Discovery of a novel O-mannosylation pathway selectively serving cadherins and protocadherins

Ida Signe Bohse Larsen[1*], Yoshiki Narimatsu[1*], Hiren Jitendra Joshi[1*], Lina Siukstaite[1], Oliver J. Harrison[2], Julia Brasch[2], Kerry Goodman[2], Lars Hansen[1], Lawrence Shapiro[2,3], Barry Honig[2,3,4], Sergey Y. Vakhrushev[1], Henrik Clausen[1], Adnan Halim[1,5,6].

[1]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark

[2]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

[3]Zuckerman Mind Brain Behavior Institute and Department of Systems Biology, Columbia University, New York, NY 10032.

[4]Howard Hughes Medical Institute.

[5]Present address: Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065, USA

[*]These authors contributed equally and are considered co-first authors.

[6]To whom correspondence should be addressed: E-mail: halim@sund.ku.dk

Key words: Glycoproteomics, O-glycosylation, glycosyltransferase, mass spectrometry, gene editing,

Abstract

The cadherin superfamily of adhesion molecules carry O-linked mannose (O-Man) glycans at highly conserved sites localized to specific β-strands of their extracellular cadherin (EC) domains. These O-Man glycans do not appear to be elongated like O-Man glycans found on α-dystroglycan, and we recently demonstrated that initiation of cadherin/protocadherin O-Man

glycosylation is not dependent on the evolutionary conserved POMT1/POMT2 enzymes that initiate O-Man glycosylation on β-dystroglycan. Here, we used a CRISPR/Cas9 genetic dissection strategy combined with sensitive and quantitative O-Man glycoproteomics to identify a homologous family of four putative protein O-mannosyltransferases encoded by the *TMTC1-4* genes, which were found to be imperative for cadherin and protocadherin O-Man glycosylation. Knockout of all four *TMTC* genes in HEK293 cells resulted in specific loss of cadherin and protocadherin O-Man glycosylation, while combined knockout of *TMTC1* and *TMTC3* resulted in selective loss of O-Man glycans on specific β-strands of EC domains, suggesting that each isoenzyme serves different functions. In addition, O-Man glycosylation of IPT/TIG domains of Plexins and hepatocyte growth factor receptor (HGFR) was not affected in TMTC KO cells, suggesting the existence of yet another O-Man glycosylation machinery. Our study demonstrates that regulation of O-mannosylation in higher eukaryotes is more complex than envisioned, and the discovery of the functions of TMTCs provide insight into Cobblestone lissencephaly caused by deficiency in TMTC3.

Significanse Statement

The large superfamily of cadherins serve essential roles in cell-cell interactions and guidance. The extracellular cadherin (EC) domains responsible for the biological functions are decorated with O-linked mannose glycans, but the functions of these O-glycans are poorly understood. We discovered a new O-mannosylation pathway orchestrated by four homologous TMTC1-4 genes that is dedicated selectively to the cadherin superfamily. Mutations in the TMTC3 gene cause Cobblestone Lissencephaly demonstrating the importance of this new type of O-mannosylation.

\Body

Introduction

Protein O-mannosylation on select Ser/Thr residues is found in yeast and metazoans, and the genetic and biosynthetic basis for this type of glycosylation has for long thought to be an evolutionarily conserved family of protein O-mannosyltransferases with seven members in yeast (Pmt1-7) and two in higher eukaryotes (POMT1/2) (1). O-Man glycosylation is the only type of O-glycosylation found in yeast, and O-Man glycans are widely found on proteins trafficking the secretory pathway much similar to O-GalNAc glycosylation in metazoans (2). In contrast, O-Man glycosylation in metazoans is found only on select proteins with β-

dystroglycan (α-DG) being the best-characterized (3–5), but with the advent of sensitive glycoproteomics strategies we recently found O-Man glycans on the large superfamily of cadherins (cdh) and protocadherins (pcdh) as well as a family of IPT/TIG domain carrying proteins including the hepatocyte growth factor receptor, HGFR (4). The O-Man glycans on cadherins are found in highly conserved positions in the β-strands of extracellular cadherin (EC) domains, and these glycosites appear highly conserved in evolution suggesting they serve important biological functions (4).

Deficiencies in enzymes catalyzing the structurally complex and diverse O-Man glycans on α-DG, including the two human POMT1 and POMT2 genes, underlie a subgroup of congenital muscular dystrophies (CMD) designated α-dystroglycanopathies, because deficient O-Man glycosylation of α-DG disrupts the interaction between the dystrophin glycoprotein complex and the extracellular matrix (5–7). Several studies have also implicated deficiency of POMT2 with E-cadherin dysfunction (8–10), although direct evidence for a role in glycosylation of cdhs and pcdhs is missing. In order to explore the functions of O-Man glycans on cdhs and pcdhs, we previously used a combinatorial gene targeting strategy in multiple cell lines and found that the two POMTs are essential for glycosylation of α-DG but not cdhs, pcdhs, and IPT/TIG domain containing proteins (11). In contrast to α-DG, we and others (4, 8, 10, 11) also found that O-Man glycans on the latter proteins were not elongated, suggesting that the biosynthesis of these distinct classes of proteins were different. We therefore predicted the existence of a new type of O-Man glycosylation machinery in higher eukaryotes (11).

Here, we report the discovery of a novel homologous family of putative O-Man glycosyltransferases encoded by the four *TMTC1-4* genes. The TMTC proteins were previously shown to be ER located and predicted to serve in Ca^{2+} regulation and protein folding (12, 13). Moreover, bi-allelic mutations in *TMTC3* were reported to cause Cobblestone lissencephaly (14), and family linkage and association analysis points to a role of TMTC2 in hearing loss (15). In murine models, *TMTC3* knockout (KO) results in early neonatal death, and *in vitro* studies have demonstrated abnormal TGF-β signaling in embryonic fibroblasts from these mice (16).

We used combinatorial gene knock out targeting in HEK293 cells combined with differential O-Man glycoproteomics for discovery of the roles of the *TMTC* genes in O-Man glycosylation of cdhs and pcdhs, and we validated this by analysis of recombinant expressed secreted cdh construct in *TMTC* mutant cell lines. We present strong evidence that the *TMTC* genes encode distinct O-Man glycosyltransferases that cooperatively glycosylate distinct

regions in the EC domains of cdhs and pcdhs. We also present evidence that the TMTCs do not glycosylate the IPT/TIG domain containing proteins, and therefore we predict the existence of yet another undiscovered O-Man glycosylation pathway dedicated to IPT/TIG domains. The severe phenotype associated with deficiency of TMTC3 (14) indicates that O-Man glycans on the cadherin superfamily serve important biological functions.

