EXPANDING THE *O*-FUCOSYLATION REPERTOIRE: DISCOVERY OF NOVEL PROTEIN *O*-FUCOSYLTRANSFERASES AND PROFILING THE *O*-FUCOSE PROTEOME OF POFUT1 AND POFUT2

by

HUILIN HAO

(Under the Direction of Robert S. Haltiwanger)

ABSTRACT

Domain-specific *O*-fucosylation is an unusual type of glycosylation, where fucose is directly attached to the serine or threonine residues of proteins via an *O*-linkage. *O*-fucosylated proteins play critical roles in a wide variety of biological events, with the most prominent being the Notch receptors. To date, only two types of domain-specific *O*-fucosylation have been identified in mammals: Protein *O*-fucosyltransferase (POFUT) 1-mediated *O*-fucosylation of Epidermal Growth Factor-like (EGF) repeats and POFUT2-mediated *O*-fucosylation of Thrombospondin Type 1 Repeats (TSRs). Both enzymes recognize a folded domain structure with specific *O*-fucose consensus sequences. Recently, a novel domain-specific *O*-fucosylation was identified on the Elastin Microfibril Interface (EMI) domain of Multimerin-1 (MMRN1) with an undefined POFUT. In Chapter 2, we present evidence that FUT10 and FUT11 are indeed POFUTs responsible for modifying EMI domains; thus, we renamed them POFUT3 and POFUT4, respectively. Similar to POFUT1 and POFUT2, POFUT3 and POFUT4 require folded EMI for efficient modification, function in the ER, and participate in a non-canonical ER quality control pathway.

Glycoproteomic analysis of intact *O*-fucosylated peptides remains a key challenge due to several inherent obstacles, including low abundancy and lack of oxonium ions for confident assigning. In Chapter 3, we labeled POFUT1/2 target proteins with an *O*-fucose-specific tag, Fuc6nyl*, which greatly improved the mass spectrometric analysis of glycopeptides by generating high-intensity, *O*-fucose specific diagnostic ions and increasing charge density of precursor ions (Fuc6nyl* does not label POFUT3/4 target proteins). Leveraging Fuc6nyl*, we developed a glycoproteomic workflow for profiling the *O*-fucose proteome of POFUT1/2, which efficiently enriched and identified fucosylated peptides, covering both predicted *O*-fucose sites and unexpected *O*-fucose sites. We envision this workflow will expand the *O*-fucose proteome of POFUT1/2 and broaden our target pool for research.

Together these findings advance our knowledge of *O*-fucosylation by unveiling two novel POFUTs and providing tools for identifying previously unknown targets for POFUT1 and 2. Our understanding of these newly identified POFUTs and protein targets is still in its infancy; the work presented here provides an entry point to this emerging field of *O*-fucosylation.

INDEX WORDS: *O*-fucosylation, POFUT1, POFUT2, POFUT3, POFUT4, FUT10, FUT11, glycoproteomic, 6-Alkynyl Fucose

EXPANDING THE *O*-FUCOSYLATION REPERTOIRE: DISCOVERY OF NOVEL PROTEIN *O*-FUCOSYLTRANSFERASES AND PROFILING THE *O*-FUCOSE PROTEOME OF POFUT1 AND POFUT2

by

HUILIN HAO

B.S., Shandong University, China, 2013

M.S., Shandong University, China, 2016

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

EXPANDING THE *O*-FUCOSYLATION REPERTOIRE: DISCOVERY OF NOVEL PROTEIN *O*-FUCOSYLTRANSFERASES AND PROFILING THE *O*-FUCOSE PROTEOME OF POFUT1 AND POFUT2

by

HUILIN HAO

Major Professor:	Robert S. Haltiwanger
Committee:	Hang Yin
	Lance Wells
	Michael Tiemeyer

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2023

DEDICATION

To my parents for their love and support. To my husband, Shen Chen, and our ferret, Lingling Hao. Especially to my guinea pig, Qiqi Hao, who passed away this summer.

ACKNOWLEDGEMENTS

Without hesitation, I first thank my supervisor Dr. Robert Haltiwanger for his mentor. I am grateful for the freedom he gave us for trying the things we want, even if it's difficult and take a lot of time. Without him, none of this would be possible. I'd like to thank my current committee members Dr. Hang Yin, Dr. Lance Wells, Dr. Michael Tiemeyer, and my previous committee members Dr. Nancy Manley and Dr. Michael Pierce for their kind mentor and helpful suggestions on my projects. I thank all the current and past lab members for the great time we had together, especially thank Hideyuki Takeuchi, who trained me during my rotation and guided me into the glycobiology world.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Fucosylated glycan structures are broadly observed in a wide variety of organisms and play important roles in numerous physiological and pathological processes, including ABO blood group histocompatibility, immune modulation, and cellular signal transduction. Two unique features distinguish fucose from other six-carbon sugars: an L-configuration and the absence of a hydroxyl group on the C-6 carbon. Fucose is typically found as terminal or core structures on glycans added by Golgi-localized fucosyltransferases (FUTs), but it can also be directly linked to proteins through an *O*-linkage to the hydroxyl groups of serines or threonines in proteins with specific domains. This domain-specific *O*-fucosylation is mediated by ER-localized protein O-fucosyltransferases (POFUTs). A growing body of research has demonstrated the biological importance of *O*-fucosylation. In this review, I will present an overview of our current knowledge of *O*-fucosylation, integrating the latest findings, and with a particular focus on its biological functions and molecular mechanisms.

1.2 GDP-fucose synthesis and transport

All FUTs utilize guanosine diphosphate (GDP)-fucose as a donor to transfer fucose to acceptor molecules. GDP-fucose is synthesized in the cytosol and then transported into the Golgi or endoplasmic reticulum (ER), where it is used by Golgi-localized FUTs or POFUTs, respectively.

In mammals, GDP-fucose is synthesized via two pathways: the de novo pathway and the salvage pathway (Figure 1.1). The de novo pathway produces GDP-fucose from mannose or glucose, while the salvage pathway generates GDP-fucose from exogenous fucose or fucose salvaged from glycoconjugate degradation.

1.2.1 De novo pathway

In the de novo pathway, GDP-mannose that is derived from either mannose or glucose, is converted into GDP-fucose. This transformation involves three reactions catalyzed by two enzymes and includes a keto-containing intermediate. GDP-mannose 4,6-dehydratase (GMD) catalyzes the initial reaction, converting GDP-mannose into GDP-4-keto-6-deoxymannose (1). This intermediate is epimerized to GDP-4-keto-6-deoxygalactose and then reduced to yield GDP-fucose, with both reactions catalyzed by GDP-L-fucose synthase (FX), a dual-functional enzyme acting as both an epimerase and a reductase (2).

1.2.2 Salvage pathway

In the salvage pathway, the source of fucose is provided by exogenous fucose from the diet or fucose released from the lysosomal degradation of glycoproteins or glycolipids (3, 4). Free fucose in the cytosol is firstly phosphorylated by L-fucose kinase (Fucokinase), forming fucose-1-phosphate, and is then converted to GDP-fucose by Fucose-1-phosphate guanylytransferase (5, 6). Exogenous fucose must be transported across cell membranes to present in the cytosol, the mechanism behind this transportation remained largely unknown for decades (7). Recently, Ng et al., identified GLUT1 (encoded by *SLC2A1*) as a cell membrane fucose transporter using an

siRNA screen of ~140 annotated transporter genes in HCT116 cells (8). They also demonstrated that macropinocytosis plays a role in cellular fucose uptake.

1.2.3 GDP-fucose transport and cellular GDP-fucose pool

GDP-fucose synthesized in the cytosol is then transported to the Golgi and ER. The major Golgi GDP-fucose transporter in mammals is SLC35C1 (9, 10). Lu et al. recently showed that *SLC35C1* knockout in HEK293T cells led to a significant decrease, without complete loss, of *O*-fucosylation of Epidermal Growth Factor-like (EGF) repeats and thrombospondin type 1 repeats (TSRs) in the ER, implicating crosstalk between the Golgi GDP-fucose pool and ER GDP-fucose pool (11). Additionally, *Slc35c1* knockout in mice is embryonic lethal with skeletal defects typical of impaired Notch signaling, and phenotypes that overlap with those observed in *Pofut1* knockout mice embryos. SLC35C2, a homologue of SLC35C1, is predicted to be a GDP-fucose transporter. However, *SLC35C2* knockout in HEK293T cells had no impact on Golgi fucosylation or ER *O*-fucosylation, and no developmental defects were observed in *Slc35c2* null mice, indicating it is not a GDP-fucose transporter. Although an ER GDP-fucose transporter has been identified in Drosophila, its human ortholog, SLC35B4, is localized in the Golgi and has been demonstrated as a UDP-xylose/GlcNAc transporter (12). The mammalian ER GDP-fucose transporter remains to be identified.

The salvage pathway is believed to contribute only a minor portion to the cellular GDP-fucose pool, based on studies in the 1970s that utilized exogenous ^{3}H-fucose as a metabolic tracer (13, 14). However, oral fucose supplementation has proven to be an effective therapy for SLC35C1-CDG (LAD II) and GFUS-CDG patients with defects in GDP-fucose transportation or de novo synthesis (CDG: congenital disorders of glycosylation) (15). This suggests that the salvage

pathway has the capacity to enhance GDP-fucose production, thus compensating for deficiencies in the transport or de novo synthesis of GDP-fucose. Recent studies revealed that GDP-fucose originating from exogenous sources, salvaged from degradation, or de novo synthesized, is stored in separate and distinct pools (16, 17). These pools communicate with each other to tightly maintain the total GDP-fucose pool at a constant level. Feedback inhibition of GMD and Fucokinase likely plays a role in this regulatory mechanism, as GDP-fucose is a potent competitive inhibitor for both enzymes, allowing efficient fine-tuning of individual pool sizes (18-20).

1.3 Fucosyltransferases (FUTs)

FUTs catalyze the transfer of a fucose residue from the donor substrate GDP-fucose to acceptor molecules, including oligosaccharides, glycoproteins, glycolipids, and glycoRNAs (21-24). To date, 13 FUTs have been identified in mammals based on conserved sequences that are involved in GDP-fucose binding (Figure 1.2). These FUTs are classified into α1-2, α1-3/4, α1-6 FUTs, and protein *O*-FUTs (POFUTs), based on their distinct fucose linkages. While POFUTs are ER-localized soluble proteins, all other FUTs are type II transmembrane proteins localized in the Golgi.

1.3.1 Golgi-localized FUTs

The Golgi-localized FUTs include FUT1-9, all of which transfer fucose to glycan structures as terminal or core modifications. FUT1, FUT2, and Sec1 (non-functional protein encoded by a pseudogene) are α1-2 FUTs that belong to the CAZy glycosyltransferase family GT-11 (25, 26). FUT1 and FUT2 are responsible for synthesizing the H antigens, which serve as fucosylated

precursors for the A and B oligosaccharide structures in the ABH blood group system (27, 28). They exhibit distinct tissue expression patterns and acceptor specificities. FUT1 (H) is primarily expressed on erythrocyte membranes and vascular endothelium, with a preference for modifying type II structures, whereas FUT2 (Se) is predominantly expressed in epithelial cells and exocrine secretions and is more active on type II structures. FUT3-7 and FUT9 are α1-3/4 FUTs that belong to the CAZy glycosyltransferase family GT-10, all of which have α1-3 activity, except FUT3 and FUT5, which also possess α1-4 activity (29-34). These FUTs are involved in the synthesis of the complex series of Lewis antigen epitopes (i.e., Lewis^{x}, Lewis^{y}, Lewis^{a}, Lewis^{b}, Sialyl-Lewis^{x}, Sialyl-Lewis^{a}) with distinct acceptor specificities. The biological roles of the ABH and Lewis antigens have been previously reviewed in detail (35). FUT8 is the only α1-6 FUT that identified in mammals, belonging to the CAZy glycosyltransferase family GT-23 (36, 37). FUT8 catalyzes *N*-glycan core fucosylation by transferring a fucose moiety to the innermost GlcNAc residue of the chitobiose unit. Core fucosylation plays critical roles in the regulation of antibody-dependent cellular cytotoxicity (ADCC) and serves as a biomarker for various types of cancer (38, 39).

Additionally, based on sequence homology to known α1-3/4 FUTs, FUT10 and FUT11 were postulated to be α1-3 FUTs. However, as will be detailed in Chapter 2, substantial experimental evidence confirms that these enzymes function as POFUTs rather than α1-3 FUTs.

1.3.2 ER-localized POFUTs

Unlike the Golgi-localized FUTs that transfer fucose to glycan structures, POFUTs add fucose directly to the hydroxyl residues of serines or threonines of proteins through an *O*-linkage. In mammals, only two POFUTs have been identified: POFUT1 (CAZy family GT-65) and

POFUT2 (CAZy family GT-68) (40-43). POFUT1 contains a C-terminal KDEL-like sequence that retains it in the ER. While POFUT2 does not contain the ER retention sequence, it is proposed to retain in the ER by association with other protein that contain retention sequences. Both POFUTs mediate domain-specific *O*-fucosylation and are highly specific for their respective substrates (44).

POFUT1

The unique fucose-protein linkage was initially identified in the urinary-type plasminogen activator (urokinase) using mass spectrometry (45, 46). This *O*-fucose monosaccharide was located on the single Epidermal Growth Factor-like (EGF) domain of urokinase within a sequence $C^2XXGG[S/T]C^3$ (where C^2 and C^3 are the second and third cysteine in the EGF, S/T is the modified site). EGF is a small protein domain with ~40 amino acids, featuring six highly conserved cysteines that form three disulfide bonds (Figure 1.3). Soon after, more proteins with *O*-fucose monosaccharide or elongated tetrasaccharide were identified, all containing one or more EGF repeats, including t-PA (47), coagulation factor VII (48), factor XII (49), and factor IX (50). These discoveries sparked an intensive investigation to identify the enzyme responsible for this modification. The Spellman group developed an enzymatic assay to test *O*-fucosyltransferase activity (51). Interestingly, using extracts from Chinese hamster ovary (CHO) cells as the enzyme source, they observed activity when employing the complete EGF domain from human factor VII (bacterially expressed, unmodified) as the acceptor substrate, but not with synthetic peptides containing the consensus sequence. This indicates that the enzyme requires a properly folded EGF structure for its activity, which sets it apart from other *N*-glycan-modifying FUTs. Using this complete EGF domain from human factor VII (bacterially expressed) as bait,

the enzyme (O-FucT-1) was successfully purified from CHO cells by the same group (52). The sequence of O-FucT-1 was used in a screening of a human cDNA library and led to the identification of the human *POFUT1* gene (53). Homologues in mice (*Pofut1*), *Drosophila* (*Ofut1*), and *C. elegans* (*pofut1*) were also identified. The *POFUT1* gene is highly conserved in mammals and shows ubiquitous expression in all tissues examined, indicating the potential important biological roles of *O*-fucosylation.

Our knowledge on POFUT1-mediated EGF *O*-fucosylation has largely expanded as more protein targets have been uncovered and increased research efforts on exploring the biological functions. A refined consensus sequence $C^2XXXX[\underline{S/T}]C^3$ was proposed to include a broader set of *O*-fucose sites (54-56). Database searches with this consensus sequence in the context of an EGF repeat identified 87 putative protein targets for POFUT1 (splice variants are not included) (Table 1.1) (23). However, the *O*-fucosylation status of many of these proteins is yet to be confirmed.

Heterozygous mutations of *POFUT1* in human causes a rare skin condition known as Dowling-Degos Disease, characterized by skin pigmentation abnormalities (57). *Pofut1* knockout in mice is embryonic lethal with multiple severe developmental defects, particularly in somitogenesis. These phenotypes are similar to those observed in *Notch1* knockout mice embryos, indicating Notch as the primary biological target of POFUT1 in embryogenesis (58, 59). The Notch family proteins possess more predicted sites for POFUT1 modification than any other proteins and are the most extensively studied POFUT1 targets (60). Notch is a key regulator in a variety of developmental processes (61). *O*-fucosylation of Notch is essential for its proper functioning in processes like the Notch-dependent regulation of lymphopoiesis and myelopoiesis (62). The extracellular domain of Notch is heavily decorated with *O*-fucose glycans (63). A substantial number of *O*-fucose glycans are elongated to GlcNAcβ1-3Fucose disaccharide by the Fringe

family of β3-*N*-acetylglucosaminyltransferases, and can be further elongated to a tetrasaccharide, Siaα2-6Galβ1-4GlcNAcβ1-3Fuc (54, 63-66). Interestingly, *O*-fucose glycans on distinct EGF repeats play varying roles in the regulation of Notch activity. *O*-fucose residues on EGF8 and EGF12 of NOTCH1 directly interact with ligands, Delta-like 1 and Jagged1, and their binding is further enhanced by Fringe modifications (67-69). However, Fringe modifications on EGF6 and EGF36 inhibit NOTCH1 activation by Jagged1 (70). Through the regulation of POFUT1, Fringe, and several other *O*-glycosyltransferases associated with Notch, cells can dynamically fine-tune Notch activity to accommodate environmental changes.

POFUT2

The discovery of *O*-fucose on thrombospondin type 1 repeats (TSRs) occurred later than the identifying of *O*-fucose on EGF domains. Similar to EGF, TSR is a small protein domain that contains six cysteines forming three disulfide bonds, albeit with distinct bonding patterns (Figure 1.3). The *O*-fucose on TSRs was initially identified as Glc-Fuc disaccharides on thrombospondin-1 by Hofsteenge et al., while mapping sites for *C*-mannosylation (71). An initial consensus sequence was proposed as $C^1SX[S/T]C^2G$, which has been further refined to $C^1XX[S/T]C^2$ (group 1 TSRs) and $C^2XX[S/T]C^3$ (group 2 TSRs) through the comparison of additional TSR *O*-fucose sites identified later and site-mutagenesis studies of TSR structures (72-74). Note that group 1 TSRs and group 2 TSRs have different disulfide bonding patterns (group 1: C^1-C^5, C^2-C^6, C^3-C^4; group 2: C^1-C^4, C^2-C^5, C^3-C^6), but both can be *O*-fucosylated (75, 76). Tryptophan frequently appears upstream of the consensus sequence. While it does not appear to be necessary for *O*-fucosylation, tryptophan residues intercalate with the conserved arginine residues in TSRs, forming a characteristic tryptophan-arginine ladder that forms the core of the

TSR fold (77). Tryptophan also serves as a potential site for *C*-mannosylation (consensus sequence WXXW) that provides further support for the folding of TSR structures (77, 78). An enzyme distinct from POFUT1 was proposed for adding *O*-fucose to TSRs, since purified, recombinant POFUT1 was unable to catalysis this reaction (44). Recombinant *Drosophila* Ofut2, a homologue of *Drosophila* Ofut1, was shown to have the capability to *O*-fucosylate TSRs (41). Homologues of *Ofut2* were also identified in the human (*POFUT2*), mouse (*Pofut2*), and *C. elegans* (*pofut2*) genomes.

Like POFUT1, POFUT2 is ER localized and requires a folded TSR structure for modification. The fact that both POFUT1 and POFUT2 can discern folded substrate structures has led to the hypothesis that they participate in a non-canonical quality control process for protein folding, which will be discussed in more detail later. *Pofut2* knockout in mice is embryonic lethal (79). These embryos die in mid-gastrulation with extensive mesoderm differentiation, suggesting that POFUT2 is a key modulator of gastrulation. Database searches with POFUT2 *O*-fucose consensus sequence in the context of a TSR identified 49 putative protein targets (splice variants are not included) (Table 1.2) (23). Notably, most of these targets are components of the extracellular matrix, with over half belonging to the A-Disintegrin and Metalloproteinase with ThromboSpondin Type-1 motifs (ADAMTS) family. *Adamts9* knockout in mice leads to early embryonic lethality with phenotypes mirroring those observed in *Pofut2* knockout embryos (80). This strongly suggests that the gastrulation defects observed in *Pofut2* knockout embryos are primarily due to the loss of *O*-fucosylation of ADAMTS9, emphasizing that *O*-fucosylation is essential for the biological function of POFUT2 protein targets. The *O*-fucose on TSR can be elongated with a glucose by β3-glucosyltransferase (B3GLCT). While no human mutations in

POFUT2 are identified, mutations in *B3GLCT* lead to a human developmental disorder named Peters Plus Syndrome (PTRPLS), which will be discussed in more detail later (81).

1.4 *O*-fucosylation: physiological and pathological significance

1.4.1 Regulation of lymphoid and myeloid cell development by Notch O-fucosylation

As previously mentioned, Notch is a key regulator in various developmental processes, frequently influencing cell fate decisions at multiple stages as cells differentiate toward specific cell types. In hematopoiesis, proper Notch signaling is essential for the survival, proliferation, and cell fate determination at various stages of lineage commitment, including myeloid cell production in bone marrow, T cells development in thymus and marginal zone (MZ) B cells development in spleen (Figure 1.4). *O*-fucose glycans incorporated to Notch extracellular domain by POFUT1 and further elongated by Fringe modulate the strength of Notch signaling, thus playing critical roles in these Notch-regulated programs.

Throughout adult life, blood cell production is maintained by a group of pluripotent hematopoietic stem cells (HSCs) located in the bone marrow. HSCs generate multipotent progenitor populations (MPPs) with increased lineage-specific potentials. Within this MPP pool, lymphoid-primed MPPs (LMPP or MPP4) subsequently give rise to common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs), which are committed to the lymphoid lineage (T, B, and natural killer cells) and the myeloid lineage (neutrophils, monocytes, erythrocytes, and more), respectively (82). Notch signaling influences this decision by favoring a lymphoid over myeloid lineage outcome (83, 84). Conditional knockout *Pofut1* in mice HSCs within bone marrow leads to reduced T lymphopoiesis in thymus and reduced marginal-zone B cells production in spleen, whereas myeloid cells in bone marrow were increased (85).

Restoration of Notch1 signaling rescued these phenotypes. The *Pofut1* knockout in HSCs completely abrogated the binding of cell surface Notch with Delta ligands and suppressed the activation of downstream Notch target genes. In another study, the selective deletion of the Notch ligand DLL4 in bone-producing osteocalcin (Ocn)-expressing cells, which highly express DLL4 in the bone marrow, results in a decrease in intrathymic T cell precursors and the production of mature T cells (86). These findings suggest that Notch *O*-fucosylation is essential for maintaining hematopoietic lineage homeostasis by promoting lymphoid development and suppressing overt myelopoiesis. Interestingly, pan-Notch inhibition through the conditional deletion of *RBP-Jκ* led to more severe phenotypes than the deletion of *Pofut1* in HSCs, suggesting other *O*-glycans on Notch may contribute to this Notch-regulated process. The roles of other Notch *O*-glycans in hematopoiesis were recently reviewed in detail (62).

T cell maturation takes place in the thymus, where lymphocyte precursors released from the bone marrow migrate through blood vessels and colonize the thymus as thymus-seeding progenitors (TSPs). Upon entering the thymus, these cells initiate a differentiation process that includes crucial events such as T cell lineage commitment and the choice between αβ and γδ T-cell lineages, as they migrate from the corticomedullary junction towards the outer cortical zone. During this migration, the progenitor cells traverse a network of thymic stromal cells that express Notch ligands, initiating Notch signaling in the progenitor cells and guiding them through the differentiation programs. The requirement of Notch signaling in T cell development is well-documented. TSPs possess robust potential to generate non-T cell lineages, including B cells, myeloid cells, and dendritic cells (DC) (87). NOTCH1-DLL4 signaling compels these cells to commit to a T cell fate while simultaneously suppressing their potential to adopt other lineages. Disrupting Notch signaling in HSCs through conditional deletion of *Notch1* or *RBPJκ* blocks

early-stage T cell development, causing T cell progenitors to differentiate into B cells and myeloid cells in the thymus (88-91). By using fucose analogs (6-alkynyl fucose and 6-alkenyl fucose) that incorporate into Notch and specifically inhibit Delta-induced Notch signaling in an *in vitro* OP9 co-culture system, the differentiation of bone marrow stem cells into T cell progenitors was completely abolished, indicating the important roles of *O*-fucose in Notch-regulated T cell lineage commitment. Interestingly, overexpression of Lfng in the thymus, which enhances Notch binding with Delta ligands, unexpectedly inhibits Notch activation in TSPs (90). This is because when Lfng is overexpressed in CD4/CD8 double-negative (DN) 3 and CD4/CD8 double-positive (DP) cells, it transforms them into 'super-competitors' for the limited intrathymic niches supplying Delta ligands (92). This limits the access of early T-cell progenitors towards these ligands, which are needed for the required Notch signaling for their differentiation. NOTCH2-DLL1 signaling is required for the development of marginal zone (MZ) B cells in the spleen (93, 94). Conditional delete *Pofut1* in HSCs reduced marginal-zone B cells production in spleen (85). Although neither Lunatic Fringe (Lfng) nor Manic Fringe (Mfng) is indispensable for MZ B cell development, the loss of both Fringes significantly impairs the production of MZ B cells in spleen (95). Therefore, Lfng and Mfng cooperatively enhance the NOTCH2-DLL1 signaling to promote the development of MZ B cells.

1.4.2 Cancer with altered O-fucosylation

Notch signaling plays diverse roles during development and tissue homeostasis. A growing number of cancers are linked to dysregulated Notch signaling, where Notch can act as either an oncogene or a tumor suppressor, depending on the cellular context (96). As a crucial regulator of Notch signaling, altered *O*-fucosylation is commonly observed in cancer. Here, we highlight

prototypical cancers where altered *O*-fucosylation plays a role in their pathological mechanisms. This alteration can involve changes in POFUT1 levels or activity, or mutations in Notch that impact its *O*-fucosylation.

POFUT1 is localized in the chromosomal region 20q11.21, which is frequently amplified in colorectal cancer (CRC) (97). This copy number variations (CNVs) in CRC results in elevated POFUT1 expression in tumors compared to non-tumor adjacent tissues (98). The high expression of POFUT1 in CRC from stage I is associated with the metastatic process. ShRNA-mediated knockdown of POFUT1 downregulates NOTCH1 signaling, leading to decreased cell proliferation, migration, and induced apoptosis in CRC cells *in vitro*, as well as suppression of CRC tumor growth and transplantation *in vivo* (99). In less frequent CRC cases with no chromosomal amplification and POFUT1 overexpression, gain-of-function mutations in POFUT1 have been identified, potentially contributing to tumor progression (99). POFUT1 is overexpressed in hepatocellular carcinoma (HCC), and its high expression correlates with poor prognosis in HCC patients (100, 101). Elevated POFUT1 levels accelerated the proliferation and migration of HCC cells and increased their binding to DLL1, consequently enhancing Notch signaling activity. Overexpression of POFUT1 is also observed in glioblastoma (102, 103), gastric cancer (104), oral cancer (105), and breast cancer (106), correlating with increased Notch activity and aggressive phenotypes, such as increased cell proliferation, invasion, metastasis, or larger tumor size. Interestingly, in cancer types where Notch functions as a tumor suppressor, such as in muscle-invasive bladder cancer (MIBC), low levels of POFUT1 mRNA are associated with poor survival in MIBC patients after radical cystectomy (107, 108). These findings highlight the potential of POFUT1 as a diagnostic marker and therapeutic target for these cancers.

In anaplastic large cell lymphoma (ALCL), the Notch pathway ranks among the pathways that are most enriched in mutations through gene-set enrichment analysis. NOTCH1 p.T311P (eliminate O-fucose site on EGF8) and p.T349P (eliminate O-fucose site on EGF9) missense mutations were found in patients' tumor samples by whole-exome sequencing (109). Interestingly, overexpression of these mutations in HEK293T cells led to increased proliferation compared to cells with wild-type NOTCH1. While the detailed mechanism remains unclear, this suggests a correlation between alterations in specific *O*-fucose sites on NOTCH1 and cancer. Recently, by using glycoproteomic site mapping and cell-based Notch assays, nine cancer-associated Notch variants were investigated to determine their impact on Notch *O*-fucosylation, ligand binding and signaling activity (110). Two variants led to a gain of function in NOTCH1(G310R and T311P), six led to a loss of function (G310R, T311P, G347S, T349P, D464N, and A465T), and one had minimal effects (G230R). This suggests that point mutations in Notch can alter its *O*-fucosylation, consequently affecting Notch signaling activity and being linked to cancer-related processes.

1.4.3 Peters plus syndrome (PTRPLS)

Failure to add glucose to *O*-fucose on TSRs causes human Peters plus syndrome (PTRPLS, MIM #261540), an autosomal recessive congenital disorder of glycosylation (CDG) characterized by Peters anomaly of the eye (anterior eye chamber segment dysgenesis), craniofacial defects (widened forehead and cleft lip/palate), short stature, brachydactyly, and developmental delay (81). PTRPLS is caused by loss-of-function mutations in β3-glucosyltransferase (B3GLCT), an enzyme responsible for extending *O*-fucose monosaccharide to the glucose-β1,3-fucose disaccharides on TSRs (111, 112). PTRPLS-like patients only share a subset of phenotypes

observed in PTRPLS patients, and some of them also have missense mutations in B3GLCT (111). In contrast to PTRPLS mutations that lead to a complete loss of enzymatic activity, PTRPLS-like mutations retained enzymatic activity with only minor destabilizing effects (113). This partially explained the milder phenotypes observed in PTRPLS-like patients.

Over half of the putative POFUT2 targets are members of the ADAMTS/ADAMTS-like family, suggesting that these proteins could be the primary biological targets for B3GLCT. These secreted ADAMTS proteases and non-catalytic members of the family (ADAMTS-like proteins) are important components of the extracellular matrix (ECM), and are implicated in diverse biological events, including embryogenesis and angiogenesis (114, 115). Numerous congenital disorders caused by mutations in *ADAMTS/ADAMTS-like* genes display overlapping phenotypes with PTRPLS patients. For instance, mutations in *ADAMTSL2* cause human Geleophysic dysplasia 1 (GPHYSD1, MIM #231050), where patients exhibit overlapped phenotypes such as short stature and short hands and feet (116, 117). Mutations in *ADAMTS10* cause Weill-Marchesani syndrome 1 (WMS, MIM #277600), with overlapped phenotypes including short stature, brachydactyly and eye anomalies (118). These suggest that impaired *O*-fucosylation of the ADAMTS/ADAMTS-like family is likely the primary biological mechanism responsible for the phenotypes observed in PTRPLS patients.

POFUT2 and B3GLCT are demonstrated to mediate a noncanonical ER quality-control process for the secretion of TSR-containing proteins (molecular details will be discussed later) (119). POFUT2 is required for the secretion of all ADAMTS/ADAMTS-like proteins tested to date [ADAMTS6 (120), ADAMTS9 (121, 122), ADAMTS10 (120), ADAMTS13 (119), ADAMTS17 (123), ADAMTS20 (121), ADAMTSL1 (119), ADAMTSL2 (119)], whereas the impact of B3GLCT on secretion varied among proteins. For instance, secretion of ADAMTS20,

ADAMTSL1, and ADAMTSL2 is profoundly reduced with B3GLCT knockdown/out (119, 121). In contrast, ADAMTS9 and ADAMTS13 secretion is moderately reduced by 20%, and ADAMTS17 secretion is unaffected (119, 121-123). Considering the early embryo lethality observed in *Pofut2*-null mice and the less severe phenotypes in PTRPLS patients, it implies that the pathology of PTRPLS is due to defects in a specific subset of POFUT2 targets that are sensitive to the loss of extended glucose.

1.4.4 Other biological processes where O-fucosylation may play a role

In addition to the biological events discussed above, *O*-fucose glycans have been identified in numerous other biologically important proteins, where they play crucial roles in modulating protein function. POFUT1-mediated *O*-fucosylation of Agrin, a key regulator of postsynaptic differentiation at the neuromuscular junction (NMJ), has been demonstrated to determine its acetylcholine receptor clustering activity (124). The *O*-fucose moiety on TSR3 of BAI1, added by POFUT2, directly interacts with the RTN4 receptor, enabling high-affinity interactions that regulate neuronal development (125). Mutating the *O*-fucose site within the EGF domain of Multimerin-1, an essential platelet component supporting platelet adhesion and thrombus formation, resulted in a complete loss of its secretion, implying an as-yet-unidentified role for POFUT1 in platelet functions (126). We envision that a deeper profiling of the *O*-fucose proteome will uncover more biological significance of *O*-fucosylation in the future.

1.5 Molecular mechanisms of how *O*-fucose regulates protein functions

1.5.1 O-fucosylation forms direct contacts with binding partners

Out of the 36 EGF repeats in the extracellular domain of mouse NOTCH1, 20 contain the *O*-fucose consensus sequence, and 17 have been confirmed to be modified at high stoichiometries (70). Similar results were observed in *drosophila* NOTCH and mouse NOTCH2 (127, 128). These heavily decorated *O*-fucose glycans are required for the proper function of Notch. Through the examination of individual *O*-fucose effects on NOTCH1 activity via site-directed mutagenesis, it was found that *O*-fucose on EGF8 and EGF12 had the most significant impact on the binding of NOTCH1 with DLL1 and JAG1 ligands (70). Similar results were obtained with NOTCH2 (127). *In vitro* assays using purified human NOTCH1 EGF11-13 also demonstrated that the addition of *O*-fucose enhanced its binding to DLL1 and JAG1 ligands (69).

Evidence confirming the direct interaction of *O*-fucose on NOTCH1 EGF8 and EGF12 with ligands is provided by two co-crystal structures: one between a portion of NOTCH1 ligand-binding domain (EGF11-13) and a portion of DLL4 (67), and the other between the NOTCH1 ligand-binding domain (EGF8-12) and a portion of JAG1 (Figure 1.5) (68). Both co-crystal structures showed direct contacts between *O*-fucose on EGF12 with the backbone and side-chain residues of both ligands. The NOTCH1-JAG1 structure also showed a direct contact of *O*-fucose on EGF8 with JAG1. Unlike many other glycan modifications, *O*-fucose residues display thermal mobility comparable to amino acids, acting as ‘surrogate amino acids’ to form specific and essential contacts with residues on binding partners, thereby facilitating ligand binding. This direct intermolecular interaction generated by *O*-fucose glycans in protein-protein interactions was also observed in the co-crystal structure of BAI1 with RTN4 receptor (Figure 1.5) (125).

1.5.2 O-fucosylation generates intramolecular interactions that stabilize protein domains and facilitate secretion

The appropriate pairing of cysteines to form correct disulfide bonds is a crucial step during protein folding in the secretory pathway, and this presents a particular challenge for proteins that contain cysteine-rich domains such as EGFs and TSRs. It provides a strategic rationale for cells to evolve dedicated quality-control mechanisms to meet the specialized requirements for the folding of these proteins. As mentioned above, both POFUT1 and POFUT2 only modify properly folded domain structures and localized in the ER. These led to the hypothesis that the ER-localized POFUTs participate in a non-canonical ER quality control system for the folding and secretion of proteins containing cysteine-rich domains.

In co-crystal structures of mouse POFUT1 with EGF domains from Notch, the binding interface between POFUT1 and EGF showed a high degree of complementarity, with approximately one-third of the EGF's surface area being deeply embedded within the groove of POFUT1 (129). The EGF region that contains *O*-fucose consensus motif $C^2XXXX[S/T]C^3$ is buried at the bottom and the hydroxyl group of serine or threonine to be modified is precisely oriented to the position necessary for the fucose transfer. Similarly, in co-crystal structures of Caenorhabditis elegans POFUT2 (CePOFUT2) with human TSR1, nearly half of the TSR1 is embedded within the groove of CePOFUT2 (72). These 'hand-in-glove' complementary interactions between enzymes and their respective protein domain structures elegantly explain why the enzyme necessitates a folded structure and a specific consensus sequence for modification.

The addition of an *O*-fucose to EGF or TSR domain stabilizes the folded structures and protects them from unexpected unfolding due to environmental factors. *O*-fucosylated EGF repeats and TSRs showed a significantly reduced unfolding rate in reductive unfolding assay compared to

their unmodified counterparts (119, 130). Additionally, the refolding of unmodified TSRs was accelerated in the presence of POFUT2 and GDP-fucose *in vitro*, whereas the addition of POFUT2 alone did not have the same effect (119). This implies that the transferred *O*-fucose, rather than POFUT2, contributes to the acceleration of the folding rate. Recently, through a combination of MD simulations, X-ray crystallography, and NMR, it was demonstrated that the *O*-fucose glycans on TSR3 of THBS1 generate intramolecular interactions with adjacent amino acids and shield the C2-C6 disulfide bond to protect it from reduction (131). This stabilizing effect generated by the *O*-fucose glycans drives the domain into an energy well, preventing the domains from re-entry into the folding cycle, and consequently facilitating protein folding and secretion. In line with this, several studies have shown that disruptions in *O*-fucosylation lead to defects in protein secretion. As mentioned previously, knockdown or knockout of *POFUT2* in cells eliminates the secretion of all ADAMTS/ADAMTS-like proteins tested to date. Knockdown of *drosophila Ofut1* led to reduced cell surface expression of NOTCH and its accumulation in the ER. Interestingly, this secretion defect can be partially rescued by an enzymatically inactive form of POFUT1 (R240A mutant), suggesting that POFUT1 might possess chaperone activity independent of its enzymatic function (132). However, mice carrying POFUT1 with point mutations at the equivalent position exhibited severe defects in somite formation, resembling those observed in *Pofut1* null mice (133). Additionally, the POFUT1 R240A mutant showed significantly reduced effectiveness in rescuing Notch signaling in *POFUT1* knockout U2OS cells when compared to the wildtype POFUT1 (134). The chaperone activity of POFUT1 may depend on the specific context within different organisms and can yield distinct biological functions. *POFUT1* knockout in different cellular contexts led to varying outcomes in reducing cell surface Notch levels. For instance, *POFUT1* knockout in HEK293T

cells resulted in ~60% reduced surface expression of NOTCH1, whereas *Pofut1* knockout in mouse embryonic stem cells displayed a similar level of cell surface NOTCH compared to the wild type (130, 135). The variation observed may stem from the distinct regulation of other chaperones that compensate for the folding and secretion defects in Notch within different cellular contexts.

1.6 Inhibitors/modulators of *O*-fucosylation

To date, no inhibitors for POFUT1 or POFUT2 have been identified. As discussed above, POFUT1 and POFUT2 play crucial roles in diseases, including various types of cancer. Having specific inhibitors or modulators for these enzymes would hold significant therapeutic potential and serve as valuable research tools. A virtual compound library screening with available POFUT1/2 structures could serve as a starting point for identifying potential inhibitors.

While not directly targeting POFUT1/2, specific fucose analogs like 2-fluoro-fucose (2-FF) and 6-alkynyl fucose (6-AF) are pan-inhibitors of cellular fucosylation as they inhibit the de novo synthesis of GDP-fucose (136, 137). Furthermore, 6-AF can enter the salvage pathway and be converted into GDP-6-AF, which is well-tolerated by POFUT1 and POFUT2 for incorporation into their respective protein targets (138-140). The incorporated 6-AF can alter the functions of proteins, as exemplified in the case of Notch, where the incorporation of 6-AF selectively inhibits DLL ligands-mediated Notch signaling but not JAG ligands (139). Therefore, fucose analogs could serve as valuable tools for *O*-fucosylation research and hold therapeutic potential for their capacity to modulate the functions of *O*-fucosylated proteins.

1.7 Goals of this dissertation

Given the biological importance of *O*-fucosylation discussed above, identification of proteins where the *O*-fucose plays a role in regulating their functions, as well as understanding how *O*-fucose contributes to their function, would be of great value. Although extensive research has been done on prominent *O*-fucosylated proteins such as Notch and ADAMTS family proteins, further investigations into other proteins would deepen our understanding of *O*-fucosylation. These potential *O*-fucosylated proteins might originate from a novel domain-specific *O*-fucosylation event or represent unexpected substrates for POFUT1 or 2.

In this dissertation, we expand the *O*-fucosylation repertoire by unveiling two novel POFUTs and providing tools for identifying previously unknown targets for POFUT1 and 2. In Chapter 2, a new type of domain-specific *O*-fucosylation is discussed. We provide evidence indicating that FUT10 and FUT11 are POFUTs responsible for modifying protein EMI domains, thus we renamed them POFUT3 and POFUT4, respectively. In Chapter 3, we show the development of an *O*-fucose-specific tag, Fuc6nyl*, designed for the specific labeling of POFUT1/2 targets and enhancing the mass spectrometric analysis of *O*-fucosylated peptides derived from them. Leveraging Fuc6nyl*, a glycoproteomic workflow was developed for profiling of the *O*-fucose proteome of POFUT1/2.

