right_R	AAAC ttacacaaaggagtatccag C
Genomic DNA	M935V6_fwd	aaaGGATCC acacaaaaacaacaactatatacatgaagg
Genomic DNA	M935V6_rev	aaaGCGGCCGC tttctctgtgtaataaagcccttaaatatc
Sequencing	M935seq	Ttggcaagctctgccaatag
Geneblock	M935 Wobble	ggagaacccaatattcaacaatatgcacaaaatatgggccttccccagaaccgt atcattttctcacctgtggctcctaaGgaAgaAcaCgtcaggagaggtcagctg gctgatgtctgcctggatactccCCtCtgCaatggacacaccacagggatgg CCgttctctgggcaggaacacccatggtgactatgccagg
Mutagenesis	C921Y_wobble_F	GAGGTCAGCTGGCTGATGTCTaCCTGGATACTCCCCTC TGCAATGGACACACCACAGGGATGGatGTTCTCTGGGC AGGAACACC
Genotyping	M935_screen_F	atgtggttttagggactttgtgagctc
Genotyping	M935_screen_R	gagaggatggtgccaagtattcaggc
RT-PCR	mOGT_solid_ex11_fwd	ggaatatcccagaagcaatagcttcttac
RT-PCR	mOGT_solid_end_rev	ggctgactcggtgacttcaacaggc
RT-PCR - sequencing	mOGT_ex18_seqF	tcctgatggaggtgacaatc

058 **Supplementary table 2: Sequences of qPCR primers.**

Primer	Sequence (5' -> 3')
Actb Forward	AGATCAAGATCATTGCTCCTCCT
Actb Reverse	ACGCAGCTCAGTAACAGTCC
Gapdh Forward	GGAGAGTGTTTCCTCGTCCC
Gapdh Reverse	*ACTGTGCCGTTGAATTTGCC*
18S Forward	*CTCAACACGGGAAACCTCAC*
18S Reverse	*CGCTCCACCAACTAAGAACG*
Ogt Forward	*CCCCCTGAGCCCTTCAAAAC*
Ogt Reverse	*TCGTTGGTTCTGTACTGTCGG*
Oga Forward	*TGCAGTGGTTAGGGTGTCG*
Oga Reverse	*AGCAAACGCTGGAACTCTCC*
HCFC1 Forward	*GTGACCCAGGTGGTGCTAAA*
HCFC1 Reverse	*TTGACAGCAGAGACGGTGAC*
Oct4 Forward	*GGTGGAACCAACTCCCGAGG*
Oct4 Reverse	*ACCTTTCCAAAGAGAACGCCC*

059

060 **Supplementary information 1: ImageJ colony analysis macro**

```
061 OG =  getTitle();
062 n =  roiManager("count")
063 for  (i=1; i<=n; i++) {
064  roiManager("Select", i-1);
065  run("Duplicate...", "duplicate");
066  run("Tiff...");
067  run("Clear Outside");
068  rename(i);
069  selectWindow(OG);}
070 for  (i=1; i<=n; i++) {
071  selectWindow(i);
072  setAutoThreshold("MaxEntropy");
073  setOption("BlackBackground", true);
074  run("Convert to Mask");}
```

```
        roiManager("reset");
for     (i=1; i<=nImages-1;i++){
        roiManager("reset")
        selectWindow(i);
        run("Analyze Particles...", "size=80-Infinity circularity=0.01-1.00 display exclude clear add");
        saveAs("Results");}
while   (nImages>0) {
        selectImage(nImages);
        close();}
roiManager("reset");
```

Supplementary information 2: ImageJ colony analysis macro

```
//Clears ROIs and results as well as sets measurement parameters (integrated means sum of
pixels)
run("Clear Results");
roiManager("reset");
run("Set Measurements...", "area mean standard min integrated display redirect=None
decimal=3");

//Dialog boxes to choose different thresholding approaches and number of channels in image
channel2 = false;
channel3 = false;
test_mode = false;
image_selection = false
manual_selection = true
ind_nuclei = true
//min_nuclear_size = 60
//max_nuclear_size = 100
//circularity = 0.3
threshold_types = newArray("Local", "Global");

global_thresholds = newArray("Default", "Huang", "Intermodes",
"Isodata", "IJ_IsoData", "Li", "MaxEntropy","Mean", "MinEror",
"Minimum", "Moments", "Otsu", "Percentile","RenyiEntropy",
"Shanbhag", "Triangle", "Yen");

local_thresholds = newArray("Phansalkar", "Otsu", "Bernsen", "Contrast", "Mean", "Median",
"MidGrey", "Niblack", "Sauvola");

Dialog.create("Settings");
Dialog.addCheckbox("Select which images to include (otherwise all processed)",
image_selection);
Dialog.addCheckbox("Manually select region of interest", manual_selection);
Dialog.addMessage("Number of non-DAPI channels (default 1):");
Dialog.addCheckbox("2", channel2);
Dialog.addCheckbox("3", channel3);
Dialog.addChoice("Thresholding type", threshold_types);
Dialog.addString("Total result file name", "Results");
```

```
Dialog.show();


image_selection = Dialog.getCheckbox();
manual_selection = Dialog.getCheckbox();
channel2= Dialog.getCheckbox();
channel3 = Dialog.getCheckbox();
threshold_types = Dialog.getChoice();
output_filename = Dialog.getString();


if (threshold_types == "Global"){
	Dialog.create("Thresholding");
	Dialog.addMessage("Thresholding method:");
	Dialog.addChoice("Thresholding method", global_thresholds);
	Dialog.show();
	threshold_method = Dialog.getChoice();
}

if (threshold_types == "Local"){
	Dialog.create("Thresholding");
	Dialog.addMessage("Thresholding method (only select try all in test mode):");
	Dialog.addChoice("Thresholding method", local_thresholds);
	Dialog.show();
	threshold_method = Dialog.getChoice();
}

File.setDefaultDir(getDirectory("Choose Folder This is in"))

//Establishes folders to take images from and place masks in
input = File.getDefaultDir + "\\RAW_Input\\";
output = File.getDefaultDir;
output_masks = File.getDefaultDir + "\\Masks\\";