Results

Bioinformatic identification of the *TMTC1-4* genes. Our finding that the *POMT1/T2* genes were dispensible for O-mannosylation of many proteins (11) prompted a search for additional enzymes. We first tested the possibility that candidate enzyme(s) responsible for cdh/pcdh O-mannosylation were already classified as glycosyltransferases (GTs) in the CAZy database albeit without known function, and knock out of six such genes in combinations predicted to cover potential redundant isoenzymes in individual GT families did not affect O-mannosylation (GLT1D1/GTDC1, GLT8D1/GLT8D2, KDELC1/KDELC2) (not shown). We next hypothesized that the candidate enzyme(s) would resemble exisiting protein mannosyltransferases in CAZy GT39 and GT98 families with respect to overall multi-transmembrane domain structure but without significant sequence similarity. We performed a broad search considering criteria common for Pmts/POMTs, and the DPY19L genes responsible for C-mannosylation (PMID: 28202721), including ER localization, general topology and domain structure features. We identified the TMTC genes based on a conserved domain from yeast Pmts. Yeast Pmts match the PMT_2 superfamily domain hmm model (cl21590), and specifically the PMT_2 PFAM model (pfam13231), however, searching for these models directly in the human proteome only identified POMTs. Using the Conserved Domain Architecture Retrieval Tool (17) , the complete set of metazoan proteins in refseq that match PMT_2 were retrieved, and the *TMTC* genes with C-terminal TPR domain invariant presence of DUF1736 were identified. The TMTCs have been reported to be located in the ER (12), and we have identified both GalNAc-type and O-Man O-glycans in CHO cells that supports luminal orientation of the C-terminal TPR domain (11, 18) https://glycodomain.glycomics.ku.dk. The PMT_2 fold belongs to a larger family of folds whose structure is similar to the ArnT arabinosyltransferase identified in bacteria classified in CAZy GT83 (19, 20). Structural alignments of TMTCs and ArnT yielded good models and indicated that the catalytical residues are conserved in the first ER lumen loop. Using human TMTC3 protein sequence without the TPR repeats as source in the I-TASSER prediction tool (21), the best aligned structure with a TM–score (metric for measuring the similarity of two

proteins) of 0.782 were the ArnT glycosyltransferases from *Cupriavidus metallidurans,* whereas the second best TM-score of 0.675 were the oligosaccharyltransferase from the *Archaeoglubus fulgidus.* We therefore considered the TMTCs as good candidates and proceeded with a knockout strategy to test this hypothesis (11).

Knockout of TMTC1-4 in HEK293 selectively impairs O-Man glycosylation of cdhs/pcdhs. We previously generated HEK293 cells with KO of *COSMC* and *POMGNT1* (HEK293^{SC}), which results in a double "SimpleCell" line with truncated O-Man and O-GalNAc glycans suitable for Lectin Weak Affinity Chromatography (LWAC) enrichment of O-Man glycopeptides with the ConA lectin as well as O-GalNAc with the VVA lectin (Fig. 1A) (4, 11, 22). β-DG contains a central mucin-like domain with O-Man glycans in the N-terminal region and O-GalNAc glycans in the C-terminal region, and is therefore a useful probe for detection of substitution of O-Man with O-GalNAc glycans and exploration of common substrates (4, 11, 18), but also serving as a sensitive control for expression and Golgi processing of β-DG in the panel of isogenic KO cells.

In the HEK293^{SC} background we first simultaneously targeted *TMTC1* and *TMTC3* to obtain HEK293$^{SC/TMTC1,3}$ using the CRISPR/cas9 strategy and subsequently targeted *TMTC2* and *TMTC4* to generate HEK293$^{SC/TMTC1,2,3,4}$ (Fig. 1A). We used our previously established comparative quantitative O-glycoproteomics workflow based on differential labeling of tryptic digests from isogenic cells with stable dimethyl isotopes to probe global O-Man (23, 24). Total digests from isogenic HEK293^{SC} and HEK293$^{SC/TMTC1,2,3,4}$ cells were labeled with light and medium labels, mixed, and glycopeptides enriched by LWAC. Following LC-MS/MS, medium/light ratios (M/L) were calculated for individual glycopeptides to quantify the relative changes of O-Man glycosylations between HEK293^{SC} and HEK293$^{SC/TMTC1,2,3,4}$ cells.

We identified and quantified O-Man glycopeptides derived from β-DG, 27 members of the cdh and pcdh family, 5 proteins containing IPT domains (Plexins and HGFR), and 13 additional proteins (Table S1). The identified α-DG O-Man glycopeptides were equally present in HEK293^{SC} and HEK293$^{SC/TMTC1,2,3,4}$ cells (Fig. 1B, Dataset S1), which demonstrates that the POMT1/2 directed O-mannosylation is unaffected by KO of *TMTC1-4*. Unaffected O-Man glycosylation was also observed on glycopeptides derived from plexins, HGFR and additional proteins. In striking contrast, nearly all O-Man glycopeptides derived from members of the cdh/pcdh family were absent in HEK293$^{SC/TMTC1,2,3,4}$ cells. We quantified 79 unique O-Man glycopeptides from cdh/pcdhs of which 77 were >100 fold less abundant in the HEK293$^{SC/TMTC1,2,3,4}$ cell line compared to HEK293^{SC} (Fig. 1B, Dataset S1). Representative

members of classical/type I cadherins, desmosomal cadherins, clustered and non-clustered protocadherins as well as unconventional/ungrouped cadherins were all identified with O-Man glycopeptides that were >100 fold less abundant. Interestingly, two quantified O-Man glycopeptides from cdh11 and cdh13 were found equally abundant. Nevertheless, these results clearly indicate that the *TMTC1-4* genes play an important role for glycosylation of cdh and pcdhs.

We next performed the same comparative analysis with HEK293$^{SC/TMTC1,3}$ cells, and identified 34 O-Man glycoproteins of which 20 were members of the cdh/pcdh family (Fig. 1C, Dataset S2). We quantified 47 unique O-Man glycopeptides from cdh/pcdhs of which 15 were >100 fold less abundant in the HEK293$^{SC/TMTC1,3}$ cell line, and 32 peptides showed no substantial change compared to HEK293^{SC}. Interestingly, essentially all the identified O-Man glycosites that were affected by the *TMTC1,3* KO were located on the G-strands of EC domains, while the unaffected glycosites were localized to EC B-strands in EC domains of Desmocollin-2, E-, P-, N-, T-cadherin, and cadherin-11.