1.8 Tables and figures

Table 1.1 List of putative human gene targets of POFUT1

UNIPROT ID	Gene	Protein	Consensus/ total	Subcellular location		
O00468	AGRIN	Agrin	2/4	Extracellular matrix	Cell membrane	
Q6UW56	ATRAID	All-trans retinoic acid-induced differentiation factor	1/1		Cell membrane	Nucleus
Q9NPY3	CD93	Complement component C1q receptor	1/5		Cell membrane	
P48960	CD97	Adhesion G protein-coupled receptor E5	1/5	Extracellular matrix	Cell membrane	
Q9NYQ6	CELSR1	Cadherin EGF LAG seven-pass G-type receptor 1	2/8		Cell membrane	
Q9HCU4	CELSR2	Cadherin EGF LAG seven-pass G-type receptor 2	2/7		Cell membrane	
Q9NYQ7	CELSR3	Cadherin EGF LAG seven-pass G-type receptor 3	2/8		Cell membrane	
P0CG37	CFC1	Cryptic protein	1/1	Extracellular matrix	Cell membrane	
P0CG36	CFC1B	Cryptic family protein 1B	1/1	Extracellular matrix		
Q8WYK1	CNTNAP5	Contactin-associated protein-like 5	1/2		Cell membrane	
P82279	CRB1	Protein crumbs homolog 1	8/19	Extracellular matrix	Cell membrane	
Q5IJ48	CRB2	Protein crumbs homolog 2	8/15	Extracellular matrix	Cell membrane	Cytoplasm
P13611	CSPG2	Versican core protein	2/2	Extracellular matrix		
O60494	CUBN	Cubilin	4/7		Cell membrane	Endosome Lysosome
P80370	DLK1	Protein delta homolog 1	3/6		Cell membrane	Cytoplasm
Q6UY11	DLK2	Protein delta homolog 2	1/6		Cell membrane	
O00548	DLL1	Delta-like protein 1	4/8		Cell membrane	
Q9NYJ7	DLL3	Delta-like protein 3	2/6		Cell membrane	
Q9NR61	DLL4	Delta-like protein 4	5/8		Cell membrane	
Q8NFT8	DNER	Delta and Notch-like epidermal growth factor-related receptor	6/10		Cell membrane	
O43854	EDIL3	EGF-like repeat and discoidin I-like domain-containing protein 3	1/3	Extracellular matrix		
P01133	EGF	Pro-epidermal growth factor	1/9		Cell membrane	
Q9UHF1	EGFL7	Epidermal growth factor-like protein 7	1/2	Extracellular matrix		
Q63HQ2	EGFLAM	Pikachurin	2/3	Extracellular matrix		
Q14246	EMR1	Adhesion G protein-coupled receptor E1	4/6		Cell membrane	
Q9UHX3	EMR2	Adhesion G protein-coupled receptor E2	1/5		Cell membrane	
Q5T1H1	EYS	Protein eyes shut homolog	11/27	Extracellular matrix		Cytoplasm Cytoskeleton Cell projection
P08709	F7	Coagulation factor VII	1/2	Extracellular matrix		
P00740	F9	Coagulation factor IX	1/2	Extracellular matrix		
P00748	F12	Coagulation factor XII	1/2	Extracellular matrix		
Q14517	FAT1	Protocadherin Fat 1	2/5		Cell membrane	Nucleus
Q9NYQ8	FAT2	Protocadherin Fat 2	1/2		Cell membrane	Golgi
Q8TDW7	FAT3	Protocadherin Fat 3	3/4		Cell membrane	
Q6V0I7	FAT4	Protocadherin Fat 4	5/6		Cell membrane	
P23142	FBLN1	Fibulin-1	1/9	Extracellular matrix		
Q53RD9	FBLN7	Fibulin-7	1/3	Extracellular matrix		
P35556	FBN2	Fibrillin-2	1/47	Extracellular matrix		
Q75N90	FBN3	Fibrillin-3	1/44	Extracellular matrix		
Q14520	HABP2	Hyaluronan-binding protein 2	1/3	Extracellular matrix		
Q04756	HGFAC	Hepatocyte growth factor activator	2/2	Extracellular matrix		
P78504	JAG1	Protein jagged-1	11/16		Cell membrane	
Q9Y219	JAG2	Protein jagged-2	9/16		Cell membrane	
Q07954	LRP1	Prolow-density lipoprotein receptor-related protein 1	5/22		Cell membrane	Cytoplasm Nucleus Golgi
Q9NZR2	LRP1B	Low-density lipoprotein receptor-related protein 1B	4/14		Cell membrane	
Q14767	LTBP2	Latent-transforming growth factor beta-binding protein 2	1/20	Extracellular matrix		
O75095	MEGF6	Multiple epidermal growth factor-like domains protein 6	1/27	Extracellular matrix		
Q7Z7M0	MEGF8	Multiple epidermal growth factor-like domains protein 8	2/5		Cell membrane	
Q96KG7	MEGF10	Multiple epidermal growth factor-like domains protein 10	2/15		Cell membrane	Cell projection
A6BM72	MEGF11	Multiple epidermal growth factor-like domains protein 11	2/14		Cell membrane	
Q13201	MMRN1	Multimerin-1	1/1	Extracellular matrix		
O14594	NCAN	Neurocan core protein	2/2	Extracellular matrix		
Q92832	NELL1	Protein kinase C-binding protein NELL1	1/5	Extracellular matrix		Cytoplasm Nucleus
Q14112	NID2	Nidogen-2	1/5	Extracellular matrix		
P46531	NOTCH1	Neurogenic locus notch homolog protein 1	20/36		Cell membrane	Nucleus

Q04721	NOTCH2	Neurogenic locus notch homolog protein 2	20/36		Cell membrane	Nucleus
Q7Z3S9	NOTCH2NL	Notch homolog 2 N-terminal-like protein A	5/6	Extracellular matrix		Cytoplasm
Q9UM47	NOTCH3	Neurogenic locus notch homolog protein 3	14/34		Cell membrane	Nucleus
Q99466	NOTCH4	Neurogenic locus notch homolog protein 4	18/29		Cell membrane	Nucleus
Q6UXH9	PAMR1	Inactive serine protease PAMR1	1/1	Extracellular matrix		
Q5VY43	PEAR1	Platelet endothelial aggregation receptor 1	1/9		Cell membrane	
P98160	PGBM	Basement membrane-specific heparan sulfate proteoglycan core protein	3/4	Extracellular matrix		
Q96GW7	PGCB	Brevican core protein	1/1	Extracellular matrix	Cell membrane	
P04070	PROC	Vitamin K-dependent protein C	1/2	Extracellular matrix		Golgi ER
P22891	PROZ	Vitamin K-dependent protein Z	1/2	Extracellular matrix		
P78509	RELN	Reelin	2/8	Extracellular matrix		
O75093	SLIT1	Slit homolog 1 protein	2/9	Extracellular matrix		
O94813	SLIT2	Slit homolog 2 protein	3/7	Extracellular matrix		
O75094	SLIT3	Slit homolog 3 protein	3/9	Extracellular matrix		
Q8TER0	SNED1	Sushi, nidogen and EGF-like domain-containing protein 1	10/15	Extracellular matrix		
Q96GP6	SREC2	Scavenger receptor class F member 2	1/7		Cell membrane	
Q9NY15	STAB1	Stabilin-1	3/16		Cell membrane	
Q8WWQ8	STAB2	Stabilin-2	6/17		Cell membrane	Cytoplasm
Q6UWL2	SUSD1	Sushi domain-containing protein 1	2/3		Cell membrane	
Q4LDE5	SVEP1	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	4/9	Extracellular matrix	Cell membrane	Cytoplasm
Q9UKZ4	TEN1	Teneurin-1	1/8		Cell membrane	Cytoplasm Nucleus
Q9NT68	TEN2	Teneurin-2	2/8		Cell membrane	Golgi ER
Q6N022	TEN4	Teneurin-4	2/8		Cell membrane	Cytoplasm Nucleus
P35590	TIE1	Tyrosine-protein kinase receptor Tie-1	1/3		Cell membrane	
P00750	TPA	Tissue-type plasminogen activator	1/1	Extracellular matrix		
P49746	TSP3	Thrombospondin-3	1/3	Extracellular matrix		

P07911	UMOD	Uromodulin	3/3	Extracellular matrix	Cell membrane	
Q5DID0	UMODL1	Uromodulin-like 1	1/3		Cell membrane	Cytoplasm
P00749	UROK	Urokinase-type plasminogen activator	1/1	Extracellular matrix		
Q6EMK4	VASN	Vasorin	1/1	Extracellular matrix	Cell membrane	
Q5GFL6	VWA2	von Willebrand factor A domain-containing protein 2	2/2	Extracellular matrix		
Q8N2E2	VWDE	von Willebrand factor D and EGF domain-containing protein	3/7	Extracellular matrix		
Q9Y5W5	WIF1	Wnt inhibitory factor 1	2/5	Extracellular matrix		

Table 1.2 List of putative human gene targets of POFUT2

UNIPROT ID	Gene	Protein	Consensus/ total	Subcellular location		
Q9UHI8	ADAMTS1	A disintegrin and metalloproteinase with thrombospondin motifs 1	3/3	Extracellular matrix		
O95450	ADAMTS2	A disintegrin and metalloproteinase with thrombospondin motifs 2	2/4	Extracellular matrix		
O15072	ADAMTS3	A disintegrin and metalloproteinase with thrombospondin motifs 3	2/4	Extracellular matrix		
O75173	ADAMTS4	A disintegrin and metalloproteinase with thrombospondin motifs 4	1/1	Extracellular matrix		
Q9UNA0	ADAMTS5	A disintegrin and metalloproteinase with thrombospondin motifs 5	2/2	Extracellular matrix		
Q9UKP5	ADAMTS6	A disintegrin and metalloproteinase with thrombospondin motifs 6	3/5	Extracellular matrix		
Q9UKP4	ADAMTS7	A disintegrin and metalloproteinase with thrombospondin motifs 7	5/8	Extracellular matrix		
Q9UP79	ADAMTS8	A disintegrin and metalloproteinase with thrombospondin motifs 8	2/2	Extracellular matrix		
Q9P2N4	ADAMTS9	A disintegrin and metalloproteinase with thrombospondin motifs 9	12/15	Extracellular matrix		ER
Q9H324	ADAMTS10	A disintegrin and metalloproteinase with thrombospondin motifs 10	3/5	Extracellular matrix		
P58397	ADAMTS12	A disintegrin and metalloproteinase with thrombospondin motifs 12	6/8	Extracellular matrix		
Q76LX8	ADAMTS13	A disintegrin and metalloproteinase with thrombospondin motifs 13	7/8	Extracellular matrix		
Q8WXS8	ADAMTS14	A disintegrin and metalloproteinase with thrombospondin motifs 14	2/4	Extracellular matrix		
Q8TE58	ADAMTS15	A disintegrin and metalloproteinase with thrombospondin motifs 15	3/3	Extracellular matrix	Cell membrane	
Q8TE57	ADAMTS16	A disintegrin and metalloproteinase with thrombospondin motifs 16	6/6	Extracellular matrix		
Q8TE56	ADAMTS17	A disintegrin and metalloproteinase with thrombospondin motifs 17	4/5	Extracellular matrix		
Q8TE60	ADAMTS18	A disintegrin and metalloproteinase with thrombospondin motifs 18	4/5	Extracellular matrix		
Q8TE59	ADAMTS19	A disintegrin and metalloproteinase with thrombospondin motifs 19	4/5	Extracellular matrix		
P59510	ADAMTS20	A disintegrin and metalloproteinase with thrombospondin motifs 20	11/15	Extracellular matrix		
Q8N6G6	ADAMTSL1	ADAMTS-like protein 1	8/9	Extracellular matrix		
Q86TH1	ADAMTSL2	ADAMTS-like protein 2	6/7	Extracellular matrix		
P82987	ADAMTSL3	ADAMTS-like protein 3	8/10	Extracellular matrix		
Q6UY14	ADAMTSL4	ADAMTS-like protein 4	2/6	Extracellular matrix		
Q6ZMM2	ADAMTSL5	ADAMTS-like protein 5	1/1	Extracellular matrix		
O14514	BAI1	Adhesion G protein-coupled receptor B1	4/5	Extracellular matrix	Cell membrane	
O60241	BAI2	Adhesion G protein-coupled receptor B2	4/4		Cell membrane	
O60242	BAI3	Adhesion G protein-coupled receptor B3	4/4		Cell membrane	
P13671	C-6	Complement component C6	1/3	Extracellular matrix		
Q8IUL8	CILP2	Cartilage intermediate layer protein 2	1/1	Extracellular matrix		
P29279	CTGF	CCN family member 2	1/1	Extracellular matrix		
O00622	CYR61	CCN family member 1	1/1	Extracellular matrix		
Q96RW7	HMCN1	Hemicentin-1	6/6	Extracellular matrix	Cell membrane	Cytoplasm
B1AKI9	ISM1	Isthmin-1	1/1	Extracellular matrix		
P48745	NOV	CCN family member 3	1/1	Extracellular matrix		Cytoplasm Cell junction
O95428	PPN	Papilin	4/5	Extracellular matrix		
P27918	PROP	Properdin	4/7	Extracellular matrix		
Q13591	SEM5A	Semaphorin-5A	2/7		Cell membrane	
Q9P283	SEM5B	Semaphorin-5B	2/5		Cell membrane	
Q9HCB6	SPON1	Spondin-1	5/6	Extracellular matrix		
A2VEC9	SSPO	SCO-spondin	10/24	Extracellular matrix		
Q9UPZ6	THS7A	Thrombospondin type-1 domain-containing protein 7A	4/15	Extracellular matrix	Cell membrane	
Q9C0I4	THS7B	Thrombospondin type-1 domain-containing protein 7B	4/18		Cell membrane	
Q9NS62	THSD1	Thrombospondin type-1 domain-containing protein 1	1/1	Extracellular matrix	Cell membrane	Endosome membrane Cell junction
Q6ZMP0	THSD4	Thrombospondin type-1 domain-containing protein 4	3/6	Extracellular matrix		
P07996	TSP1	Thrombospondin-1	3/3	Extracellular matrix		ER
P35442	TSP2	Thrombospondin-2	3/3	Extracellular matrix		Basement membrane
O95388	WISP1	CCN family member 4	1/1	Extracellular matrix		
O76076	WISP2	CCN family member 5	1/1	Extracellular matrix		
O95389	WISP3	Cellular communication network factor 6	1/1	Extracellular matrix		Mitochondrion

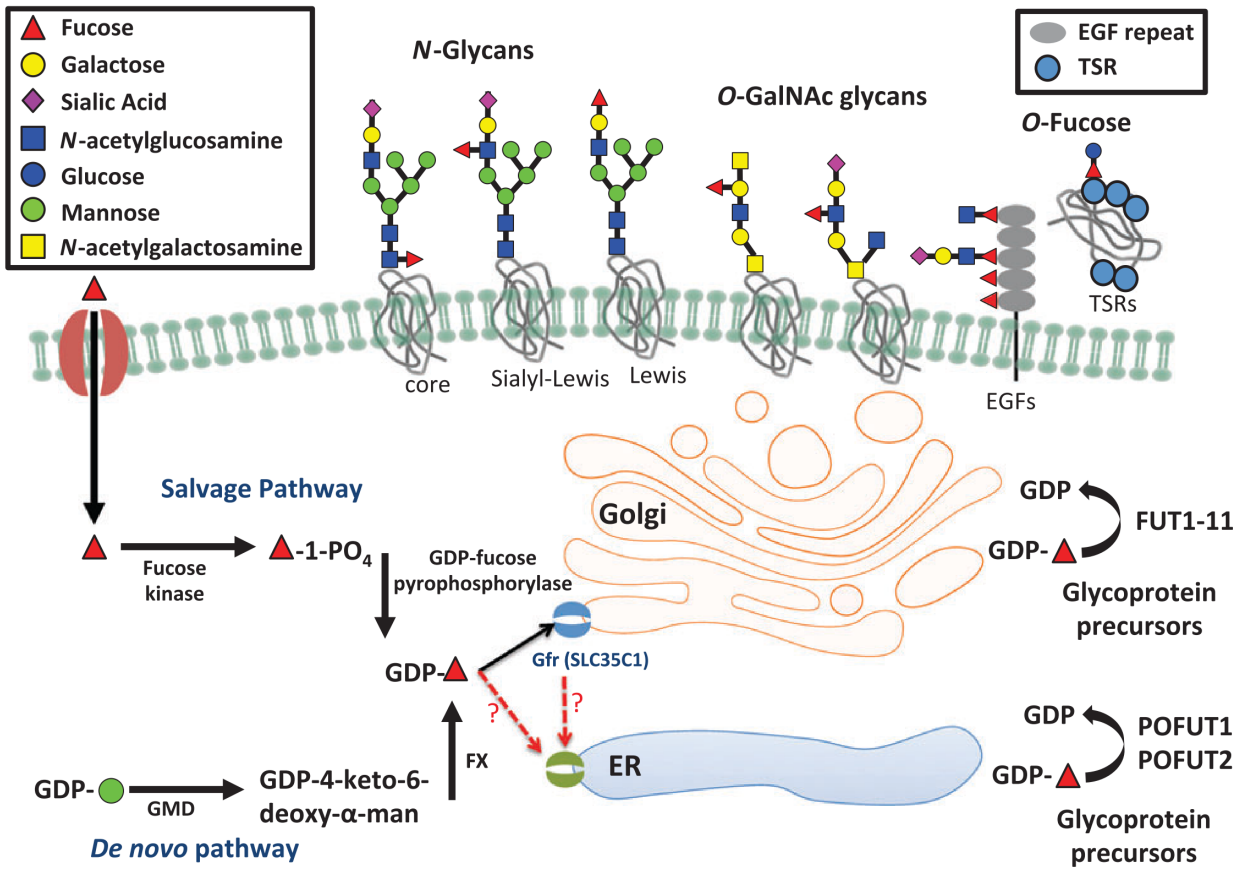


Figure 1.1 GDP-fucose synthesis pathways and transport. Major synthesis steps and involved enzymes for the de novo pathway and salvage pathway are shown. Synthesized GDP-fucose in the cytosol is transported to Golgi by the Golgi GDP-fucose transporter SLC35C1, while the ER GDP-fucose transporter is still undefined. Reproduced from (23) with permission.

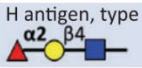
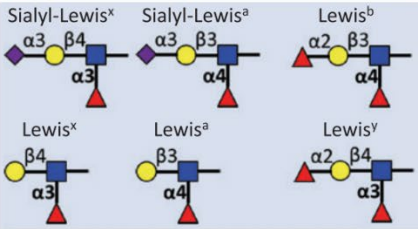
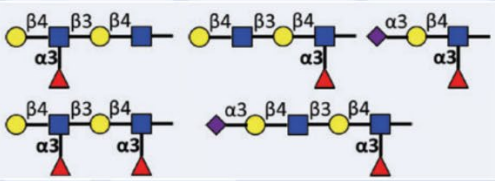
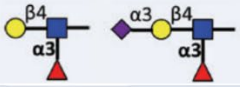
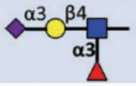
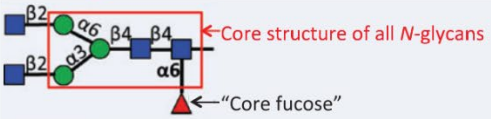
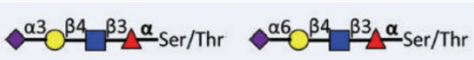
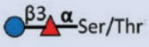
Common Name(s)	Abbreviation	Representative Major Product(s)
H blood group α2fucosyltransferase	[a]FUT1	
Secretor (Se) blood group α2fucosyltransferase	[a]FUT2	
Fuc-TIII α3/4fucosyltransferase Lewis blood group fucosyltransferase	[a]FUT3	
Fuc-TIV α3fucosyltransferase ELAM-1 ligand fucosyl transferase	[a]FUT4	
Fuc-TV α3fucosyltransferase	[a]FUT5	
Fuc-TVI α3fucosyltransferase	[a]FUT6	
Fuc-TVII α3fucosyltransferase	[a]FUT7	
Fuc-TVIII α6fucosyltransferase	[b]FUT8	
Fuc-TIX α3fucosyltransferase	[a]FUT9	
Fuc-TX α3fucosyltransferase	[c]FUT10	Unknown
Fuc-TXI α3fucosyltransferase	[c]FUT11	Unknown
Protein *O*-fucosyltransferase 1	[d]POFUT1 / FUT12	
Protein *O*-fucosyltransferase 2	[e]POFUT2 / FUT13	

Figure 1.2 List of 13 known human fucosyltransferases with respective acceptor specificity. [a]Fucosyltransferases that add fucose to glycan structures on *N*-glycans, mucin *O*-glycans, glycolipids, and (or) glycoRNAs. [b]Fucosyltransferase that only adds core fucose to *N*-glycans. [c]Fucosyltransferases with undefined acceptor specificity. [d]Modifications specific to EGF repeats with a POFUT1 O-fucose consensus sequence $C^2XXXX[\underline{S/T}]C^3$. [e]Modifications specific to TSRs with a POFUT2 O-fucose consensus sequence $C^{1/2}XX[\underline{S/T}]C^{2/3}$. Reproduced from (23) with permission.

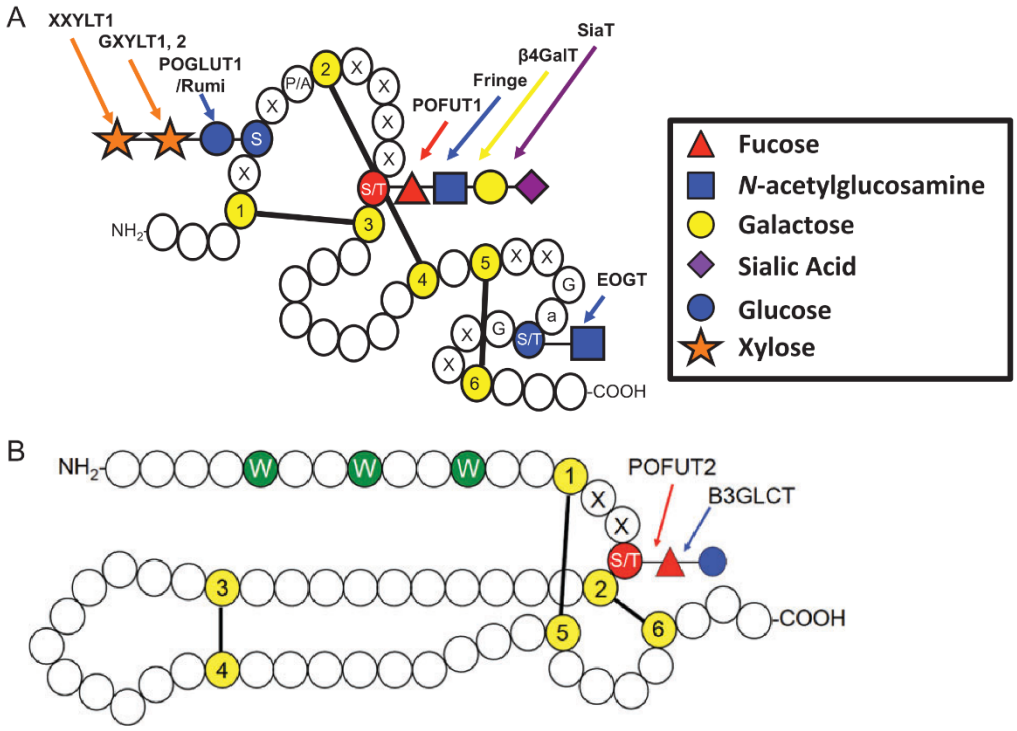


Figure 1.3 Cartoons showing the key features of EGF repeats and TSRs.
(A) A single EGF repeat with key *O*-glycan modifications. White circle, amino acids; yellow circle with numbers, numbered conserved cysteines linked with disulfide bonds; blue circle with S or S/T, *O*-glucose site for POGLUT1 modification and *O*-GlcNAc site for EOGT modification, respectively; red circle with S/T, *O*-fucose site for POFUT1 modification. Specific consensus sequences for POFUT1, POGLUT1, and EOGT are indicated. (B) A single TSR with *O*-fucose modification. Disulfide bonds are indicated. *C*-Mannose sites are shown in green circle with W, *O*-fucose site is shown in red circle with S/T. Specific consensus sequences for POFUT2 is indicated. Reproduced from (23) with permission.

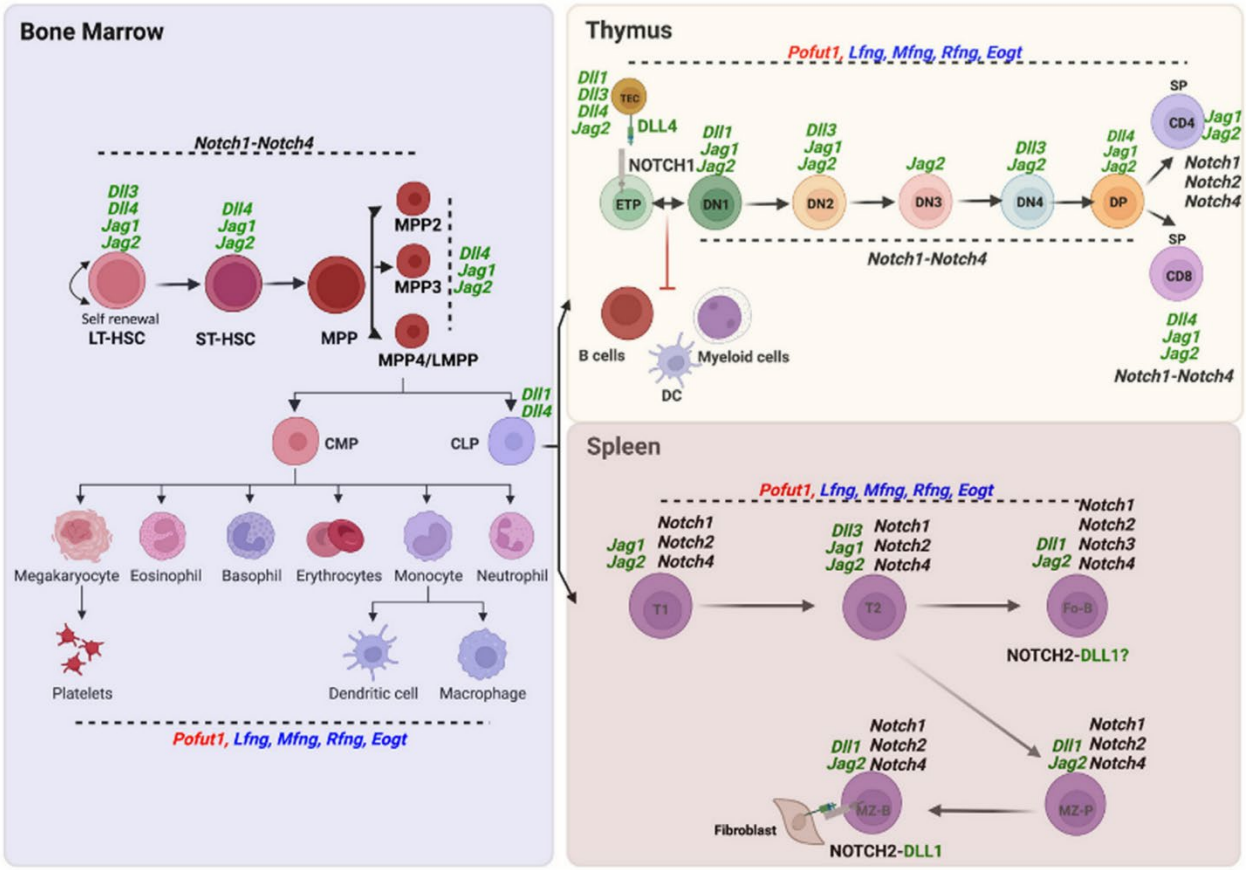


Figure 1.4 Notch receptors, Notch ligands and Notch-modifying O-glycosyltransferases in lymphopoiesis and myelopoiesis. The genes whose deletion results in developmental alterations in lymphopoiesis and myelopoiesis are shown in three hematopoietic compartments: bone marrow, thymus, and spleen. NOTCH1-DLL4 in thymus and NOTCH2-DLL1 in spleen are shown as proteins. Reproduced from Stanley, P. and Tanwar, A., 2022, CC-BY license (62).

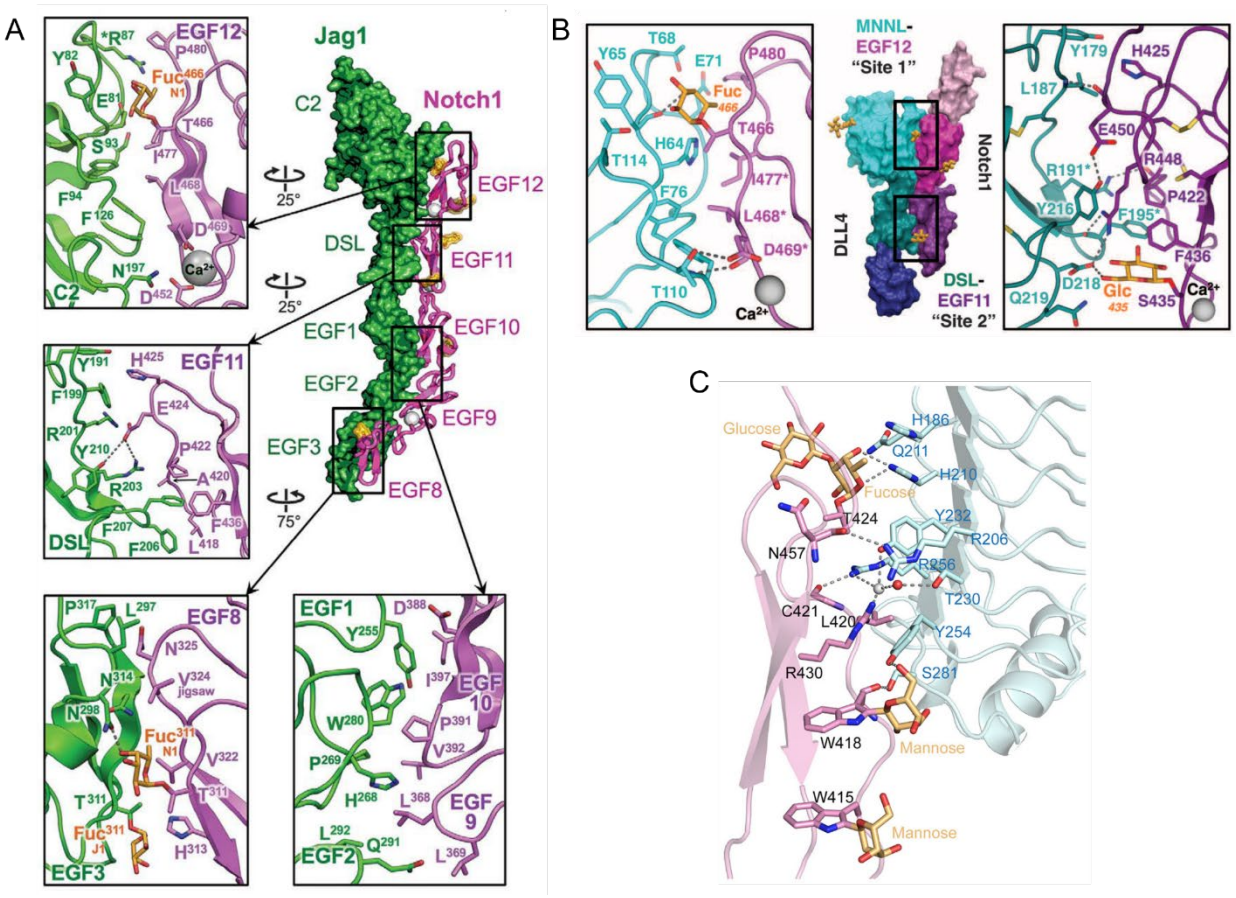


Figure 1.5 *O*-fucose residues form direct contacts with binding partners. (A) Co-crystal structure of NOTCH1 EGF8-12 and JAG1 N-EGF3. Reproduced from (68) with permission. (B) Co-crystal structure of NOTCH1 EGF11-13 and DLL4 N-EGF1. Reproduced from (67) with permission. (C) Co-crystal structure of RTN4R and BAI1. Reproduced from (125) with permission.

CHAPTER 2

FUT10 AND FUT11 ARE NOVEL PROTEIN *O*-FUCOSYLTRANSFERASES THAT MODIFY PROTEIN EMI DOMAINS

Hao H, Yuan Y, Ito A, Eberand BM, Cielesh M, Larance M, Haltiwanger RS. Manuscript in preparation. To be submitted to PNAS

2.1 Abstract

Domain-specific *O*-fucosylation plays critical roles in modulating the biological functions of numerous proteins, with the most prominent targets being the NOTCH receptor and ADAMTS family proteins. To date, only two types of domain-specific *O*-fucosylation have been characterized: Protein *O*-fucosyltransferase (POFUT) 1-mediated *O*-fucosylation of Epidermal Growth Factor-like (EGF) repeats and POFUT2-mediated *O*-fucosylation of Thrombospondin Type 1 Repeats (TSRs). Both POFUT1 and 2 require folded domain structures and a specific consensus sequence for modification. In a recent platelet proteome study by Houlahan et al., a novel domain-specific *O*-fucosylation was found on the Elastin Microfibril Interface (EMI) domain of Multimerin-1 (MMRN1), a major platelet protein that supports platelet adhesion and thrombus formation. Yet the enzyme responsible for this modification remains undefined. Here we show the identification of two novel protein *O*-fucosyltransferases, POFUT3 and POFUT4 (formerly FUT10 and FUT11, respectively) responsible for this modification. Of all the thirteen human fucosyltransferases, only POFUT3 and 4 demonstrated significant EMI binding in Alphafold2 screens. Robust *O*-fucosylation activity was observed for both POFUT3 and 4 in our *in vitro* assays using purified, recombinant POFUT3/4 and EMI substrates combined with mass spectrometric analysis. Much like POFUT1 and 2, POFUT3 and 4 function in the endoplasmic reticulum (ER) and require properly folded EMI structures for efficient modification. Eliminating EMI *O*-fucosylation by either mutating *O*-fucose sites or expression in *FX* KO cells lacking GDP-fucose synthesis reduced protein secretion. These data suggest that akin to POFUT1 and 2, the EMI-modifying POFUT3 and 4 participate in a non-canonical ER quality control pathway for EMI domains.

2.2 Introduction

Fucosylated glycan structures are frequently found on cell membranes and extracellular matrix proteins in mammals. They are involved in a broad range of physiological and pathological processes, such as selectin-mediated leukocyte adhesion, ABO blood group histocompatibility, cellular signal transduction, host-microbe interactions, and tumor metastasis (21-23, 141-143). Fucosyltransferases (FUTs) are the enzymes that transfer fucose residues from guanosine diphosphate (GDP)-fucose to acceptor molecules, including glycoproteins, glycolipids, and oligosaccharides (21). Thirteen *FUT* genes have been identified in the human genome. Based on the linkage of fucose addition, these FUTs are classified into α1-2, α1-3/4, α1-6 FUTs, and protein *O*-FUTs (POFUTs). α1-2 FUTs (CAZy GT11 family) includes FUT1, FUT2, and Sec1 (non-functional protein encoded by a pseudogene) (25, 26). FUT1 and FUT2 transfer fucose in α1-2 linkage to the terminal galactose of lactosamine to make the H antigens (27, 28). α1-3/4 FUTs (CAZy GT10 family) includes FUT3-7 and FUT9, all of which have α1-3 activity, except FUT3 and FUT5, which also possess α1-4 activity. α1-3/4 FUTs are involved in the synthesis of the complex series of Lewis antigen epitopes and have distinct acceptor specificity (29-34). Additionally, based on sequence homology to known α1-3/4 FUTs, FUT10 and FUT11 (CAZy GT10 family) were postulated to be α1-3 FUTs, but a comprehensive investigation is needed to elucidate their function and acceptor specificity (144-148). The only α1-6 FUT that identified in mammals is FUT8 (CAZy GT23 family). It catalyzes *N*-glycan core fucosylation that occurs at the innermost GlcNAc moiety of the chitobiose unit (36, 37).

In contrast to the Golgi-localized FUT1-9 that add fucose to glycan structures, POFUTs catalyze the direct transfer of fucose to proteins via *O*-linkage to the hydroxyl groups of Serines or Threonines in the endoplasmic reticulum (ER) (40, 41, 43). To date, only two POFUTs have

been identified in mammals, POFUT1 (CAZy GT65 family) and POFUT2 (CAZy GT68 family). Both enzymes require the recognition of a specific consensus sequence within a certain cysteine-rich domain for modification: POFUT1 mediates *O*-fucosylation of Epidermal Growth Factor-like (EGF) repeats containing the consensus sequence $C^2XXXX[S/T]C^3$ (where C^2 and C^3 are the second and third cysteine in the EGF) and POFUT2 adds *O*-fucose to Thrombospondin Type 1 Repeats (TSRs) with the consensus sequence $C^1XX[S/T]C^2$ in group 1 TSRs and $C^2XX[S/T]C^3$ in group 2 TSRs (41, 53). Both POFUT1 and POFUT2 are highly specific for their substrates due to the complementary features displayed between the interface of enzyme binding pockets and the three-dimensional (3D) fold of their respective substrates (72, 129). Disruption of *O*-fucosylation is associated with many human congenital disorders and tumorigenesis of various types of cancer (57, 100, 149). The Notch receptor family is the major biological target of POFUT1 (58, 60). Knockout of *Pofut1* in mice is embryonic lethal with multiple developmental defects similar to *Notch1* knockout (58). Members of the A-Disintegrin and Metalloproteinase with ThromboSpondin Type-1 motifs (ADAMTS) family constitutes almost half of the POFUT2 targets. Knockout of *Pofut2* in mice causes early embryo lethality with phenotypes similar to *Adamts9* mutants (79, 80), suggesting that *O*-fucosylation is essential for the biological function of these protein targets.

Recent evidence has demonstrated that *O*-fucose glycans modulate protein function through both intermolecular and intramolecular interactions (43). In co-crystal structures of portions of the NOTCH1 ligand-binding domain with its ligands [NOTCH1-DLL4 (67) and NOTCH1-JAG1 (68)], *O*-fucose residues on NOTCH1 EGF8 and 12 were found to directly interact with the ligands. Mutations of these *O*-fucose sites result in reduced NOTCH1-ligand binding and signaling (68, 70). In crystal structure of BAI1/RTN4-receptor complex, the *O*-fucosylation of

TSR3 on BAI1 directly interacts with the RTN4 receptor, and this interaction regulates neuronal development (125). Additionally, *O*-fucose forms intramolecular interactions that stabilize folded EGF repeats and TSRs (130, 131). This stabilizing effect drives the domain into an energy well, preventing the re-entry of the domain into the folding cycle and thereby aiding in protein folding and secretion. The requirement of both POFUT1 and POFUT2 to modify only properly folded substrates, along with the stabilizing effects of *O*-fucose, suggests that the ER-localized POFUTs may function in a non-canonical quality control pathway designed to ensure the efficient folding of proteins containing cysteine-rich domains such as EGF repeats and TSRs (119, 130).

In addition to EGF repeats and TSRs, a recent platelet proteome study identified a novel *O*-fucose site on the Elastin Microfibril Interface (EMI) domain of Multimerin-1 (MMRN1), a major platelet protein that supports platelet adhesion and thrombus formation (Figure 2.1A) (126). Similar to EGF repeats and TSRs, the EMI domain is also a highly conserved cysteine-rich domain with three disulfide bonds (Figure 2.1B). However, unlike EGF repeats and TSRs which are frequently found embedded within proteins as tandem repeats, the EMI domain is consistently singular and located at the N-terminus of proteins (150). Proteins containing the EMI domain constitute the EMI domain endowed (EDEN) superfamily (151). This superfamily can be subdivided into three groups, with the EMILIN/Multimerin family being the most numerous one, including EMILIN-1, EMILIN-2, EMILIN-3, Multimerin-1, and Multimerin-2. All EMILIN/Multimerin family members have the capability to form homotrimers (Figure 2.1C) and assemble into supramolecular multimers (152). The function of EMI domain remains largely unknown, but it is proposed to be involved in protein-protein interactions and participates in multimerization (150, 153-156). The recently identified EMI *O*-fucose is located within a

partially conserved motif ($C^1XXXX[S/T]X$) among all human EMI domains, suggesting a potential role for *O*-fucose in EMI domain functionality (Figure 2.1C).

As the EMI *O*-fucose locates within a domain context and motif distinct from the classic *O*-fucosylation by POFUT1 and POFUT2, we hypothesize that an as-yet-unidentified POFUT is responsible for this modification. Here we show that neither POFUT1, nor POFUT2, is responsible for the *O*-fucosylation of MMRN1 EMI domain. Instead, we identified FUT10 and FUT11 as the POFUTs that modify protein EMI domains *in vitro* and in cells; thereby, we renamed them POFUT3 and POFUT4, respectively. Much like POFUT1 and 2, both POFUT3 and 4 function in the ER and rely on properly folded EMI structures for efficient modification. Eliminating EMI *O*-fucosylation by either mutating the *O*-fucose sites or expression in *FX* KO cells lacking GDP-fucose synthesis led to a substantial reduction in protein secretion, suggesting that akin to POFUT1 and 2, POFUT3 and 4 participate in a non-canonical ER quality control pathway for EMI domains.

2.3 Results

Neither POFUT1 nor POFUT2 is responsible for addition of *O*-fucose in MMRN1 EMI domain

To confirm the novel *O*-fucose site, we performed transient transfections with a full-length human MMRN1 construct in HEK293T cells. Recombinant MMRN1 was purified from conditioned culture medium, digested with proteases, and analyzed by nano LC-MS/MS. Consistent with Houlahan et al., we observed high stoichiometry *O*-fucosylation at T216 on the N-terminal EMI domain (Figure 2.1D), as well as the classical *O*-fucosylation of the C-terminal EGF domain at T1055 (Dataset S1). Additionally, we identified a second *O*-fucose site on the EMI domain at T265, albeit with lower stoichiometry (Figure 2.1D and Dataset S1). Notably,

T216 and T265 are in close spatial proximity within the predicted 3D structure of MMRN1 EMI domain (Figure 2.1B), suggesting that *O*-fucosylation at T265 might occur as a sequential reaction following the T216 modification.

The full-length MMRN1 is poorly expressed in HEK293T cells, likely due to its strong tendency to form large-sized multimers via the interaction of EMI domain and C1q domain. We made a construct subcloned from MMRN1 with the first 290 amino acids covering the N-terminal EMI domain for better expression (Figure 2.1A). To investigate if POFUT1 or POFUT2 is the responsible enzyme for *O*-fucosylation of the EMI domain, we performed transient transfections with the N-terminal EMI construct in wild-type (WT), *POFUT1* KO (130), *POFUT2* KO (80), and *FX* KO HEK293T cells (157). *FX* encodes the enzyme required for GDP-fucose synthesis; knocking out FX eliminates all forms of fucosylation within cells. Recombinant N-terminal EMI was purified from conditioned culture medium, digested with proteases, and analyzed by nano LC-MS/MS. Extracted ion chromatograms (EICs) were generated for the *O*-fucosylated and unmodified glycoforms of peptides containing T216 or T265 site in each cell line. A lower *O*-fucosylation stoichiometry was observed in recombinant EMI compared to the full-length MMRN1, likely due to a shorter ER retention time for the smaller protein. Peptides containing T216 or T265 site were modified with *O*-fucose at similar stoichiometries in WT, *POFUT1* KO and *POFUT2* KO cells, but not in *FX* KO cells (Figure 2.2A and Dataset S1). Our positive controls showed loss of *O*-fucose on EGF2, 3 and 5 of mNOTCH1 EGF1-5 in *POFUT1* KO cells, and loss of *O*-fucose on TSR1, 2 and 3 of THBS1 TSR1-3 in *POFUT2* KO cells (Figure 2S1, A and B, and Dataset S1), confirming that the KO cells cause loss of *O*-fucose on known substrates of POFUT1 and POFUT2, respectively. The loss of *O*-fucose in all recombinant proteins in *FX* KO cells confirms that the modification on EMI domain is a fucose (Figure 2.2A,

Figure 2S1, A and B, and Dataset S1). Equivalent levels of EMI *O*-fucosylation were also observed in full-length MMRN1-transfected WT and *POFUT2* KO HEK293T cells (Figure 2S2, A and B, and Dataset S1). These data show that neither POFUT1 nor POFUT2 is responsible for addition of *O*-fucose in the EMI domain, strongly suggesting that a novel POFUT exists that modifies the EMI domain.

Additional evidence supporting this hypothesis was derived from metabolic labeling of 6-alkynyl fucose (6-AF). Our prior work demonstrated that POFUT1 and POFUT2 can efficiently incorporate 6-AF onto their respective substrates (138). Metabolic labeling experiments revealed robust incorporation of 6-AF by POFUT1 onto EGF2, 3, and 5 of mNOTCH1 EGF1-5, and by POFUT2 onto TSR2 and 3 of THBS1 TSR1-3 (Figure 2S1, C and D, and Dataset S1). However, no 6-AF incorporation was detected in EMI. In contrast, the complete absence of *O*-fucose and fucosylation of *N*-glycans on N-terminal EMI, confirming that neither POFUT1 nor POFUT2 are responsible for their modification (Figure 2.2B and Dataset S1). Loss of fucose on *N*-glycans is consistent with previous results showing that 6-AF depletes cellular GDP-fucose levels (140, 158).

FUT10 and FUT11 are POFUTs that add *O*-fucose to MMRN1 EMI domain

The only known FUTs without a known acceptor are FUT10 and 11 (23). To examine whether FUT10 and FUT11 can modify MMRN1 EMI domain, we performed enzymatic assays using purified, recombinant FUT10/11 with non-fucosylated EMI substrates. FUT10 and FUT11 were expressed and purified from HEK293F cells using FUT10/11 expression constructs that are designed for glycoenzyme secretion (Figure 2S3C) (159). Recombinant N-terminal EMI substrates were purified from HEK293F cells treated with 6-AF to remove all fucose glycans as

described above (Figure 2S3, A and B). EMI substrates were incubated with FUT10 or FUT11 in the presence of GDP-fucose for different durations at 37ºC. Reactions were quenched by precipitation with acetone, followed by digestion with protease and analysis using nano LC-MS/MS. EICs were generated and quantified to evaluate the *O*-fucosylation status of EMI T216 and T265 sites. We observed that both FUT10 and FUT11 can independently add fucose to EMI substrates, exhibiting time-dependent modification on both T216 and T265 sites (Figure 2.4, A and B, and Figure 2S4). Remarkably, FUT11 displayed higher efficiency in comparison to FUT10. FUT11-mediated reactions rapidly reached saturation within 2 hours for both T216 and T265 sites. In contrast, FUT10 displayed a slower reaction, gradually approaching saturation around 4 hours for T216 site, while maintaining a near-linear trend for T265 site (Figure 2.4B).

N-terminal EMI contains multiple *N*-glycans, accounting for ~20 kDa of its molecular size (Figure 2.1A and Figure 2S5). The EMI substrates used in enzymatic assays lack *N*-glycan fucosylation due to the 6-AF treatment, making them good substrates for *N*-glycan modifying FUTs (Figure 2.2B). Notably, no changes in *N*-glycan fucosylation on EMI were detected in the enzymatic assays mentioned above (Figure 2S6).

Both FUT10 and FUT11 apparently do not require any divalent metal ions for activity, as they were active when assayed without metal ions. However, their activities are significantly enhanced in the presence of divalent metal ions including $MnCl_2$, $MgCl_2$, and $CaCl_2$, where $MnCl_2$ showed the highest degree of activation (Figure 2S7). Interestingly, while EDTA does not affect FUT10 activity, it enhances FUT11 activity for reasons that remain unclear.

These findings clearly show that both FUT10 and FUT11 function as POFUTs, responsible for modifying EMI domains. We designate them as POFUT3 and POFUT4, respectively.

Kinetic analysis of POFUT3 and POFUT4

The GDP-Glo Glycosyltransferase assay is a common method for determining kinetic parameters of glycosyltransferases. Yet, we encountered a challenge in quenching the activity of POFUT4 without impacting the luciferases required for luminescence signal generation, possibly due to the strong affinity between POFUT4 with EMI, which enables the enzyme to remain active even in the presence of detergent. As a solution, we used mass spectrometric analysis to quantify the fucosylation stoichiometry and subsequently convert it into kinetic parameters. Considering the inherent ionization suppression effects of glycopeptides, the calculated parameters could potentially underestimate the actual values.