//conditional to either show or hide analysis
if (test_mode == true){
	setBatchMode(false);
}
else {
	setBatchMode(true);
}

//Makes result file (.csv)
var resultFileLine;

resultFileLineMod("init", "FileName", true);

```

```
resultFileLineMod("append", "Nuclear Area", true);
resultFileLineMod("append", "Nuclear Mean Signal C1", true);
resultFileLineMod("append", "Nuclear StDev Signal C1", true);
resultFileLineMod("append", "Nuclear Sum Intensity C1", true);
resultFileLineMod("append", "Non-Nuclear C1 Area", true);
resultFileLineMod("append", "Non-Nuclear C1 Mean Signal", true);
resultFileLineMod("append", "Non-Nuclear StDev Signal C1", true);
resultFileLineMod("append", "Non-Nuclear C1 Sum Intensity", true);

if (channel2 == true){
	resultFileLineMod("append", "Nuclear Mean Signal C2", true);
	resultFileLineMod("append", "Nuclear StDev Signal C2", true);
	resultFileLineMod("append", "Nuclear Sum Intensity C2", true);
	resultFileLineMod("append", "Non-Nuclear C2 Area", true);
	resultFileLineMod("append", "Non-Nuclear C2 Mean Signal", true);
	resultFileLineMod("append", "Non-Nuclear StDev Signal C2", true);
	resultFileLineMod("append", "Non-Nuclear C2 Sum Intensity", true);
}

if (channel3 == true){
	resultFileLineMod("append", "Nuclear Mean Signal C3", true);
	resultFileLineMod("append", "Nuclear StDev Signal C3", true);
	resultFileLineMod("append", "Nuclear Sum Intensity C3", true);
	resultFileLineMod("append", "Non-Nuclear C3 Area", true);
	resultFileLineMod("append", "Non-Nuclear C3 Mean Signal", true);
	resultFileLineMod("append", "Non-Nuclear StDev Signal C3", true);
	resultFileLineMod("append", "Non-Nuclear C3 Sum Intensity", true);
}

resultFileLineMod("writeFile", output + output_filename + ".csv", false);

//function for making and writing to result file
function resultFileLineMod(command, parameter, addSeparator){
	resultFileSeparator=", ";
	if (command=="init"){
		resultFileLine=""+parameter;
		if (addSeparator)	{
			resultFileLine=resultFileLine+""+resultFileSeparator;	}	}
	else if (command=="append")	{
		resultFileLine=resultFileLine+""+parameter;
		if (addSeparator)	{
			resultFileLine=resultFileLine+""+resultFileSeparator;	}	}
	else if (command=="report")	{
		print(resultFileLine);}
	else if (command=="writeFile")	{
		File.append(resultFileLine, parameter);	}
}
```

```
216
217 //function to saves images (in this context overlay of ROI)
218 function make_overlay(image, selection_index){
219         selectImage(image);
220         //name = getTitle();
221         roiManager("deselect");
222         roiManager("select", selection_index);
223         roiManager("draw");
224         return image
225         //saveAs("Tiff", output_masks + substring(name, 0, lengthOf(name)-5) + s);
226 }
227
228 function title_from_ID(imageID){
229         selectImage(imageID);
230         return getTitle();
231 }
232
233 function select_ROI(){
234         setBatchMode(false);
235         waitForUser("Select region of interest");
236         roiManager("add");
237         roiManager("select", 0);
238         run("Crop");
239         run("Clear Outside");
240         roiManager("reset");
241         resetMinAndMax();
242         if (test_mode == false){
243                 setBatchMode(true);
244         }
245 }
246
247 function export_overlay(){
248         overlay_DAPI = title_from_ID(overlay_DAPI_ID);
249         overlay_C2 = title_from_ID(overlay_C2_ID);
250         if (channel3 == true){
251                 overlay_C3 = title_from_ID(overlay_C3_ID);
252                 overlay_C4 = title_from_ID(overlay_C4_ID);
253                 run("Merge Channels...", "c1=["+overlay_DAPI+"] c2=["+overlay_C2+"]
254 c3=["+overlay_C3+"] c4=["+overlay_C4"] create");
255                 saveAs("Tiff", output_masks + Original + s);
256         }
257         else if (channel2 == true){
258                 overlay_C3 = title_from_ID(overlay_C3_ID);
259                 run("Merge Channels...", "c1=["+overlay_DAPI+"] c2=["+overlay_C2+"]
260 c3=["+overlay_C3+"]  create");
261                 saveAs("Tiff", output_masks + Original + s);
262         }
```

```
263         else{
264             run("Merge Channels...", "c1=["+overlay_DAPI+"] c2=["+overlay_C2+"] create");
265             saveAs("Tiff", output_masks + Original + s);
266         }
267 }
268
269 //Segmentation of nuclei, global threshold approach not optimised, Phansalkar local best
270 function nuclear_mask(DAPI){
271         selectImage(DAPI);
272         run("Duplicate...", " ");
273         //Raw image for mask output
274         for_export = getImageID();
275         if (threshold_types == "Local"){
276             selectImage(DAPI);