We identified two differentially regulated O-Man glycopeptides (four O-Man glycosites) >100 fold less abundant in HEK293$^{SC/TMTC1,3}$ cells for classical/type I cadherins, all of which were localized to the EC3 G-strand of N-cdh (Fig. 1C, Dataset S2). A single differentially regulated O-Man glycosite, localized in the loop between EC3 B and C-strands, was identified for Cadherin EGF LAG seven-pass G-type receptor 2. The same trend was observed for pcdhs; 8 glycopeptides were identified from pcdh10, 16, 17, α11, β2 and γB5, all of which had O-Man glycans on G-strands and were >100 fold less abundant in HEK293$^{SC/TMTC1,3}$ cells. In contrast, all O-Man glycosites (15 glycopeptides; pcdh7, 9, 17, α11, γA11, FAT1, 2 and 3) mapped to pcdh B-strands showed no change in relative abundance between in HEK293^{SC} and in HEK293$^{SC/TMTC1,3}$ cells. However, we note that four pcdh glycopeptides with overlapping O-Man glycosylations in B-strand/loop regions were identified as differentially regulated in HEK293$^{SC/TMTC1,3}$ cells, indicating that the addition of O-Man glycans on loops adjacent to B-strands is dependent on TMTC1 and/or TMTC3 isoenzymes. Taken together, these results indicate that individual *TMTC* genes have distinct roles for glycosylation of the two opposite regions of EC domains (B and G strands) on cdhs and pcdhs.

The quantitative O-Man glycoproteomic analysis revealed additional proteins targeted for O-mannosylation by the TMTC1-4 family. Protein disulfide-isomerase A3 (PDIA3) was previously also identified with several O-Man glycans (4, 11) and all quantified PDIA3 O-Man glycopeptides were >100 fold less abundunant in HEK293$^{SC/TMTC1,2,3,4}$ and HEK293$^{SC/TMTC1,3}$ cells compared to HEK293^{SC} (Fig. 1C, Dataset S1 and S2). In addition, one O-Man

glycopeptide from the extracellular region of TGF-β receptor type-1 was found to be differentially regulated in HEK293$^{SC/TMTC1,3}$ cells, and this observation is discussed further below.

We previously demonstrated that O-mannosylation of IPT/TIG domain containing proteins (Plexins and HGFR) were not dependent on POMT1/2 , and we predicted that these would then be controlled by the enzyme(s) dedicated to cdh/pcdh glycosylation. Surprisingly, the quantitative analysis of HEK293$^{SC/TMTC1,2,3,4}$ and HEK293$^{SC/TMTC1,3}$ cells undertaken here revealed unaffected O-Man glycosylation within the IPT/TIG domains of these proteins. HGFR was identified with five unique O-Man glycopeptides from IPT/TIG domains, all of which were equally abundant between HEK293^{SC} and HEK293$^{SC/TMTC1,2,3,4}$ cells (Fig. 1B, Dataset S1). We quantified 18 unique glycopeptides from plexin-A1, A2, A3 and B2, of which 15 were localized within IPT/TIG domains. All O-Man glycans mapped to plexin IPT/TIG domains were equally abundant (Fig. 1B, Dataset S1 and S2) in HEK293^{SC} and HEK293$^{SC/TMTC1,2,3,4}$ cells.

Interestingly, we also identified 3 unique O-Man glycopeptides from plexin-B2, covering two O-Man glycosites localized outside the IPT/TIG domains that were >100 fold less abundant in HEK293$^{SC/TMTC1,2,3,4}$ cells (Dataset S1). These O-Man glycosites in plexin-B2 were unaffected in the HEK293$^{SC/TMTC1,3}$ (Dataset S2), indicating that TMTC2 and/or TMTC4 play selective roles for glycosylation of other proteins than cdhs and pcdhs.

TMTCs selectively control O-mannosylation of recombinant expressed E-cadherin. We previously demonstrated efficient expression and O-mannosylation of a secreted E-cdh ectodomain construct in HEK293 wild-type and POMT1/T2 KO cells (11). Here, we expressed the same His-tagged E-cdh ectodomain fragment in HEK293^{SC}, HEK293$^{SC/TMTC1,3}$, and HEK293$^{SC/TMTC1,2,3,4}$ cells to evaluate the role of TMTCs on E-cdh production, secretion, and O-mannosylation. E-cdh was expressed and secreted equally in all three cell lines and purified (Fig. S1). Label free mass spectrometric analysis of tryptic digests from HEK293^{SC} revealed seven O-Man glycosites on EC2-EC4 with relatively high stoichiometry (30-70% site occupancy) (Fig. 2). In contrast, E-cdh expressed in HEK293$^{SC/TMTC1,2,3,4}$ produced essentially no detectable O-Man glycopeptides (with in the S/N level). However, E-cdh expressed in HEK293$^{SC/TMTC1,3}$ cells had normal O-Man glycans at two residues (Ser^{287} and Thr^{511}) localized to the B-strand of EC2 and EC4 (Fig. 2), while the O-Man glycopeptides derived from the G-strands of EC2-EC4 were absent. This correlates well with the global analysis of O-Man cdh/pcdh glycoproteins in HEK293$^{SC/TMTC1,3}$, and further confirms that the TMTC genes play

different roles in glycosylation of the cdh/pcdh family.

Discussion

Our study presents compelling evidence that the four *TMTC* genes encode protein O-mannosyltransferases dedicated primarily to the cadherin superfamily, and that the individual enzymes serve distinct roles in decorating the common EC domains with O-Man glycans at specific regions. O-Man glycans on these proteins were only recently discovered (4, 8), and while our knowledge of the functions of the conserved glycans in cdhs and pcdhs is still limited, the recent findings that bi-allelic mutations of *TMTC3* cause Cobblestone lissencephaly (14), clearly indicates that this novel type of O-mannosylation has critical functions. This is further supported by the finding that mice deficient in *tmtc3* suffer early postnatal death (16). TMTCs are predicted to serve distinct non-redundant functions, which is in excellent agreement with our finding that KO of *TMTC1* and *TMTC3* eliminated O-Man glycans found on G strands but not B strands of EC domains, while KO of all four was required to eliminate glycans completely. Clearly further work is needed to decipher the specific roles of the individual TMTC enzymes, but the existence of a distinct gene family dedicated to O-mannosylation of cdhs and pcdhs was a surprising finding obtained through genetic dissection that now opens for wider studies of the mechanism and biological functions.