We performed kinetic analyses of POFUT3 and POFUT4 with varied concentrations of acceptor substrate (non-fucosylated N-terminal EMI). EICs were generated for the *O*-fucosylated and unmodified glycoforms of peptides containing T216 or T265 site for each sample. The relative abundance of fucosylation was quantified from EICs and then converted into product concentration, specifically EMI with transferred fucose. The kinetic parameters for the fucosylation at T216 and T265 sites were determined independently by fitting the curve into Michaelis-Menten equation. Similar kinetic curve patterns for both T216 and T265 sites were observed in both enzymes, suggesting a shared catalytic mechanism for modifying both sites (Figure 2.4A and Figure 2S8). Notably, the T216 site exhibits higher Km and Vmax values (Table 2.1). Given the *O*-fucose stoichiometry at these two sites on full-length MMRN1, EMI domain can accommodate two fucose residues simultaneously (Figure 2.1D). The differences in kinetic parameters may result from a sequential addition of fucose residues. The initial fucose presence on the substrate when adding the second can alter the catalytic environment. Surprisingly, the overall activity of POFUT3 is significantly higher compared to POFUT4, with

Km and Vmax values over five times greater than POFUT4 (T216 Km, 36.81 μM vs. 6.738 μM; T216 Vmax, 280.4 nmol/min/mg vs. 48.92 nmol/min/mg) (Figure 2.4A and Table 2.1). At lower EMI substrate concentrations (0-2 μM), POFUT4 exhibited higher activity than POFUT3, but the reaction rate decreased rapidly as the substrate concentration increased (Figure 2.4B). This explains the higher activity of POFUT4 over POFUT3 in our time-dependent enzymatic assay, as the enzyme and EMI substrate concentrations we used fall within this range (Figure 2.3, A and B). This is also consistent with our observation that POFUT4 shows a high affinity for the EMI domain, potentially resulting in a low turnover number at steady state. We additionally performed a GDP-Glo Glycosyltransferase assay for POFUT3 to validate the kinetic properties obtained via mass spectrometry. As expected, we observed a similar kinetic curve pattern with a comparable Km value but a higher Vmax value (Km, 47.22 μM; Vmax, 482.0 nmol/min/mg) (Figure 2S9 and Table 2S1).

Both POFUT3 and POFUT4 are independently capable of adding *O*-fucose to EMI domains in cells and are the sole enzymes responsible for this modification

Our *in vitro* data suggests that both POFUT3 and POFUT4 can add *O*-fucose to the EMI domain. To determine whether both enzymes participate in the *O*-fucosylation of MMRN1 EMI domain within cells, we generated CRISPR/Cas9-mediated knockouts of *POFUT3*, *POFUT4*, or both genes in HEK293T cells. Successful biallelic gene knockouts were confirmed by genomic DNA sequencing for all *POFUT3*, *POFUT4* single knockouts (KOs) and *POFUT3/4* double knockouts (DKOs). These cells were transiently transfected with the N-terminal EMI construct. Conditioned culture medium containing the secreted recombinant protein was collected, digested with proteases, and analyzed by nano LC-MS/MS. Knocking out *POFUT3* led to a ~60%

reduction in the *O*-fucose stoichiometry at the T216 site and a ~90% reduction at the T265 site, while knocking out POFUT4 led to a ~50% reduction at the T216 site and a ~60% reduction at the T265 site. Knocking out both *POFUT3* and *POFUT4* completely eliminated the *O*-fucosylation at both T216 and T265 sites (Figure 2.5, A and B, and Figure 2S10). Expression of either POFUT3 or POFUT4 in the DKO cells fully restored the *O*-fucosylation at both T216 and T265 sites and even drove the reaction to completion (Figure 2.5, C and D, and Figure 2S11). These findings indicate that both POFUT3 and POFUT4 are independently capable of adding *O*-fucose to EMI domains within cells and are the sole enzymes responsible for modifying EMI domains.

In contrast to the significant changed in *O*-fucosylation, there was no notable differences in the fucosylation of *N*-glycans on N-terminal EMI between WT and *POFUT3/4* DKO cells, even when co-transfected with POFUT3 or POFUT4, indicating that POFUT3 and 4 are not modifying *N*-glycans (Figure 2S12).

POFUT3 and POFUT4 only modify folded EMI structures and participate in a non-canonical ER quality control pathway for EMI domains

All EGF-modifying *O*-glycosyltransferases (POFUT1, POGLUT1, 2, and 3, EOGT) and the TSR-specific *O*-fucosyltransferase (POFUT2) have been shown to only recognize and modify properly folded structures in contrast to linear peptides (44, 52, 53, 160-162). Thus, we proposed that the EMI-modifying POFUT3 and POFUT4 may also share this requirement. To examine whether POFUT3 and POFUT4 only modify folded EMI structures, we performed assays with folded versus unfolded EMI domains, with POFUT2 and THBS1 TSR3 as positive controls (Figure 2S13). Recombinant non-fucosylated N-terminal EMI was denatured by reducing and

alkylating the disulfide bonds and used as the unfolded EMI substrates (Figure 2S13A). POFUT3 or POFUT4 enzymes were incubated with folded or unfolded EMI substrates at 37ºC for varying time periods. Reaction products were digested and analyzed by nano LC-MS/MS. The modifying efficiency of enzymes were monitored by quantification of *O*-fucosylation levels at EMI T216 and T265 sites. In line with our prior study, denaturation completely eliminates the ability of TSR3 to serve as a substrate for POFUT2 (Figure 2.6A). Interestingly, similar behavior for both POFUT3 and POFUT4 was observed. Compared to the folded EMI, the unfolded EMI was a poor substrate for both POFUT3 and POFUT4 (Figure 2.6A and Figure 2S14). This suggests that akin to other *O*-glycosyltransferases that modify EGF repeats or TSRs, POFUT3 and POFUT4 are capable of distinguishing between folded and unfolded EMI structures and require only the folded forms for modification.

POFUT1 and POFUT2 are localized to the ER and engage in a non-canonical ER quality control pathway for the folding and stabilization of EGF repeats and TSRs, respectively (119, 130, 131). In contrast, all other FUTs are localized to the Golgi (23). Previous proteomic studies of protein subcellular localization have indicated the ER localization of POFUT3 and POFUT4 in rat liver and U2OS cell line (163, 164). It is reasonable to propose that akin to POFUT1 and 2, POFUT3 and 4 could also be ER-localized and function in a quality control pathway for proteins containing EMI domains. To examine if POFUT3 and POFUT4 function in the Golgi, we performed transient transfections with the N-terminal EMI construct in WT and *SLC35C1* KO HEK293T cells (11). *SLC35C1* encodes the Golgi GDP-fucose transporter. Loss of SLC35C1 leads to the depletion of the Golgi GDP-fucose pool, thereby inhibiting the activity of Golgi-localized FUTs. As expected, nearly all fucosylation of *N*-glycan on EMI N136 site was lost in *SLC35C1* KO cells compared to WT cells, confirming the KO status of the cells (Figure 2.6B).

In contrast to *N*-glycan fucosylation, equivalent *O*-fucosylation levels on EMI T216 and T265 sites were observed in WT and *SLC35C1* KO cells. This suggests that POFUT3 and POFUT4 function in the ER rather than Golgi.

A known function of *O*-fucose glycans is to aid in protein folding and promote secretion (119, 130, 131). Mutating the T216 *O*-fucose site is known to reduce MMRN1 secretion (126). To examine the effects of the T265 *O*-fucose site, we mutated either T265 site alone or both T216 and T265 sites on full-length MMRN1 construct (Figure 2S15A). The distinct effects of individual *O*-fucose on MMRN1 secretion were compared. Consistent with Houlahan et al., the T216A mutation showed a ~50% reduction in MMRN1 secretion, and the T1055A mutation showed a complete loss of secretion. Mutating both T216 and T265 on EMI domain led to an over 50% reduction in MMRN1 secretion, while mutating T265 alone resulted in a ~15% reduction (Figure 2.6, C and D, and Figure 2S15B). This indicates that both T216 and T265 *O*-fucose glycans play a role in MMRN1 secretion, with T216 having a more significant effect.

We also performed secretion assays in *FX* KO cells, which lack GDP-fucose synthesis (157). To eliminate the impact of *O*-fucosylation on the EGF repeat, we used the N-terminal EMI instead of full-length MMRN1. In line with our previous studies, eliminating *O*-fucosylation of EGF repeats or TSRs by expression in *FX* KO cells led to a near complete loss in the secretion of NOTCH1 EGF1-18 and AdamTS9 TSR2-8 (Figure 2.6, E-H, and Figure 2S16) (121, 130). We observed a substantial ~70% decrease in the secretion of N-terminal EMI in *FX* KO cells compared to WT cells, highlighting the role of EMI *O*-fucosylation in its secretion, similar to its contributions to proteins containing EGF repeats or TSRs. Notably, a substantial ~70% reduction in N-terminal EMI protein levels was observed in the lysate of *FX* KO cells, while NOTCH1

EGF1-18 and AdamTS9 TSR2-8 mostly remained present or accumulated within the cells. This indicates that EMI undergoes degradation when there is a fucosylation defect.

2.4 Discussion

Domain-specific *O*-fucosylation plays critical roles in modulating the biological functions of numerous proteins, with the most prominent targets being the NOTCH receptor and ADAMTS family proteins. In addition to EGF repeats and TSRs that can be modified with *O*-fucose, a new domain-specific *O*-fucosylation was recently discovered within the EMI domain (126). In the present study, we report POFUT3 and POFUT4 (formerly FUT10 and FUT11, respectively) are the enzymes responsible for this modification. Both enzymes exhibited strong *O*-fucose transfer activity both *in vitro* and within cells. Similar to POFUT1 and POFUT2, POFUT3 and POFUT4 rely on properly folded EMI structures for efficient modification. Finally, our findings indicate that POFUT3 and POFUT4 localized in the ER and participate in a non-canonical ER quality control pathway for EMI domains.

FUT10 and 11 have been annotated as a1,3-fucosyltransferases (see Q6P4F1 and Q495W5 in UniProt). They were originally identified in the human genome for sharing sequence homology with Drosophila *FucTB* (148). The function and acceptor substrate for the fly FucTB has not yet been defined. It is a homolog of FucTA which has well-established core a1-3 FUT activity, but such activity was not detected for FucTB (165, 166). Both FUT10 and FUT11 belong to the CAZy GT-10 family, along with the invertebrate and plant core a1-3 FUTs and all terminal Golgi a1-3/4 FUTs (26). The Golgi a1-3/4 FUTs (FUT3-7 and FUT9) share five conserved peptide motifs (145). Motif VI and V are involved in the recognition and binding of GDP-fucose donor, and they are well-conserved in FUT10 and FUT11. Motif I-III are involved in the recognition of acceptor substrate, whereas FUT10 and FUT11 show limited sequence homology

with other a1-3/4 FUTs in these motifs. Especially motif II, instead of [F/I/V]HH[R/W][E/D] that present in all a1-3/4 FUTs for lactosamine acceptor recognition, FUT10 and FUT11 present a highly conserved FYGTD in the equivalent position, indicating that they have a distinct acceptor substrate (145). Numerous studies have attempted to explore the acceptor specificity of FUT10 and FUT11. Mollicone et al. reported core a1-3 FUT activity for FUT10 and FUT11 using crude lysates of FUT10/11 transfected COS7 as enzyme sources and tested with a library of synthetic glycan acceptors through a radioactive-based assay (145). Kumar et al. reported that FUT10 is involved in an a1-3 FUT activity for the synthesis of Lewis X (Le^X) epitope (144). They used crude lysates of COS1 cells that overexpress FUT10 as enzyme sources. Although no activity was detected with synthetic glycan substrates, an increase in Le^X-containing *N*-glycans was detected in HPLC analysis when crude lysates of Neuro2a cells were used as glycoprotein substrates. While it is conceivable that FUT10/11 may add a small amount of fucose to *N*-glycans when tested with large quantities of glycan substrates, it is more likely that the observed activity in crude lysates indirectly results from protein targets modified by FUT10/11. Supporting this, we did not detect any fucose addition to *N*-glycans on the N-terminal EMI in our *in vitro* enzymatic assay (Figure 2S6), while there was a robust addition of *O*-fucose (Figure 2.3). Moreover, there were no significant changes in *N*-glycan fucosylation when *POFUT3* and *POFUT4* were knocked out in cells (Figure 2S12). In contrast to the monoexonic a1-3/4 FUTs, the genomic structure of both FUT10 and FUT11 is polyexonic. Phylogenetic analysis indicates that FUT10 and FUT11 are ancient enzymes that originated at about 830 MYA (Million Years Ago), clearly distinct from the a1-3/4 FUTs (about 450 MYA) (25, 145). POFUT1 and POFUT2 are also ancient enzymes (about >1000 MYA) and have a polyexonic structure (167). These

clues reinforce our conclusion that FUT10 and FUT11 (POFUT3 and POFUT4, respectively) function as POFUTs, rather than a1-3/4 FUTs.

Similar to POFUT1 and 2, POFUT3 and 4 add fucose to MMRN1 EMI domain within a motif $C^{1}XXXX[S/T]X$ that is highly conserved across EMI domains (Figure 2.1C). Interestingly, a second fucose was found in close proximity to the primary fucose site in space, and many other EMI domains also possess a threonine at this equivalent site (Figure 2.1, B and C). Examining other EMI domain-containing proteins with this motif, such as EMILIN1, multimerin-2 (MMRN2), and EMI domain-containing protein 1 (EMID1), would provide more information and refine the precision of the consensus sequence. Based on the fucosylation stoichiometry of the two sites, it appears that EMI can accommodate two fucose residues simultaneously (Figure 2.1D). It would be intriguing to understand the spatial orientation of these two fucose residues and the catalytic mechanisms involved in their addition to EMI.

Finally, we demonstrated that POFUT3 and POFUT4 function in the ER, require folded EMI structures for efficient modification, and assist in protein secretion. This evidence indicates that these enzymes play a role in the quality control of EMI domains, ensuring they are properly folded for secretion. *O*-fucose glycans are known to form intramolecular interactions that stabilize EGF repeats and TSRs, facilitating protein folding and secretion by preventing the re-entry of these domains into the folding cycle. Further work will be needed to verify if the two fucose residues added by POFUT3/4 have a similar effect on EMI domains. Furthermore, *O*-fucose glycans can form direct intermolecular interactions to facilitate protein-protein interactions, regulating processes like Notch-ligand binding. In the context of MMRN1, it is proposed that the N-terminal EMI domain interacts with the C-terminal C1q domain for multimerization, indicating a potential role for *O*-fucosylation in these processes (150).

We are just beginning to understand the biological function of POFUT3/4 and the proteins they modified. POFUT3 and 4 were reported to have important functions in vertebrate development (147). They are expressed ubiquitously in mouse embryos throughout development and display distinct expression profiles. Knocking down POFUT4 in zebrafish embryos resulted in malformations of the posterior trunk and tail. POFUT3 has also been reported to be required for the maintenance of mouse embryonic stem (ES) cells and neural stem cells (144). Identifying the protein targets of POFUT3/4 that are involved in regulating these biological processes would provide a full understanding of the biological role of POFUT3/4 in embryonic development and stem cell maintenance.

2.5 Experimental Procedures

Cell culture

HEK293T WT, *POFUT1* KO, *POFUT2* KO, *POFUT3* KO, *POFUT4* KO, *POFUT3/4* DKO, *FX* KO and *SLC35C1* KO cells were cultured at 37°C with 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, GE Healthcare Life Sciences) supplemented with 10% (v/v) bovine calf serum (BCS, VWR), 100 units/mL penicillin, and 100 µg/mL streptomycin (Lonza). HEK293F cells were cultured in Freestyle293^{TM} medium (Gibco) with 100 units/mL penicillin and 100 µg/mL streptomycin. HEK293T cells were purchased from ATCC. HEK293F cells were generously provided by Dr. Kelley Moremen at the University of Georgia.

Plasmids and mutagenesis

The mammalian expression plasmids for MMRN1 analysis including pcDNA3.1-hMMRN1 WT-Myc-His_6, pcDNA3.1-hMMRN1 T216A-Myc-His_6, and pcDNA3.1-hMMRN1 T1055A-Myc-

His_6 were described previously (126). For the generation of pcDNA3.1-hMMRN1 T265A-Myc-His_6, and pcDNA3.1-hMMRN1 T216A/T265A-Myc-His_6, the T265A mutation was introduced by standard PCR-based mutagenesis using CloneAmp HiFi PCR Premix (Takara Bio Inc.) with mutagenic primers AI452 and AI464 using the parental plasmid cDNA3.1-hMMRN1 WT-Myc-His_6 (for T265A mutatant) or pcDNA3.1-hMMRN1 T216A-Myc-His_6 (for T216A/T265A double mutant) as template. PCR products were treated with DpnI to remove the parental plasmid before transformed into DH5α-competent cells (Invitrogen).

pcDNA4-N-terminal EMI-Myc-His_6 encoding the first 290 amino acids covering the N-terminal EMI domain of hMMRN1 was subcloned from pcDNA3.1-hMMRN1 WT-Myc-His_6 using CloneAmp HiFi PCR Premix with primers AI364 and AI365. PCR products were ligated to pcDNA4/TO/Myc-His_6 expression vector using T4 Polynucleotide Ligase (NEB).

pSecTag2-mNOTCH1 EGF1-5-Myc-His_6, pSecTag2-mNOTCH1 EGF1-18-Myc-His_6, pSecTag2-hTHBS1 TSR1-3-Myc-His_6 and pSecTag2-hAdamTS9 TSR2-8-Myc-His_6 plasmids were described previously (55, 121, 138). pGEn2-GFP, pGEn2-GFP-hPOFUT2, pGEn2-GFP-hFUT10 and pGEn2-GFP-hFUT11 plasmids were described previously and generously provided by Dr. Kelly Moremen (159).

For generating pcDNA4-full length hFUT10 (POFUT3)-Myc-His_6 for the rescue assays: the 5'-signal peptide portion (99 bp) of hFUT10 was amplified by PCR using Platinum SuperFi II DNA Polymerase (Thermo Fisher) with primers AI456 and AI457 using pUC57-hFUT10 5' side (made by gene synthesis, Genscript) as template. The hFUT10 coding region was amplified by PCR using CloneAmp HiFi PCR Premix with primers AI458 and AI459 using pGEn2-GFP-hFUT10 as template. The backbone vector portion was amplified by PCR using Platinum SuperFi II DNA Polymerase with primers AI454 and AI455 using pcDNA4/TO/Myc-His_6

expression vector as template. Infusion reaction was performed using In-Fusion HD Cloning Kits (TAKARA).

For generating pcDNA4-full length hFUT11 (POFUT4)-Myc-His_6 for the rescue assays: the 5'-signal peptide portion (93 bp) of hFUT11 was amplified by PCR using Platinum SuperFi II DNA Polymerase (Thermo Fisher) with primers AI460 and AI461 using pUC57-hFUT11 5' side (made by gene synthesis, Genscript) as template. The hFUT11 coding region was amplified by PCR using CloneAmp HiFi PCR Premix with primers AI462 and AI463 using pGEn2-GFP-hFUT11 as template. The backbone vector portion was amplified by PCR using Platinum SuperFi II DNA Polymerase with primers AL454 and AL455 using pcDNA4/TO/Myc-His_6 expression vector as template.

All plasmids were confirmed by sequencing. Sequences of all primers are listed in *SI Appendix*, Table S2.

Production of proteins used for mass spectral analysis of *O*-fucosylation

HEK293T WT, *POFUT1* KO, *POFUT2* KO, *FX* KO or *SLC35C1* KO cells (6 x 10^6 cells per plate) were seeded in 10-cm dishes in 8 mL DMEM with 10% bovine calf serum and 1% penicillin/streptomycin. Cells were cultured overnight to reach a confluency of 70%. Each plate of cells was transiently transfected with 10 µg of pcDNA3.1-hMMRN1-Myc-His_6, 5 µg of pcDNA4-N-terminal EMI-Myc-His_6, 10 µg of pSec-mNOTCH1 EGF1-5-Myc-His_6, or 10 µg of pSec-THBS1 TSR1-3-Myc-His_6 in 8 mL Opti-MEM (Thermo Fisher, 31985088) using 60 µL PEI (1 µg/µL stock). For metabolic labeling of cells with 6-alkynyl fucose (6-AF, generously provided by Dr. Peng Wu at the Scripps Research Institute), cells were transfected in 8 mL Opti-MEM containing 100 µM 6-AF or equal volume of DMSO. Two days later, media from 3-8

plates of cells was combined. Recombinant proteins were purified using Ni-NTA (Qiagen) affinity chromatography with 600 µL agarose and eluted with 600 µL of 250 mM imidazole in Tris buffered saline, pH 7.5 (TBS). Eluted purified proteins were stored at -20°C until use.

For testing the *POFUT3/4* KO cells, HEK293T WT, *POFUT3* KO, *POFUT4* KO or *POFUT3/4* DKO cells (1 x 10^6 cells per well) were seeded in 6-well plates and cultured overnight for attachment. Cells were transiently transfected with 1 µg of pcDNA4-N-terminal EMI-Myc-His_6 in 1.2 mL Opti-MEM using 6 µL PEI (1 µg/µL stock). Two days later, media was collected and stored at -20°C until use. For POFUT3/4 rescue assay coupled with mass spectrometry analysis, HEK293T WT and *POFUT3/4* DKO cells (1 x 10^6 cells per well) were seeded in 6-wells plates followed by transiently transfected with 1 µg of pcDNA4-N-terminal EMI-Myc-His_6 with 0.2 µg of pcDNA4-full length hPOFUT3-Myc-His_6, pcDNA4-full length hPOFUT4-Myc-His_6, or empty vector in 1.2 mL Opti-MEM using 7.2 µL PEI (1 µg/µL stock). Cells were cultured for two days. Culture media was collected and stored at -20°C until use.

Glycoproteomic site mapping of *O*-fucosylated peptides

Purified proteins (300-600 µL of elution of Ni-NTA purification) or 300 µL of conditioned culture medium were precipitated with 3X volumes of cold acetone overnight at -20°C. After centrifuging at 12,700 rpm for 15 minutes at 4°C, the pellets were denatured and reduced using 50 µL of reduction buffer containing 8 M urea, 0.4 M ammonium bicarbonate and 10 mM TCEP, incubated at 60°C for 10 minutes. Alkylation was performed by adding 25 µL of 100 mM iodoacetamide in 50 mM Tris, pH 8, and incubating in the dark for 40 minutes at room temperature. Samples were then diluted with 225 µL of water and digested with 1 µg of trypsin (Thermo Fisher, 90057) in 37°C water bath overnight. For THBS1 TSR1-3 samples, 1 µg of

chymotrypsin (Thermo Fisher, 90056) was added after trypsin digestion, and samples were incubated for 3 hours at 37°C water bath. Samples were then acidified with 5% formic acid (FA), sonicated for 20 minutes, and desalted with a C18 ZipTip (Millipore, ZTC18S960). Samples [1-10 µL, ~5 ng peptides per run in 10% acetonitrile (ACN) and 0.1% aqueous FA] were analyzed by nano LC-MS/MS using an EASY-nLC 1200 System (Thermo Fisher) coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher). Peptides were loaded via autosampler and preconcentrated onto an Acclaim PepMap-100 75µm X 2 cm nanoViper C18 pre-column with loading buffer, 5% ACN and 0.1% aqueous FA (Solvent A). Peptides were gradient eluted onto a C18 EasySpray PepMap RSLC C18 analytical column (50 µm X 15 cm, Thermo Fisher) at a constant flow rate of 300 nL/min using a 30 minute gradient and a 61 minute instrument method (for purified protein samples) or 90 minute gradient and a 120 minute instrument method (for whole-medium digested samples) method. The gradient profile was as follows [min:% solvent B (80% ACN and 0.1% aqueous FA)]: 30 minute gradient 0:0, 25:50, 28:98, 31:98, 34:2, 37:2, 40:98, 43:98, 46:2, 49:2, 52:98, 55:98, 58:2, 61:2; 90-minute gradient 0:0, 90:50, 93:98, 96:98, 99:2, 102:2, 105:98, 108:98, 111:2, 114:2, 117:98, 120:98. The instrument method used an MS1 resolution of 70,000, an AGC target of 1e6, and a mass range from 400 to 2,000 m/z. Dynamic exclusion was enabled for 6 s. Only charge states 2-6 were permitted for fragmentation. MS2 scans were acquired in the Orbitrap at a resolution of 17,500, an isolation window of 1.2 m/z, and an AGC target of 1e5. The HCD fragmentation was set with fixed collision energy of 27%. Data analysis was performed with Byonic (Protein Metrics, v4.1.10). Search parameters included fully specific cleavage specificity at the C-terminal site of R and K for trypsin with two missed cleavages allowed, C-terminal site of R, K, F, W, Y and L for trypsin/chymotrypsin double digestion with five missed cleavages allowed. Mass tolerance was set at 10 ppm for precursors

and 0.1 Dalton for fragments. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation (common 1), asparagine oxidation (common 1), asparagine deamidation (common 1), N-terminal acetylation (rare 1), and tryptophan hexosylation (common 2) were set as variable modifications with a total common max of 3, rare max of 2. Glycans were set as variable modifications (common1) using a customed *O*-glycan search space including Fuc(1), HexNAc(1)Fuc(1), HexNAc(1)Hex(1)Fuc(1), HexNAc(1)Hex(1)Fuc(1)NeuAc(1), Hex(1)Fuc(1), Hex(1), Hex(1)Pent(1), Hex(1)Pent(2), and HexNAc(1). For 6-AF metabolic labeled samples, 6-AF incorporated glycopeptides were searched with an additional 9.9800 m/z to account for the chemical modification. Peptides with *N*-glycans were searched with the internal *N*-glycan search list of Byonic (52 common biantennary). To make the extracted ion chromatograms (EICs) for a given peptide, the ions of each glycoform were extracted from the MS1 spectrum using their respective m/z (observed m/z from the peptide with the highest Byonic score, mass tolerance: ± 0.005 for purified protein samples, ± 0.003 for whole-medium digested samples), then overlaid to compare the relative ion intensity. The EICs were smoothed using Gauss algorithm. All EICs were manually generated and analyzed using Thermo Xcalibur software.

POFUT2, 3 and 4, and N-terminal EMI, expression and purification

Recombinant POFUT2, POFUT3 and POFUT4 (GFP tagged, lacking transmembrane domain) were expressed and purified from HEK293F cells. HEK293F cells were maintained in Freestyle293^{TM} medium to a density of 3 x 10^6 cells/mL. For transfection, 200 x 10^6 of cells were calculated and centrifuged at 800 rpm for 3 minutes. Cell pellets were resuspended in 50 mL of 9:1 medium (Freestyle293^{TM} medium: EX-Cell® medium (Sigma Aldrich), v/v), and

transiently transfected with 200 μg of pGEn2-GFP-hPOFUT2, pGEn2-GFP-hFUT10, pGEn2-GFP-hFUT11, or pGEn2-GFP plasmids (4 μg plasmid/mL) with 450 μL PEI (1 μg/μL stock). After 24 hours, Valproic acid (Sigma Aldrich) was supplemented to cells in 50 mL of 9:1 medium to a final concentration of 2.2 mM. 3 days later, culture medium was collected and filtered through 0.45 μm filters (Millipore).

For generating non-fucosylated EMI substrates for enzymatic assays, 200 x 10^6 of HEK293F cells were resuspended in 50 mL of 9:1 medium containing 200 μM 6-AF, and transiently transfected with 200 μg of pcDNA4-N-terminal EMI-Myc-His_6 plasmid (4 μg plasmid/mL) with 450 μL PEI (1 μg/μL stock). After 24 hours, Valproic acid was diluted to 2.2 mM in 50 mL of 9:1 medium containing 200 μM 6-AF and added to transfected cells. After 3 days, culture medium was collected and filtered through 0.45 μm filters.

Recombinant proteins were purified using Ni-NTA affinity chromatography with 1 mL Ni-NTA agarose and eluted with 3 mL of 250 mM imidazole in Tris buffered saline, pH 7.5 (TBS). Eluted proteins were buffer exchanged to 10% glycerol in 50 mM HEPES, pH 7 with 10 kDa molecular weight cutoff centrifugal filters (Amicon) using manufacturer's recommended procedures and stored at -80°C until use. Protein concentration was measured with NanoDrop (Thermo Fisher). The extinction coefficient values for individual proteins were calculated based on their respective sequences on ProtParam tool in ExPASy (168). The purity of proteins was verified by SDS-PAGE and Coomassie blue staining.

Enzymatic assay for POFUT activity

For acquiring the time-dependent EMI modifying profiles for POFUT3 and POFUT4, 0.1 μM of recombinant POFUT3, POFUT4, or GFP was incubated in 50 μL reaction mixtures containing

100 μM GDP-fucose, 0.5 μM purified non-fucosylated N-terminal EMI, and 50 mM HEPES, pH 7. Reactions were incubated at 37°C for indicated times. For acquiring the substrate concentration-dependent kinetics of POFUT3 and POFUT4, 50 nM of recombinant POFUT3 or POFUT4 was incubated in 30 μL reaction mixtures containing 100 μM Ultra-pure GDP-fucose (Promega), 0.3 mM $MnCl_2$ with 0.156 μM, 0.313 μM, 0.625 μM, 1.25 μM, 2.5 μM, 5 μM, 10 μM, 20 μM or 30 μM of purified non-fucosylated N-terminal EMI in 50 mM HEPES, pH 7. Reactions were incubated at 37°C for 15 min and stopped by adding 500 μL of cold acetone. For negative control samples, reaction mixtures containing 1.25 μM EMI substrate but without enzyme were precipitated with acetone first, followed by adding POFUT3 or POFUT4 to a final concentration of 50 nM. The acetone-precipitated proteins were reduced, alkylated, digested with trypsin, and analyzed with nano LC-MS/MS using the procedures described earlier. Data analysis with Byonic, EIC generation and quantification were performed as described earlier. The conversion of the relative abundance of fucosylation to product concentrations was performed using the following equation:

Product concentration = (Fuc%-Fuc% $_{background}$) X EMI substrate concentration (Fuc%: relative abundance of fucosylation; Fuc% $_{background}$: relative abundance of fucosylation in negative control samples).

The kinetic curve was plotted using product formation rate per enzyme amount (nmol/min/mg) with EMI substrate concentration (μM). The kinetic parameters were determined with nonlinear Michaelis-Menten least squares (ordinary) fitting in Prism 7. All reactions were performed in biological triplicates with three batches of purified enzymes.

EMI unfolding assays

20 μg of purified recombinant unmodified N-terminal EMI (prepared as described above) or THBS1 TSR3 (bacterially expressed) was denatured in 100 μL of buffer containing 0.4 M ammonium bicarbonate, 8 M urea with or without 5 mM TCEP (Thermo Fisher) and incubated at 50°C for 15 minutes. After cooling down to room temperature, 20 μL of 200 mM iodoacetamide or water was added. Samples were incubated at room temperature for 40 minutes in dark. Samples were then buffer exchanged to 100 μL of 10% glycerol in 50 mM HEPES, pH 7 using 10 kDa molecular weight cutoff Zeba spin column (Thermo Fisher) according to manufacturer's recommended procedures. Protein concentrations were measured by Nano-drop with extinction coefficient number of 10.6% (calculated with ProtParam tool in ExPASy using the protein sequence of N-terminal EMI). Proteins were stored at -80°C until use. Unmodified human THBS-1 TSR3 was expressed and purified from *Escherichia coli* BL21 strain (Invitrogen) as previously described (44). Unfolded and folded proteins were verified by SDS-PAGE and Coomassie blue staining.

Enzymatic assays with folded or unfolded substrates were performed in 50 μL reaction containing 0.1 μM of purified enzyme (recombinant POFUT3/4 or POFUT2), 0.5 μM folded or unfolded substrate (N-terminal EMI or THBS1 TSR3), 100 μM GDP-fucose and 50 mM HEPES, pH 7. Reactions were incubated at 37°C for 0, 1, 2, or 4 hours before stopped with 500 μL of cold acetone. Reaction products were reduced, alkylated, digested with trypsin (for N-terminal EMI) or double-digested with trypsin/chymotrypsin (for THBS1 TSR3), and analyzed with nano LC-MS/MS using the procedures described earlier. All reactions were performed in biological triplicates with three batches of purified enzymes.

Western blot-based secretion assay

HEK293T WT or *FX* KO cells (1 x 10^6 cells per well) were seeded in 6-well dishes and incubated overnight to reach a confluency of 80%. Medium was changed to 1 mL Opti-MEM before transfection. Transiently transfection was performed using PEI (6 µL PEI per 1 µg plasmid) with 2 µg/well pcDNA3.1-hMMRN1 WT-Myc-His_6, 2 µg/well pcDNA3.1-hMMRN1 T216A-Myc-His_6, 2 µg/well pcDNA3.1-hMMRN1 T265A-Myc-His_6, 2 µg/well pcDNA3.1-hMMRN1 T216A/T265A-Myc-His_6, 2 µg/well pcDNA3.1-hMMRN1 T1055A-Myc-His_6, 0.5 µg/well pcDNA4-N-terminal EMI-Myc-His_6, 1 µg/well pSec-mNOTCH1 EGF1-18-Myc-His_6, 1 µg/well pSec-hAdamTS9 TSR2-8-Myc-His_6, or empty vector, together with 0.1 µg/well IgG plasmid as secretion control. Cells were incubated for 48 hours before collecting medium samples. 100 µL of medium were precipitated with 500 µL of cold acetone incubating overnight at -20°C. After centrifuging at 12,700 rpm for 15 minutes at 4°C, pellets were resuspended in 12 µL of 2X denaturing sample buffer (contains 0.04% SDS, 200 mM 2-mercaptoethanol, 20% glycerol in 100 mM Tris/HCl, pH 6.8) by sonication for 10 minutes and boiling at 105°C for 5 minutes. For collecting lysate samples, transfected cells were freeze-thawed with one cycle at -20°C to break down long DNA chains. Cell pellets were lysed with 200 µL of RIPA buffer (Thermo Fisher, 89900), incubated on ice for 15 minutes and centrifuged at 12,700 rpm for 5 minutes at 4°C. Cleared supernatants were collected. 10 µL of cell lysate was used for analysis. Samples were loaded onto 4-20% SDS-PAGE (Bio-Rad), transferred to a nitrocellulose membrane. Membranes were blocked with 5% nonfat milk (Bio-Rad) for 30 minutes at room temperature, followed by incubation with anti-Myc antibody (Clone 9E10, Invitrogen, 1:2500) at 4°C overnight. Membranes were then incubated with IDRye 800-conjugated goat anti-mouse IgG antibody (LI-COR, 1:2500) and IDRye 680-conjugated goat anti-human IgG antibody (LI-

COR, 1:2500) for 1 hour at room temperature. The Western blot bands were visualized and quantified using Odyssey System (LI-COR).

CRISPR /Cas9-mediated genome editing

CRISPR-Cas9 HEK293T knockouts of *FX*, *SLC35C1*, *POFUT3*, *POFUT4*, and double knockout of *POFUT3/4* were generated by Dr. Youxi Yuan.

Analysis of *N*-glycans on N-terminal EMI by PNGase F digestion

6 X 10^6 of HEK293T cells were seeded and grown to 70% confluence in 10-cm dishes. Cells were transiently transfected with 5 μg of pcDNA4-N-terminal EMI-Myc-His_6 or 2 μg of pSecTag2-mLfng-Myc-His_6 using PEI (6 μL per 1 μg of DNA, 1 μg/μL stock) in 8 mL of Opti-MEM (Thermo Fisher, 31985088). After 2 days, cell medium from 3-5 plates of cells was combined. Recombinant proteins were purified using Ni-NTA (Qiagen) affinity chromatography with 600 μL agarose and eluted with 600 μL of 250 mM imidazole in Tris buffered saline, pH 7.5 (TBS). 50 μL of Ni-NTA elution was precipitated with 250 μL of cold acetone and incubated at -20ºC for 2 hours. Samples were centrifuged at 12,700 rpm for 15 minutes at 4ºC. Protein pellets were resolubilized with 20 μL of reduction buffer containing 1% SDS, 10 mM TCEP, and 0.4 M NH_4HCO_3 with vigorous shaking. Proteins were boiled at 105ºC for 5 minutes and cooled to room temperature. 10 μL of 100 mM Iodoacetamide (Sigma Aldrich) was added and samples were incubated in dark for 40 minutes. 170 μL of 1% NP-40/TBS was added to dilute SDS to a final concentration of 0.1%. 1.5 μL of PNGase F (20 units/μL, Lectez Bio) was added to samples and incubated at 37ºC for over 6 hours. Samples were precipitated with 1 mL of cold acetone and incubated at -20ºC for 2 hours. Protein pellets were obtained by centrifuging at 12,700 rpm for

15 minutes at 4ºC, followed by resolubilizing with 12 µL of 2X denaturing sample buffer (contains 0.04% SDS, 200 mM 2-mercaptoethanol, 20% glycerol in 100 mM Tris/HCl, pH 6.8) by sonication for 10 minutes and boiling at 105°C for 5 minutes. Proteins were analyzed with Western Blot probed with anti-Myc antibody (Clone 9E10, Invitrogen, 1:2500) as previously described.

Metal ion screening assay

0.1 µM of recombinant POFUT3 or POFUT4 was incubated in 50 µL reaction mixtures containing 100 µM GDP-fucose, 0.5 µM purified non-fucosylated N-terminal EMI, and varied concentrations of divalent metal ions ($MnCl_2$, $MgCl_2$ or $CaCl_2$, dissolved in water) or 5 mM of EDTA in 50 mM HEPES, pH 7. Reactions were incubated at 37°C for 4 hours and stopped by adding 250 µL of cold acetone. Proteins were incubated with acetone at -20°C for 2 hours and centrifuged at 12,700 rpm at 4ºC for 15 minutes. Protein pellets were reduced, alkylated, digested with trypsin, and analyzed with nano LC-MS/MS using the procedures described earlier. Data analysis with Byonic, EIC generation and quantification were performed as described earlier.

GDP-Glo glycosyltransferase assay

50 nM of recombinant POFUT3 was incubated in 20 µL reaction mixtures containing 100 µM Ultra-pure GDP-fucose (Promega), 0.3 mM $MnCl_2$ with 0.156 µM, 0.313 µM, 0.625 µM, 1.25 µM, 2.5 µM, 5 µM, 10 µM, 20 µM or 30 µM of purified non-fucosylated N-terminal EMI in 50 mM HEPES, pH 7. Reactions were incubated at 37°C for 15 min and samples were immediately put on ice. According to the manufacturer's guidelines, adding the nucleotide detection reaction

buffer (NDR, Promega) supplied with the GDP Glo Enzyme (Promega) should terminate enzyme activity, followed by a one-hour incubation period before luminescence detection. However, we observed that POFUT3 still exhibited weak activity in the presence of the NDR buffer. We modified the protocol to immediately detect luminescence after adding the NDR buffer (with GDP Glo enzyme). The GDP standard (Promega) used in parallel showed a strong linear relationship with an R-square (r^2) value over 0.999 (data not shown), indicating that the luminescence is directly proportional to the GDP concentration within the reactions. Produced GDP was quantified based on GDP standards and converted to desired units (nmol/min/mg). The kinetic parameters were determined with nonlinear Michaelis-Menten least squares (ordinary) fitting in Prism 7. All reactions were performed in biological triplicates with three batches of purified enzymes.

Acknowledgments

We thank Kelley Moremen (University of Georgia) for providing the GFP-FUT10/11 expressing plasmids. We thank Peng Wu (Scripps Research Institute) for providing the 6-alkynyl fucose. We thank all the current and past members of the R.S.H. laboratory for technical advice and critical comments on this manuscript, especially Kelvin Luther for helpful discussions on enzyme kinetics, Megumi Takeuchi and Steven Berardinelli for technical assistance on mass spectrometry. We thank the valuable assistance provided by ChatGPT (version 3.5) for language and writing refinement during the preparation of this manuscript. This work was supported by NIH grant R35GM148433.

Data Availability

Dataset 1 includes MS2 spectra used for the EIC generation in this paper was uploaded to a server:
https://1drv.ms/f/s!AqYxR6fJYhp0j2xErR2Y81DO77MQ?e=1Sc95B

2.6 Main Text Tables and Figures

Table 2.1 N-terminal EMI concentration-dependent kinetic analysis of POFUT3 and POFUT4

Kinetic parameter	POFUT3		POFUT4	
	T216	T265	T216	T265
Vmax (nmol/min/mg)	280.4	73.58	48.92	5.786
Km (μM)	36.81	17.47	6.738	0.999

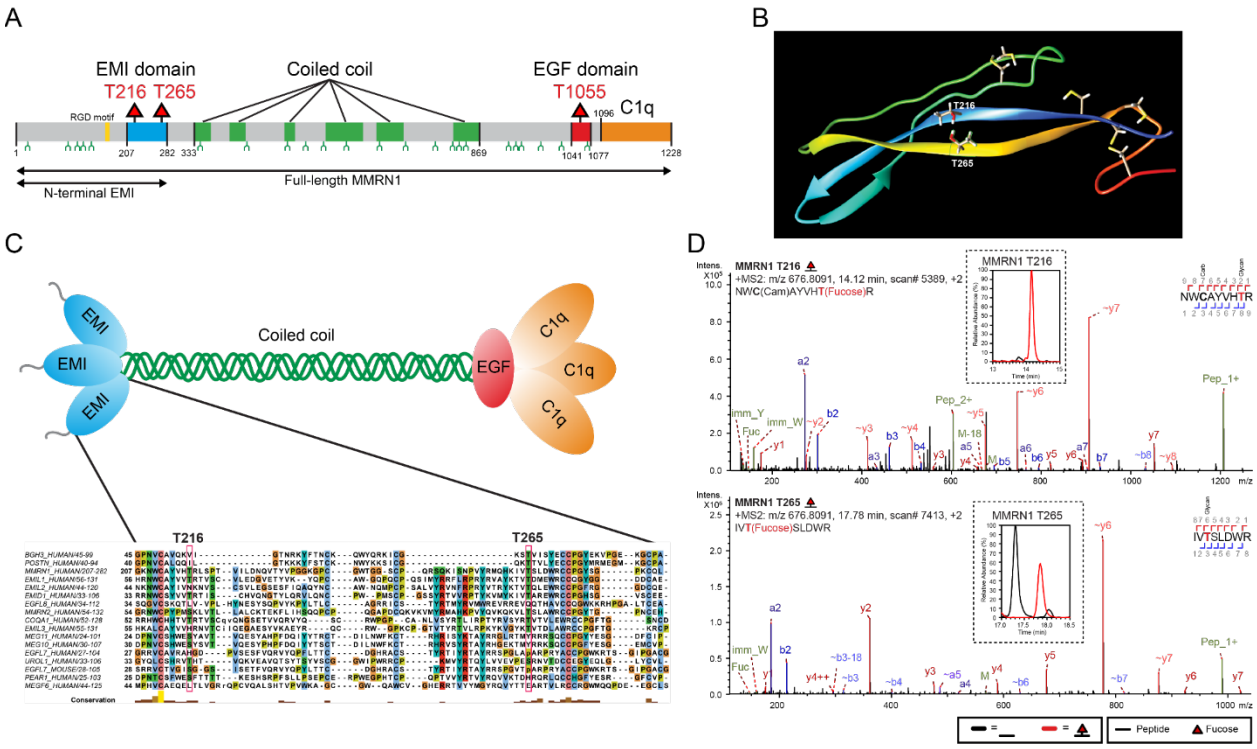


Figure 2.1 Two novel *O*-fucose sites were identified on the EMI domain of MMRN1. (*A*) Domain organization of human MMRN1 protein. The positions of *O*-linked fucosylation and *N*-glycan sequons are shown as red triangles and green branches, respectively. Protein domains are illustrated: yellow - RGD motif (cell attachment site); blue – EMI domain; green – coiled-coil region; red – EGF-like domain; orange – C1q domain. Black arrows indicate the expression constructs used in this study. (*B*) AlphaFold structure of the EMI domain from MMRN1. Data from Dr. Mark Larance. (*C*) Cartoon of the homotrimer MMRN1 structure with key domains shown. Insert – protein sequence alignment of all human EMI domains. Sequence alignment data from Dr. Mark Larance. (*D*) HCD-MS/MS spectra of fucosylated peptides ^{209}NWCAYVHTR^{217} and ^{263}IVTSLDWR^{270} that contain the novel T216 and T265 *O*-fucose sites in MMRN1 EMI domain. Extracted ion chromatograms (EICs) of different glycoforms were inserted for each peptide: red lines, *O*-fucose modified; black lines, unmodified. Spectra for corresponding unmodified peptides used for EIC generation are in Dataset S1.