277             //blur to reduce effect of uneven nuclear staining
278             run("Median...", "radius=1.5");
279             run("Duplicate...", " ");
280             pDAPI = getImageID();
281             selectImage(DAPI);
282             setAutoThreshold("MinError dark");
283             setOption("BlackBackground", false);
284             run("Convert to Mask");
285             run("Create Selection");
286             roiManager("add");
287             selectImage(pDAPI);
288             //stretches histogram across total range of possible pixel values
289             run("Enhance Contrast...", "saturated=0 equalize");
290             run("Duplicate...", " ");
291             roiManager("select", 0)
292             run("Clear Outside");
293             run("Select None");
294             roiManager("reset");
295             run("Duplicate...", " ");
296             temp_threshold = getImageID();
297             setAutoThreshold("Otsu dark no-reset");
298             //run("Threshold...");
299             setOption("BlackBackground", false);
300             run("Convert to Mask");
301             run("Create Selection");
302             roiManager("add");
303             if (test_mode == true){
304             waitForUser("debug");
305             }
306             close();
307             selectImage(pDAPI);
308             roiManager("select", 0);
309             run("Clear Outside");
```

```
                roiManager("reset");
                run("Select None");
                run("8-bit");
                run("Auto Local Threshold", "method="+threshold_method+" radius=15
parameter_1=0 parameter_2=0 white");
                if (test_mode == true){
                waitForUser("debug");
                }
                run("Invert");
                run("Erode");
                run("Erode");
                run("Dilate");
                run("Dilate");
                //run("Watershed");
                if (test_mode == true){
                waitForUser("debug");
                }
                run("Create Selection");
                roiManager("add");
                close();
                close();
                }
        else {
                selectWindow(DAPI);
                setAutoThreshold(""+threshold_method+" white");
                //run("Threshold...");
                setOption("BlackBackground", false);
                run("Convert to Mask");
                run("Dilate");
                run("Dilate");
                run("Erode");
                run("Erode");
                run("Invert");
                run("Create Selection");
        }
        if (test_mode == true){
                waitForUser("debug");
        }
        return make_overlay(for_export, 0);
        if (test_mode == true){
                waitForUser("debug");
        }
}

//makes conservative i.e. maximal mask of given channel and adds as ROI
function channel_masks(channel){
        selectImage(channel);
```

```
357             run("Select None");
358             run("Duplicate...", "duplicate");
359             mask_image = getImageID();
360             selectImage(mask_image);
361             setAutoThreshold("MinError dark");
362             //run("Threshold...");
363             setOption("BlackBackground", false);
364             run("Convert to Mask");
365             run("Erode");
366             run("Erode");
367             run("Erode");
368             run("Dilate");
369             run("Dilate");
370             run("Dilate");
371             run("Create Selection");
372             roiManager("Add");
373             close();
374     }
375
376     function measure_parameters(channel){
377             //duplicate nuclear roi as roiManager("combine") replaces instead of adding
378             roiManager("select", 0)
379             roiManager("Add")
380             channel_masks(channel);
381             selectImage(channel);
382             nuclear_and_TFs = newArray(1,2);
383             //Make a new ROI representing DAPI signal AND CX signal (i.e. cytoplasm and nuclear
384     signal in one ROI from nuclear and C1 signal)
385             roiManager("select", nuclear_and_TFs);
386             roiManager("Combine");
387             //Make a new ROI representing non-nuclear CX signal (i.e. cytoplasm CX from exclusive
388     OR of nuclear and CX+nuclear signal)
389             //Total signal now has index 1, rest same (i.e. orignal nuclear at 0, channel at 2)
390             nuclear_and_total = newArray(0, 1);
391             roiManager("select", nuclear_and_total);
392             roiManager("XOR");
393             roiManager("Add");