The presented results demonstrate that the TMTCs selectively control O-mannosylation of cdhs and pcdhs, which strongly indicates that they serve as O-mannosyltransferases predicted to use Dol-P-Man as donor substrate similar to the POMTs and DPY19s isoenzymes (25–27). Current assays demonstrating enzyme activities of the human POMTs and yeast Pmts have been limited to total cell lysates and overexpression in cell systems using short acceptor peptide substrates (28, 29). Presenting direct evidence for activities of the TMTC enzymes, which are multi-transmembrane proteins predicted to use relatively complex acceptor substrates with defined structural folds like the EC domain, therefore represents a considerable challenge, one which has to be addressed in future studies. Several findings, however, provide supporting evidence to our genetic deconstruction and quantitative glycoproteomics approach which strongly suggests that TMTCs are enzymes. The TMTCs share predicted overall structure with the POMT, ArnT, and DPY19 proteins, and the first luminal loop contains a conserved acidic DD/DE motif essential for POMT/Pmt activity (Fig. 3) (30). Moreover, a homozygous mutation of His67Asp close to the DD/DE motif in this loop of TMTC3 cause Cobblestone lissencephaly (Fig. 3) (14).

TMTC3 has previously been reported to be involved in ER stress response and interact with

PDIA3 and pcdhs (12, 14), and here we demonstate that these are O-mannosylated by the TMTC dependent pathway although the function is still unclear (Fig. 1C, Dataset S2). In addition, two O-Man glycans previously localized on a specific β-strand of the TGF-β receptor type-1 (4), were also found to be differentially regulated in HEK293$^{SC/TMTC1,3}$ cells. Abnormal TGF-β signaling has been linked to *tmtc3* in a murine KO model (16), and it is possible that loss of O-mannosylation in mice deficient in *tmtc3* may underlie the phenotype. Finally, TMTC1 and TMTC2 have been proposed to be ER resident proteins involved in calcium homeostasis (12), and this may not be inconsistent with a role in glycosylation of cdhs and pcdhs as these are major Ca^{2+} dependent adhesion molecules.

We did find two O-Man glycopeptides, one derived from cdh11 (EC2 G strand) and the other from cdh13 (EC2 B strand) that were not affected by KO of TMTCs (Fig. 1B). It is currently unclear why these appear to be O-mannosylated independently of the TMTCs in contrast to all other cdh and pcdh glycosites, however, we note that both cdhs are less related in sequence to classical cdhs, with cdh11 classified as a type II cadherin and cdh13 being truncated and a glycosylphosphatidylinositol (GPI) anchored protein. Moreover, the larger group of Plexins, HGFR and other proteins that we previously found were O-mannosylated independently of POMT1/2 (11), were also found here to be O-mannosylated independently of the TMTC pathway. We were unable to identify specific characteristics for the proteins that appear to be O-mannosylated independently of both the POMTs and the TMTCs. It is formally possible that the POMTs and TMTCs share some overlap in substrate specificities that will require knock out of both groups of genes to elcuidate, but we predict the existence of yet a novel O-mannosylation capacity selectively serving IPT/TIG domains and HGFR, and hence a total of three distinct pathways for O-mannsoylatin in higher eukaryotes (Fig. 4).

It is important to stress that several reports have implicated the POMTs in E-cdh glycosylation and function (8, 9). While we do not have explanations for these seemingly contradictory results, we note that most of the data inferring this relationship relies on antibodies generated to O-Man glycopeptides. In particular, the key evidence that loss of POMT1/T2 function results in loss of O-mannosylation of E-cdh is drawn primarily from antibody binding data, but the antibodies seem to have highly restricted reactivity with O-glycopeptides (10), and it is unclear if they react with O-Man glycans on cdhs and pcdhs. Further work is needed to confirm this, but our previous studies clearly demonstrate that POMTs are not required for O-Man glycosylation in HEK293 cells, and a truncated E-cdh construct can be expressed and secreted normally with O-Man glycans in HEK293 without POMT1 and/or POMT2 (11). The present results show that the *TMTC* genes are required for

O-mannosylation of cdhs and pcdhs, and interestingly KO of *TMTC*s in HEK293 cells did not appear to affect expression, secretion and immediate stability of recombinant expressed E-cdh (Fig. S1). We did not explore the N-glycans on these constructs, but a study has suggested that POMT directed O-mannosylation interacted with N-glycosylation (9), and further studies may be needed to clarify this.

POMTs and Pmts are part of the family of GT-C fold enzymes classified in CAZy GT39, while DPY19s are classified in GT98. We propose that the TMTCs be classified as a novel CAZy GTXXX (Number upon publication) family. The C-terminal TPR domains of TMTCs have only been found on one other glycosyltransferase, the nucleocytoplasmic O-GlcNAc-transferase, OGT (31), where it serves in protein-protein interactions and is proposed to function in guiding selection of substrates and the kinetic efficiency (31–33). TPR domains present protein-binding surfaces that can bind chemically distinct peptides and promote formation of multiprotein complexes (32), and it is tempting to speculate that the TPRs of TMTCs drive the distinct substrate specificity for the EC folds of cdh and pcdh. In this respect truncating mutations in the TPR region of TMTC3 were also found to result in Cobblestone lissencephaly (14), but future studies need to address the role of the TPRs in detail. Cobblestone lissencephaly is a severe brain malformation with cortical dysplasia, irregular borders between white and grey matter, brainstem hypoplasia, and cystic cerebellar dysplasia (34). It is found associated with mutations in a large number of genes of which most are involved in the elaborate POMT1/T2 O-mannosylation pathway of β-DG (35), including *POMT1, POMT2, POMGNT1, POMGNT2, B4GAT1, B3GALNT2, ISPD, TMEM5, LARGE, FKRP and FKTN* or associated with the basememt membrane including *LAMB1*, *LAMC3* and *COL4A1* (14). The reported cases of Cobblestone lissencephaly caused by deficiency in TMTC3 appear to differ from deficiencies in the genes involved in the classical β-DG O-mannosylation pathway by not being characterized by eye and muscle phenotypes. In contrast, mutations in LAMB1 has also been reported to cause Cobblestone lissencephaly without eye and muscle involvement (36).

In conclusion, we represent strong evidence for a novel O-mannosylation pathway dedicated to cdhs and pcdhs orchestrated by the four *TMTC1-4* genes. Preliminary evidence moreover suggests that the individual TMTCs serve distinct functions in glycosylation of different sites in cdh and pcdh EC domains, indicating that glycosylation of these important cell adhesion proteins may be differentially regulated by expression of different TMTC repertoires. The

TMTC directed O-mannosylation is clearly biologically important and our findings now open for more detailed studies.