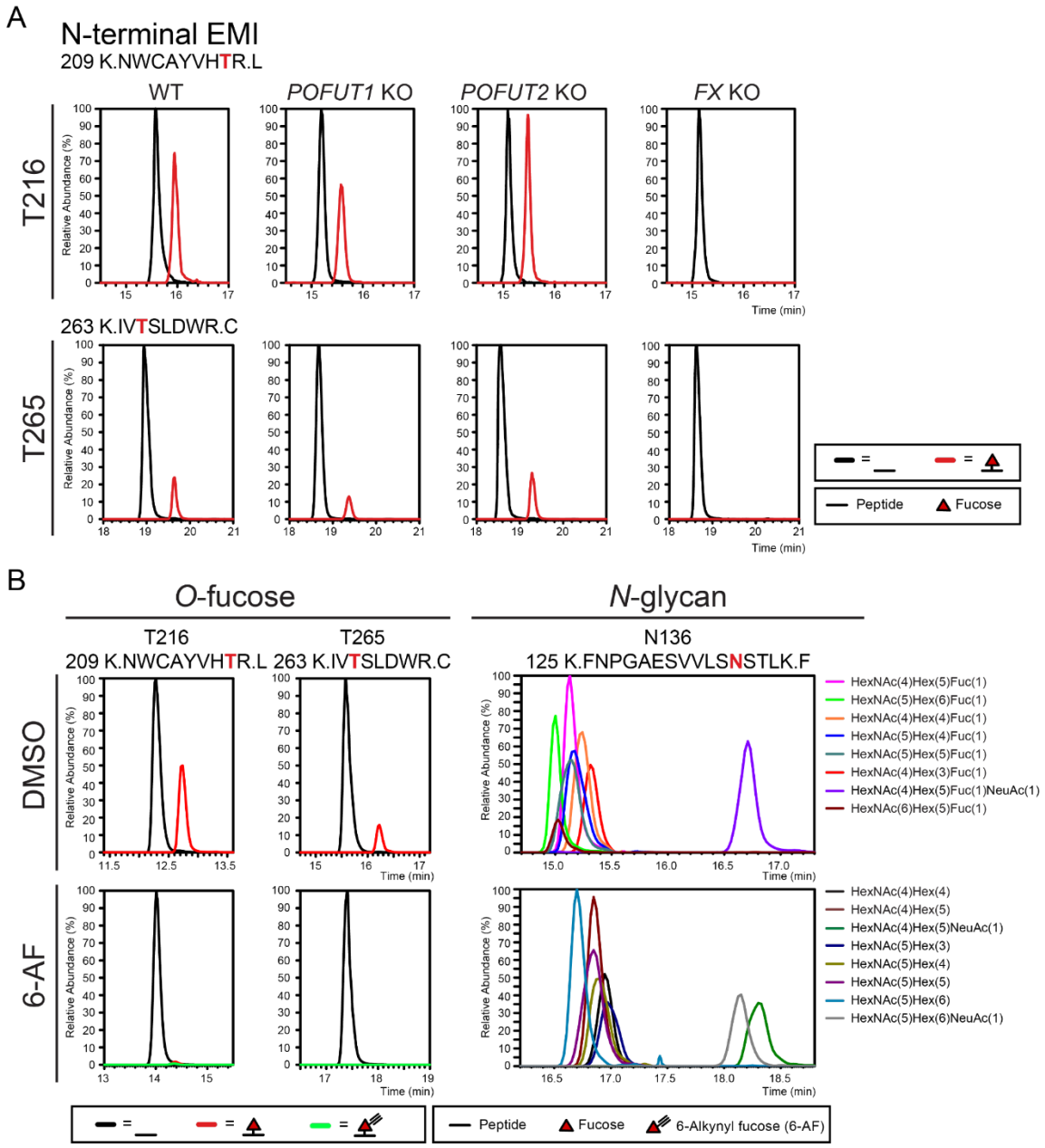


Figure 2.2 Neither POFUT1 nor POFUT2 is responsible for the *O*-fucosylation of MMRN1 EMI domain. (*A*) EICs of different glycoforms of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI produced in HEK293T WT, *POFUT1* KO, *POFUT2* KO, or *FX* KO cells. EICs of positive controls (mNOTCH1 EGF1-5 for POFUT1-mediated EGF *O*-fucosylaton, THBS1 TSR1-3 for POFUT2-mediated TSR *O*-fucosylation) are in *SI Appendix*, Figure 2S1, A and B. Spectra for the corresponding ions are in Dataset S1. (*B*) EICs of different glycoforms of peptides containing T216 or T265 *O*-fucose sites, or N136 *N*-glycan site from N-terminal EMI produced in HEK293T WT cells incubated with DMSO or 6-alkynyl fucose (6-AF). EICs of positive controls (mNOTCH1 EGF1-5 and THBS1 TSR1-3 for 6-AF incorporation) are in are in *SI Appendix*, Figure 2S1 C and D. Spectra for the corresponding ions are in Dataset S1.

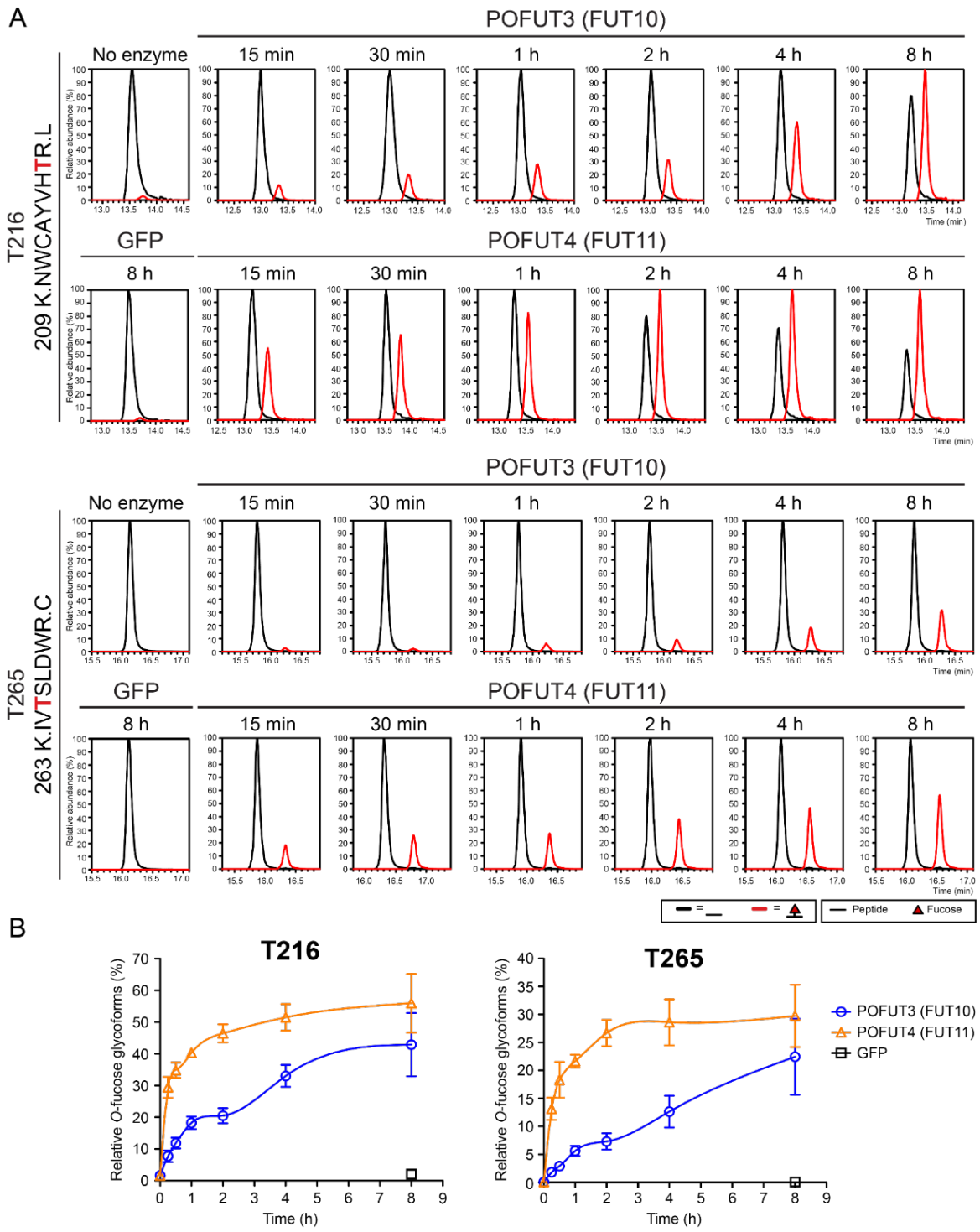


Figure 2.3 FUT10 and FUT11 are POFUTs that are responsible for the *O*-fucosylation of MMRN1 EMI domain. (*A*) 0.1 μM of purified POFUT3, POFUT4, or GFP (negative control) was incubated with 0.5 μM non-fucosylated N-terminal EMI and 100 μM of GDP-fucose for indicated time periods. Reaction products were reduced, alkylated, digested with trypsin, and analyzed by nano LC-MS/MS. EICs of different glycoforms of peptides containing T216 or T265 *O*-fucose site on N-terminal EMI were generated. (*B*) Relative abundances of *O*-

fucosylation at T216 and T265 sites for each reaction were calculated from the EICs in (*A*) and plotted as time-dependent curves. Data presented as mean with standard deviation (SD) from biological triplicates using three batches of purified enzymes (Figure 2S4).

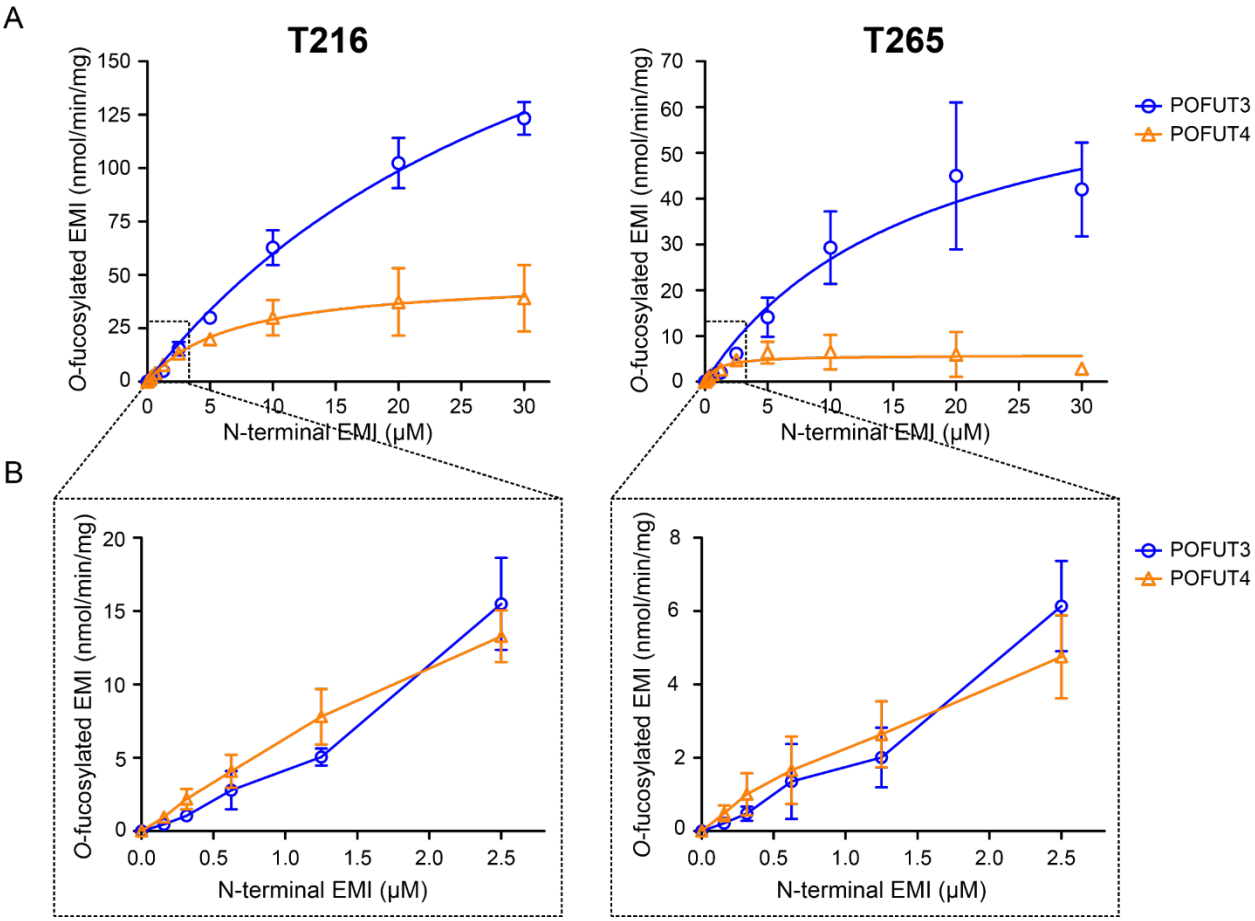


Figure 2.4 Substrate concentration-dependent kinetics of POFUT3 and POFUT4. (*A*) 50 nM of purified POFUT3/4 was incubated with varied concentration of non-fucosylated N-terminal EMI, 100 μM of ultra-pure GDP-fucose and 0.3 mM $MnCl_2$ for 15 minutes. Reaction products were reduced, alkylated, digested with trypsin, and analyzed by nano LC-MS/MS. *O*-fucosylation stoichiometry on T216 site and T265 site was quantified from EICs and converted into product concentration. Kinetic analysis was performed with nonlinear Michaelis-Menten fitting in Prism 7. (*B*) Data points with substrate concentrations ranging from 0 to 2.5 μM in (*A*) were zoomed in to show the distinct kinetic profiles of POFUT3 and 4 with low substrate concentration. Data presented as mean with SD from biological triplicates using three batches of purified enzymes (Figure 2S8).

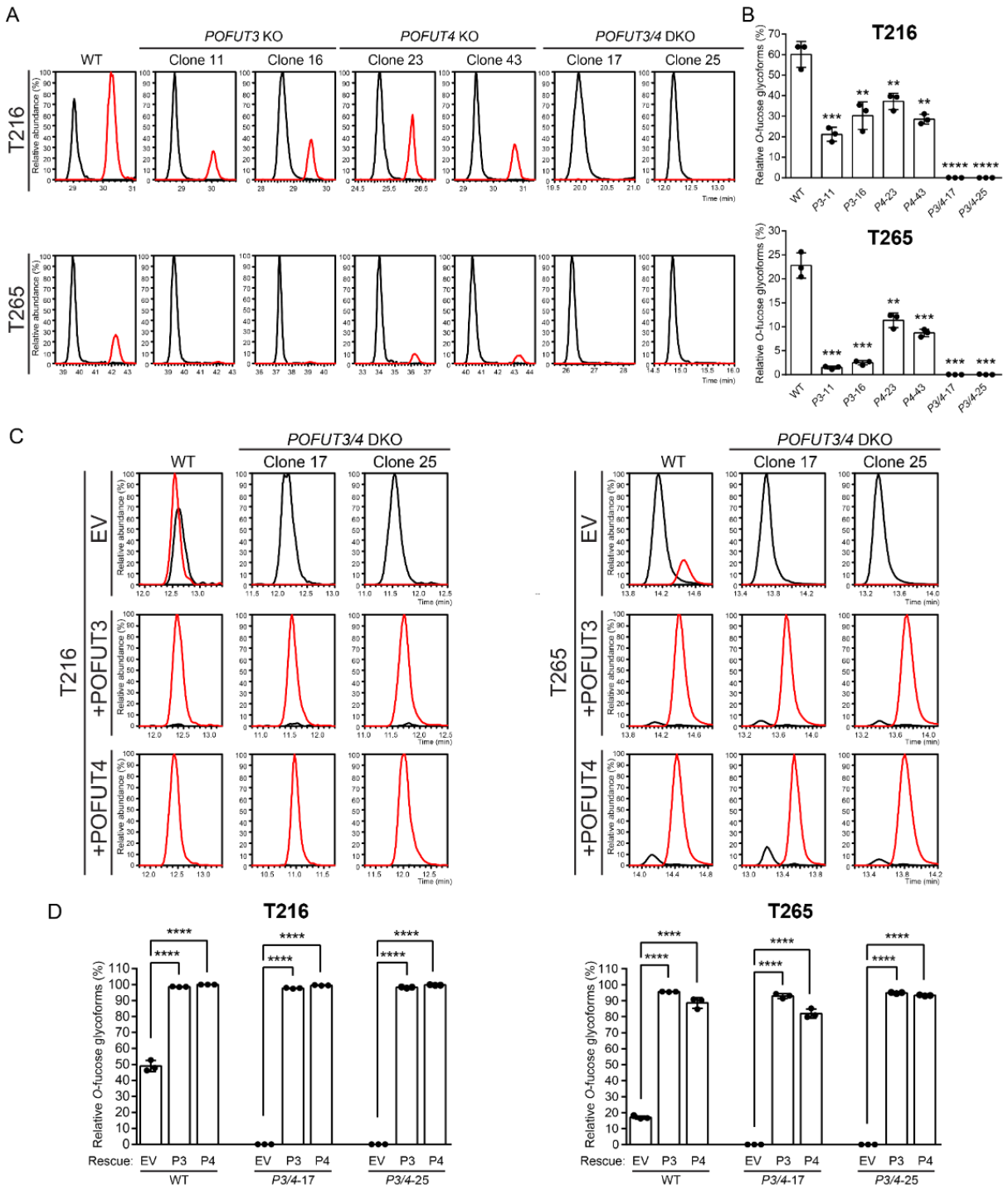


Figure 2.5 EMI *O*-fucosylation in *POFUT3* and *POFUT4* SKO and DKO cells. (*A*) EICs of different glycoforms of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI produced in HEK293T WT, *POFUT3* KO, *POFUT4* KO, or *POFUT3/4* DKO cells. (*B*) Relative abundances of *O*-fucosylated glycoforms in (*A*) were quantified. *P3*-11, *POFUT3* KO cells-clone 11; *P3*-16, *POFUT3* KO cells-clone 16; *P4*-23, *POFUT4* KO cells-clone 23; *P4*-43, *POFUT4* KO cells-clone 43; *P3/4*-17, *POFUT3/4* DKO cells-clone 17; *P3/4*-25, *POFUT3/4* DKO cells-clone 25. Statistical analysis was performed with unpaired t test in Prism 7. **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$ compared with control (WT cells). (*C*) EICs of different

glycoforms of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI produced in HEK293T WT or *POFUT3/4* DKO cells that co-transfected with plasmids encoding POFUT3, POFUT4 or empty vector (EV). (*D*) Quantified relative abundances of *O*-fucosylated glycoforms in (*C*). EV, empty vector; P3, POFUT3 plasmid; P4, POFUT4 plasmid. Statistical analysis was performed with unpaired t test in Prism 7. ****, $p < 0.0001$ compared with control (empty vector). All data was shown as mean with SD from biological triplicates of three individual transfections (Figure 2S10 and Figure 2S11).

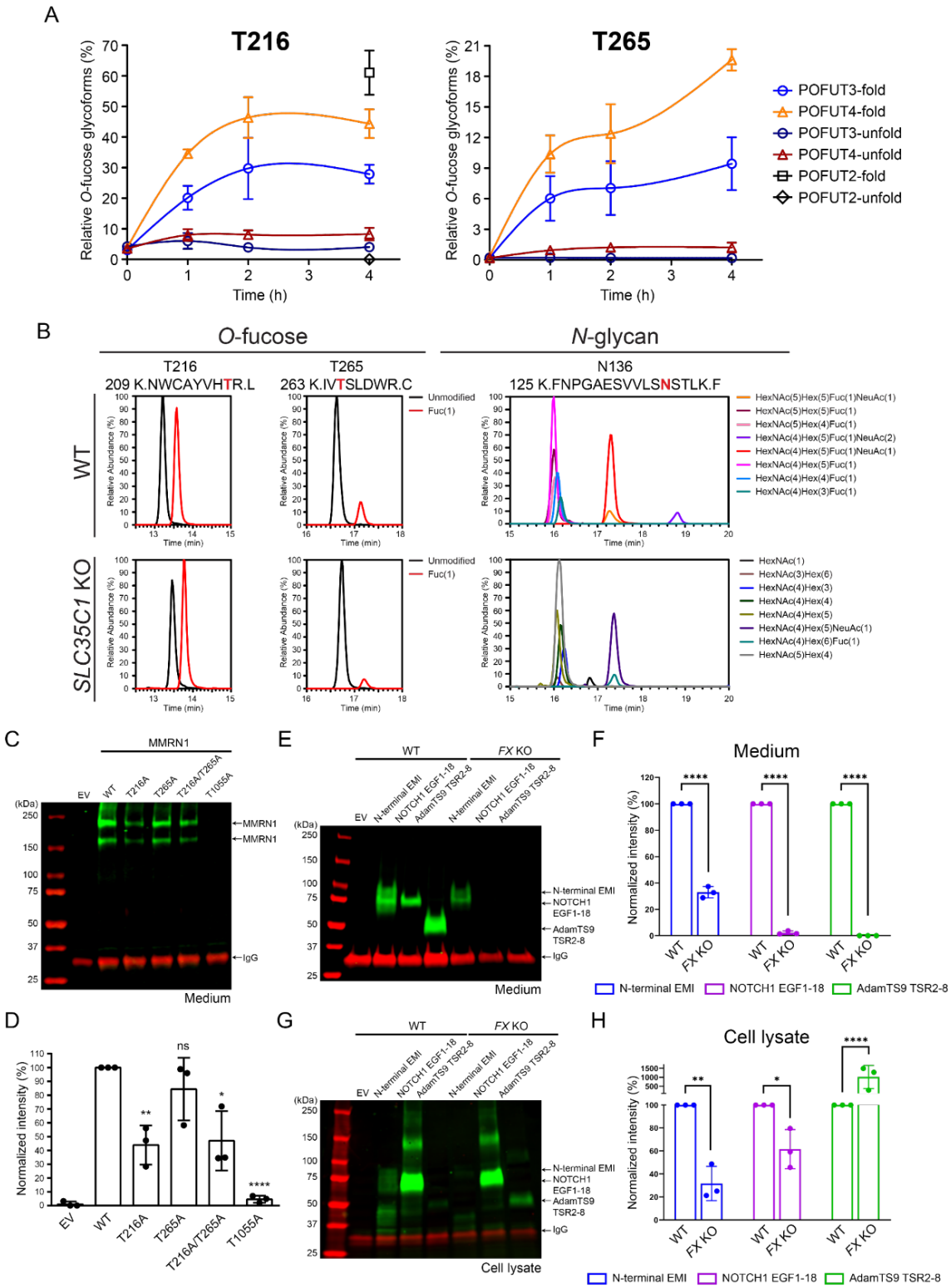


Figure 2.6 POFUT3 and POFUT4 require folded EMI structures for modification and function in the ER, participating in a non-canonical ER quality control pathway for EMI domains. (*A*) POFUT enzymatic assays with folded and unfolded substrates (N-terminal EMI or THBS1 TSR3) for POFUT3, POFUT4, or POFUT2 (positive control). EICs of the reaction products were generated and quantified. Relative abundances of *O*-fucosylated glycoforms were

calculated and plotted as time-dependent curves. Data was shown as mean with SD from biological triplicates using three batches of purified enzymes (Figure 2S14). (*B*) POFUT3 and POFUT4 function in the ER rather than Golgi. EICs of different glycoforms of peptides containing T216 or T265 *O*-fucose sites, or N136 *N*-glycan site from N-terminal EMI produced in HEK293T WT or *SLC35C1* KO cells were generated. Spectra for the corresponding ions are in Dataset S1. (*C*) HEK293T cells were transfected with plasmids encoding Myc-tagged MMRN1 WT, MMRN1 T216A, MMRN1 T265A, MMRN1 T216A/T265A, MMRN1-T1055A, or empty vector (EV) and IgG (secretion control). 2-days culture medium was collected and analyzed by Western Blot probed with anti-Myc and anti-human IgG antibodies. (*D*) The bar graph shows quantified band intensity of MMRN1 normalized with IgG band intensity obtained in panel (*C*). (*E-H*) HEK293T WT or *FX* KO cells were transfected with plasmids encoding Myc-tagged N-terminal EMI, NOTCH1 EGF1-18, AdamTS9 TSR2-8, or empty vector (EV) and IgG (secretion control). Cells were cultured for 2 days. Culture medium and cell lysates were analyzed by Western Blot probed with anti-Myc and anti-human IgG antibodies, presented in panel (*E*) and (*G*), respectively. Bar graphs show quantified band intensity of protein normalized with IgG bands in culture medium and cell lysates, presented in panel (*F*) and (*H*), respectively. All data was shown as mean with SD from biological triplicates of three individual transfections (Figure 2S15 and Figure 2S16).

2.7 Supporting Information Tables and Figures

Contents

Table 2S1. N-terminal EMI concentration-dependent kinetic analysis of POFUT3

Substrate	POFUT3	
	Km (μM)	Vmax (nmol/min/mg)
N-terminal EMI	47.22	482.0

Table 2S2. List of primers sequences

Name		Primer sequence
pcDNA3.1-hMMRN1 T265A-Myc-His$_6$ and pcDNA3.1-hMMRN1 T216A/T265A-Myc-His$_6$		
5’	AI452	5’-ATAAAATTGTCGCCTCATTGGATTGGAGGTGCTGTCC-3’
3’	AI464	5’-ATCCAATGAGGCGACAATTTTATGTTGCATCCTATAGACAGG-3’
pcDNA4-N-terminal EMI-Myc-His$_6$		
5’	AI364	5’-ATATATAAGCTTGCCACCATGAAGGGGGCAAGATTATTTG-3’
3’	AI365	5’-ATCTAGCTCGAGCAAACTTTGCTGTTCCTGGG-3’
pcDNA4-full length hFUT10 (POFUT3)-Myc-His$_6$ and pcDNA4-full length hFUT11 (POFUT4)-Myc-His$_6$		
5’	AI456	5’-TGGCTAGTTAAGCTTGCCACCATGGTGAGAATC-3’
3’	AI457	5’-CTCGAACTTGCCCAGCTCCACCATCACCTGCAG-3’
5’	AI458	5’-CTGGGCAAGTTCGAGCG-3’
3’	AI459	5’-CCCTCTAGACTCGAGGTCCTTAAAGACCAGGCC-3’
5’	AI454	5’-CTCGAGTCTAGAGGGCCCTTC-3’
3’	AI455	5’-AAGCTTAACTAGCCAGCTTGGG-3’
5’	AI460	5’-TGGCTAGTTAAGCTTGCCACCATGGCCGCCGGC-3’
3’	AI461	5’-CCCGCCGGCCTCCCTCTCGGCCACGCTGCCGTG-3’
5’	AI462	5’-AGGGAGGCCGGCGGGGAG-3’
3’	AI463	5’-CCCTCTAGACTCGAGGAGATGTTGCCTCTTCATGAAGATTTCATGTAGGTAATC-3’

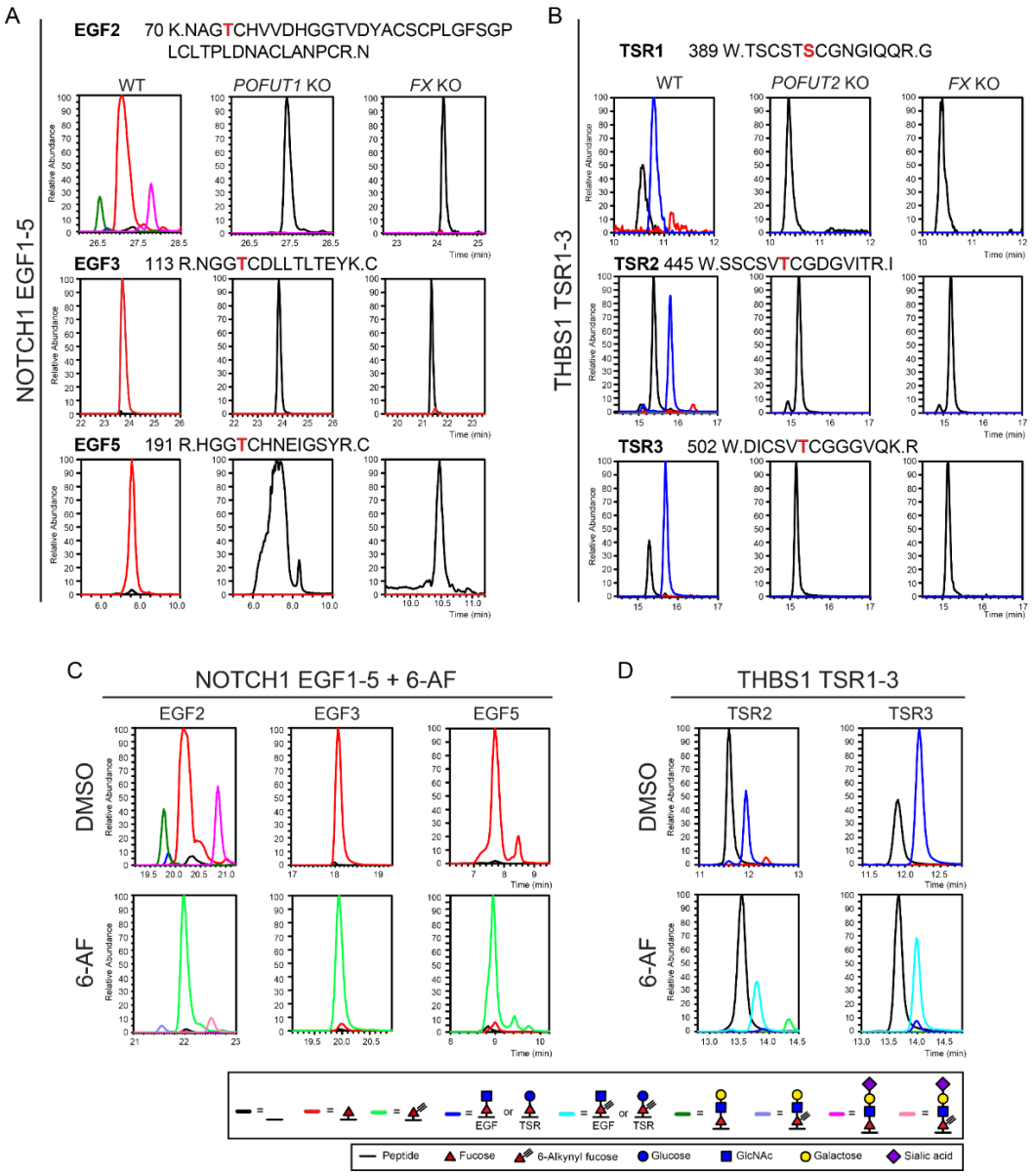


Figure 2S1. Mass spectrometric analysis of peptides from transfected mNOTCH EGF1-5 and THBS1 TSR1-3 as positive controls for the data shown in Figure 2.2. (*A*) EICs of different glycoforms of peptides from EGF2, 3, and 5 of mNOTCH EGF1-5 expressed and purified in HEK293T WT, *POFUT1* KO, or *FX* KO cells, demonstrating the loss of *O*-fucose on known POFUT1 substrate when knocking out *POFUT1* or *FX*. (*B*) EICs of different glycoforms of peptides from TSR1, 2, and 3 of THBS1 TSR1-3 expressed and purified in HEK293T WT, *POFUT2* KO, or *FX* KO cells, demonstrating the loss of *O*-fucose on known POFUT2 substrate when knocking out *POFUT2* or *FX*. (*C*) EICs of peptides from EGF2, 3, and 5 of mNOTCH EGF1-5 expressed and purified in HEK293T WT that metabolic labeled with 6-alkynyl fucose (6-AF) or equal volume of DMSO, showing the incorporation of 6-AF to known POFUT1 substrate. (*D*) EICs of peptides from TSR1, 2, and 3 of THBS1 TSR1-3 expressed and purified in HEK293T WT that metabolic labeled with 6-AF or equal volume of DMSO, demonstrating the incorporation of 6-AF to known POFUT2 substrate. Spectra for the corresponding ions are in Dataset S1.

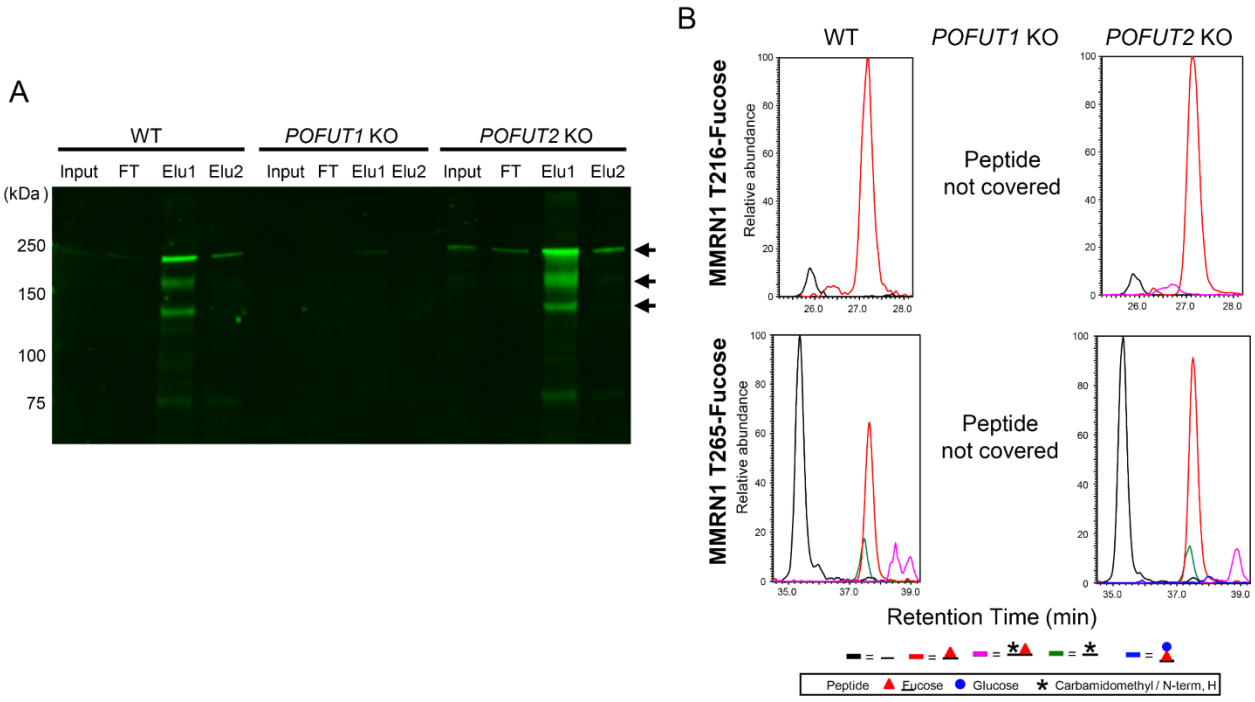


Figure 2S2. Mass spectrometric analysis of peptides from full-length MMRN1 that expressed and purified from HEK293T WT, *POFUT1* KO, and *POFUT2* KO cells. (*A*) HEK293T WT, *POFUT1* KO or *POFUT2* KO cells were transfected with a plasmid encoding MMRN1. After 2 days, MMRN1 was purified from conditioned media using Ni-NTA agarose as described in Materials and Methods. MMRN1 was analyzed by Western Blot probed with anti-Myc and anti-His antibodies. FT: flowthrough from Ni-NTA purification; Elu1: first elution; Elu2: second elution. (*B*) MMRN1 from panel (*A*) was reduced, alkylated, digested with trypsin, and analyzed by mass spectrometry. EICs of different glycoforms of the peptides containing T216 or T265 *O*-fucose sites were generated.

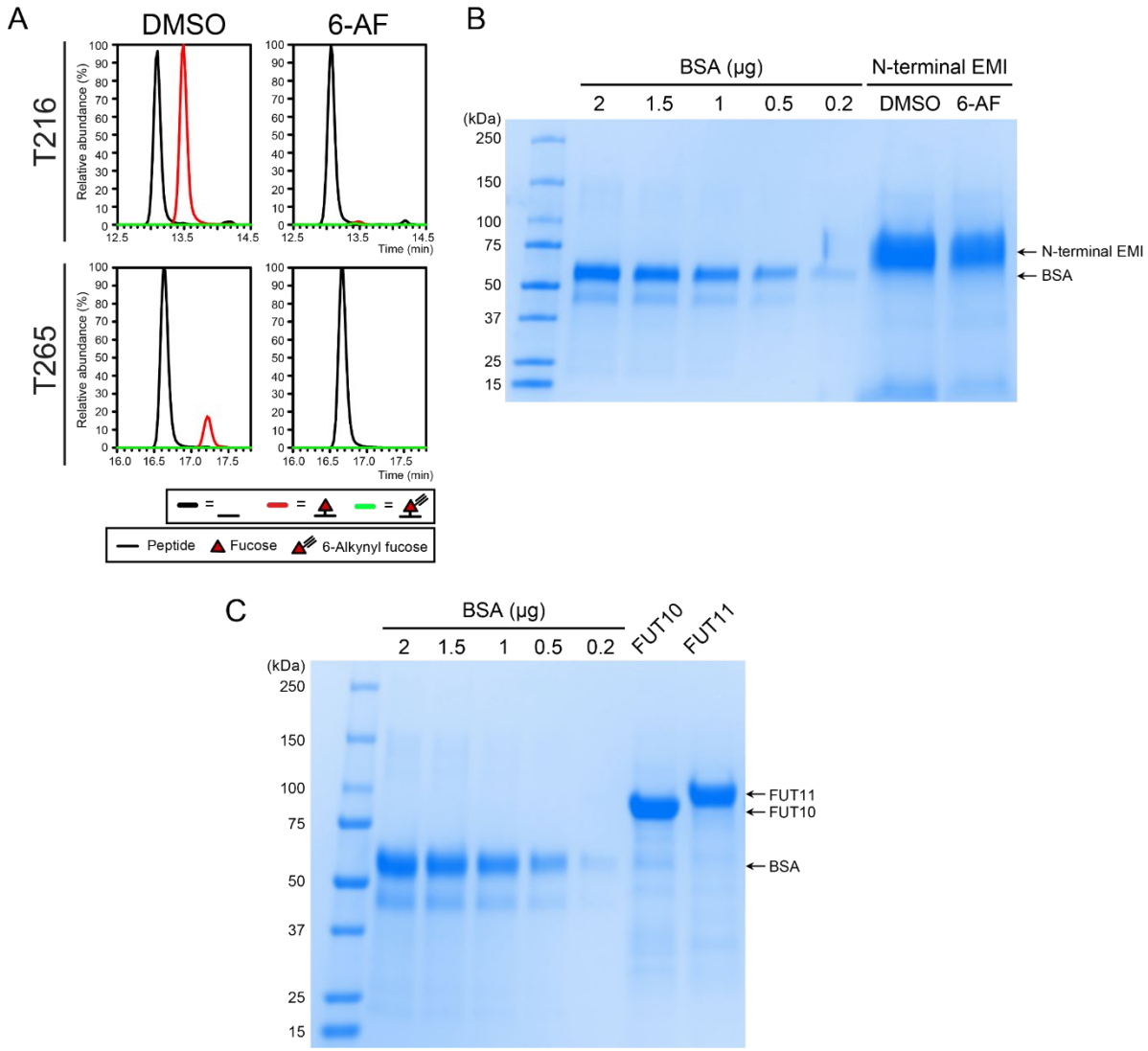
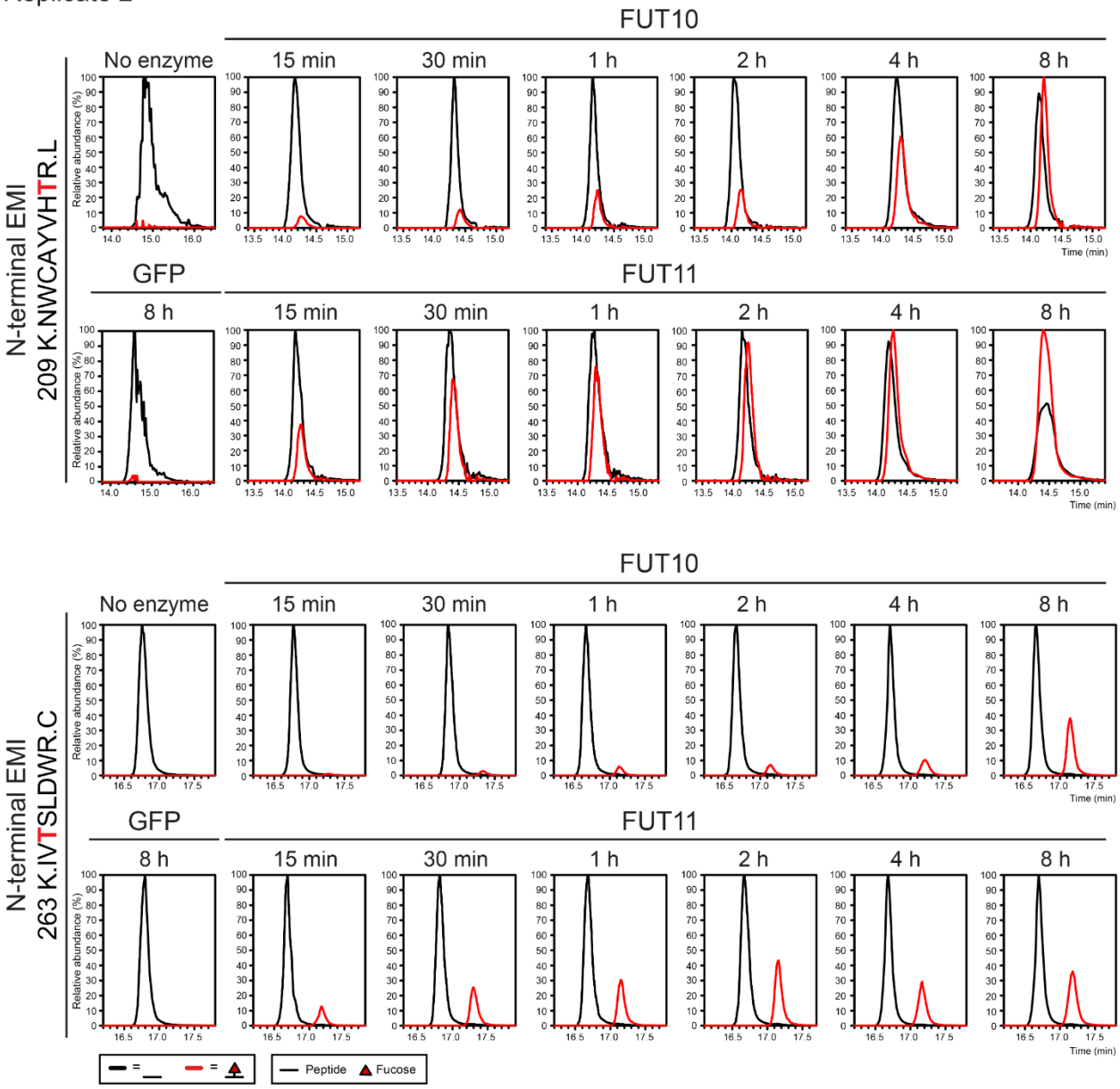


Figure 2S3. Purified, recombinant FUT10, FUT11, and non-fucosylated N-terminal EMI used for the enzymatic assays. (*A*) EICs of different glycoforms of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI produced in HEK293F cells that incubated with 6-AF or equal volume of DMSO, demonstrating the non-fucosylated status of N-terminal EMI purified from 6-AF treated HEK293F cells. 300 µL of conditioned culture medium was used for mass spectrometric analysis. (*B*) The N-terminal EMI produced in panel (*A*) was purified by Ni-NTA agarose as described in Materials and Methods. The purity of protein was verified by Coomassie blue staining. 35 µL Ni-NTA elution (6% of total purified proteins from a 100 mL cell culture) was used for testing. (*C*) Coomassie blue staining of FUT10 and FUT11 expressed and purified from HEK293F cells as described in Materials and Methods. 10 µL Ni-NTA elution (2% of total purified proteins produced from a 100 mL cell culture) was used for testing.