394             roiManager("show all without labels");
395             //result of XOR at index 3 i.e. infered cytoplasm
396             nuclear_and_cytoplasmic = newArray(0,3);
397             roiManager("select", nuclear_and_cytoplasmic);
398             selectImage(channel);
399             roiManager("Measure");
400             //as only relevant ROI selected, index of cytoplasm results is 1
401             if (test_mode == true){
402                     waitForUser("debug");
403             }
```

```
404             resultFileLineMod("append", getResult("Mean", 0), true);
405             resultFileLineMod("append", getResult("StdDev", 0), true);
406             resultFileLineMod("append", getResult("IntDen", 0), true);
407             resultFileLineMod("append", getResult("Area", 1), true);
408             resultFileLineMod("append", getResult("Mean", 1), true);
409             resultFileLineMod("append", getResult("StdDev", 1), true);
410             resultFileLineMod("append", getResult("IntDen", 1), true);
411             run("Clear Results");
412             overlay = make_overlay(channel, 3);
413             roiManager("select", newArray(1,2,3));
414             roiManager("delete");
415             return overlay
416
417     }
418     //function which ties analysis together
419     function nuclear_vs_non(){
420             Original = getTitle();
421             Original_ID = getImageID();
422             run("Duplicate...", "duplicate");
423             Duplicate = getTitle();
424             run("Split Channels");
425             selectWindow("C1-" + Duplicate);
426             DAPI = getImageID();
427             selectWindow("C2-" + Duplicate);
428             pC2 = getImageID();
429             if (channel2 == true){
430                     selectWindow("C3-" + Duplicate);
431                     pC3 = getImageID();
432
433             }
434             if (channel3 == true){
435                     selectWindow("C4-" + Duplicate);
436                     pC4 = getImageID();
437             }
438
439             resultFileLineMod("init", Original, true);
440             overlay_DAPI_ID = nuclear_mask(DAPI);
441             //measures nuclear area     roiManager("Measure");
442
443             roiManager("Measure");
444             resultFileLineMod("append", getResult("Area", 0), true);
445             run("Clear Results");
446             run("Select None");
447             overlay_C2_ID = measure_parameters(pC2);
448
449             if (channel2 == true){
450                     overlay_C3_ID = measure_parameters(pC3);
```

```
451                 }
452
453         if (channel3 == true){
454                 overlay_C4_ID = measure_parameters(pC4);
455                 }
456         export_overlay();
457         print("Analysis Complete: " + Original + s);
458 }
459
460
461 function keep_image(){
462         setBatchMode(false);
463         keep_image_bool = getBoolean("Keep image?");
464         if (test_mode == false){
465                 setBatchMode(true);
466         }
467         if(keep_image_bool == true){
468                 if (manual_selection == true){
469                         select_ROI();
470                 }
471                 return true;
472         }
473         else{
474                 return false;
475         }
476 }
477
478 function bio_formats_open(in_dir, out_dir, filename){
479         run("Bio-Formats Macro Extensions");
480         seriesToOpen = newArray;
481         sIdx = 0;
482         path = in_dir+filename;
483         print(path);
484         Ext.setId(path);
485         Ext.getSeriesCount(seriesCount);
486         for(s = 1; s <= seriesCount; s++){
487         run("Bio-Formats Importer", "open=[" + path + "] autoscale color_mode=Composite
488 rois_import=[ROI manager] view=Hyperstack stack_order=XYCZT series_"+s+"");
489         if (image_selection == true){
490                 if (keep_image() == true){
491                         nuclear_vs_non();
492                 }
493                 else{
494                         close();
495                 }
496         }
497         else{
```

```
                if (manual_selection == true){
                        select_ROI();
                }
                nuclear_vs_non();
        }
        run("Close All");
        resultFileLineMod("writeFile", output + output_filename + ".csv", false);
        roiManager("reset")
        }
}

list = getFileList(input);

for (i = 0; i < list.length; i++){
        if (test_mode == false){
                setBatchMode(true);
        }
      bio_formats_open(input, output, list[i]);
}

setBatchMode(false);
```