Experimental procedures

Precise gene targeting of glycogenes in HEK293 cells. HEK293^{SC} with double KO of *COSMC* and *POMGNT1* (9) were maintained in Dulbecco's Modified Eagle Medium supplemented with 10 % FBS, 2 mM L-glutamine. Gene targeting was performed as previously described (37), briefly, 20-nucleotide guide sequences targeting human *TMTC1-4* were designed using a Desktop Genetics-Horizon Discovery algorithm. The single-guide RNAs were co-transfected with the GFP-tagged Cas9 plasmid using Lipofectamine 3000 according to the manufacturers instructions (Thermo Fisher Scientific). 24 h after transfection GFP positive cells were enriched by FACS, and following 1 week of culture cells were single-sorted in 96-well plates. KO clones with frameshift mutations were identified by Indel-Detection-by-Amplicon-Analysis (IDAA) (39) with gene specific primers (Table S2). Clones were selected with frameshift mutations that result in premature stop codons. All genes were targeted in the center of an exon, and clones were selected with small introduced indels limited to the particular exon. All selected clones were confirmed by Sanger sequencing of 200-300 bp of the target regions.

Sample preparation for differential glycoproteomics. Packed cell pellets (0.5 mL) were lysed and digested with either trypsin or chymotrypsin, and subsequently labeled with dimethyl stable isotopes as described (11, 23). Briefly, proteolytic digests were labeled by $NaBH_3CN$ and formaldehyde (COH_2) (light label) or $NaBH_3CN$ and deuterated formaldehyde (COD_2) (medium label) and finally mixed in 1:1 ratio. Following N-glycan removal by PNGase F (Roche), the differentially labeled peptides were resuspended in ConA loading buffer (20 mM Tris-HCl, pH 7.4, 300 mM NaCl, 1 mM $CaCl_2$/$MgCl_2$/$MnCl_2$/$ZnCl_2$). O-Man glycopeptides were separated from non-glycosylated peptides by Lectin Weak Affinity Chromatography (LWAC) using a 2.8 m column packed in-house with ConA-conjugated agarose beads as previously described (11). The glycopeptide-containing fractions were purified by in-house packed Stage tips (Empore disk-C18, 3M) prior to LC-MS/MS analysis.

Recombinant expression. A secreted His-tagged E-cdh construct was transiently expressed using PEI and conditioned media collected 3 days post-transfection as previously described (11). Secreted E-cdh was purified from 20 mL on 200uL of Ni-NTA Sepharose beads as described (11), analysed by NUPAGE, and subjected to trypsin digestions (10 µg E-cdh) and LC-MS/MS.

nLC-MS/MS analyses. Samples were analyzed essentially as previously described (11). Briefly, a set up composed of an EASY-nLC 1000 (Thermo Fisher Scientific, Bremen, Germany) was interfaced via a nanoSpray Flex ion source to an LTQ-Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific). The EASY-nLC 1000 was operated using a single analytical column setup packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch, Ammersbach, Germany). The LC gradient duration was 120 min at 200 nL/min. The mobile phases were composed of solvent A (H_2O) and solvent B (acetonitrile); both solvents containing 0.1% formic acid (v/v); the LC gradient was 2–20% B for 95 min followed by 20–80% B for 10 min and finally 80% B for 15 min. Tryptic peptides from purified E-cad were separated using a 60 min LC gradient method; the LC gradient was 2-25% B for 35 min followed by 25-80% B for 10 min and finally 80% B for 15 min.
Precursor MS1 scan (m/z 355–1700) was acquired in the Orbitrap at a resolution setting of 120,000 (AGC: 4.0e5; Injection time: 100 ms) followed by Orbitrap HCD-MS/MS (Resolution: 50,000; AGC: 5.0e4; Injection time: 75 ms) and ETD-MS/MS (Resolution: 60,000; AGC: 1.0e5; Injection time: 150 ms) of multiply charged precursors (z = 2-7) in the MS1 spectrum; a minimum MS1 signal threshold of 50,000 ions was used for triggering data-dependent fragmentation events. To improve fragmentation, ETD supplemental activation (ETcid = 25%) was used in all analyses described above.

Data analyses. Data processing used Proteome Discoverer 1.4 software (Thermo Fisher Scientific) with minor modifications as outlined below (11). Raw data files (.raw) were processed using either the SequestHT or the MS Amanda (40) node and searched against the the human proteome (January, 2013) downloaded from the UniProtKB database (http://www.uniprot.org/). In all cases, the precursor mass tolerance was set to 5 ppm and fragment ion mass tolerance to 0.02 Da. Carbamidomethylation on Cys was used as a fixed modification, oxidation of Met and hexose modification of Ser and Thr residues were used as variable modifications. Enzyme specificity (full or semi) was trypsin or chymotrypsin; a maximum of two missed cleavage sites were tolerated. Final results were filtered for high-confidence ($p < 0.01$) identifications only. For dimethyl labeled samples, glycopeptide medium/light ratios were determined using the Event Detector Node and the Precursor Ion Node of the Proteome Discoverer workflow as previously described (24). Relative quantification on the total proteolytic digests from $HEK293^{SC}$/$HEK293^{SC/TMTC1,2,3,4}$ and $HEK293^{SC}$/$HEK293^{SC/TMTC1,3}$ was performed to determine the technical/biological variability of the assay. A ±10-fold change was estimated for the assay since >97% of the quantified

peptides showed <10-fold relative change in abundance (Fig. S2). We therefore chose to use >100-fold change as a stringent cut-off value when categorizing differentially regulated glycopeptides in this assay.

Figure legends

Fig. 1. Genetic dissection and differential quantitative glycoproteome analysis identify TMTC1-4 as directing O-Man glycosylation of cdhs and pcdhs. (*A*) Graphic depiction of gene editing in HEK293 double SimpleCells (SC) with truncated O-Man and O-GalNAc glycans targeting POMTs or TMTCs. For differential O-Man glycoproteome analyses paired dimethyl labeled digests from HEK293^{sc} and either HEK293$^{SC/POMT1,T2}$, HEK293$^{SC/TMTC1,2,3,4}$, or HEK293$^{SC/TMTC1,3}$ were performed. (*B*) Scatter plot of dimethyl labeled O-Man glycopeptide ratios (HEK293^{SC} / HEK293$^{SC/TMTC1,2,3,4}$) expressed on a $\log_{10}$ scale showing loss of O-Man glycopeptides (>100 fold change) derived from cdhs/pcdhs and other proteins, but not glycopeptides derived from IPT/TIG domains or known POMT1/2 substrates (β-DG, KIAA1549, SUCO) (<10 fold change) (11) in HEK293 cells with KO of TMTC1-4. Median ratios are plotted for unique O-Man glycopeptides. Box represents the inter quartile range (IQR) and vertical line represents the median value within each protein group. (*C*) Scatter plot of dimethyl labeled O-Man glycopeptide ratios (HEK293^{SC} / HEK293$^{SC/TMTC1,3}$) showing selective loss of O-Man glycopeptides derived from EC G-strands (red dots; >100 fold change) of cdhs and pcdhs, but not from B-strands (green dots; <10 fold change) in HEK293 cells with KO of TMTC1/3. White dots represent overlapping O-Man glycosylations on B-strand/loop regions. Additional proteins with TMTC1/3 dependent O-Man glycopeptide identifications are indicated (bottom right); TGF-beta receptor type-1 (TGFR1), Latrophilin-3 (LPHN3), Melanoma inhibitory activity protein 3 (MIA3) and Protein disulfide-isomerase A3 (PDIA3).