Replicate 2



Continued on next page

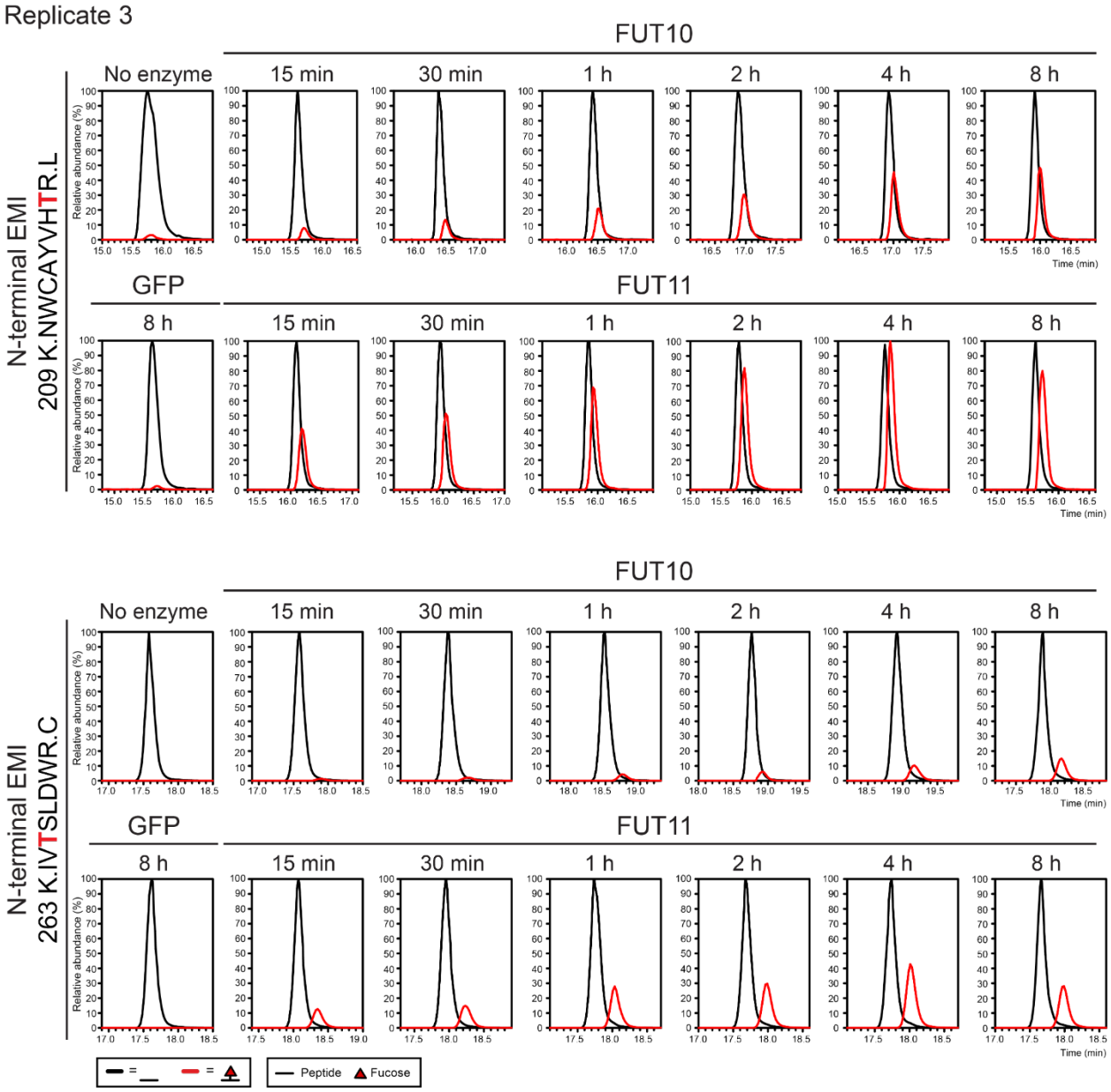


Figure 2S4. Biological triplicates for the data presented in Figure 2.3. Enzymatic assays using 0.5 µM non-fucosylated N-terminal EMI, 100 µM of GDP-fucose, and 0.1 µM of purified POFUT3, POFUT4, or GFP (negative control). Reaction products were analyzed with nano LC-MS/MS and EICs of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI were generated as described in Materials and Methods. In addition to the one replicate used in Figure 2.3A, two additional biological replicates of data included for the statistical analysis in Figure 2.3B were presented.

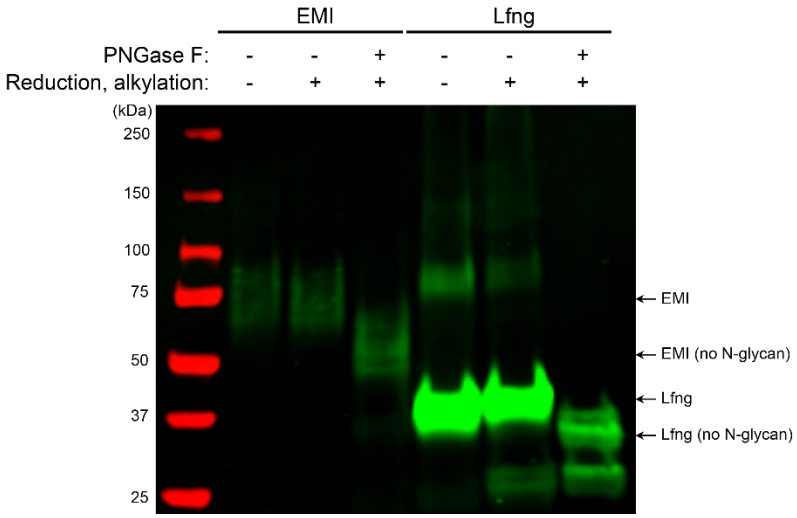


Figure 2S5. N-terminal EMI contains multiple *N*-glycans. N-terminal EMI and Lfng (contains one *N*-glycan, used as a positive control) were expressed and purified from HEK293T cells. Proteins were reduced, alkylated, treated with PNGase F, and analyzed by Western Blot probed with anti-Myc antibodies as described in *SI Appendix*, Supplementary Methods.

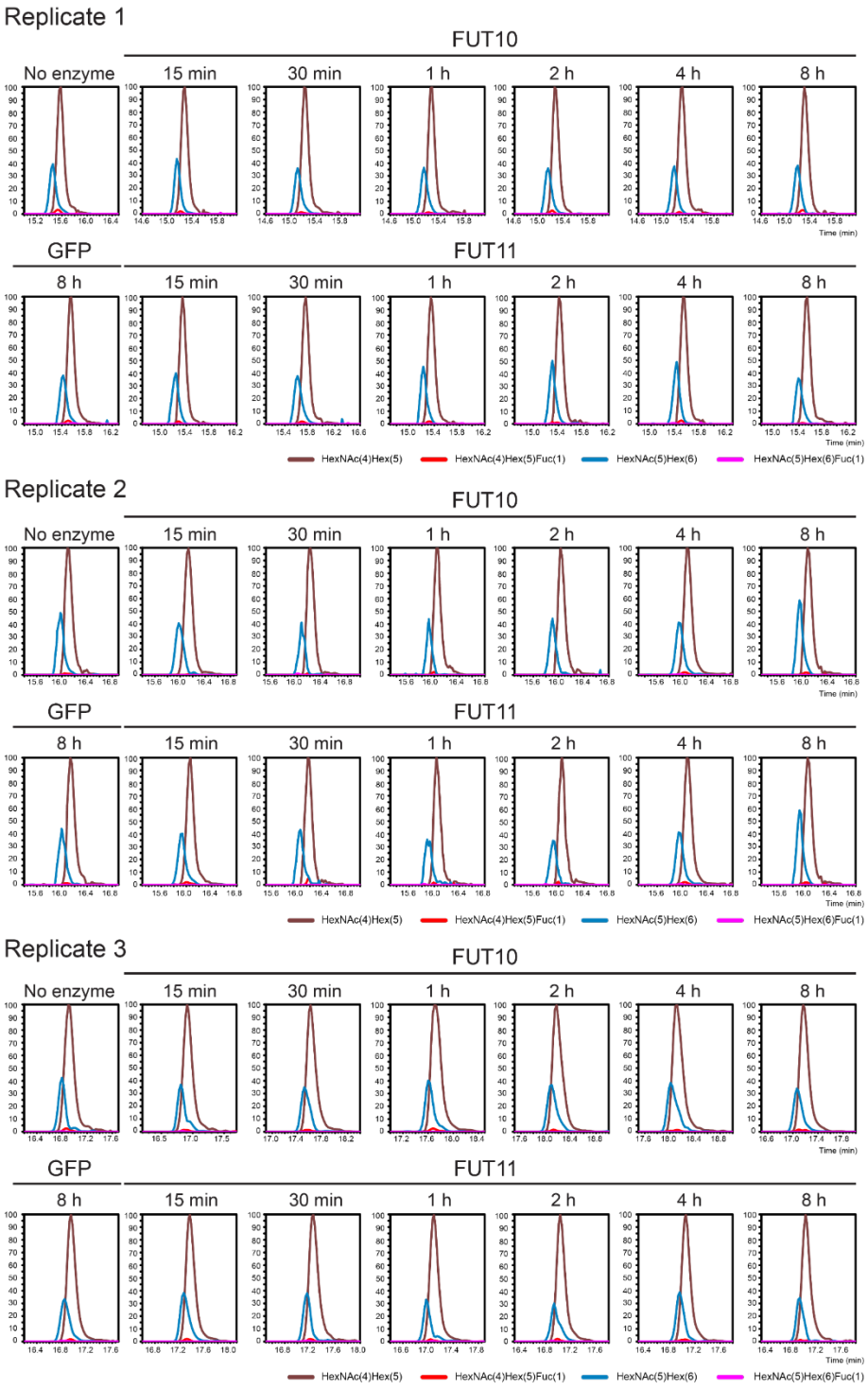
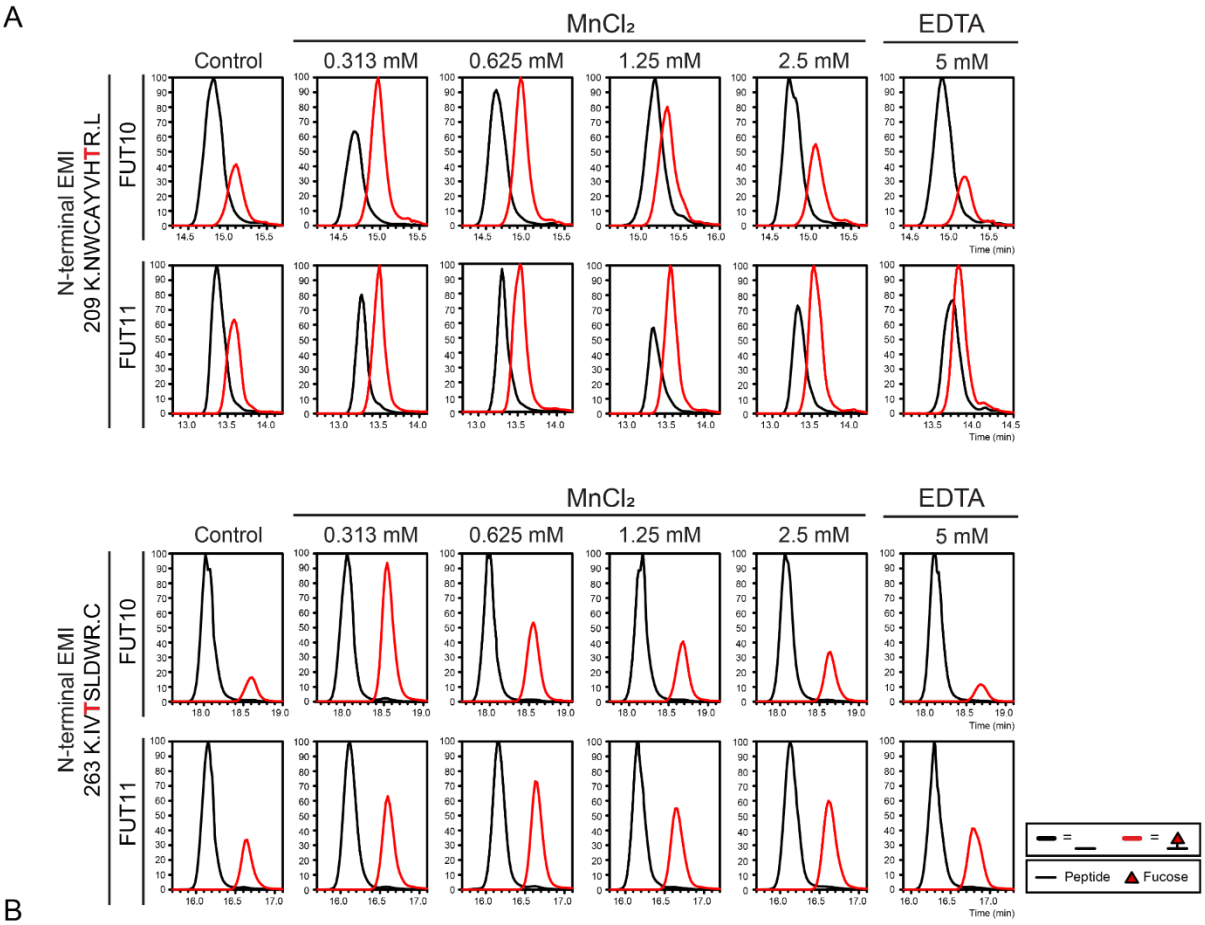


Figure 2S6. POFUT3 and POFUT4 do not modify *N*-glycans in enzymatic assays. Mass spectrometric analysis of peptides ^{125}FNPGAESVVLSNSTLK^{140} containing a complex *N*-glycan site in N-terminal EMI that was produced in the enzymatic assays shown in Figure 2.3. EICs of *N*-glycosylated peptides with or without fucosylation were generated. Biological triplicates of data using three batches of purified enzymes were presented.



Continued on next page

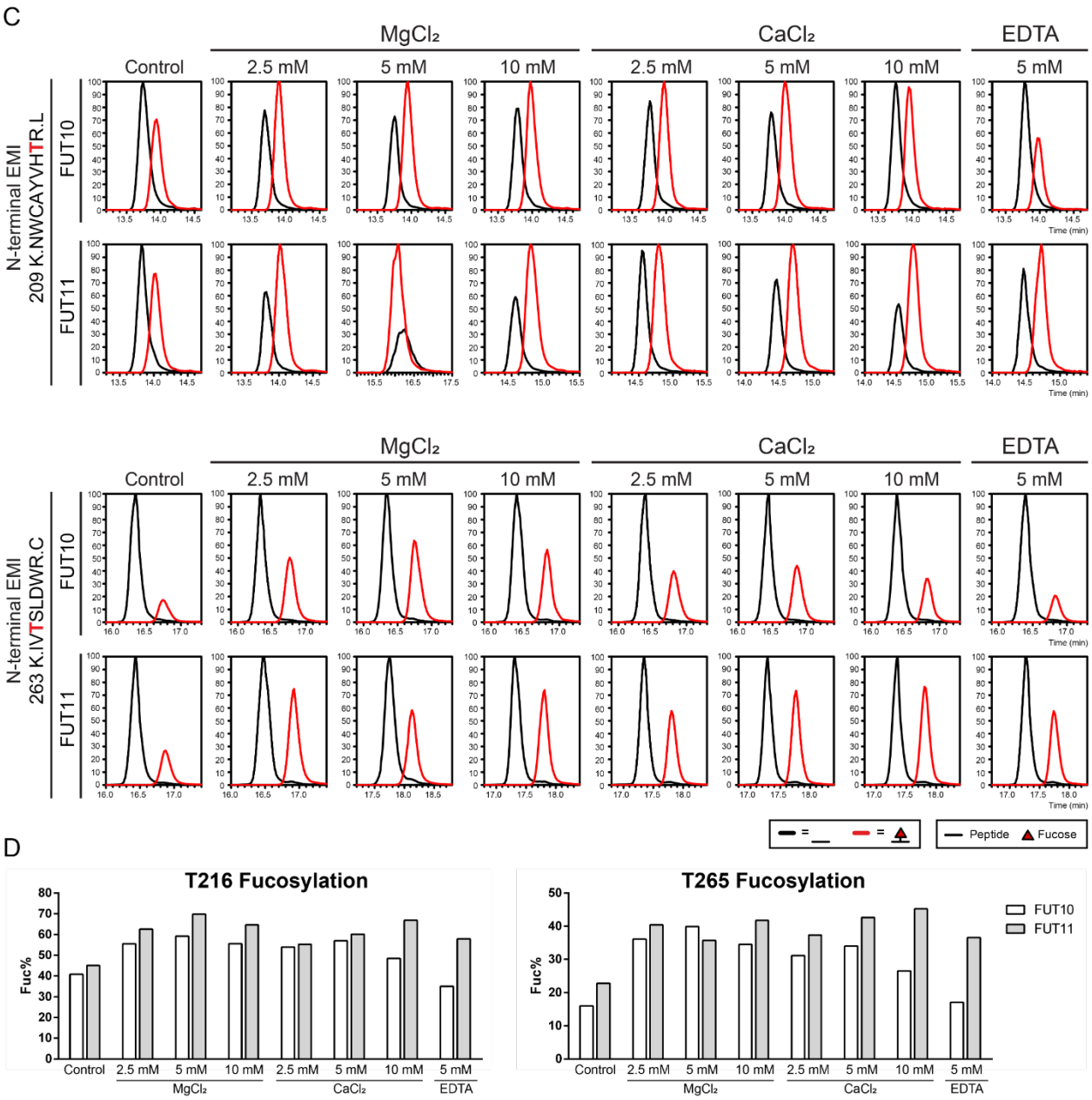
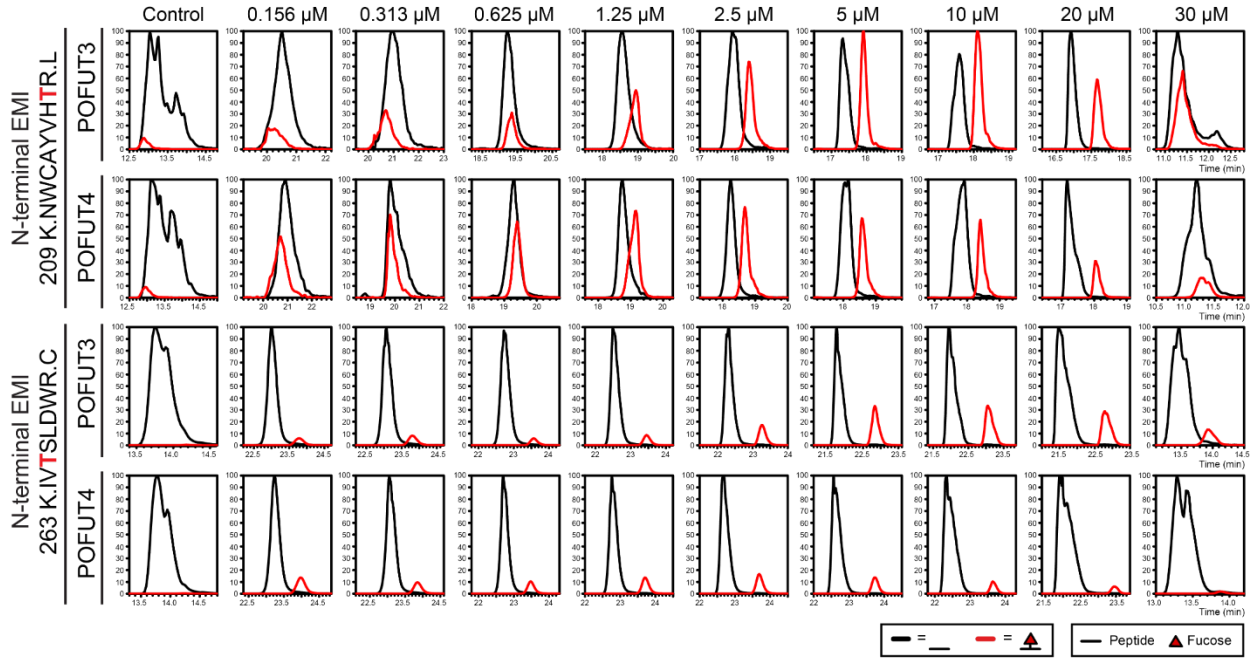
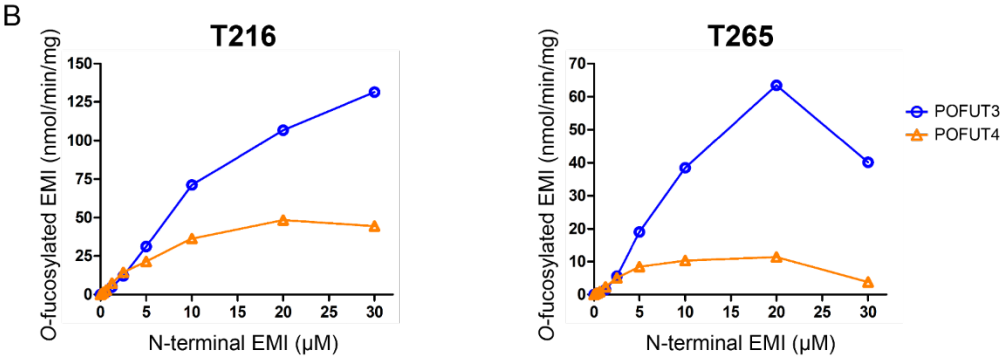


Figure 2S7. The activities of POFUT3 and POFUT4 were enhanced by divalent metal ions. Enzymatic assays of POFUT3 and POFUT4 with N-terminal EMI were supplied with varied concentrations of $MnCl_2$ (*A*), $MgCl_2$ (*C*), or $CaCl_2$ (*C*). Reaction products were analyzed with nano LC-MS/MS. EICs of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI were generated and quantified in (*B*) and (*D*). Experimental details can be found in *SI Appendix*, Supplementary Methods.

A



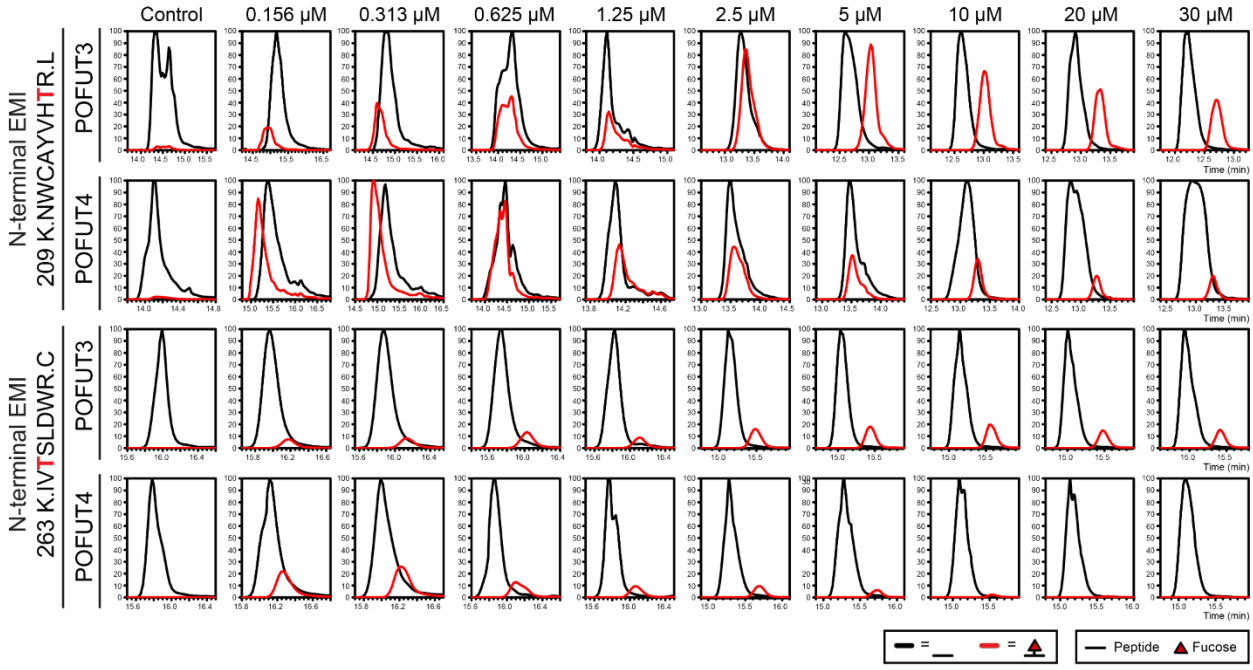
B



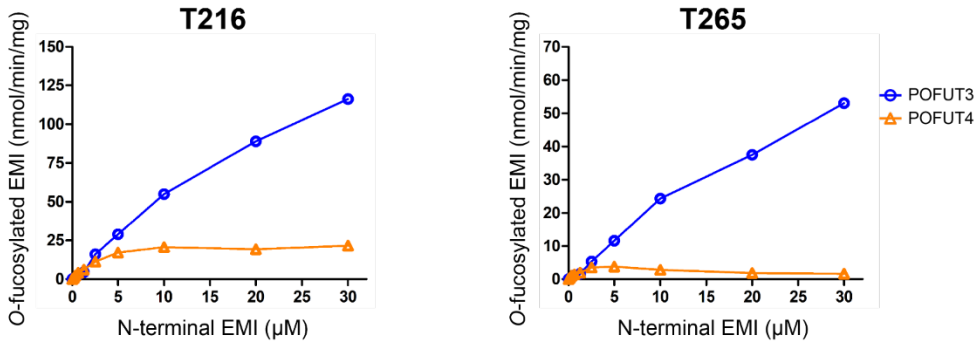
Continued on next page

C

Replicate 2



D



Continued on next page

Replicate 3

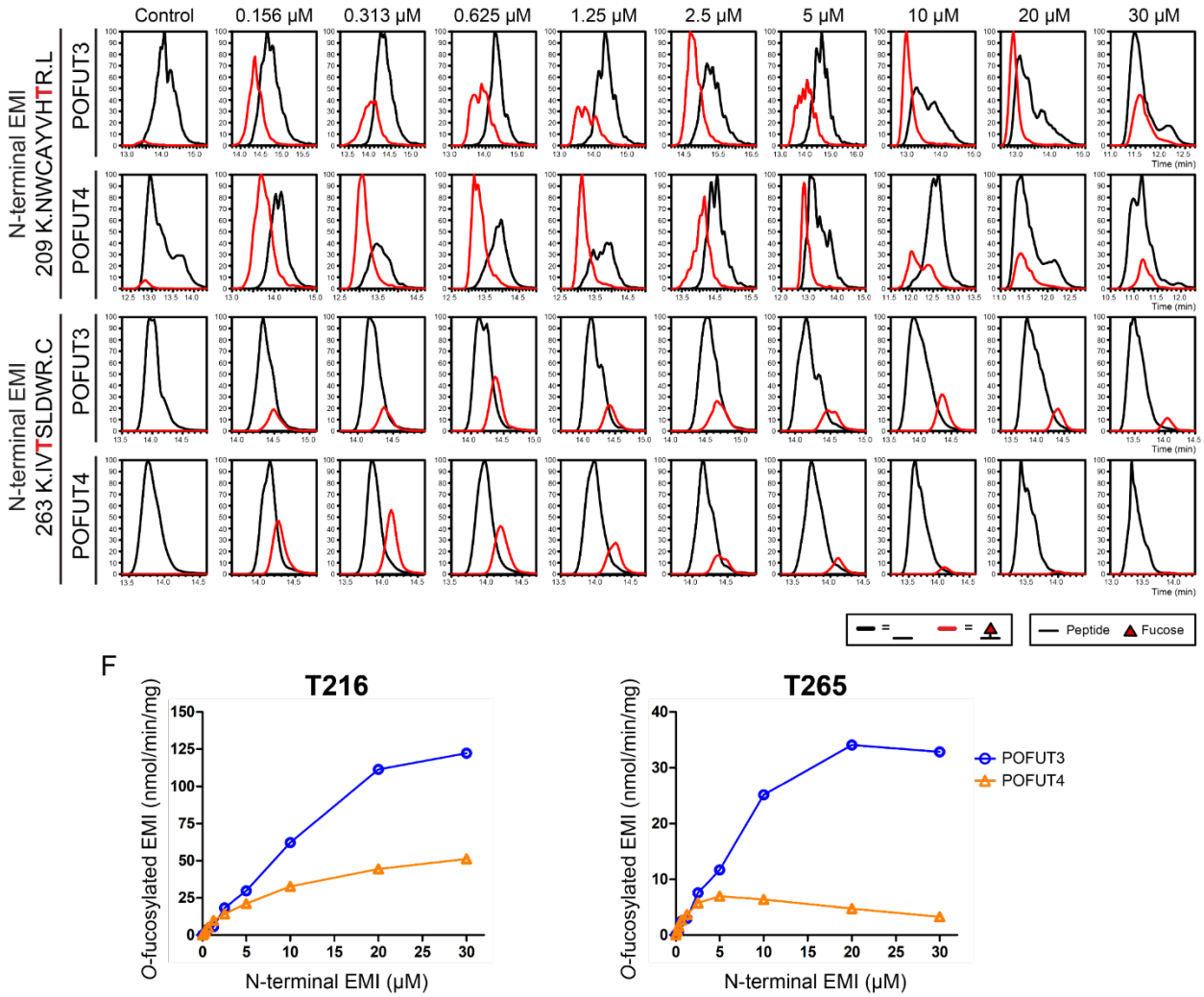


Figure 2S8. Biological triplicates for the data presented in Figure 2.4. Kinetic analysis of POFUT3 and POFUT4 with varied concentrations of N-terminal EMI substrates. Reaction products were analyzed using a mass spectrometric method as described in Materials and Methods. Biological triplicates of data included for the kinetics analysis in Figure 2.4 were presented as EICs in (*A*), (*C*), and (*E*), and plotted kinetic curves in (*B*), (*D*), and (*F*).

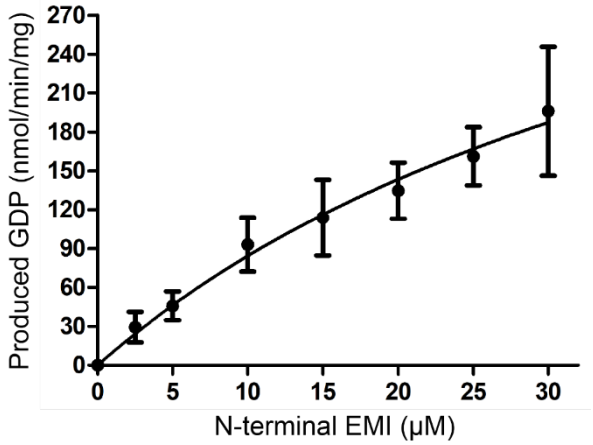


Figure 2S9. Substrate concentration-dependent kinetics of POFUT3 measured by GDP-Glo Glycosyltransferase assay. 50 nM of purified POFUT3 was incubated with varied concentration of non-fucosylated N-terminal EMI, 100 μM of ultra-pure GDP-fucose and 0.3 mM $MnCl_2$ for 15 minutes. GDP-Glo Glycosyltransferase assays were performed to determine the kinetic parameters (*SI Appendix*, Table S1). The curve was generated with nonlinear Michaelis-Menten fitting in Prism 7. Experimental details can be found in *SI Appendix*, Supplementary Methods.

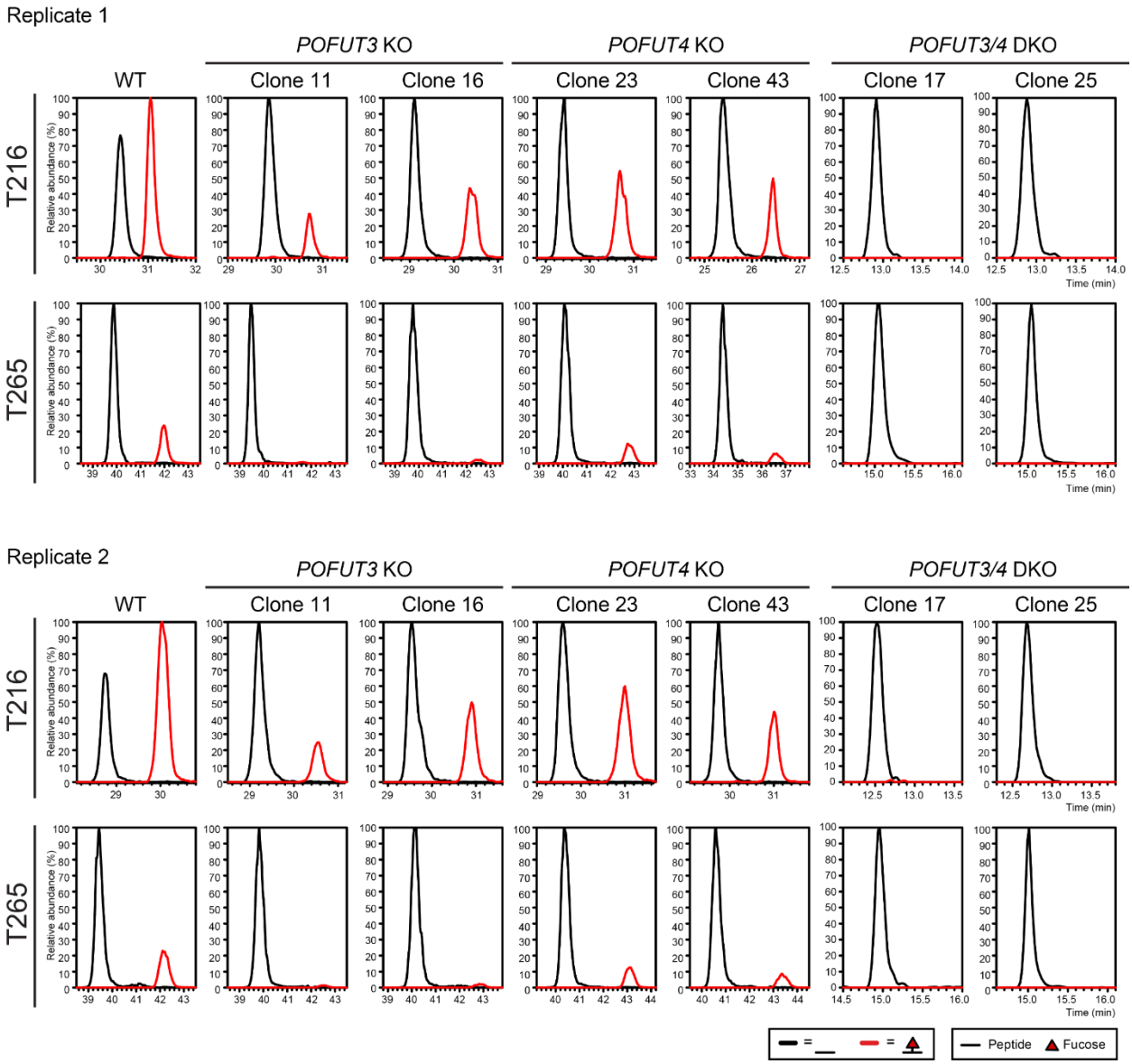


Figure 2S10. Biological triplicates for the data presented in Figure 2.5, A and B. N-terminal EMI was expressed in HEK293T WT, *POFUT3* KO, *POFUT4* KO, or *POFUT3/4* DKO cells. 300 µL of conditioned culture medium were analyzed with nano LC-MS/MS as described in Materials and Methods. EICs of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI were generated. In addition to the one replicate used in Figure 2.5A, two additional biological replicates of data included for the statistical analysis in Figure 2.5B were presented.

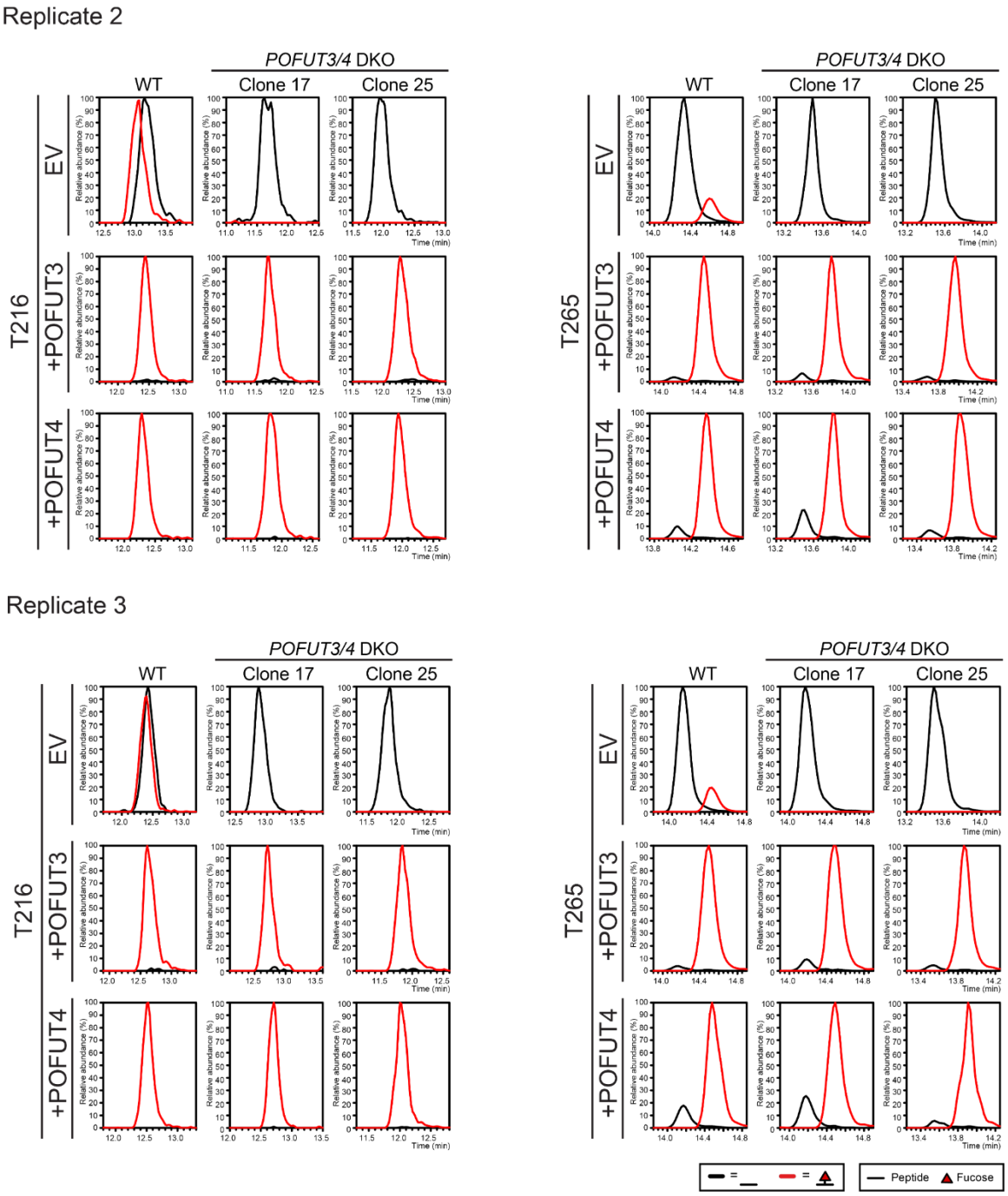


Figure 2S11. Biological triplicates for the data presented in Figure 2.5, *C* and *D*. N-terminal EMI was transfected in HEK293T WT or *POFUT3/4* DKO cells that co-transfected with plasmids encoding POFUT3, POFUT4 or empty vector (EV). 300 µL of conditioned culture medium were analyzed with nano LC-MS/MS as described in Materials and Methods. EICs of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI were generated. In addition to the one replicate used in Figure 2.5C, two additional biological replicates of data included for the statistical analysis in Figure 2.5D were presented.

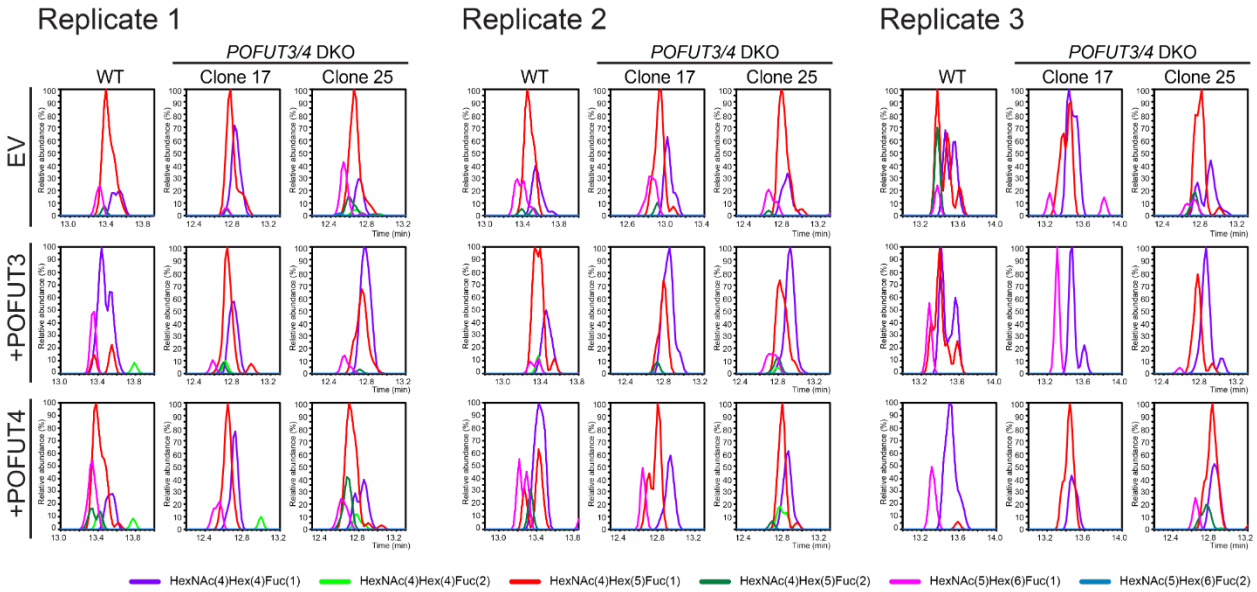


Figure 2S12. POFUT3 and POFUT4 do not modify *N*-glycans in cells. Mass spectrometric analysis for *N*-glycan fucosylation using the data obtained in Figure 2.5, C and D, and Figure 2S11. EICs of different glycoforms of peptides ^{125}FNPGAESVVLSNSTLK^{140} modified with fucosylated *N*-glycans were generated. Biological triplicates of data using three batches of purified enzymes were presented.