Fig. 2. Summary of O-Man glycosites identified on recombinant E-cdh expressed in HEK293 mutant cell lines. Top panel illustrates schematic representation of E-cadherin domain organization (top) with known O-Man glycosylation sites (green circles). Selected β-strands are indicated by black arrows for each EC domain and all O-Man and O-GalNAc glycosites identified previously are shown with positions and numbering relative to EC β-strands and amino acids of E-cdh (UniProtKB: P12830). The lower three panels represent O-Man glycosites identified on the recombinantly expressed E-cdh in HEK293^{SC}, HEK293$^{SC/TMTC1,3}$ and HEK293$^{SC/TMTC1,2,3,4}$, respectively. The stoichiometry (site occupancy) of O-Man glycans was calculated from the liquid chromatography (LC) peak areas for each peptide (with and

without O-Man glycans) and is indicated above each glycosylation site (%). White circles represent loss of O-Man glycosylation sites (not detectable). The canonical EC β-strand arrangement is shown (bottom left) with B- and G-strands indicated in green.

Fig. 3. Graphic depiction of the predicted membrane organization of TMTCs, POMTs, Arn-T and DPY19s. (*A*) Sequence alignment of a segment from the first luminal loop of human TMTCs and POMTs. A conserved acidic amino acid motif DD/DE in the TMTCs, POMTs, and Arn-T (30) is indicated by a blue circle. Single amino acid mutations in TMTC3 causing Cobblestone lissencephaly are indicated by a red circle. (*B*) The multi-transmembrane organization of TMTCs is based on the predicted structure of TMTC3 (14), (*C*) POMT1 and POMT2 is adapted from (41), (*D*) the Arn-T structure is based on the crystal structure from *Cupriavidus metallidurans* (20), and (*E*) the structure of DPY19Ls is adapted from the *C.elegans* DPY19 (25)

Fig. 4. A proposed model for differential genetic regulation of O-mannosylationof different classes of proteins. O-Mannosylation in higher eukaryotes is predicted to be controlled by at least three distinct enzyme families. The classical POMT1/T2 control β-DG and KIAA1549, while the TMTCs control the cadherin superfamily. O-Man glycans on β-DG and potentially KIAA1549 are elongated by different core structures, while those on other protein do not appear to be elongated (11). A novel O-Man glycosylation capacity dedicated to Plexins and IPT/TIG domains is predicted to exist.

References

1. Lommel M, Strahl S (2009) Protein O-mannosylation: Conserved from bacteria to humans. *Glycobiology* 19(8):816–828.

2. Neubert P, et al. (2016) Mapping the O-mannose glycoproteome in Saccharomyces cerevisiae. *Mol Cell Proteomics* 15(4):1323–1337.

3. Barresi R, Campbell KP (2006) Dystroglycan: from biosynthesis to pathogenesis of human disease. *J Cell Sci* 119(2):199–207.

4. Vester-Christensen MB, et al. (2013) Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proc Natl Acad Sci USA* 110:21018–21023.

5. Praissman JL, et al. (2016) The Functional O-Mannose Glycan on α-Dystroglycan Contains a Phospho-ribitol Primed for Matriglycan Addition. *Elife* 5:e14473

6. Live D, Wells L, Boons GJ (2013) Dissecting the molecular basis of the role of the O-mannosylation pathway in disease: α-dystroglycan and forms of muscular dystrophy. *Chem BioChem* 14(18):2392–2402.

7. Wells L (2013) The o-mannosylation pathway: Glycosyltransferases and proteins implicated in congenital muscular dystrophy. *J Biol Chem* 288(10):6930–6935.

8. Lommel M, et al. (2013) Protein O-mannosylation is crucial for E-cadherin-mediated cell adhesion. *Proc Natl Acad Sci USA* 110(52):21024–9.

9. Carvalho S, et al. (2016) O–mannosylation and N-glycosylation: two coordinated mechanisms regulating the tumour suppressor functions of E-cadherin in cancer. *Oncotarget* 5:1–16.

10. Bartels MF, et al. (2016) Protein O-Mannosylation in the Murine Brain: Occurrence of Mono-O-Mannosyl Glycans and Identification of New Substrates. *PLoS One* 11(11):e0166119.

11. Larsen ISB, et al. (2017) Mammalian O-mannosylation of Cadherins and Plexins is Independent of Protein O-mannosyltransferase 1 and 2. *J Biol Chem*:jbc.M117.794487.

12. Sunryd JC, et al. (2014) TMTC1 and TMTC2 are novel endoplasmic reticulum Tetratricopeptide repeat-containing adapter proteins involved in calcium homeostasis. *J Biol Chem* 289(23):16085–16099.

13. Racapé M, et al. (2011) The involvement of SMILE/TMTC3 in endoplasmic reticulum stress response. *PLoS One* 6(5):e19321.

14. Jerber J, et al. (2016) Biallelic Mutations in TMTC3, Encoding a Transmembrane and TPR-Containing Protein, Lead to Cobblestone Lissencephaly. *Am J Hum Genet* 99:1–9.

15. Runge CL, et al. (2016) Association of *TMTC2* With Human Nonsyndromic Sensorineural Hearing Loss. *JAMA Otolaryngol Neck Surg* 142(9):866.

16. Yun EJ, Vu TH (2012) mSmile is necessary for bronchial smooth muscle and alveolar myofibroblast development. *Anat Rec* 295(1):167–176.

17. Geer LY, Domrachev M, Lipman DJ, Bryant SH (2002) CDART : Protein Homology by Domain Architecture CDART: Protein Homology by Domain Architecture. *Genome Res* 12:1619–1623.

18. Steentoft C, et al. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* 32(10):1478–1488.

19. Tavares-Carreón F, Fathy Mohamed Y, Andrade A, Valvano MA (2016) ArnT proteins that

catalyze the glycosylation of lipopolysaccharide share common features with bacterial N-oligosaccharyltransferases. *Glycobiology* 26(3):286–300.

20. Petrou VI, et al. (2016) Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* 351(6273).

21. Yang J, et al. (2014) The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12(1):7–8.

22. Steentoft C, et al. (2011) Mining the O-glycoproteome using zinc-finger nuclease–glycoengineered SimpleCell lines. *Nat Methods* 8(11):977–982.

23. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJR (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* 4(4):484–494.

24. Schjoldager KT, et al. (2015) Deconstruction of O-glycosylation-GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep* 16(12):1713–1722.

25. Buettner FFR, Ashikov A, Tiemann B, Lehle L, Bakker H (2013) C. elegans DPY-19 Is a C-Mannosyltransferase Glycosylating Thrombospondin Repeats. *Mol Cell* 50(2):295–302.

26. Willer T, Amselgruber W, Deutzmann R, Strahl S (2002) Characterization of POMT2, a novel member of the PMT protein O-mannosyltransferase family specifically localized to the acrosome of mammalian spermatids. *Glycobiology* 12(11):771–783.

27. Jurado LA, Coloma A, Cruces J (1999) Identification of a human homolog of the Drosophila rotated abdomen gene (POMT1) encoding a putative protein O-mannosyl-transferase, and assignment to human chromosome 9q34.1. *Genomics* 58(2):171–80.

28. Bausewein D, Engel J, Jank T, Schoedl M, Strahl S (2016) Functional similarities between the protein O-mannosyltransferases pmt4 from bakers' yeast and human POMT1. *J Biol Chem* 291(34):18006–18015.

29. Manya H, et al. (2004) Demonstration of mammalian protein O-mannosyltransferase activity: coexpression of POMT1 and POMT2 required for enzymatic activity. *Proc Natl Acad Sci USA* 101(2):500–505.

30. Lommel M, Schott A, Jank T, Hofmann V, Strahl S (2011) A conserved acidic motif is crucial for enzymatic activity of protein O-mannosyltransferases. *J Biol Chem* 286(46):39768–39775.

31. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O (2011) Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu*

Rev Biochem 80:825–858.

32. Zeytuni N, Zarivach R (2012) Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. *Structure* 20(3):397–405.

33. Lubas WA, Hanover JA (2000) Functional Expression of O-linked GlcNAc Transferase. *J Biol Chem* 275(15):10983–10988.

34. Van Der Knaap MS, et al. (1997) Magnetic resonance imaging in classification of congenital muscular dystrophies with brain abnormalities. *Ann Neurol* 42(1):50–59.

35. Vuillaumier-Barrot S, et al. (2012) Identification of Mutations in TMEM5 and ISPD as a Cause of Severe Cobblestone Lissencephaly. *Am J Hum Genet* 91(6):1135–1143.

36. Radmanesh F, et al. (2013) Mutations in LAMB1 cause cobblestone brain malformation without muscular or ocular abnormalities. *Am J Hum Genet* 92(3):468–474.

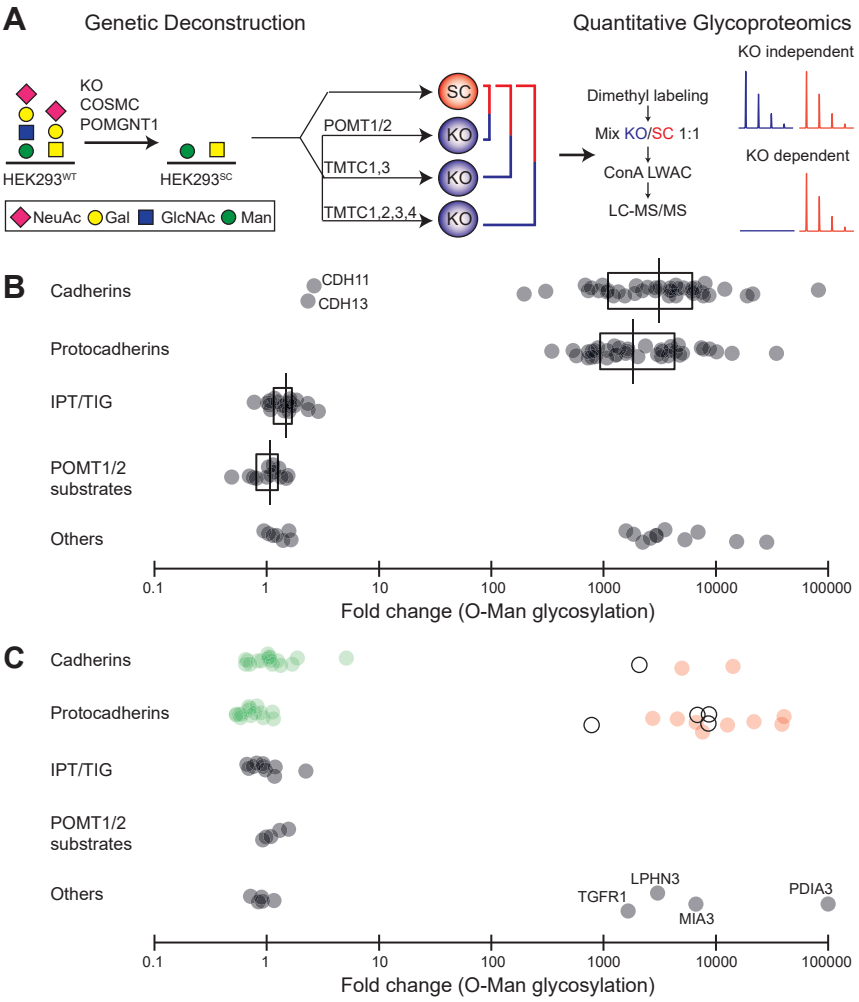
37. Lonowski LA, et al. (2017) Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat Protoc* 12(3):581–603.

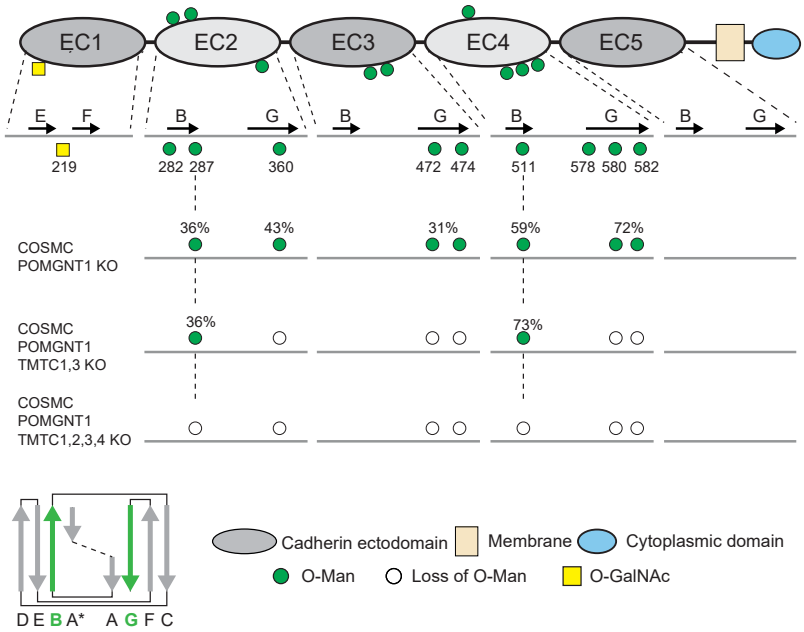
38. Baldi L, Hacker DL, Meerschman C, Wurm FM (2012) Protein Expression in Mammalian Cells. 801(4):13–26.