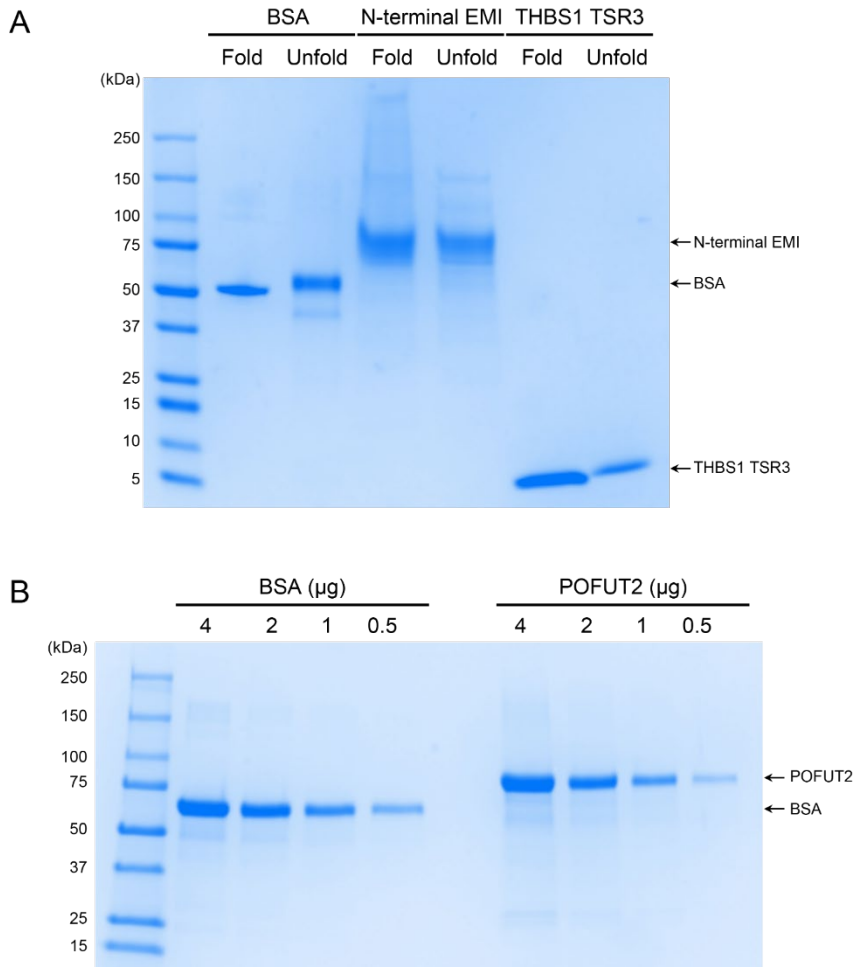
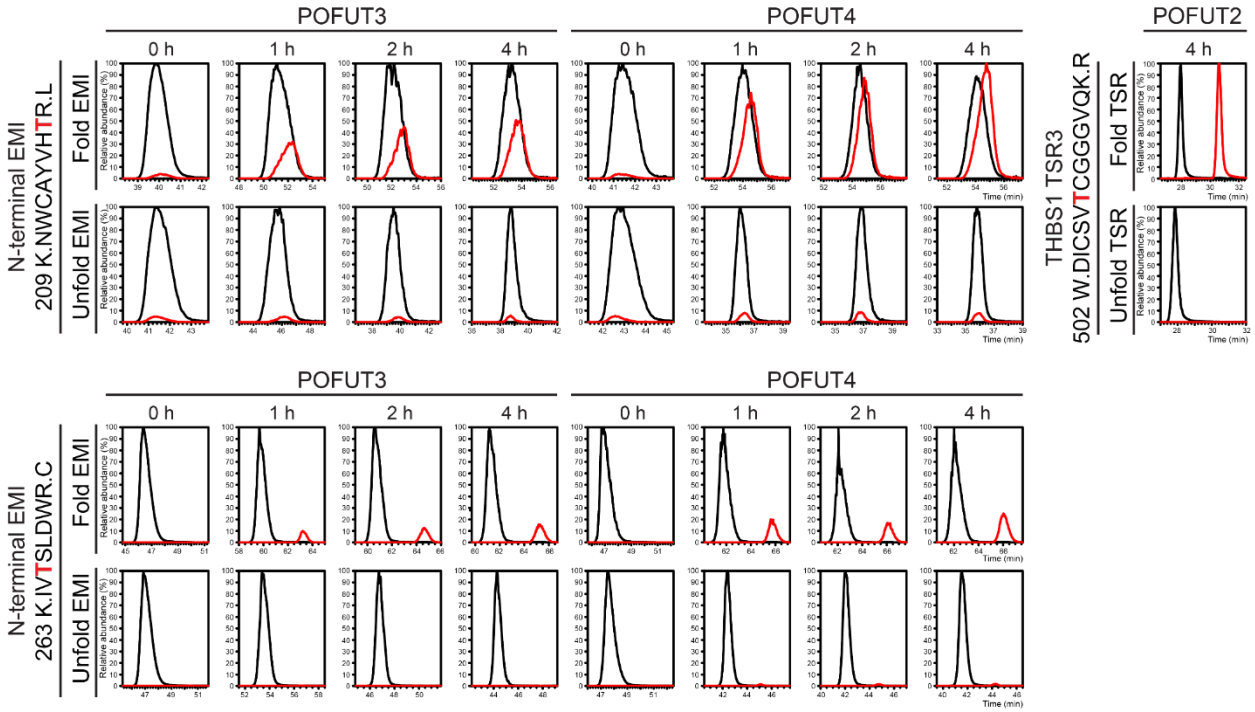
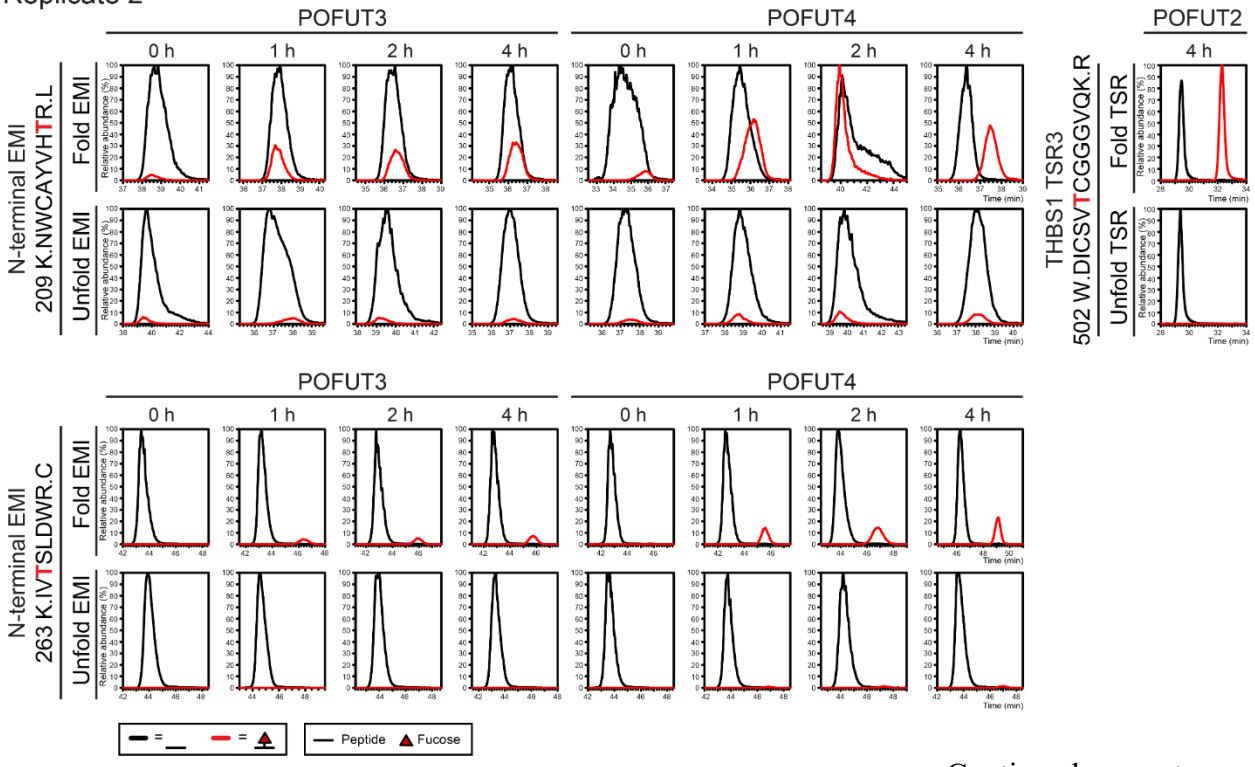


Figure 2S13. Folded and unfolded substrates (N-terminal EMI and THBS1 TSR3) and purified POFUT2 used for EMI unfolding assays that presented in Figure 2.6A. (*A*) Non-fucosylated N-terminal EMI and bacterial produced THBS1 TSR3 were unfolded as described in in Materials and Methods. 5 µg of proteins were mixed with non-denaturing sample buffer (contains 0.04% SDS, 20% glycerol in 100 mM Tris/HCl, pH 6.8) and loaded for Coomassie blue staining. Fold BSA standard: 2 µg BSA mixed with non-denaturing sample buffer; unfold BSA standard: 2 µg BSA mixed with denaturing sample buffer (contains 0.04% SDS, 200 mM 2-mercaptoethanol, 20% glycerol in 100 mM Tris/HCl, pH 6.8) and boiled at 105°C for 5 minutes. (*B*) Purity of the recombinant and purified POFUT2 was verified by Coomassie blue staining.

Replicate 1



Replicate 2



Continued on next page

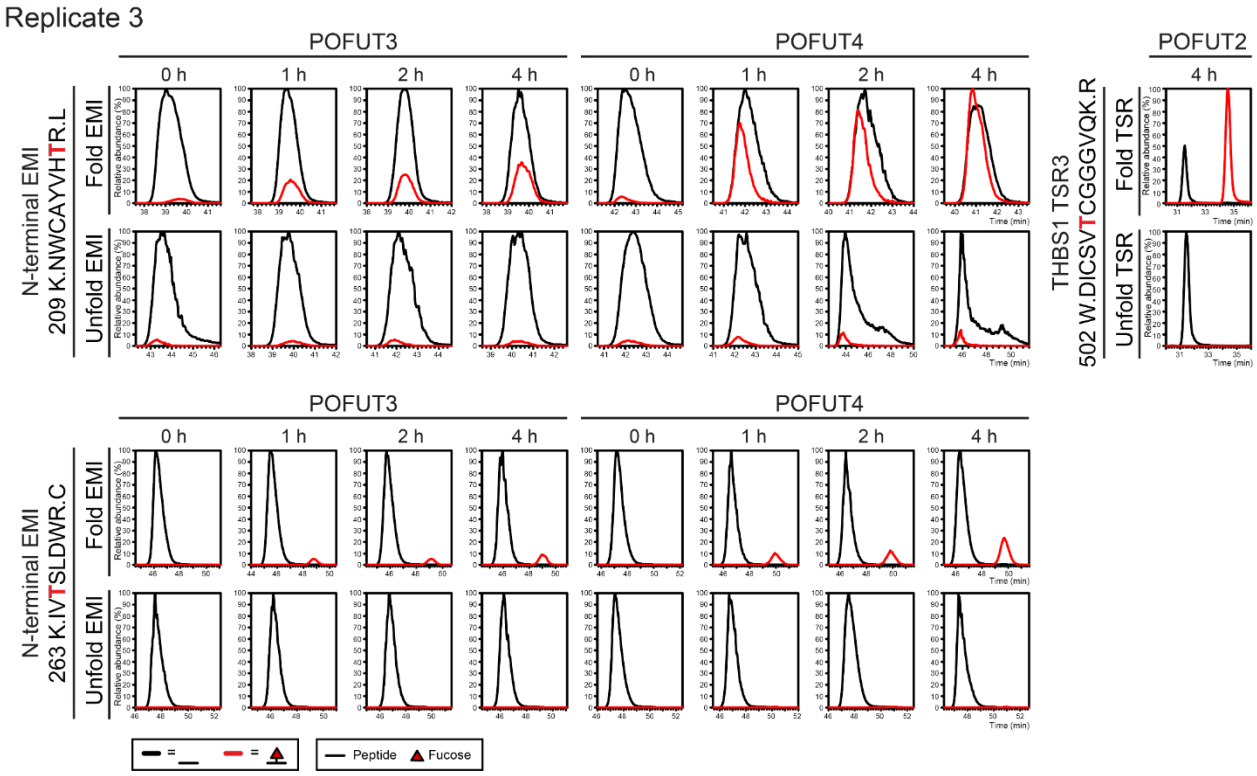


Figure 2S14. Biological triplicates for the data presented in Figure 2.6A. Enzymatic assays of POFUT3 and POFUT4 with folded or unfolded N-terminal EMI substrates. POFUT2 with folded or unfolded THBS1 TSR3 was used as a positive control. Reaction products were analyzed with nano LC-MS/MS and EICs of peptides containing the T216 or T265 *O*-fucose site from N-terminal EMI were generated as described in Materials and Methods. Biological triplicates of data included for the curve plotted in Figure 2.6A were presented.

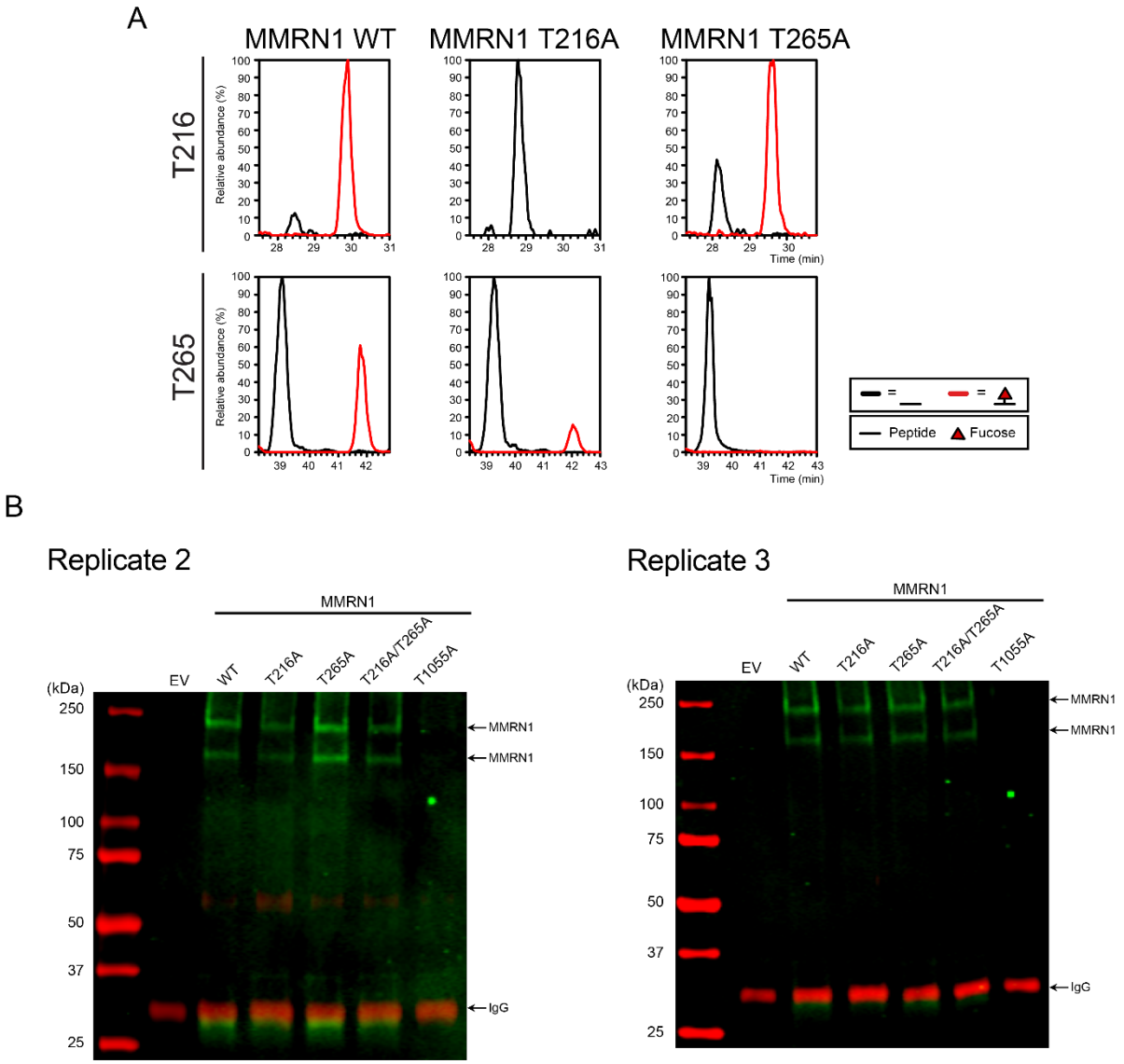


Figure 2S15. Comparing MMRN1 WT with MMRN1 T216A, MMRN1 T265A, MMRN1 T216A/T265A, and MMRN1-T1055A mutants. (*A*) Full-length MMRN1 WT, T216A mutant, and T265A mutant were expressed and purified from HEK293T cells. The *O*-fucosylation status on the T216 and T265 sites on EMI domain was analyzed by nano LC-MS/MS as described in Materials and Methods. (*B*) Secretion assays of MMRN1 WT, MMRN1 T216A, MMRN1 T265A, MMRN1 T216A/T265A, and MMRN1-T1055A in HEK293T cells. In addition to the one replicate used in Figure 2.6C, two additional biological replicates of data included for the statistical analysis in Figure 2.6D were presented.

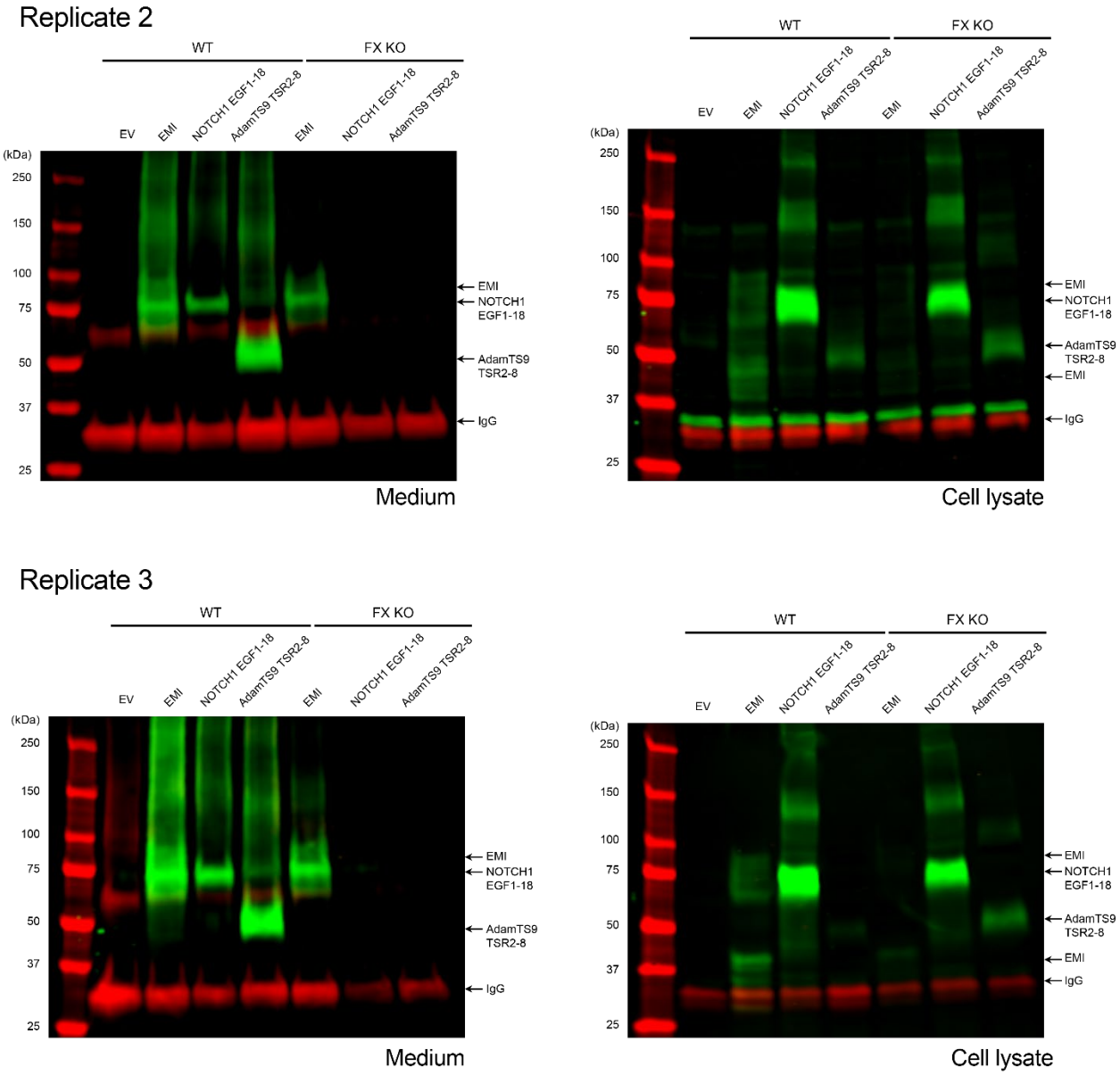


Figure 2S16. Biological triplicates for the data presented in Figure 2.6, E-H. HEK293T WT or FX KO cells were transfected with N-terminal EMI, NOTCH1 EGF1-18, AdamTS9 TSR2-8, or empty vector (EV) and IgG (secretion control). Cells were incubated for 2 days. Culture medium and cell lysates were analyzed by Western Blot probed with anti-Myc and anti-human IgG antibodies. In addition to the one replicate used in Figure 2.6, *E* and *G*, two additional biological replicates of data included for the statistical analysis in Figure 2.6, *F* and *H* were presented.

CHAPTER 3

PROFILING OF *O*-FUCOSE PROTEOME OF POFUT1 AND POFUT2 USING 6-ALKYNYL FUCOSE

Hao H, Liu Z, Wu P, Haltiwanger RS. Manuscript in preparation. To be submitted to Molecular & Cellular Proteomics

3.1 Abstract

Domain-specific *O*-fucosylation by Protein *O*-fucosyltransferase 1 and 2 (POFUT1 and 2) is a rare type of *O*-glycan in mammals. It plays critical roles in numerous biological events by modulating the functions of proteins it modified, including the Notch receptor and ADAMTS family proteins. Systematic identification of intact *O*-fucosylated peptides at the proteomic level remains a key challenge, primarily because the *O*-fucose subpopulation constitutes a very small fraction of all *O*-glycans in cells. Furthermore, unlike other glycan modifications that generate oxonium ions as signatures of glycopeptide spectra in mass spectrometric analysis, *O*-fucose is frequently observed as a neutral loss, complicating the confident assigning of glycopeptides. Here, through metabolic labeling with 6-alkynyl fucose and bioorthogonal conjugation, we 'click on' *O*-fucosylated peptides with an *O*-fucose-specific diagnostic tag, Fuc6nyl*. The conjugated Fuc6nyl* displayed unique mass spectrometric features by generating high-intensity, *O*-fucose-specific diagnostic ions that greatly improved the confident assigning of glycopeptides. Using Fuc6nyl* conjugation coupled with a peptide-centric enrichment strategy, we developed a glycoproteomic workflow for the profiling of the *O*-fucose proteome of POFUT1/2, named Direct detection of Fucose Analog-Tagged peptides (DidFAT). In a proof-of-principle experiment with expressed ADAMTS9 in *MGAT1* KO HEK293F cells lacking complex/hybrid *N*-glycans, we observed robust enrichment of peptides from all fucosylated Thrombospondin Type 1 repeats (TSRs) using DidFAT, with over 90% of the spectra containing the *O*-fucose-specific diagnostic ions. Furthermore, we confidently identified a handful of endogenous proteins with canonical *O*-fucose consensus sequences, including Versican and ADAMTS1. Additionally, we unexpectedly discovered *O*-fucose sites in peptides located within an unannotated Epidermal Growth Factor-like (EGF) repeat of Agrin and within a noncanonical

consensus sequence of Fibrillin-2 (and Fibrillin-1). We envision that DidFAT will provide a much deeper coverage of *O*-fucose proteome of POFUT1/2 and discover more novel targets to study.

3.2 Introduction

Fucose is a common component of glycan structures in mammals. The fucose residues typically occur as terminal modifications on glycan structures. These include core fucosylation of *N*-glycans, in which fucose is added to the innermost GlcNAc moiety of the chitobiose unit, as well as terminal fucosylation on Lewis epitopes and H antigens on the branches of *N*-glycans and Mucin-type *O*-glycans (21, 23). In addition, fucose can be directly linked to proteins via an *O*-linkage to the hydroxyl groups of Serine (S) or Threonine (T) residues located within a specific protein domain (43).

To date, three types of domain-specific *O*-fucosylation are identified in mammals: *O*-fucosylation of Epidermal Growth Factor-like (EGF) repeats containing the consensus sequence $C^2XXXX[\underline{S/T}]C^3$ (where C^2 and C^3 are the second and third cysteine in the EGF repeat, $\underline{S/T}$ is the site of modification) mediated by Protein *O*-fucosyltransferase 1 (POFUT1), *O*-fucosylation of Thrombospondin Type 1 repeats (TSRs) containing the consensus sequence $C^1XX[\underline{S/T}]C^2$ in group 1 TSRs and $C^2XX[\underline{S/T}]C^3$ in group 2 TSRs mediated by POFUT2 (40, 41, 53), and the recently discovered *O*-fucosylation of Elastin Microfibril Interface (EMI) domains by POFUT3 and POFUT4 with a consensus sequence that is yet to be refined (Chapter 2). A growing body of literature demonstrates that *O*-fucosylation plays critical roles in a plethora of physiological and pathological processes. Human heterozygous mutations in *POFUT1* lead to a rare skin condition characterized by pigmentation abnormalities known as Dowling-Degos Disease (57). While no

human *POFUT2* mutations have been reported to date, knockout of either *Pofut1* or *Pofut2* in mice is embryonic lethal (58, 79). The Notch family proteins contain the most predicted *O*-fucose sites of POFUT1 (60). Loss of *Pofut1* in mice embryos causes developmental defects mirroring those seen in *Notch1* knockout, suggesting Notch as the primary biological target of POFUT1 in embryogenesis (58). Nearly half of the predicted POFUT2 targets belong to the A-Disintegrin and Metalloproteinase with ThromboSpondin Type-1 motifs (ADAMTS) family. Knockout of *Pofut2* in mice embryos causes phenotypes mirroring *Adamts9* mutants, suggesting POFUT2-mediated *O*-fucosylation is essential for its biological function in early embryogenesis (80). Moreover, dysregulated *O*-fucosylation is associated with tumorigenesis and poor prognosis in numerous cancers (100, 149).

Database searches with POFUT1 consensus sequence in the context of an EGF repeat reveal that 87 proteins are predicted to be modified by POFUT1 (23). Similarly, using POFUT2 consensus sequence in the context of a TSR, 49 proteins are predicted to be modified by POFUT2. While the presence of consensus sequences increases the likelihood for *O*-fucosylation, it does not necessarily guarantee that these sites will be modified. The *O*-fucose modification status of most of the putative targets remains unconfirmed, yet they may have important roles in the biological processes regulated by POFUT1 and POFUT2. Furthermore, proteins with domains other than EGF repeats or TSRs can be *O*-fucosylated, as exemplified by the recently identified *O*-fucosylation of EMI domains, raising the possibility that proteins outside the list can be modified by POFUT1 and POFUT2 (Chapter 2) (126). This underscores the necessity to develop an unbiased method to confirm these putative targets and to identify novel targets outside the list.

We previously demonstrated that through metabolic labeling, peracetylated 6-alkynyl fucose (6-AF) is passively taken up by cells, converted into GDP-6-AF in the salvage pathway, and

transported into the endoplasmic reticulum (ER) (138-140). In the ER, GDP-6-AF is efficiently used by POFUT1/2 and incorporated into their respective protein targets. Note that 6-AF is not incorporated into EMI domains by POFUT3/4; instead, it functions as an inhibitor, preventing the addition of *O*-fucose on EMI domains (Chapter 2). Through this process, a chemical reporter tag (the alkynyl group) is incorporated into the *O*-fucose proteome of POFUT1/2. This enables the conjugation of functionalized biotin probes using biorthogonal chemistry, thus allowing for the detection and enrichment of labeled POFUT1/2 targets.

Leveraging 6-AF as a tool, we aim to build a glycoproteomic workflow for the in-depth profiling of intact *O*-fucosylated peptides, thereby enabling the mapping of the *O*-fucose proteome of POFUT1/2. While the mass spectrometry (MS)-based glycoproteomic methodologies are well-established across the community, the unambiguous assignments of intact *O*-fucosylated peptides at the proteome level still presents several challenges (169). In contrast to other *O*-glycosylation types like *O*-GalNAc and *O*-GlcNAc, *O*-fucose glycans represent a significantly lower proportion of all *O*-glycans (170). Furthermore, most of the predicted targets of POFUT1/2 are large membrane or extracellular proteins with multiple domains, posing a challenge for the direct detection of fucose modifications on peptides using conventional methods, primarily due to ion suppression from abundant nonglycosylated peptides (23, 171). Additionally, unlike other glycan modifications that generate oxonium ions as signatures of glycopeptide spectra, *O*-fucose is frequently observed as a neutral loss, adding complexity to the confident assignment of glycopeptides. Technical improvements are needed to ensure the efficient enrichment of *O*-fucosylated species and a robust MS-based approach for the unambiguous identification of *O*-fucosylated peptides.

Herein, we report a glycoproteomic workflow based on 6-AF metabolic labeling for the profiling of *O*-fucose proteome of POFUT1/2, named Direct detection of Fucose Analog-Tagged peptides (DidFAT). Inspired by the Direct Detection of Biotin-containing Tags (DiDBiT) strategy, we performed proteolytic digestion of proteins prior to enrichment to increase the yield of enrichment and enable the direct detection of modified peptides in MS analysis (172). A cleavable biotin enrichment handle (diazo biotin azide) was conjugated to 6-AF-labeled proteins through bioorthogonal click chemistry. Chemical cleavage specifically released the labeled peptides from enrichment matrix while leaving a signature residue on the peptides. Interestingly, this residue displays unique mass spectrometric features by generating high-intensity, *O*-fucose-specific diagnostic ions, which greatly improves glycoproteomic analysis. Robust enrichment of *O*-fucosylated peptides were observed in a proof-of-principle experiment with expressed ADAMTS9 in the secretome of *MGAT1* KO HEK293F cells, with over 90% of peptide-spectrum matches (PSMs) containing the diagnostic ions. A handful of endogenous proteins were identified with fucose on their predicted sites. Moreover, unexpected *O*-fucose sites were identified on Fibrillin-2 in a noncanonical consensus sequence, and Agrin, which may present an unannotated EGF domain.

3.3 Results

Metabolic labeling of POFUT1/2 targets in cell secretome

6-alkynyl fucose (6-AF) was previously shown to be able to efficiently incorporate into POFUT1 and POFUT2 targets, such as NOTCH1 and THBS1, respectively (138-140). However, a portion of 6-AF can be incorporated into the core-fucosylation sites on *N*-glycans (138, 173). We hypothesized that the *O*-fucose proteome of POFUT1/2 could be obtained through metabolic

labeling of cells with 6-AF, followed by the removal of all *N*-glycans (Figure 3.1A). As most putative targets of POFUT1/2 are membrane or secreted proteins within the extracellular matrix, we started by evaluating the incorporation efficiency of 6-AF in cell secretome. 6-AF was peracetylated to enhance its cellular permeability. HEK293T cells were treated with peracetylated 6-AF for 2 days, the conditioned culture media was collected, concentrated, desalted, and used as 6-AF-labeled secretome. PNGase F digestions were performed to remove the *N*-glycans. To assess the incorporation efficiency of 6-AF and visualize the labeling pattern of 6-AF-incorporated proteins, secretome samples were derivatized with diazo biotin azide using Copper (I)-catalyzed azide-alkyne cycloaddition (CuAAc or Click chemistry) and analyzed on a streptavidin blot.

Cells transiently transfected with ADAMTS9 TSR2-8 and Lfng were used as positive controls to monitor the *O*-fucose glycan labeling and PNGase F digest efficiency, respectively. ADAMTS9 TSR2-8 is a known substrate of POFUT2 that contains multiple TSRs with *O*-fucose sites, while Lfng possesses a single complex *N*-glycan with core-fucose (Figure 3.1B). Secretome from untransfected 6-AF-treated cells showed robust labeling compared to the DMSO-treated control cells (Figure 3.1C). Both ADAMTS TSR2-8 and Lfng that expressed in 6-AF-treated cells showed robust labeling. The labeling of Lfng was completely eliminated by PNGase F treatment, accompanied by a reduction in its molecular weight, confirming the successful removal of 6-AF-labeled *N*-glycans. Notably, cleaving off the *N*-glycans in secretome by PNGase F resulted in a remarkable reduction of over 90% in the 6-AF labeling of endogenous proteins, suggesting that the *O*-fucose proteome exists in scarce quantities.

'Click on' the POFUT1/2 targets with an *O*-fucose-specific diagnostic tag (Fuc6nyl*)

Enrichment is necessary for identifying the low-abundance POFUT1/2 targets within a complex protein mixture such as the cell secretome. We derivatized the alkyne moieties in 6-AF with a cleavable biotin enrichment handle (diazo biotin azide) using click chemistry, thus enabling the selective release of *O*-fucosylated moieties from enrichment matrix (Figure 3.3A). The diazo biotin linker that conjugated to 6-AF can be cleaved using sodium dithionite, leaving a residual tag on 6-AF (Fuc6nyl*). To evaluate the performance of this residual tag in mass spectrometry, we performed a comparative mass spectrometric analysis of 6-AF-labeled peptides with and without click/cleave (Figure 3.2A). Myc-tagged ADAMTS9 TSR2-8 and Lfng were expressed in HEK293T cells metabolically labeled with 6-AF. Recombinant proteins from conditioned culture media were captured using anti-Myc agarose, followed by on-bead click conjugation with diazo biotin azide or mock conjugation with DMSO. Captured proteins were eluted off anti-Myc agarose using urea, followed by reduction, alkylation, and trypsin digestion. Digested peptides were subjected to sodium dithionite cleavage before analysis by nano LC-MS/MS. We noticed that the peptides conjugated with Fuc6nyl* consistently possessed one additional charge of their most abundant precursor ions compared to the unconjugated peptides, either mono fucosylated or elongated with Glc (Figure 3.2B and Figure 3S1). We propose the amino group within the conjugated Fuc6nyl* moiety could obtain an additional proton, thus leading to a higher precursor charge state in MS1 spectra (Figure 3.3A). Note that the conjugated biotin linker may introduce additional dynamics to the attached protein. A higher elution level of ADAMTS9 TSR2-8 in the click/cleave sample compared to the mock sample was observed, while Lfng showed consistent elution levels (Figure 3S2). This potentially explains the greater intensity of precursor ions observed for Fuc6nyl* conjugated peptides. When comparing different glycoforms of identical

peptides in extracted ion chromatograms (EICs), we observed a higher relative abundance of labeled peptides in the Fuc6nyl* conjugated samples in comparison to the unconjugated samples (Figure 3.2B). This suggests that the conjugation of Fuc6nyl* facilitated the detection of labeled peptides in mass spectrometric analysis.

We serendipitously observed that the conjugated Fuc6nyl* residues led to unique signature ions in MS2 spectra after fragmentation by higher energy collision dissociation (HCD). An abundant fragment ion at 335.1355 m/z, along with a second abundant ion at 136.0762 m/z were detected in the HCD spectra of peptides from ADAMTS9 TSR2-8, mono fucosylated or elongated with GlcNAc (Figure 3.3B). These signature ions were reliably present in spectra of Fuc6nyl* conjugated peptides throughout our analyses. We propose that the signature ion at 335.1355 m/z is the result of a proton accepted by the amino group within the Fuc6nyl* residue, followed by a neutral loss of the fucose moiety from the peptide backbone during HCD (Figure 3.3A). The smaller signature ion at 136.0762 m/z is likely a secondary fragment derived from the primary ion at 335.1355 m/z. Interestingly, these signature ions were predominantly detected in the MS2 spectra of conjugated *O*-fucosylated peptides. In contrast, only a handful of spectra from the conjugated *N*-glycosylated peptides contain those ions, and their intensity was markedly low, likely due to the dissipation of kinetic energy by cleaving off terminal sugars like sialic acid and galactose (Figure 3.3C). We envision those *O*-fucose-specific diagnostic ions will lead to more confident assignment of *O*-fucosylated peptides in glycoproteomic analysis.

Above all, these findings indicate that the conjugation of Fuc6nyl* introduces additional benefits in both the enrichment and glycoproteomic analysis, including aiding peptide release from the enrichment matrix, increased charge intensity of precursor ions, increased relative abundance of

labeled peptides for easy detection, and generating unique diagnostic ions for confident glycopeptide assignment.

Build a 6-AF based glycoproteomic workflow: Direct **d**etection of **F**ucose **A**nalog-**T**agged **peptides (DidFAT)**

Our aim was to build a dedicated glycoproteomic workflow for the unambiguous identification of *O*-fucosylated proteins/peptides in the secretome. The major challenge is the very-low abundance of *O*-fucose proteome that decreases the chances of identifying them by MS analysis (170). Taking inspiration from the DiDBiT method by Schiapparelli et al., we employed a peptide-centric strategy that enriches biotin-labeled peptides instead of proteins to enhance enrichment yield and enable the unambiguous assignment of *O*-fucosylated peptides (172). 6-AF labeled secretome samples were conjugated with diazo biotin azide via click chemistry, followed by in-solution trypsin digestion to break down the entire secretome into peptides. (Figure 3.4A) The resulting peptides conjugated with biotin tags were captured using NeutrAvidin beads, selectively released with sodium dithionite, and analyzed by nano LC-MS/MS for peptide identifications. To monitor the enrichment and cleavage efficiency, we performed a protein-centric capture in parallel. Biotinylated proteins were enriched by NeutrAvidin beads and selectively released into elution with sodium dithionite. The eluted proteins were subjected to trypsin digestion for protein identification by nano LC-MS/MS for protein identification. Additionally, they were also loaded onto a streptavidin blot to evaluate the efficiency of each step in the workflow.

First, we validated the DidFAT workflow through a proof-of-principle experiment using transfected positive controls, ADAMTS9 TSR2-8 and Lfng. To minimize the background

interference from *N*-glycan labeling, we used *MGAT1* KO HEK293F cells which lack complex and hybrid *N*-glycans, as evidenced by the absence of 6-AF labeling of Lfng expressed in these cells and the consistent labeling patterns of secretome samples after PNGase F treatment (Figure 3.4, B and C). We observed robust labeling of ADAMTS9 TSR2-8 expressed in 6-AF-treated cells (Figure 3.4B). In addition, a distinct labeling pattern of 6-AF-incorporated endogenous proteins in untransfected cells was observed. Efficient NeutrAvidin enrichment was confirmed by the absence of all biotinylated proteins including the Myc-tagged ADAMTS9 TSR2-8 in the flowthrough, while the unlabeled Myc-tagged Lfng remained (Figure 3.4C). Efficient protein release from the beads was confirmed by the loss of biotin signal and the detection of Myc-tagged ADAMTS9 TSR2-8 in the elution. Excluding sodium dithionite from the elution buffer prevented successful protein release, confirming the specificity of chemical cleavage.

The peptide-centric DidFAT strategy identified 3497 PSMs from ADAMTS9 protein, covering all TSRs containing *O*-fucose consensus sequences (Figure 3.4D). Over 90% of the PSMs identified were modified with Fuc6nyl* and displayed diagnostic ions in their spectra, as most of the PSMs were also identifiable using Byonic search with MS2 Peak Filtering, which exclusively retained MS2 spectra containing peaks at 335.1355 m/z or 136.0762 m/z. In contrast, the protein-centric capture method identified 544 PSMs for ADAMTS9 protein using equivalent amount of secretome input, with the majority of identified PSMs remaining unmodified. These results indicate that DidFAT dramatically enhanced the identification of *O*-fucosylated peptides compared to conventional methods, highlighting the benefits from Fuc6nyl* conjugation and peptide-centric capture strategy.

Endogenous POFUT1/2 targets identified in secretome using DidFAT

A total number of 4033 PSMs were identified from the secretome of *MGAT1* KO HEK293F cells using the peptide-centric DidFAT strategy. In addition to the 3497 PSMs from the transfected ADAMTS9, 536 PSMs were endogenously derived from the secretome and distributed on 29 proteins (Figure 3.5 and Table 3.1). Notably, 18 of these proteins are also present in the putative POFUT1/2 target library [based on database searches with POFUT1/2 consensus sequences from (23)]. Thirteen proteins containing EGF repeats or TSRs were identified with Fuc6nyl* incorporated on their predicted sites (Table 3.2 and Figure 3S3). These modified peptides contain the canonical POFUT1/2 *O*-fucose consensus sequences, including those from Agrin (EGF4), Versican (EGF2), ADAMTS1 (TSR1 and TSR2), CCN3 (TSR1), NOTCH3 (EGF11), NELL1 (EGF3), Papillin (TSR3), Sushi (EGF6), ADAMTS3 (TSR4), and NOTCH1 (EGF35). In addition, peptides with Fuc6nyl* modification, but lacking a canonical *O*-fucose consensus sequence were also confidently identified, such as those from Agrin, Fibrillin-1, and Fibrillin-2, which will be discussed in the following section (Table 3.2 and Figure 3S3). Among the proteins identified, Agrin, Versican, and Fibrillin-2 consistently appeared on our list and had the most abundant identified PSMs, indicating that they are likely the predominant POFUT1/2 targets in the secretome. All the identified proteins were extracellular matrix-resident proteins, with the exception of NOTCH1 and NOTCH3. These membrane proteins are likely shed from the cell membrane through protease cleavage during activation, allowing them to appear in the secretome and be detected. Agrin, Fibrillin-1, Fibrillin-2, ADAMTS1, and NOTCH3 were also detected using the parallel protein-centric capture method, thus providing additional confirmation of the *O*-fucosylated status of these proteins (Figure 3.5).

The protein-centric capture method identified a total of 107 proteins including the transfected ADAMTS9 (Figure 3.5). However, only 8 of these proteins are shared with the putative POFUT1/2 target library, and the majority of the identified PSMs are unmodified. This emphasizes the necessity for extra MS analyses on mock experiments and the requirement of *stringent threshold of acceptance for protein assignment* in conventional protein-centric enrichment methods, in order to distinguish the real 'hits' from non-specific contaminants. With the DidFAT method, nearly all the identified PSMs were conjugated with Fuc6nyl* and featured diagnostic ions in their MS2 spectra, allowing for the unambiguous assignment of *O*-fucosylated peptides/proteins and minimizing the need for time-consuming validation in glycoproteomic analysis.

Noncanonical POFUT1/2 targets identified using DidFAT

Agrin and Fibrillin-2 rank among the top three proteins identified with the highest PSM number. Yet, apart from the peptide from EGF4 of Agrin that contains the POFUT1 consensus sequence, all other identified peptides from Agrin and Fibrillin-2 lack a canonical *O*-fucose consensus sequence (Figure 3.6A, 3.7A and Table 3.3).

The heparan sulphate proteoglycan Agrin is a key regulator of postsynaptic differentiation at the neuromuscular junction (NMJ). Agrin exhibits a complex domain organization, including 9 follistatin-like domains (FS) at the N-terminus, 4 EGF repeats at the C-terminus, and additional domains, such as 2 laminin EGF-like domains (LE), 3 lamin globular domains (LG), and a single SEA domain (Figure 3.6A). *O*-fucosylation of EGF4 has been confirmed by Kim M. L. et al., 2008, and is shown to determine the acetylcholine receptor (AChR) clustering activity of Agrin (124). We also detected peptides modified with Fuc6nyl* from EGF4 using DidFAT (Figure

3S3). Surprisingly, an additional *O*-fucose site was unexpectedly found at the N-terminus of Agrin. A substantial number of PSMs with mono Fuc6nyl* modification were identified from peptide ^{246}DPCSNVTCSFGSTCAR^{261}, where 'T' represents the site of Fuc6nyl* modification (Figure 3.6B). This *O*-fucose site locates between the first and second FS domain (Figure 3.6A). Upon manual examination of the surrounding sequences, we were surprised to discover an unannotated EGF repeat that begins at the end of the first FS domain and extends to overlap with part of the second FS domain. Notably, there are only three amino acids in between the fifth and sixth cysteines, possibly explaining why this EGF was not annotated in UniProt (Table 3.3).

Fibrillin-2 is a major component of connective tissue microfibrils. It contains 47 EGF repeats, and over half of these EGF repeats contain the *O*-glucose consensus sequence C^{3}XNTXGS(F/Y)XC^4 for POGLUT2/3 modification (174). Only EGF2 and EGF22 contain the *O*-fucose consensus sequence C^{2}XXXX[S/T]C^3 for POFUT1 modification (Figure 3.7A). However, the only *O*-fucose site we identified using DidFAT locates in EGF34 with a sequence motif C^{2}XXX[S/T]C^3 (Figure 3.7B). This motif differs from the canonical *O*-fucose sequence, with only three amino acids in between C2 and the modified site. All PSMs identified for peptide ^{2013}NCIDTNECVALPGSCSPGTCQNLEGSFR^{2040} from EGF34 were modified with mono Fuc6nyl* (Table 3.2). In addition to EGF34, several other EGF repeats also contain this noncanonical sequence, suggesting possible *O*-fucose modification on these EGF repeats (Table 3S1). This noncanonical *O*-fucose site on EGF 34 was also found on the equivalent position of Fibrillin-1 (Table 3.2 and Figure 3S3).

3.4 Discussion

O-fucosylation is a rare type of glycosylation in mammals. Unlike other forms of fucosylation by Golgi-localized fucosyltransferases that add fucose as core or terminal modifications, *O*-fucosylation by POFUT1/2 requires the recognition of specific protein domains for fucose addition (72, 129). The most renowned and extensively studied proteins with *O*-fucose are the Notch family proteins (60). The extracellular domain of Notch is heavily decorated with fucose glycans, which play important roles in regulating Notch-ligand binding, trafficking and cell surface expression (67, 68, 130, 175). Unlike many other glycan modifications, fucose residues display thermal mobility similar to amino acids, acting as 'surrogate amino acids' to make specific and essential contacts with the residues on binding partners. This direct intermolecular interactions generated by fucose glycans in protein-protein interaction were observed in co-crystals of portions of NOTCH1 with ligands and the co-crystal of BAI1 with RTN4-receptor (67, 68, 125). In addition, fucose residues generate intramolecular interactions with neighboring amino acids within protein domains to stabilize folded EGF repeats and TSRs, preventing them from re-entering the folding cycle (130, 131). Only properly folded domain structures are recognized and modified by POFUT1/2, suggesting that *O*-fucosylation is a component of a non-canonical quality control system designed for the efficient folding of EGF repeats/TSRs-containing proteins (119, 130). Through intensive site-mapping and funtional study of Notch and ADAMTS family members, our knowledge on the molecular mechanisms by which *O*-fucose regulates protein functions has greatly expanded. Yet, a systematic characterization of all *O*-fucosylated proteins and peptides is essential to gain deeper insights into the *O*-fucosylation process.