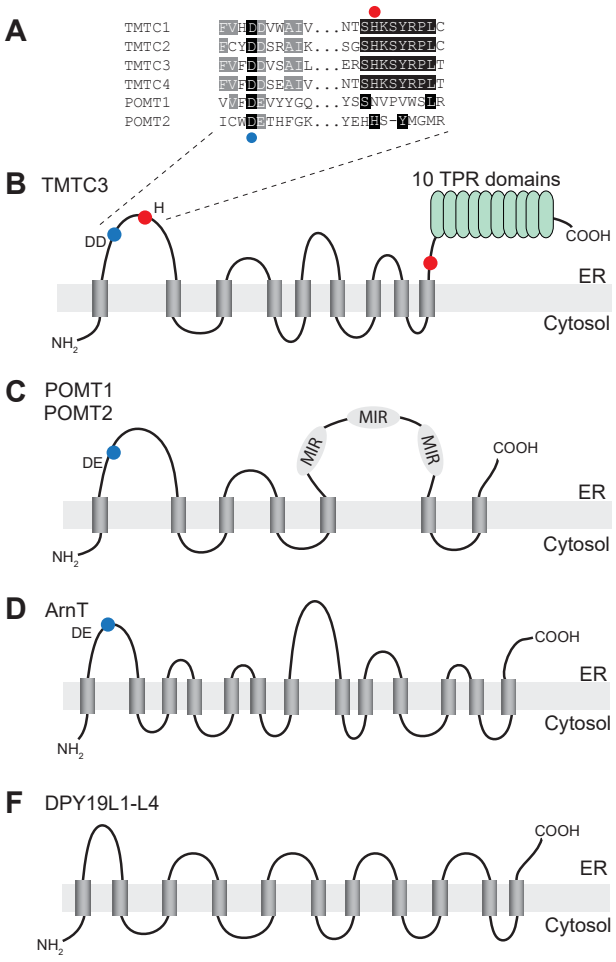
39. Yang Z, et al. (2015) Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res* 43(9):e59.

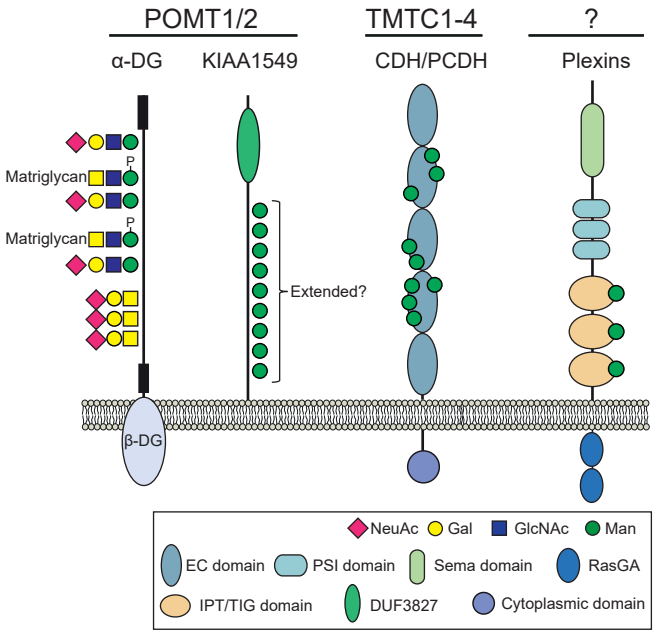
40. Dorfer V, et al. (2014) MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J Proteome Res* 13(8):3679–3684.

41. Akasaka-Manya K, Manya H, Hayashi M, Endo T (2011) Different roles of the two components of human protein O-mannosyltransferase, POMT1 and POMT2. *Biochem Biophys Res Commun* 411(4):721–725.


A
Genetic Deconstruction
Quantitative Glycoproteomics
KO
COSMC
POMGNT1
HEK293WT
HEK293SC
NeuAc Gal GlcNAc Man
POMT1/2
TMTC1,3
TMTC1,2,3,4
SC
KO
KO
KO
Dimethyl labeling
Mix KO/SC 1:1
ConA LWAC
LC-MS/MS
KO independent
KO dependent
B
Cadherins
Protocadherins
IPT/TIG
POMT1/2 substrates
Others
CDH11
CDH13
0.1
1
10
100
1000
10000
100000
Fold change (O-Man glycosylation)
C
Cadherins
Protocadherins
IPT/TIG
POMT1/2 substrates
Others
LPHN3
TGFR1
MIA3
PDIA3
0.1
1
10
100
1000
10000
100000
Fold change (O-Man glycosylation)


EC1
EC2
EC3
EC4
EC5
E
F
B
G
B
G
B
G
B
G
219
282 287
360
472 474
511
578 580 582
36%
43%
31%
59%
72%
COSMC
POMGNT1 KO
36%
73%
COSMC
POMGNT1
TMTC1,3 KO
COSMC
POMGNT1
TMTC1,2,3,4 KO
D E B A* A G F C
Cadherin ectodomain
Membrane
Cytoplasmic domain
O-Man
Loss of O-Man
O-GalNAc


A
TMTC1 FVHDDVWAIV...NTSHKSYRPLC
TMTC2 FCYDDSRAIK...SGSHKSYRPLC
TMTC3 FVFDDVSAIL...ERSHKSYRPLT
TMTC4 FVFDDSEAIV...NTSHKSYRPLT
POMT1 VMFDDVYYGQ...YSSNVPVWSLR
POMT2 ICWDDTHFGK...YEHHS-YWGMR
B TMTC3
10 TPR domains
H
DD
COOH
ER
Cytosol
NH2
C POMT1
POMT2
MIR
MIR
MIR
DE
COOH
ER
Cytosol
NH2
D ArnT
DE
COOH
ER
Cytosol
NH2
F DPY19L1-L4
COOH
ER
Cytosol
NH2


POMT1/2
TMTC1-4
?
α-DG
KIAA1549
CDH/PCDH
Plexins
Matriglycan
P
Matriglycan
P
Extended?
β-DG
NeuAc
Gal
GlcNAc
Man
EC domain
PSI domain
Sema domain
RasGA
IPT/TIG domain
DUF3827
Cytoplasmic domain