Here, an *O*-fucose-specific tag (Fuc6nyl*) was introduced to POFUT1/2 targets through 6-AF metabolic labeling coupled with bioorthogonal derivation. The conjugation of Fuc6nyl* brings additional benefits for glycoproteomic analysis, including increased precursor ion charge intensity, increased relative abundance of labeled peptides, and generating unique diagnostic ions for confident glycopeptide assignment. This prompted us to develop a glycoproteomic workflow for identifying intact *O*-fucosylated peptides at the proteome level. Using Fuc6nyl* conjugation coupled with a peptide-centric enrichment strategy, we have established the DidFAT workflow specifically designed for in-depth profiling of the *O*-fucose proteome of POFUT1/2. This approach allowed us to confirm the *O*-fucose modification of a handful of proteins with canonical *O*-fucose consensus sequences and discovered novel *O*-fucose sites on peptides that lack *O*-fucose consensus sequences. Unexpected *O*-fucose sites were identified in peptides located within an unannotated EGF repeat of Agrin and within a noncanonical consensus sequence of Fibrillin-2 (and Fibrillin-1).

Prior work in our group have provided solid evidence of the efficient incorporation of 6-AF into substrates of POFUT1/2 through metabolic labeling, demonstrating the potential of 6-AF as a specifc bioorthogonal chemical reporter for *O*-fucose glycans (138-140). It is important to note that 6-AF is a bifunctional tool that serves as both a probe and an inhibitor of fucosylation. Kizuka et al. showed that the active metabolite GDP-6-AF competitively inhibits GDP-fucose synthetase FX, leading to the depletion of cellular GDP-fucose pool (137). Metabolic labeling of cells with 6-AF might disrupt other fucosylation events, potentially affecting global cellular functions. As a result, the observed *O*-fucosylation might not accurately represent the actual state. In addition, the efficiency of 6-AF incorporation and the accessibility of click reagents can differ among individual EGF repeats/TSRs. The direct comparison of *O*-fucose abundance across

different peptides might not be accurate. It would be beneficial to examine the *O*-fucosylation status and abundance of peptides identified by DidFAT in unperturbed cells. Nevertheless, DidFAT remains an elegant and efficient workflow for examining *O*-fucosylation of predicted targets and discovering novel targets with *O*-fucose.

We observed a limited number of hits in the secretome of *MGAT1* KO HEK293F cells, contrary to our expectations of a higher yield. This is possibly due to the low abundance of *O*-fucosylated speices and the inherent dynamic range limititations of data-dependent acquisition (DDA) for collecting mass spectrometric data. Fuc6nyl*-conjugated peptides display predictable fragmentation behavior, generating specific diagnostic ions at 335.1355 and 136.0762 m/z. These unique fingerprints could be employed to trigger product-dependent (pd) HCD and/or electron transfer dissociation (ETD), boosting the collection of fucosylated spectra and/or site-localize the modifications.

HEK293F cells might not be the suitable cell line for the in-depth profiling *O*-fucosylated proteins. Applying DidFAT to cell lines that exhibit higher levels of glycoprotein expression and secretion, such as fibroblasts, platelets, and cancer cells with elevated POFUT1/2 levels, may result in a greater number of hits. In addition, many of the predicted targets for POFUT1 are membrane proteins. Applying DidFAT in cell lysates could potentially yield more hits on these proteins.

3.5 Experimental Procedures

Cell culture

HEK293T cells (ATCC) were cultured at 37 ºC with 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, GE Healthcare Life Sciences) supplemented with 10% (v/v) bovin calf serum

(BCS, VWR), 100 units/mL penicillin, and 100 µg/mL streptomycin (Lonza). *MGAT1* KO HEK293F cells (a gift from Kelley Moremen, University of Georgia, US) were cultured at 37ºC with 5% CO_2 in Freestyle293^{TM} medium (Gibico) with 100 units/mL penicillin and 100 µg/mL streptomycin.

Metabolic labeling of cell secretome

Metabolic labeling of HEK293T cells

6 X 10^6 of HEK293T cells were seeded and grown to 70% confluence in 10-cm dishes. Cell layer was washed one time with 5 mL of prewarmed phosphate-buffered saline (PBS). Cell medium was switched with 8 mL of prewarmed OPTI-MEM (Invitrogen) containing 100 µM peracetylated 6-alkynyl fucose (6-AF) or equal volume of DMSO. For transfection of cells with positive controls, each plate of cells was transiently transfected with 2 µg of pSec-h ADAMTS9 TSR2-8-Myc-His_6 (121) and 0.5 µg of pSec-mLfng-Myc-His_6 (176) using 15 µL of PEI (1 µg/µL stock) in 8 mL of OPTI-MEM containing 100 µM 6-AF or equal volume of DMSO. Three days later, media from 10 plates of cells was collected and filtered through 0.45 µm filters (Millipore). 80 mL of media was desalted and concentrated to ~5 mL using 10 kDa molecular weight cutoff centrifugal filters (Amicon) and used as secretome samples. Concentrated secretome samples were aliquoted and stored at -80 ºC until use.

Metabolic labeling of MGAT1 KO HEK293F cells

50 X 10^6 of cells were resuspended in 25 mL of Freestyle293^{TM} media containing 100 µM 6-AF or equal volume of DMSO. Three days later, cell suspension was centrifuged at 5000 rpm for 5 minutes. Culture medium was collected and filtered through 0.45 µm filters (Millipore). For transfection of cells with positive controls, 50 X 10^6 of cells were resuspended in 12.5 mL of 9:1

medium [Freestyle293^{TM} medium: EX-Cell® medium (Sigma Aldrich), v/v] containing 100 µM 6-AF or equal volume of DMSO, and transiently transfected with 12.5 µg of pSec-h ADAMTS9 TSR2-8-Myc-His_6 and 6.25 µg of pSec-mLfng-Myc-His_6 with 112.5 µL of PEI (1 µg/µL stock). After 24 hours, Valproic acid (Sigma Aldrich) was supplemented to cells to a final concentration of 2.2 mM in 12.5 mL of 9:1 medium containing 100 µM 6-AF or equal volume of DMSO. After 2 more days, cell suspension was centrifuged at 5000 rpm for 5 minutes. Culture medium was collected and filtered through 0.45 µm filters. 25 mL of media was desalted and concentrated to ~2.5 mL using 10 kDa molecular weight cutoff centrifugal filters and used as secretome samples. Concentrated secretome was aliquoted and stored at -80 ºC until use.

Click chemistry

Click reaction was performed using 100 µL of concentrated secretome sample in 200 µL reaction volume containing 100 µM diazo biotin azide (Sigma Aldrich, 2 mM stock in DMSO), Premixed THPTA:$CuSO_4$ complex [final concentration of 2 mM THPTA (Click Chemistry Tools, 25 mM stock in water):1 mM $CuSO_4$ (5 mM stock in water)], and 5 mM sodium ascorbate (50 mM stock in water, freshly prepared). Click reactions were carried out at room temperature for 2 hours and quenched by addition of 50 mM EDTA. Reaction mixtures were then precipitated by adding four volumes of methanol, one volume of chloroform, and three volumes of water, vortexed, and centrifuged at 12,700 rpm for 20 minutes at 4 ºC. The aqueous upper layer was carefully removed without disturbing the protein layer at the interface. Four volumes of methanol were added, and the mixture was centrifuged at 12,700 rpm for 15 minutes at 4 ºC. An additional wash with four volumes of methanol was performed to ensure the complete

removal of any remaining diazo biotin azide. Pellets were air-dried for 4 minutes before used in subsequent experiments.

PNGase F digestion

After biotinylation by click chemistry and methanol/chloroform precipitation, protein pellets obtained after centrifuging were resolubilized in 20 µL of reduction buffer [1% SDS, 10 mM TCEP (Thermo fisher), and 0.4 M NH_4HCO_3] with vigorous shaking. Proteins were fully denatured by boiling at 105 ºC for 5 minutes. After cooling to room temperature, 10 µL of 100 mM Iodoacetamide (Thermo fisher) was added, and proteins were incubated in dark for 40 minutes for alkylation. 170 µL of 1% NP-40/Tris-saline buffer, pH 7.6 (TBS) was added to dilute SDS to a final concentration of 0.1%. Digestion was performed with 1.5 µL of PNGase F (20 units/µL, Lectez Bio) or 1.5 µL of water (for mock digestion) and incubated in a 37 ºC water bath for over 6 hours. Digested proteins were precipitated with 1 mL of cold acetone and incubated at -20 ºC for 2 hours. Protein pellets were collected by centrifuging at 12,700 rpm for 15 minutes at 4 ºC. Pellets were air-dried for 4 minutes before resolubilized for Western Blot analysis.

Western Blot

Protein pellets were resuspended in 12 µL of 2X denaturing buffer containing 0.04% SDS, 200 mM 2-mercaptoethanol (Sigma Aldrich), 20% glycerol in 100 mM Tris/HCl, pH 6.8 with vigorous shaking, and then sonicated for 6 minutes using a bath sonicator. Samples were boiled at 105 ºC for 5 minutes. After cooling to room temperature, samples were loaded onto 4-20% SDS-PAGE (Bio-Rad), transfected to a nitrocellulose membrane, and then blocked with 5%

nonfat milk (Bio-Rad) for 30 minutes at room temperature. Membranes were incubated with primary antibodies in blocking solution overnight at 4 ºC. Membranes were washed three times in Tris-saline buffer, pH 7.6 and 0.05% Tween 20 (TBST) for 5 minutes and incubated with secondary antibodies in TBST for 1 hour at room temperature. The following antibodies were used in this study: anti-Myc antibody (primary, 1:2500, Invitrogen, clone 9E10), anti-His-tag antibody (primary, 1:2500, Bio-Rad, MCA1396), IRDye 800CW goat anti-mouse IgG antibody (secondary, 1:2500, LI-COR, 926-32210), IRDye 680RD goat anti-mouse IgG antibody (secondary, 1:2500, LI-COR, 926-68070), IRDye 800CW Streptavidin (secondary, 1:2500, LI-COR, 926-32230). Blots were visualized using Odyssey System (LI-COR).

Comparative analysis of 6-AF-labeled peptides with and without Fuc6nyl* conjugation

Concentrated secretome samples from 6-AF metabolic labeled HEK293T cells transfected with positive controls (Myc tagged- ADAMTS9 TSR2-8 and Myc tagged-Lfng) were generated as described above. 500 µL of secretome sample was mixed with 300 µL slurry of anti-c-Myc agarose (25% slurry, pre-washed one time with 1% NP-40/TBS, Thermo Fisher, #20168). 1% NP-40/TBS was added to reach a final volume of 1.9 mL. Immunoprecipitation (IP) was performed overnight at 4 ºC. Beads were collected using spin columns (VWR centrifugal filters) and washed three times with 500 µL of 1% NP-40/TBS for 15 minutes at room temperature. On-bead click reactions were performed by adding a pre-mixed click mixture to the beads (~75 µL bead volume), which included 20 µL of diazo biotin azide (2 mM stock in DMSO), premixed THPTA:$CuSO_4$ complex (32 µL: 80 µL, THPTA: 25 mM stock in water, $CuSO_4$: 5 mM stock in water), 40 µL of sodium ascorbate (50 mM stock in water, freshly prepared), and 100 µL of 1% NP-40/TBS. For samples with mock click reactions, 20 µL of DMSO was used in the click

mixture while keeping all other conditions unchanged. Click reactions were carried out at room temperature for 2 hours. Beads were washed three times with 500 µL of 1% NP-40/TBS for 20 minutes at room temperature. To elute the captured proteins, 200 µL of 8M urea in 0.4 M NH_4HCO_3 was added to the beads. Beads were rotated at room temperature for 10 minutes and subsequently incubated at 65 ºC for 10 minutes before collecting the elution. A second elution was performed using 100 µL of water, followed by rotating the beads at room temperature for 10 minutes. The two elutions were pooled and supplemented with 6 µL of TCEP (0.5 M stock), followed by incubating at 60 ºC for 10 minutes to ensure complete protein reduction. After cooling to room temperature, 150 µL of 100 mM Iodoacetamide was added, and proteins were incubated in dark for 40 minutes for alkylation. 750 µL of water was added to dilute urea to a final concentration of 1.3 M in a total volume of 1.2 mL. Proteins were subjected to digestion by adding 3 µL of trypsin (0.5 µg/µL stock, Thermo Fisher, 90057) and incubated in a 37 ºC water bath for 6 hours. Chemical cleavage was performed by adding 700 µL of freshly prepared 136 mM $Na_2S_2O_4$/3.2 M urea/136 mM Tris, pH 8.0 (final concentration: 50 mM $Na_2S_2O_4$, 2 M urea in 50 mM Tris, pH 8.0) to digested peptides, followed by incubating at room temperature for 1 hour. An equal volume of 1% formic acid (FA) was added to achieve a final FA concentration of 0.5%. Peptides were sonicated in a bath sonicator for 20 minutes and then centrifuged at 12,700 rpm for 8 minutes at 4 ºC. Supernatants were collected. Peptides were desalted with C18 Zip Tip Pipette Tips (Millipore) and analyzed with nano LC-MS/MS using the methods described later.

DidFAT protocol

Peptide-centric enrichment of biotinylated peptides

100 µL of concentrated secretome was biotinylated using click chemistry and precipitated with methanol/chloroform precipitation, following the methods described earlier. Protein pellets were resuspended in 200 µL of reduction buffer containing 8 M urea, 10 mM TCEP, and 0.4 M NH_4HCO_3 with vigorous shaking, followed by sonicating for 10 minutes in a bath sonicator. The protein solution was incubated at 65 ºC for 10 minutes. Alkylation was performed by adding 100 µL of 100 mM Iodoacetamide to a final concentration of 33.3 mM and incubating in the dark for 40 minutes. Protein solutions were then diluted with 900 µL water and digested with 4 µg of trypsin (2 µg enzyme per 100 µL secretome) in 37 ºC water bath overnight. Samples were boiled at 105 ºC for 5 minutes and cooled down to room temperature. For cleaving off *N*-glycans, 3 µL of PNGase F (20 units/µL stock, 1.5 µL enzyme per 100 µL secretome) was added, and samples were incubated for 6 hours in 37 ºC water bath. Digested samples were acidified with trifluoroacetic acid (TFA, Thermo Fisher) to 0.1% final concentration, sonicated for 15 minutes, and centrifuged at 12,700 rpm for 20 minutes at room temperature. Supernatant containing digested peptides was collected. 200 µL of 0.1% TFA was added to pellets to extract any remaining peptides. Pellets were resuspended with vigorously vortexing, sonicated for 15 minutes, and centrifuged again at 12,700 rpm for 20 minutes at room temperature. The supernatant was pooled with the previous one (total ~1.4 mL) and desalted using Sep-Pak tC18 solid-phase extraction cartridges (Waters, 50 mg capability). The cartridges were equilibrated subsequently with 3 mL of acetonitrile (ACN), 3 mL of 50% ACN/0.5% acetic acid and 3 mL of 0.1% TFA. Digested peptide mixtures from 10 samples were combined (total ~14 mL) and loaded onto the cartridges. The cartridges were washed with 3 mL of 0.1% TFA and then with

250 µL of 0.5% acetic acid. Peptides were eluted with 1.4 mL of 80% ACN/0.5% acetic acid and dried in a Speed Vac. The dried peptides were resolubilized with 1.5 mL of 1% NP-40/TBS with gentle pipetting and incubated with a 200 µL slurry of NeutrAvidin agarose (pre-washed one time with 1% NP-40/TBS, Thermo Fisher, 29200) for at least 6 hours at 4 ºC. Beads were collected using spin columns and washed three time with 500 µL of 1% NP-40/TBS for 20 minutes at room temperature to remove non-specifically bound proteins. Beads were rinsed with 500 µL of TBS twice. The biotinylated peptides captured by the beads were cleaved and eluted by incubating with 400 µL of freshly prepared 50 mM $Na_2S_2O_4$/2 M urea in 50 mM Tris, pH 8.0 for 1 hour at room temperature. The elution procedure was repeated once and with a final rinse using 200 µL of water. The peptides from three elutions were combined, and the mixture was acidified by adding an equal volume of 1% FA to achieve a final FA concentration of 0.5%. Peptides were desalted with C18 Zip Tip Pipette Tips and analyzed with nano LC-MS/MS using the methods described later.

Protein-centric enrichment of biotinylated proteins

100 µL of concentrated secretome was biotinylated using click chemistry and precipitated with methanol/chloroform precipitation, following the methods described earlier. Protein pellets were resuspended in 100 µL of reduction buffer containing 1% SDS, 10 mM TCEP, and 0.4 M NH_4HCO_3 with vigorous shaking, followed by sonicating for 10 minutes in a bath sonicator. Proteins were boiled at 105 ºC for 5 minutes and cooled down to room temperature. 50 µL of 100 mM Iodoacetamide was added, and proteins were incubated in dark for 40 minutes for alkylation. 1 mL of 1% NP-40/TBS was added to dilute SDS to a final concentration of 0.1%. For cleaving off *N*-glycans, 1.5 µL of PNGase F (20 units/µL stock, 1.5 µL enzyme per 100 µL secretome) was added and samples were incubated in 37 ºC water bath overnight. After digestion,

1% NP-40/TBS was added to samples to make a total volume of 1.95 mL. Five samples (total ~10 mL) were combined into a 15 mL Conical Centrifuge Tube (VWR). 200 µL slurry of NeutrAvidin agarose (pre-washed one time with 1% NP-40/TBS) was added and samples were rotated at 4 ºC overnight. Beads were collected using spin columns. Beads were extensively washed three times with 500 µL of 1% NP-40/TBS for 30 minutes, followed by three times with 500 µL of 1% SDS in 50 mM Tris, pH 8.0 for 30 minutes at room temperature. Beads were rinsed with 500 µL of TBS twice before elution. The biotinylated proteins captured by the beads were cleaved and eluted by incubating with 200 µL of freshly prepared 50 mM $Na_2S_2O_4$/1% SDS in 50 mM Tris, pH 8.0 for 1 hour at room temperature. The elution procedure was repeated once and with a final rinse using 100 µL of water. The three elutions were pooled, and the volume was reduced to ~100 µL using a SpeedVac. SDS was removed using an S-trap (ProtiFi), following the manufacturer's recommended protocol. Briefly, 20 µL of 10% SDS was added to 100 µL of condensed elution (contains ~4% SDS) to reach a final SDS of 5%. Then 12 µL of 27.5% phosphoric acid (Sigma Aldrich) was added to reach a final concentration of 2.5%. Sample was mixed with 792 µL of binding buffer (6X of the sample volume, ProtiFi) and loaded to an S-trap column. On-column digestion was performed with 1.5 µg of trypsin and incubated overnight in 37 ºC water bath. Digested peptides were eluted sequentially using 50 µL of 50 mM TEAB (Sigma Aldrich), followed by 50 µL of 0.2% FA, and finally 50 µL of 50% ACN in water. The three elutions were pooled, acidify with FA to a final concentration of 0.5%, and dried using a SpeedVac. Peptides were resuspended in 0.1% FA and analyzed with nano LC-MS/MS using the methods described later.

LC-MS/MS analysis

Samples were analyzed by online nanoflow LC-MS/MS using an EASY-nLC 1200 System (Thermo Fisher) coupled to a Q-Exactive Plus mass spectrometer (Thermo Fisher). Peptides desalted using C18 Zip Tip Pipette Tips or dried with a SpeedVac were diluted or resolubilized in 20 µL of 10% ACN and 0.1% FA for subsequent injection, respectively. Peptides (~8% of total amount) were loaded via autosampler isocratically into a 20 µL sample loop and pre-concentrated onto an Acclaim PepMap-100 75 µm X 2 cm nanoViper C18 pre-column with loading buffer, 5% ACN/0.1% FA (Solvent A). Subsequently, peptides were gradient eluted onto a C18 EasySpray PepMap RSLC C18 analytical column (50 µm X 15 cm, Thermo Fisher) at a constant flow rate of 0.3 µL/min using a 90 minute gradient and a 120 minute instrument method. The gradient profile was as follows [min:% solvent B (80% ACN/0.1% FA)]: 0:0, 90:50, 93:98, 96:98, 99:2, 102:2, 105:98, 108:98, 111:2, 114:2, 117:98, 120:98. The instrument method used MS1 resolution of 70,000, AGC target of 1e6, maximum IT of 100 ms, and scan range from 400 to 2000 m/z. Dynamic exclusion was enabled for 6 s. Only charge states 2-6 were permitted for fragmentation. MS2 scans were acquired at a resolution of 17,500, following HCD fragmentation with fixed collision energy of 27% after quadrupole isolation with an isolation window of 1.2 m/z. Parameters used for MS2 scans were first mass 130 m/z, AGC target of 1e5, maximum IT of 100 ms, and the spectrum data was acquired in centroid mode.

Database search and data analysis

Data analysis was performed with Byonic^{TM} (Protein Metrics, v4.1.10). The following search settings were used for different experiments.

Comparative analysis of 6-AF-labeled peptides with and without Fuc6nyl conjugation*

Byonic search was performed against protein sequences of human ADAMTS9 (UniProt accession number: Q9P2N4) and mouse Lfng (UniProt accession number: O09010). Search parameters included fully specific cleavage specificity at the C-terminal site of R and K for trypsin with two missed cleavages allowed. Mass tolerance was set at 5 ppm for MS1s and 10 ppm for HCD MS2s. Cysteine carbamidomethylation was set as fixed modification. Peptide N-terminal/histidine/lysine carbamidomethylation (common 1), peptide N-terminal dicarbamidomethyl (common 1), histidine/asparagine/tryptophan/proline oxidation (common 1), tryptophan dioxidation (common 2), asparagine/glutamine deamidation (common 1), N-terminal glutamine to pyro-glutamic acid (rare 1), N-terminal glutamic acid to pyro-glutamic acid (rare 1), N-terminal cysteine ammonia-loss (rare 1), protein N-terminal acetylation (rare 1), N-terminal histidine/lysine Delta:H(2)C(2) (common 1), and tryptophan hexosylation (common 2) were set as variable modifications with a total common max of 3, rare max of 2. Glycans were set as variable modifications (common 1) using a customed *O*-glycan search space including Hex(1), Fuc(1), Hex(1)Fuc(1), Fuc(1) 9.9800, Hex(1)Fuc(1) 9.9800, Fuc(1) 188.0700, Hex(1)Fuc(1) 188.0700, HexNAc(1), where the mono or di-fucose modifications were searched with an additional 9.9800 or 188.0700 m/z to account for the chemical modifications of 6-AF or Fuc6nyl*, respectively.

To make the extracted ion chromatograms (EICs), the ions of each peptide glycoform were extracted from the MS1 spectrum using their respective m/z. The observed m/z values from the PSMs with the highest Byonic score were used for generating EICs, with mass tolerance $\pm$ 0.005. The EICs were smoothed using Gauss algorithm. All EICs were manually generated and analyzed using Thermo Xcalibur software.

DidFAT

A pilot search was first performed against the Uniprot human proteome (downloaded July 8, 2021). For peptide-centric enrichment data, the pilot search was performed with MS/MS Filtering of 335.1360 m/z to restrict the search to spectra containing the diagnostic ion (177). For protein-centric enrichment data, the pilot search was performed without MS/MS Filtering, while keeping all other search parameters and subsequent searches identical to those used for peptide-centric enrichment data.

Search parameters included fully specific cleavage specificity at the C-terminal site of R and K for trypsin with two missed cleavages allowed. Mass tolerance was set at 10 ppm for MS1s and 0.1 Da for HCD MS2s. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation (common 1), asparagine oxidation (common 1), asparagine deamidation (common 1), peptide N-terminal acetylation (rare 1), and tryptophan hexosylation (common 2) were set as variable modifications with a total common max of 3, rare max of 1. Glycans were set as variable modifications (common 1) using a customed *O*-glycan search space including Fuc(1) 188.0700, HexNAc(1)Fuc(1) 188.0700, Hex(1)Fuc(1) 188.0700, and Hex (1), where the mono or di-fucose modifications were searched with an additional 188.0700 m/z to account for the chemical modification. MS/MS Filtering was enabled with a user-defined peak of 335.1360 m/z, requiring spectra to contain 335.1360 ± 0.02 among the top 50 peaks to be selected for the search (command: Fuc 188.07 [H+] / 335.136).

Proteins that were found in the pilot search were entered into a ‘focused database’ for glycopeptide searches. The raw file was then searched against the ‘focused database’ with user-defined Custom Peaks to assign higher score to fucosylated spectra with diagnostic ions in their spectra and allow for user-specified peak annotations (177). Search parameters included fully

specific cleavage specificity at the C-terminal site of R and K for trypsin with two missed cleavages allowed. Mass tolerance was set at 10 ppm for MS1s and 0.1 Da for HCD MS2s. Cysteine carbamidomethylation was set as fixed modification. Methionine oxidation (common 1), asparagine oxidation (common 1), asparagine deamidation (common 1), peptide N-terminal acetylation (rare 1), and tryptophan hexosylation (common 1) were set as variable modifications with a total common max of 3, rare max of 2. Glycans were set as variable modifications (common 1) using a customed *O*-glycan search space including Hex (1), HexNAc (1), and a list of Custom Peaks. CustomPeak commands were as follows: Fuc6nyl*/+334.1279 @S,T | common1 | CustomPeaks{Diag-335: 335.1355, Diag-136: 136.0762}, Fuc6nyl*Hex/+496.1807 @S,T | common1 | CustomPeaks{Diag-335: 335.1355, Diag-136: 136.0762}, Fuc6nyl*HexNAc/+537.2073 @S,T | common1 | CustomPeaks{Diag-335: 335.1355, Diag-136: 136.0762}, Fuc6nyl*HexNAcHex/+699.2601 @S,T | common1 | CustomPeaks{Diag-335: 335.1355, Diag-136: 136.0762}. Those CustomPeak commands direct the software to look for peaks at 335.1355 and 136.0762 for PSMs identified with mono, di or tri-Fuc6nyl* modification on S or T. If the spectrum contains the peaks, the score will increase; conversely, if the peaks are not detected, the score will decrease.

A third search was performed against a library of putative POFUT1/2 targets (136 proteins) with user-defined Custom Peaks. All parameters were set the same as in the focused database search described above.

Search results from the focused database and putative POFUT1/2 targets were combined. Only PSMs with Byonic score higher than 100 were chosen for analysis. All spectra identified with Fuc6nyl*-containing modifications were manually inspected and annotated.

3.6 Main text tables and figures

Table 3.1 Proteins identified in the secretome of *MGAT1* KO HEK293F cells by peptide-centric DidFAT method.

Protein name	UniProt ID	UniProt accession No.	Gene	Predict targets	Consensus/ total	Cellular component		
Fibrillin-2	FBN2_HUMAN	P35556	FBN2	YES	EGF, 1/47	Extracellular		
Agrin	AGRIN_HUMAN	O00468	AGRN	YES	EGF, 2/4	Extracellular	Cell membrane	
Versican core protein	CSPG2_HUMAN	P13611	VCAN	YES	EGF, 2/2	Extracellular		
A disintegrin and metalloproteinase with thrombospondin motifs 1	ATS1_HUMAN	Q9UHI8	ADAMTS1	YES	TSR, 3/3	Extracellular		
Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	SVEP1_HUMAN	Q4LDE5	SVEP1	YES	EGF, 4/9	Extracellular	Cell membrane	Cytoplasm
A disintegrin and metalloproteinase with thrombospondin motifs 3 (AdamTS3)	ATS3_HUMAN	O15072	ADAMTS3	YES	TSR, 2/4	Extracellular		
CCN family member 3	CCN3_HUMAN	P48745	CCN3	YES	TSR, 1/1	Extracellular		Cytoplasm Nucleus
Protein kinase C-binding protein NELL1	NELL1_HUMAN	Q92832	NELL1	YES	EGF, 1/5	Extracellular		Cytoplasm
Neurogenic locus notch homolog protein 3 (NOTCH3)	NOTC3_HUMAN	Q9UM47	NOTCH3	YES	EGF, 14/34		Cell membrane	Nucleus
Neurogenic locus notch homolog protein 1 (NOTCH1)	NOTC1_HUMAN	P46531	NOTCH1	YES	EGF, 20/36		Cell membrane	Nucleus
Fibrillin-1	FBN1_HUMAN	P35555	FBN1	NO	EGF, 0/47	Extracellular		
Thrombospondin type-1 domain-containing protein 4 (AdamTSL-6)	THSD4_HUMAN	Q6ZMP0	THSD4	YES	TSR, 3/6	Extracellular		
Papilin	PPN_HUMAN	O95428	PAPLN	YES	TSR, 4/5	Extracellular		
Triosephosphate isomerase	TPIS_HUMAN	P60174	TPI1	NO	-			Cytoplasm
Nucleolar protein 14 OS=Homo sapiens	NOP14_HUMAN	P78316	NOP14	NO	-			Nucleus
Ephrin type-B receptor 2 OS=Homo sapiens	EPHB2_HUMAN	P29323	EPHB2	NO	-		Cell membrane	
Collagen alpha-1(III) chain OS=Homo sapiens	CO3A1_HUMAN	P02461	COL3A1	NO	-	Extracellular		
Enhancer of polycomb homolog 1 OS=Homo sapiens	EPC1_HUMAN	Q9H2F5	EPC1	NO	-			Cytoplasm Nucleus
Probable G-protein coupled receptor 132	GP132_HUMAN	Q9UNW8	GPR132	NO	-		Cell membrane	
Zinc finger SWIM domain-containing protein 5	ZSWM5_HUMAN	Q9P217	ZSWM5	NO	-	Extracellular		
Laminin subunit alpha-5 OS=Homo sapiens	LAMA5_HUMAN	O15230	LAMA5	NO	-	Extracellular		
Uncharacterized protein KIAA1958	K1958_HUMAN	Q8N8K9	KIAA1958	NO	-			Unclear
Proteasome subunit beta type-11	PSB11_HUMAN	A5LHX3	PSMB11	NO	-			Cytoplasm Nucleus
Stabilin-1	STAB1_HUMAN	Q9NY15	STAB1	YES	EGF, 3/16		Cell membrane	
Prolow-density lipoprotein receptor-related protein 1	LRP1_HUMAN	Q07954	LRP1	YES	EGF, 5/22		Cell membrane	
Thrombospondin type-1 domain-containing protein 7A	THS7A_HUMAN	Q9UPZ6	THSD7A	YES	TSR, 4/15	Extracellular	Cell membrane	
CCN family member 4	CCN4_HUMAN	O95388	CCN4	YES	TSR, 1/1	Extracellular		
A disintegrin and metalloproteinase with thrombospondin motifs 15	ATS15_HUMAN	Q8TE58	ADAMTS15	YES	TSR, 3/3	Extracellular	Cell membrane	
Protein delta homolog 2	DLK2_HUMAN	Q6UY11	DLK2	YES	EGF, 1/6		Cell membrane	

Table 3.2 Glycopeptides identified in the secretome of *MGAT1* KO HEK293F cells by peptide-centric DidFAT method.

Protein name	UniProt ID	Consensus/ total	Peptide	Starting position	Predict glycosite	EGF/ TSR	Consensus	Glycans	Observed m/z	z	No. of Spectra	Byonic score	ppm err.
Agrin	O00468	EGF, 2/4	R.GPCGSRDPCSNVTCSFGSTCAR.S	240	T258	X	X	Fuc6nyl*	707.2850	4	5	221.6	0.79
			R.DPCSNVTCSFGSTCAR.S	246	T258	X	X	Fuc6nyl*	718.6157	3	12	634.6	0.36
			R.ASGHPCLNGASCVPR.E	1825	S1835	EGF4	Yes	Fuc6nyl*	480.2134	4	41	490.1	1.34
Versican core protein	P13611	EGF, 2/2	R.NGATCVDGFNTFR.C	3138	T3141	EGF2	Yes	Fuc6nyl*	598.2632	3	15	439.3	2.26
			R.NGATCVDGFNTFR.C	3138	T3141	EGF2	Yes	HexNAc-Fuc6nyl*	666.2875	3	7	237.7	4.53
			R.NGATCVDGFNTFR.C	3138	T3141	EGF2	Yes	HexNAc-Hex-Fuc6nyl*	720.3018	3	2	128.2	3.91
A disintegrin and metalloproteinase with thrombospondin motifs 1 (AdamTS1)	Q8UHI8	TSR, 3/3	R.TCGGGVQYTMR.E	574	T574	TSR1	Yes	Hex-Fuc6nyl*	575.9119	3	5	275.1	0.64
			R.SCNLEDCPDNNGK.T	807	S807	TSR1	X	Hex-Fuc6nyl*	1018.3813	2	3	106.8	0.70
			K.SCELGWQR.R	869	S869	TSR2	Yes	Fuc6nyl*	457.2036	3	9	374.2	0.28
			K.SCELGWQR.R	869	S869	TSR2	Yes	Hex-Fuc6nyl*	511.2213	3	10	326.1	0.68
CCN family member 3	P48745	TSR, 1/1	K.SCGMGFSTR.V	219	S219	TSR1	Yes	Fuc6nyl*	446.1851	3	20	322.4	-0.26
Fibrillin-2	P35556	EGF, 1/47	K.NCIDTNECVALPGSCSPGTCQNLEGSFR.C	2013	S2026	EGF34	X	Fuc6nyl*	1160.4949	3	14	742.7	3.16
Neurogenic locus notch homolog protein 3 (NOTCH3)	Q9UM47	EGF, 14/34	R.NQAICLDR.I	442	T445	EGF11	Yes	Fuc6nyl*	438.1974	3	11	282.3	-0.33
Protein kinase C-binding protein NELL1	Q92832	EGF, 1/5	R.YGGTCVAPNK.C	525	T528	EGF3	Yes	Fuc6nyl*	467.8819		8	304.6	0.67
Papilin	O95428	TSR, 4/5	K.AGPWAPCSASCGGGSQSR.S	368	S377	TSR3	Yes	Hex-Fuc6nyl*	763.9812	3	2	114.5	-1.41
Neurogenic locus notch homolog protein 1 (NOTCH1)	P46531	EGF, 20/36	R.CLNGGTCISGPR.S	1357	T1362	EGF35	Yes	Fuc6nyl*	543.2397	3	5	255.5	0.28
Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Q4LDE5	EGF, 4/9	R.NQATCVDELNSYSCK.C	1394	T1397	EGF6	Yes	Fuc6nyl*	708.6332	3	4	366.0	0.91
A disintegrin and metalloproteinase with thrombospondin motifs 3 (AdamTS3)	O15072	TSR, 2/4	K.TGPWSECSVTCGEGTEVR.Q	972	T981	TSR4	Yes	Hex-Fuc6nyl*	836.6805	3	3	324.2	3.60
Thrombospondin type-1 domain-containing protein 4 (AdamTSL-6)	Q6ZMP0	TSR, 3/6	K.QLGTTECSTTCGK.G	623	T632	X	X	Hex-Fuc6nyl*	647.2747	3	2	161.3	-1.60
Fibrillin-1	P35555	EGF, 0/47	K.CAPGTCQNLDGSYR.C	1984	T1988	EGF34	X	Fuc6nyl*	644.9377	3	2	205.9	3.54

Table 3.3 Sequences of EGF repeats in Agrin. EGF repeats with POFUT1-specific consensus sequences are in red. Predict *O*-fucose sites are in red. EGF repeats with confirmed *O*-fuosylation by DidFAT are highlighted in yellow.

	Position		C1	C2	C3	C4	C5	C6
EGF0	244-286	SRDP	CSNVT	CSFGST	CARSADGLTAS	CL	CPAT	CR
EGF1	1329-1367	PPKP	CDSQP	CFHGGT	CQDWALGGGFT	CS	CPAGRGGAV	CE
EGF2	1549-1586	GDHP	CLPNP	CHGGAP	CQNLEAGRFH	CQ	CPPGRVGPT	CA
EGF3	1588-1625	EKSP	CQPNP	CHGAAP	CRVLPEGGAQ	CE	CPLGREGTF	CQ
EGF4	1818-1857	AGHP	CTRASGHP	CLNGAS	CVPREAAYV	CL	CPGGFSGPH	CE

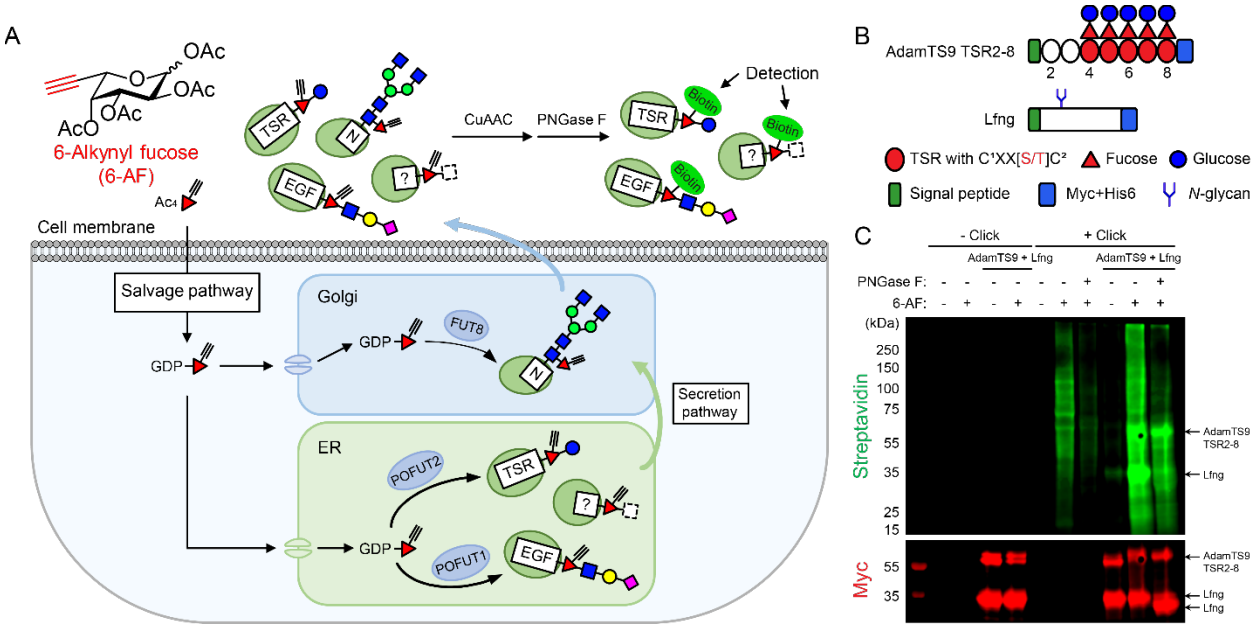
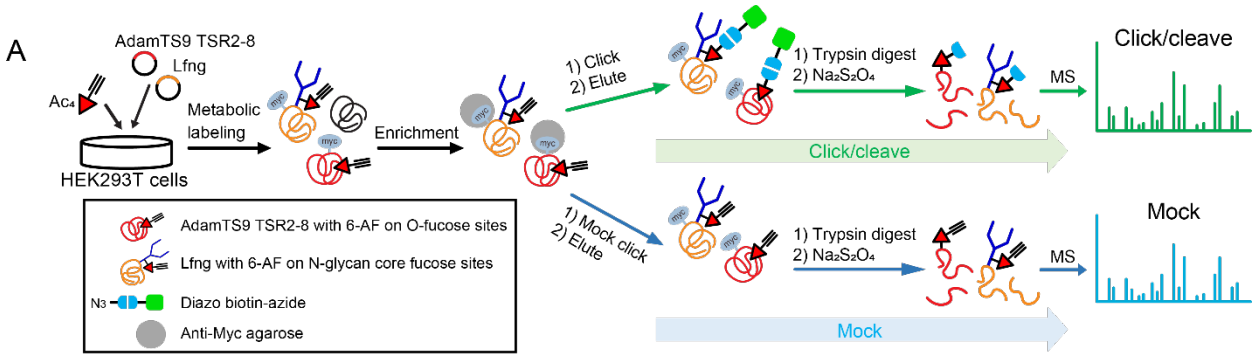


Figure 3.1 Metabolic labeling of the *O*-fucose proteome of POFUT1/2 using 6-alkynyl fucose (6-AF). *A,* Schematic of metabolic labeling using 6-AF. Exogenously added peracetylated 6-AF was taken into cells, converted to GDP-6-AF by the GDP-fucose salvage pathway in the cytosol, and transported to either the endoplasmic reticulum (ER) or the Golgi apparatus, where GDP-6-AF is incorporated into glycoproteins through *O*-fucosylation by POFUT1/2 or core fucosylation on *N*-glycans by FUT8, respectively. The 6-AF incorporated glycoproteins are then secreted into the extracellular medium, conjugated with biotin tags using click chemistry, and treated with PNGase F to remove *N*-glycans, allowing for the specific detection of biotinylated *O*-fucose. *B,* A domain map of constructs of ADAMTS9 TSR2-8 and Lfng used in this study. The five TSRs are depicted as ovals. Red ovals, TSRs containing the POFUT2-specific consensus sequence; red triangle, experimentally confirmed fucose modification; blue circle, experimentally confirmed glucose elongation; blue branch, *N*-glycan site; green square, signal peptides; blue square, C-terminal Myc-His_6 tags. *C,* HEK293T cells were treated with 6-AF or DMSO for 2 days, with or without transfection with ADAMTS9 TSR2-8 and Lfng. 6-AF-labeled proteins in the conditioned culture medium were either biotinylated by click chemistry, or without biotinylating, and analyzed by Western blot. Top panel: probed with streptavidin; bottom panel: probed with anti-Myc.


A
AdamTS9 TSR2-8
Lfng
Ac4
Metabolic labeling
HEK293T cells
Enrichment
1) Click
2) Elute
1) Trypsin digest
2) Na2S2O4
MS
Click/cleave
Click/cleave
1) Mock click
2) Elute
1) Trypsin digest
2) Na2S2O4
MS
Mock
Mock
AdamTS9 TSR2-8 with 6-AF on O-fucose sites
Lfng with 6-AF on N-glycan core fucose sites
N3
Diazo biotin-azide
Anti-Myc agarose

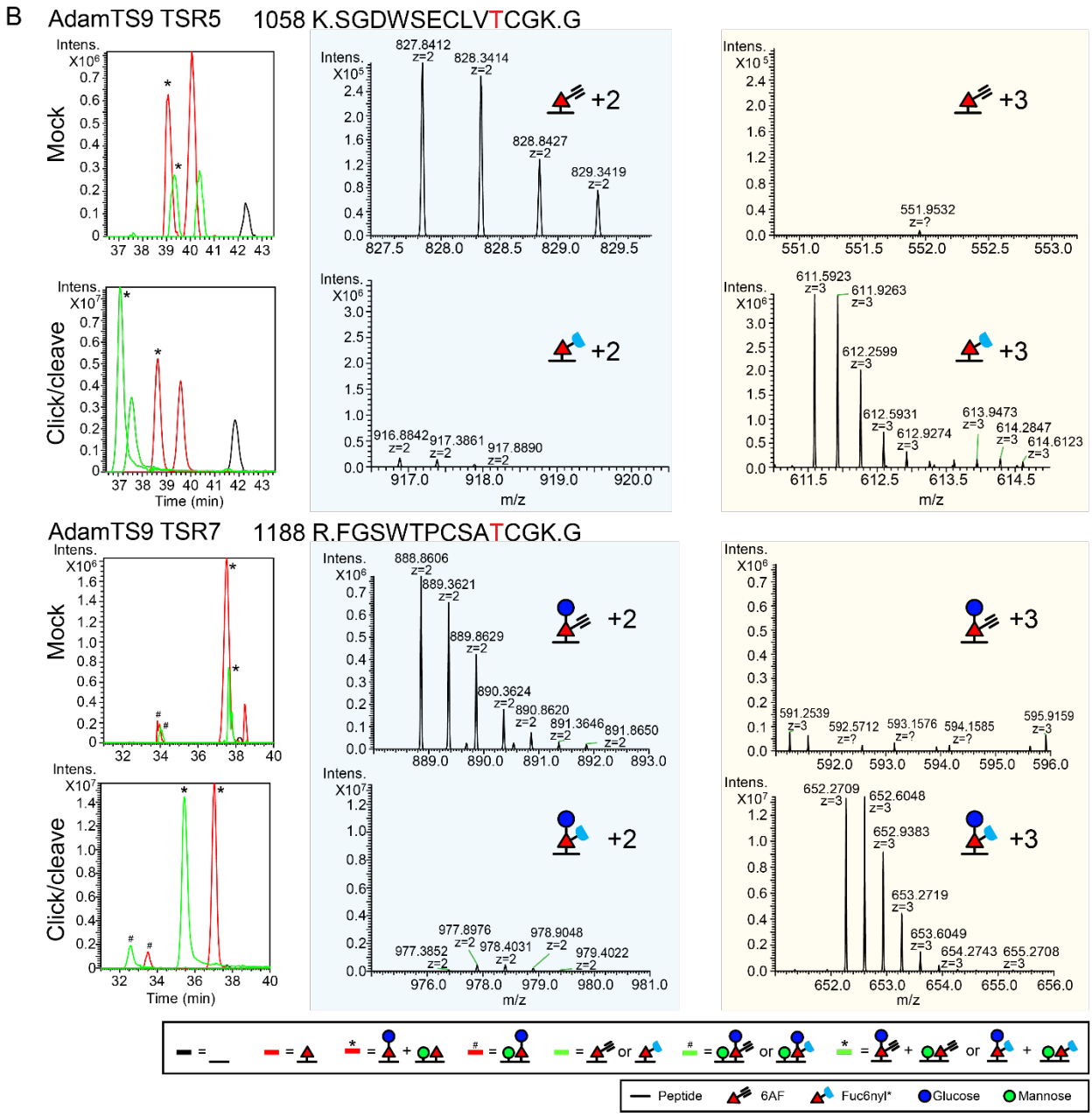

B
AdamTS9 TSR5
1058 K.SGDWSECLVTCGK.G
Mock
Click/cleave
Time (min)
+2
+3
m/z
AdamTS9 TSR7
1188 R.FGSWTPCSATCGK.G
Mock
Click/cleave
Time (min)
+2
+3
m/z
Peptide
6AF
Fuc6nyl*
Glucose
Mannose

Figure 3.2 Comparative mass spectrometric analysis of peptides with and without Fuc6nyl* conjugation. *A,* Workflow depicting the comparative analysis of 6-AF-incorporated peptides from positive controls with and without Fuc6nyl* conjugation. Myc tagged ADAMTS9 TSR2-8 and Lfng were expressed in HEK293T cells treated with 6-AF. Expressed proteins that secreted into the medium were enriched with anti-Myc-agarose and either conjugated with diazo biotin azide via click chemistry (Click/cleave) or subjected to mock conjugation with DMSO (Mock). The eluted proteins were digested with trypsin and incubated with sodium dithionite. The resulting peptides were analyzed by nano LC-MS/MS. *B,* Fuc6nyl* conjugation can increase the relative abundance of labeled peptides and impart charge to precursor ions of *O*-fucosylated peptides. Left panel: extracted ion chromatograms (EICs) of the different glycoforms of a peptide from ADAMTS9 TSR5 and TSR7 derived from *A*. EICs of other TSRs of ADAMTS9 are in Figure 3S1. Black line, unmodified; red line, Fucose (Fuc) modified; red line with *: glucose (Glc)-Fuc modified + mannose (Man)-, Fuc modified; red line with #: Man-, Glc-Fuc modified; green line: 6-AF modified or Fuc6nyl* modified; green line with *: Glc-6-AF modified + Man-, 6-AF modified or Glc-Fuc6nyl* modified + Man-, Fuc6nyl* modified; green line with #: Man-, Glc-6-AF modified or Man-, Glc-Fuc6nyl* modified. Right panel: detected precursor ions for 6-AF-incorporated peptides and Fuc6nyl* conjugated peptides, with glycoforms and charge states indicated.

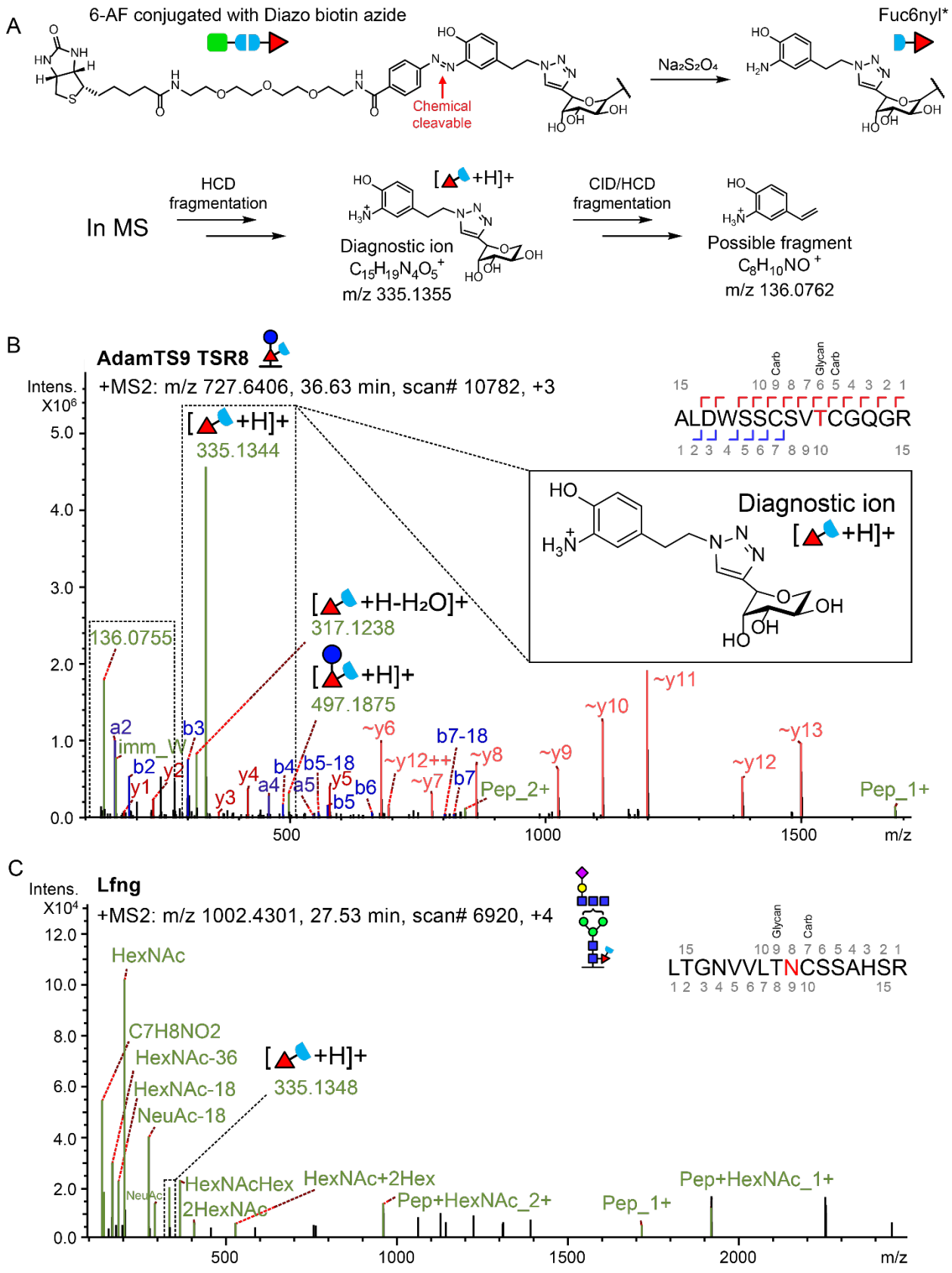

A
6-AF conjugated with Diazo biotin azide
Chemical cleavable
Na2S2O4
Fuc6nyl*
In MS
HCD fragmentation
[+H]+
Diagnostic ion
C15H19N4O5+
m/z 335.1355
CID/HCD fragmentation
Possible fragment
C8H10NO+
m/z 136.0762
B
AdamTS9 TSR8
Intens.
X10^6
+MS2: m/z 727.6406, 36.63 min, scan# 10782, +3
[+H]+
335.1344
[+H-H2O]+
317.1238
[+H]+
497.1875
136.0755
Carb
Glycan
Carb
ALDWSSCSVTCGQGR
Diagnostic ion
[+H]+
a2
imm_W
b3
b2
y2
y1
y4
b4
a4
b5-18
a5
y5
b6
b5
y3
~y6
b7-18
~y12++
~y8
~y7
b7
Pep_2+
~y9
~y10
~y11
~y12
~y13
Pep_1+
m/z
C
Intens.
X10^4
Lfng
+MS2: m/z 1002.4301, 27.53 min, scan# 6920, +4
HexNAc
C7H8NO2
HexNAc-36
HexNAc-18
NeuAc-18
[+H]+
335.1348
NeuAc
HexNAcHex
2HexNAc
HexNAc+2Hex
Pep+HexNAc_2+
Pep_1+
Pep+HexNAc_1+
Glycan
Carb
LTGNVVLTNCSSAHSR
m/z

Figure 3.3 Unique mass spectrometric behavior of Fuc6nyl* conjugated glycopeptides, featured by the generation of diagnostic ions specific to *O*-fucosylated peptides. *A,* Proposed structures and fragmentation scheme for generation of the 335.1355 and 136.0762 ions. *B,* Example HCD spectra of a Fuc6nyl* conjugation peptide from ADAMTS9 TSR8 demonstrating the presence of the diagnostic ions at 335.1344 m/z and 136.0755 m/z with high intensity. *C,* Example HCD spectra of a Fuc6nyl* conjugation peptide from Lfng indicating the low abundance of the diagnostic ion in *N*-glycosylated peptides.

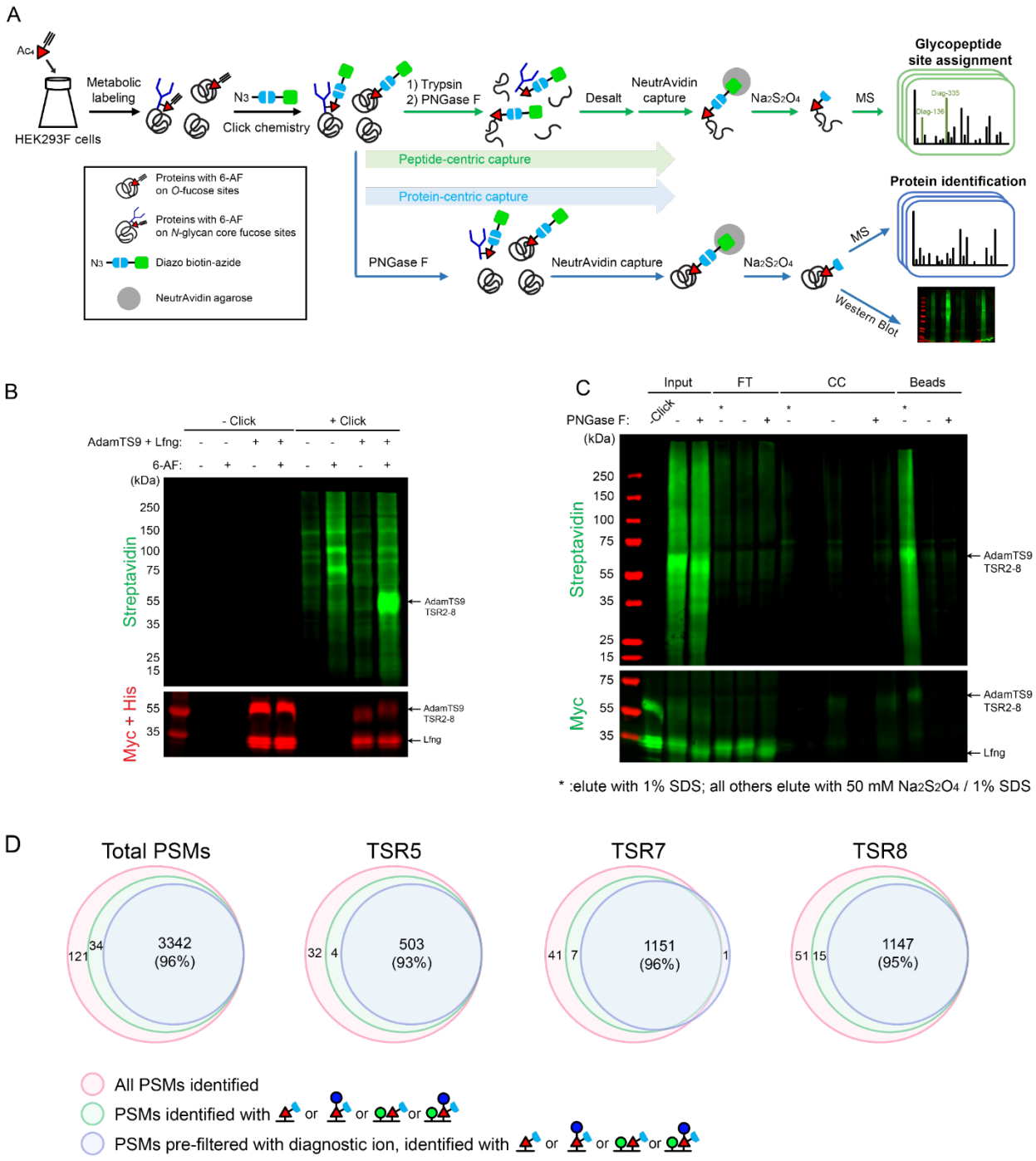


Figure 3.4 DidFAT: a glycoproteomic workflow for profiling the *O*-fucose proteome of POFUT1/2, validated with a transfected positive control. *A,* Schematic of DidFAT workflow for the preparation and analysis of *O*-fucosylated proteins/peptides in cell secretome. Peptide-centric capture methods, schematized on the top, involve metabolic labeling of secretome with 6-AF, click chemistry using diazo biotin azide, trypsin digestion, N-glycan removal using PNGase F, desalting peptides with Sep-Pak tC18 cartridges, capturing biotinylated peptides with NeutrAvidin beads, eluting peptides with sodium dithionite for nano LC-MS/MS analysis. Protein-centric capture and quality control methods, schematized on the bottom, involve a

PNGase F treatment after click chemistry, capturing of biotinylated proteins with NeutrAvidin beads, eluting proteins with sodium dithionite and either processed for nano LC-MS/MS analysis or load on a streptavidin blot for quality control. *B, MGAT1* KO HEK293F cells were treated with 6-AF or DMSO for 3 days, with or without transfection with ADAMTS9 TSR2-8 and Lfng. 6-AF-labeled proteins in the conditioned culture medium were either biotinylated by click chemistry, or without biotinylating, and analyzed by Western blot. Top panel: probed with streptavidin; bottom panel: probed with anti-Myc and anti-His. *C,* Secretome samples from 6-AF treated *MGAT1* KO HEK293F cells that transfected with ADAMTS9 TSR2-8 and Lfng, prepared as described in the bottom schematic in *A*, were analyzed by Western blot. Top panel: probed with streptavidin; bottom panel: probed with anti-Myc. *D,* Venn diagrams showing the identified peptide-spectrum matches (PSMs) in ADAMTS9 TSR2-8 using the DidFAT workflow. Total PSMs and the three TSRs with the highest PSM numbers are shown. Pink circle: all identified PSMs for a peptide; green circle: identified PSMs with Fuc6nyl* modification, either mono or disaccharide, and with or without C-mannosylation; purple circle: PSMs identified in Byonic search with MS2 Peak Filtering that only retain spectra containing fragments at 335.1355 or 136.0762, and with Fuc6nyl* modification.

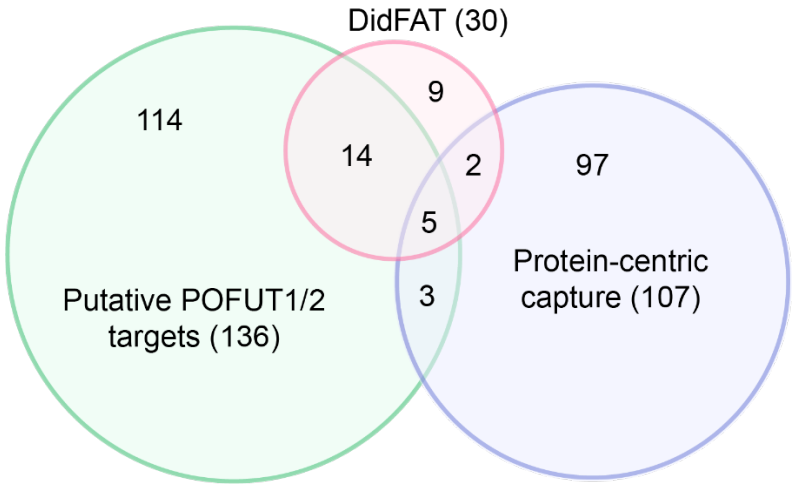


Figure 3.5 Overlap of proteins identified by the peptide-centric enrichment and protein-centric enrichment methods with the library of putative POFUT1/2 targets.

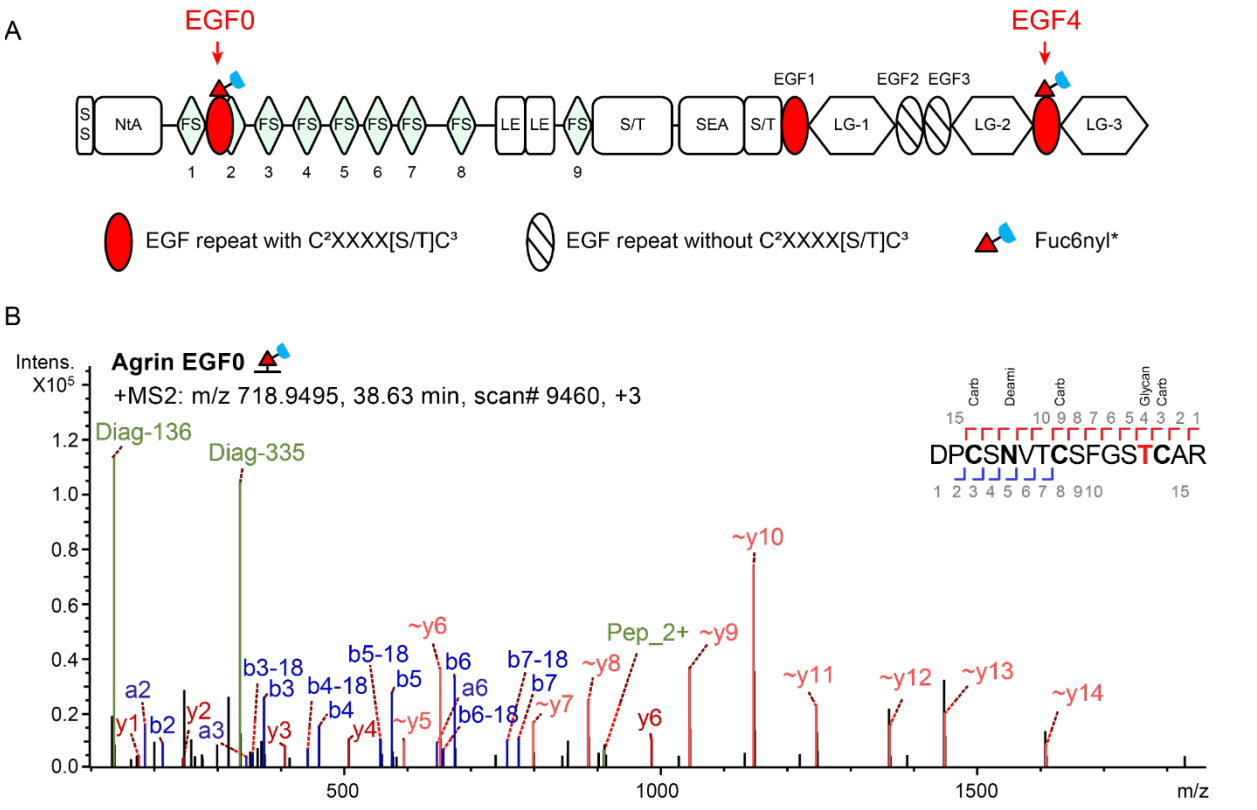


Figure 3.6 Novel *O*-fucose site identified in peptides located within an unannotated EGF repeat of Agrin. *A.* Domain organization of human Agrin protein. EGF repeats are depicted as ovals. SS, Signal sequence; NtA, N-terminal agrin; FS, Follistatin-like; LE, Laminin EGF-like; S/T, Ser/Thr-rich; SEA, Sea urchin sperm protein, enterokinase, agrin; LG, Laminin G-like. Annotations from UniProt O00468. *B.* Example HCD spectra of a Fuc6nyl* conjugation peptide from Agrin EGF0. Diagnostic ions at presented as 'Diag-335' and 'Diag-136'.

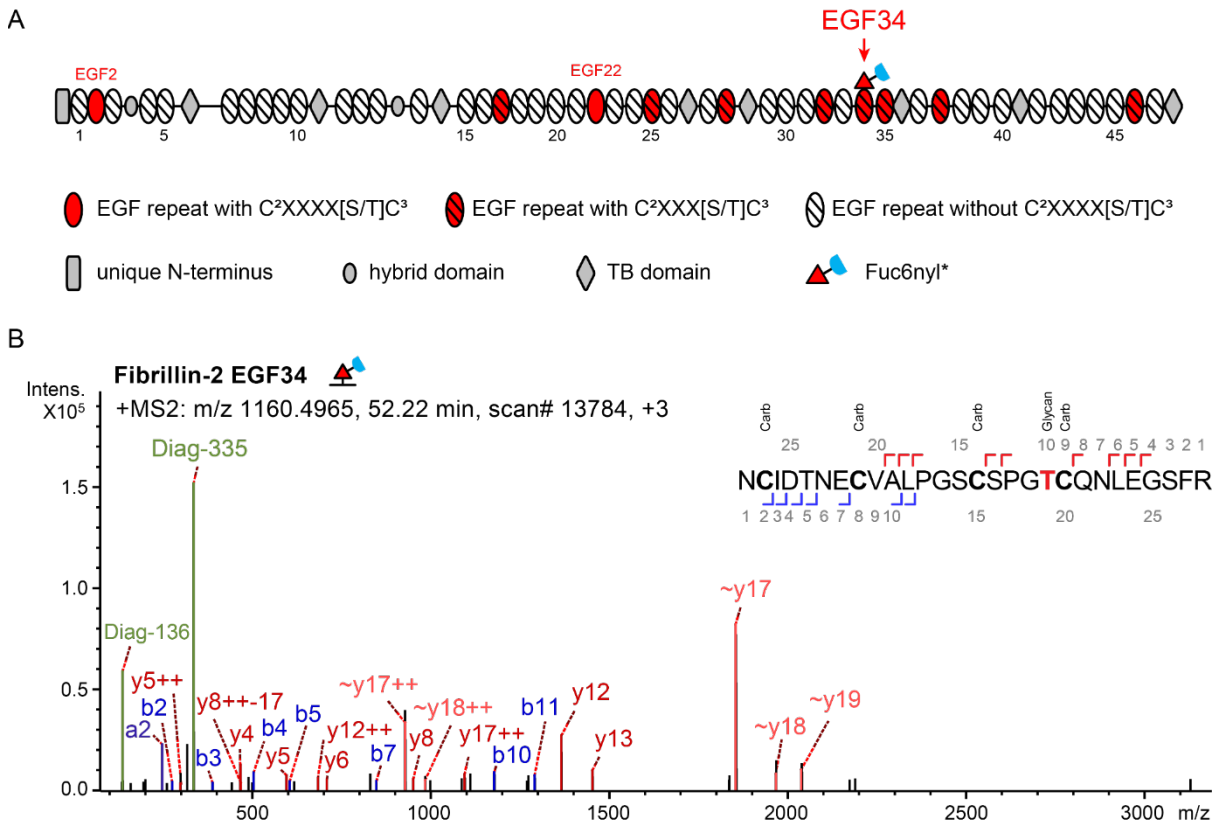


Figure 3.7 Novel *O*-fucose site identified in peptides located within a noncanonical *O*-fucose consensus sequence of Fibrillin-2 EGF34. *A.* Domain organization of human Fibrillin-2 protein. EGF repeats are depicted as ovals. Annotations from UniProt P35556. *B.* Example HCD spectra of a Fuc6nyl* conjugation peptide from Fibrillin-2 EGF34. Diagnostic ions at presented as 'Diag-335' and 'Diag-136'.

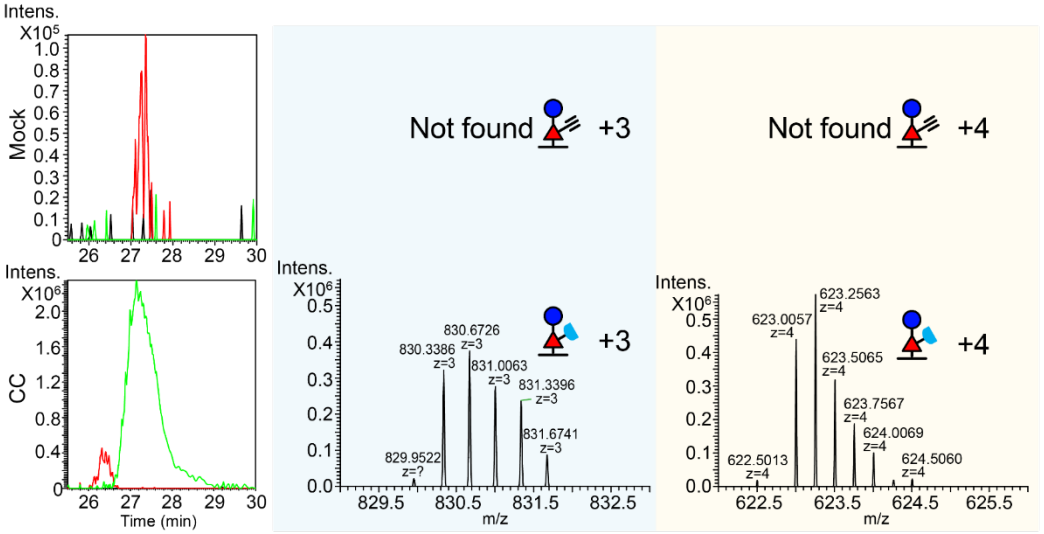
3.7 Supplemental tables and figures

Contents

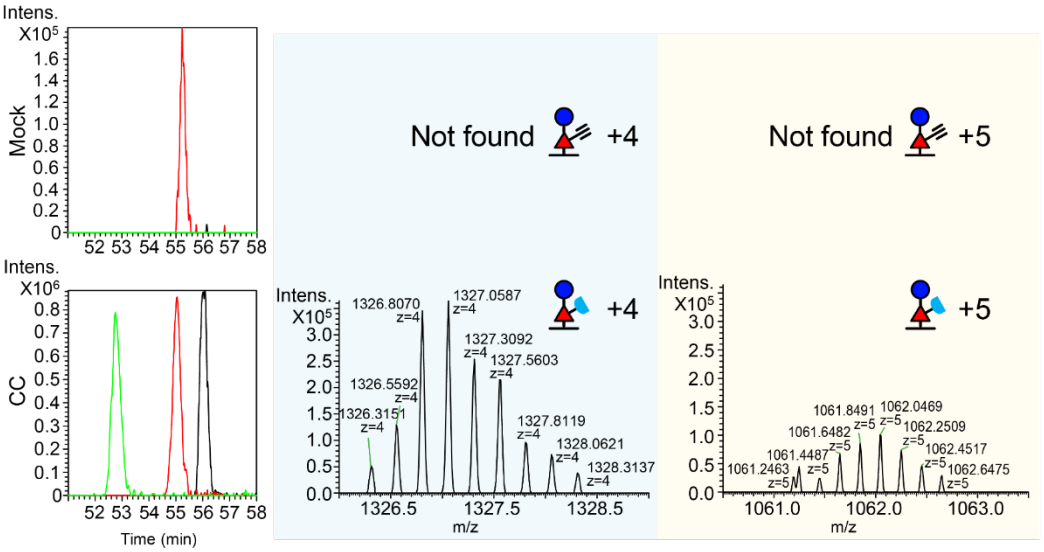
Table 3S1. Sequences of EGF repeats in Fibrillin-2. EGF repeats with canonical POFUT1 consensus sequences $C^2XXXX[S/T]C^3$ and their predicted sites are in red. EGF repeats with noncanonical POFUT1 consensus sequences $C^2XXX[S/T]C^3$ and their predicted sites are in magenta. EGF34 with confirmed *O*-fuosylation by DidFAT is highlighted in yellow.

	Position		C1	C2	C3	C4	C5	C6
EGF 1	111-142	IVPI	**C**RNS	**C**GDGF	**C**SRPNM	**C**T	**C**SSGQISST	**C**G
EGF2	145-176	SIQQ	**C**SVR	**C**MNGG**T**	**C**ADDH	**C**Q	**C**QKGYIGTY	**C**G
EGF3	176-208	GQPV	**C**ENG	**C**QNGGR	**C**IGPNR	**C**A	**C**VYGFTGPQ	**C**E
EGF4	276-317	DVDE	**C**QAIPGI	**C**QGGN	**C**INTVGSFE	**C**R	**C**PAGHKQSETTQK	**C**E
EGF5	318-359	DIDE	**C**SIIPGI	**C**ETGE	**C**SNTVGSYF	**C**V	**C**PRGYVTSTDGSR	**C**I
EGF6	494-534	TIDI	**C**KHHANL	**C**LNGR	**C**IPTVSSYR	**C**E	**C**NMGYKQDANGD	**C**I
EGF7	535-574	DVDE	**C**TSNP	**C**TNGD	**C**VNTPGSYY	**C**K	**C**HAGFQRTPTKQA	**C**I
EGF8	575-616	DIDE	**C**IQNGVL	**C**KNGR	**C**VNTDGSFQ	**C**I	**C**NAGFELTTDGKN	**C**V
EGF9	617-657	DHDE	**C**TTTNM	**C**LNGM	**C**INEDGSFK	**C**I	**C**KPGFVLAPNGRY	**C**T
EGF10	658-698	DVDE	**C**QTPGI	**C**MNGH	**C**INSEGSFR	**C**D	**C**PPGLAVGMDGRV	**C**V
EGF11	768-809	DINE	**C**ALDPDI	**C**ANGI	**C**ENLRGSYR	**C**N	**C**NSGYEPDASGRN	**C**I
EGF12	810-851	DIDE	**C**LVNRLL	**C**DNGL	**C**RNTPGSYS	**C**T	**C**PPGYVFRTETET	**C**E
EGF13	852-891	DINE	**C**ESNP	**C**VNGA	**C**RNNLGSFN	**C**E	**C**SPGSKLSSTGLI	**C**I
EGF14	955-996	DVNE	**C**EVFPGV	**C**PNGR	**C**VNSKGSFH	**C**E	**C**PEGLTLDGTGRV	**C**L
EGF15	1073-1114	DINE	**C**KAFPGM	**C**TYGK	**C**RNTIGSFK	**C**R	**C**NSGFALDMEERN	**C**T
EGF16	1115-1157	DIDE	**C**RISPDL	**C**GSGI	**C**VNTPGSFE	**C**E	**C**FEGYESGFMMMKN	**C**M
EGF17	1158-1199	DIDE	**C**ERNPLL	**C**RGG**T**	**C**VNTEGSFQ	**C**D	**C**PLGHELSPSRED	**C**V
EGF18	1200-1241	DINE	**C**SLSDNL	**C**RNGK	**C**VNMIGTYQ	**C**S	**C**NPGYQATPDRQG	**C**T
EGF19	1242-1282	DIDE	**C**MIMNGG	**C**DTQ	**C**TNSEGSYE	**C**S	**C**SEGYALMPDGRS	**C**A
EGF20	1283-1324	DIDE	**C**ENNPDI	**C**DGGQ	**C**TNIPGEYR	**C**L	**C**YDGFMASMDMKT	**C**I
EGF21	1325-1366	DVNE	**C**DLNSNI	**C**MFGE	**C**ENTKGSFI	**C**H	**C**QLGYSVKKGTTG	**C**T
EGF22	1367-1407	DVDE	**C**EIGAHN	**C**DMHA**S**	**C**LNIPGSFK	**C**S	**C**REGWIGNGIK	**C**I
EGF23	1408-1448	DLDE	**C**SNGTHQ	**C**SINAQ	**C**VNTPGSYR	**C**A	**C**SEGFTGDGFT	**C**S
EGF24	1449-1490	DVDE	**C**AENINL	**C**ENGQ	**C**LNVPGAYR	**C**E	**C**EMGFTPASDSRS	**C**Q
EGF25	1491-1531	DIDE	**C**SFQNI	**C**VFG**T**	**C**NNLPGMFH	**C**I	**C**DDGYELDRTGGN	**C**T
EGF26	1532-1572	DIDE	**C**ADPIN	**C**VNGL	**C**VNTPGRYE	**C**N	**C**PPDFQLNPTGVG	**C**V
EGF27	1650-1691	DIDE	**C**QELPGL	**C**QGGN	**C**INTFGSFQ	**C**E	**C**PQGYYLSEDTRI	**C**E
EGF28	1692-1733	DIDE	**C**FAHPGV	**C**GPG**T**	**C**YNTLGNYT	**C**I	**C**PPEYMQVNGGHN	**C**M
EGF29	1808-1849	DIDE	**C**KEIPGI	**C**ANGV	**C**INQIGSFR	**C**E	**C**PTGFSYNDLLLV	**C**E
EGF30	1850-1891	DIDE	**C**SNGDNL	**C**QRNAD	**C**INSPGSYR	**C**E	**C**AAGFKLSPNGA	**C**V
EGF31	1892-1933	DRNE	**C**LEIPNV	**C**SHGL	**C**VDLQGSYQ	**C**I	**C**HNGFKASQDQTM	**C**M
EGF32	1934-1972	DVDE	**C**ERHP	**C**GNG**T**	**C**KNTVGSYN	**C**L	**C**YPGFELTHNND	**C**L
EGF33	1973-2015	DIDE	**C**SSFFGQV	**C**RNGR	**C**FNEIGSFK	**C**L	**C**NEGYELTPDGKN	**C**I
EGF34	2016-2055	DTNE	**C**VALPGS	**C**SPG**T**	**C**QNLEGSFR	**C**I	**C**PPGYEVKSEN	**C**I
EGF35	2056-2097	DINE	**C**DEDPNI	**C**LFG**S**	**C**TNTPGGFQ	**C**L	**C**PPGFVLSDNGRR	**C**F
EGF36	2171-2212	DVNE	**C**LESPGI	**C**SNGQ	**C**INTDGSFR	**C**E	**C**PMGYNLDYTGVR	**C**V
EGF37	2213-2252	DTDE	**C**SIGNP	**C**GNG**T**	**C**TNVIGSFE	**C**N	**C**NEGFEPGPMMN	**C**E
EGF38	2253-2293	DINE	**C**AQNPLL	**C**AFR	**C**MNTFGSYE	**C**T	**C**PIGYALREDQKM	**C**K
EGF39	2294-2337	DLDE	**C**AEGLHD	**C**ESRGMM	**C**KNLIGTFM	**C**I	**C**PPGMARRPDGEG	**C**V
EGF40	2338-2379	DENE	**C**RTKPGI	**C**ENGR	**C**VNIIGSYR	**C**E	**C**NEGFQSSSSGTE	**C**L
EGF41	2449-2490	DIDE	**C**KVMPNL	**C**TNGQ	**C**INTMGSFR	**C**F	**C**KVGYTTDISGTS	**C**I
EGF42	2491-2531	DLDE	**C**SQSPKP	**C**NYI	**C**KNTEGSYQ	**C**S	**C**PRGYVLQEDGKT	**C**K
EGF43	2532-2570	DLDE	**C**QTKQHN	**C**QFL	**C**VNTLGGFT	**C**K	**C**PPGFTQHHTA	**C**I
EGF44	2571-2613	DNNE	**C**GSQPSL	**C**GAKGI	**C**QNTPGSFS	**C**E	**C**QRGFSLDATGLN	**C**E
EGF45	2614-2653	DVDE	**C**DGNHR	**C**QHG	**C**QNILGGYR	**C**G	**C**PQGYIQHYQWNQ	**C**V
EGF46	2654-2694	DENE	**C**SNPNA	**C**GSA**S**	**C**YNTLGSYK	**C**A	**C**PSGFSFDQFSSA	**C**H
EGF47	2695-2734	DVNE	**C**SSSKNP	**C**NYG	**C**SNTEGGYL	**C**G	**C**PPGYYRVGQGH	**C**V

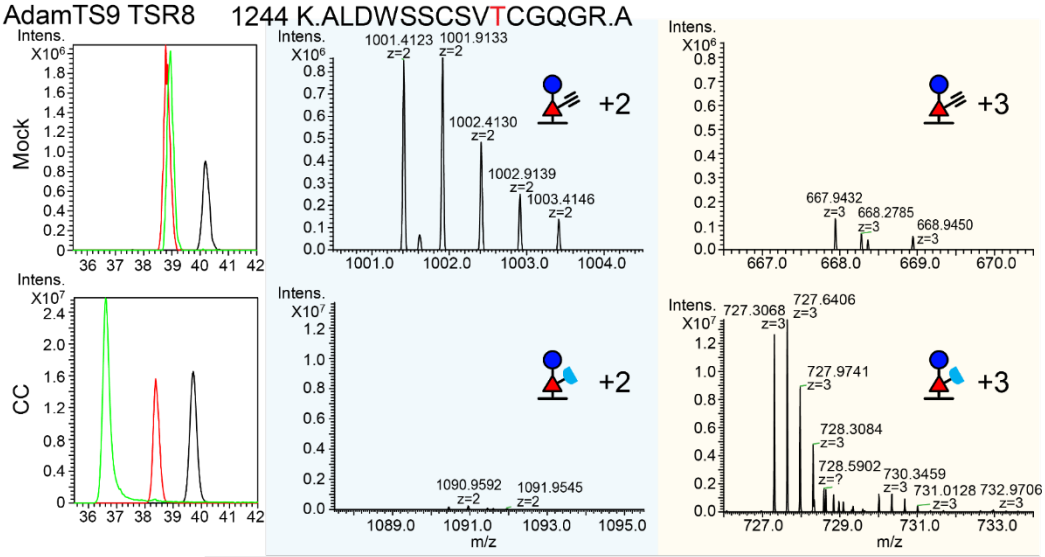
AdamTS9 TSR4 1003 R.YSAWTECSKSCDGGTQR.R



AdamTS9 TSR6 1090 R.MCDPETKPTSMQTCQQPECASWQAGPWGQCSVTCGQGYQLR.A



AdamTS9 TSR8 1244 K.ALDWSSCSVTCGQGR.A



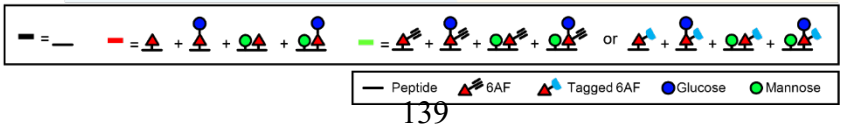


Figure 3S1. Data from TSR4, TSR6, and TSR8 of ADAMTS9 for Figure 3.2*B*. Left panel: extracted ion chromatograms (EICs) of the different glycoforms of a peptide from ADAMTS9 TSR4, TSR6, and TSR8 derived from Figure 3.2*A*. Black line, unmodified; red line, Fucose (Fuc) modified + glucose (Glc)-Fuc modified + mannose (Man)-, Fuc modified + Man-, Glc-Fuc modified; green line: 6-AF modified + Glc-6-AF modified + Man-, 6-AF modified + Man-, Glc-6-AF modified or Fuc6nyl* modified + Glc-Fuc6nyl* modified + Man-, Fuc6nyl* modified + Man-, Glc-Fuc6nyl* modified. Right panel: detected precursor ions for 6-AF-incorporated peptides and Fuc6nyl* conjugated peptides, with glycoforms and charge states indicated.

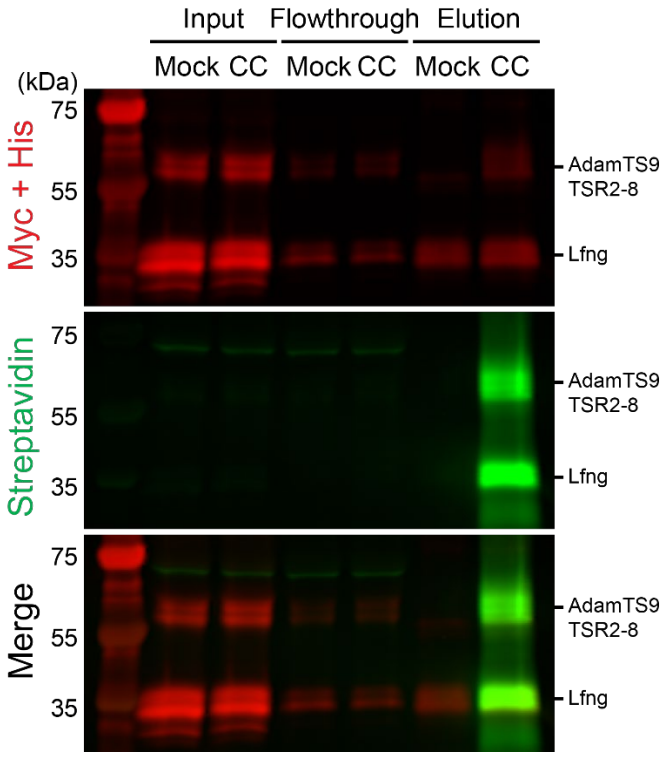


Figure 3S2. Western blot-based quality control of the experiment depicted in Figure 3.2. Myc-tagged ADAMTS9 TSR2-8 and Myc-tagged Lfng that expressed and secreted to the secretome of HEK293T cells were captured by anti-Myc agarose and underwent on-bead click chemistry (CC) or mock click reaction (Mock). Captured proteins were eluted off the beads using urea and detected with Western blot. Note that only the CC sample with on-bead click reaction showed robust signal detected by streptavidin in the elution.

Figure 3S3. Mass spectral data of peptides with Fuc6nyl* modifications identified in the secretome of *MGAT1* KO HEK293F cells by DidFAT.

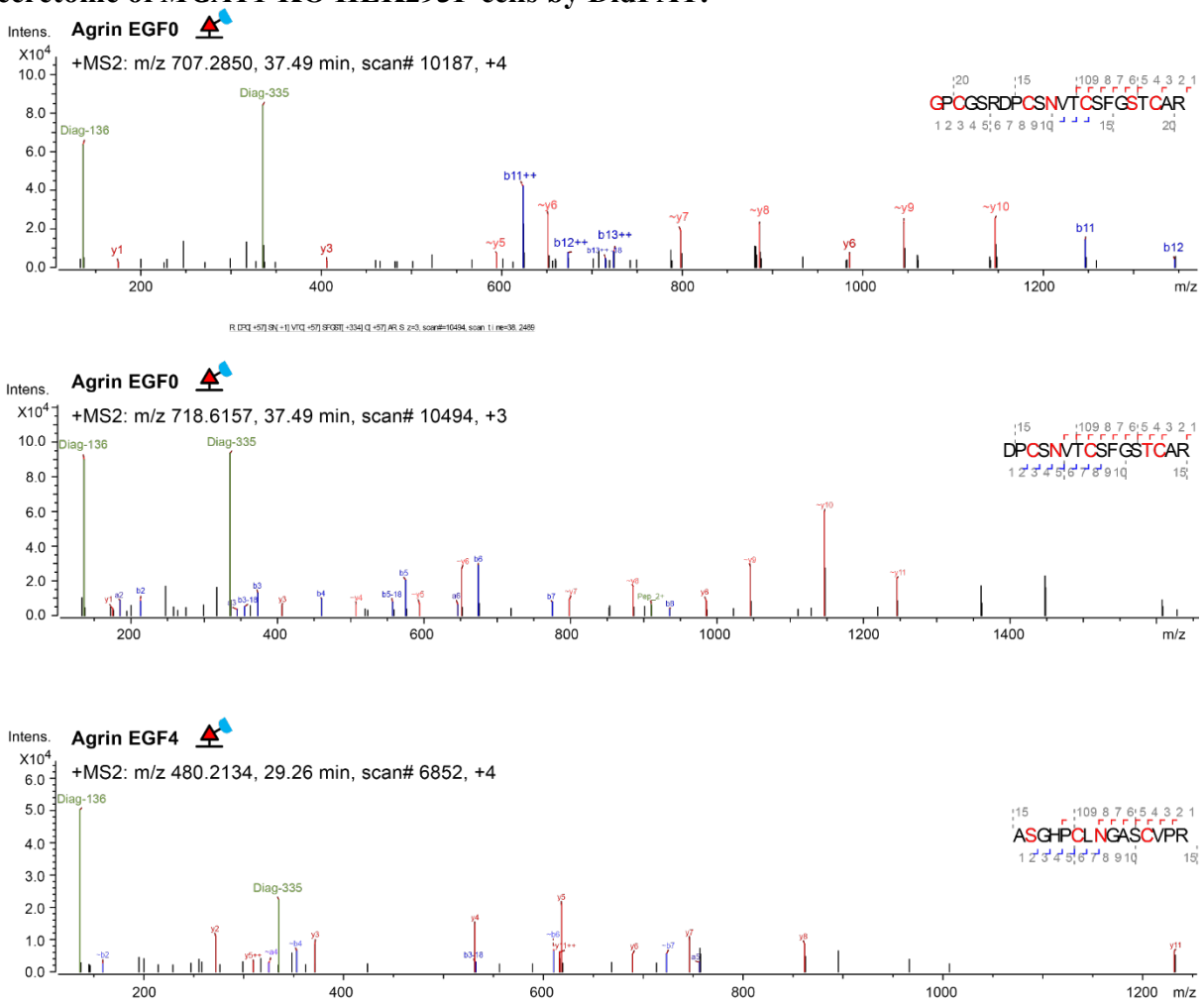


Figure 3S3A. Spectra from Agrin.

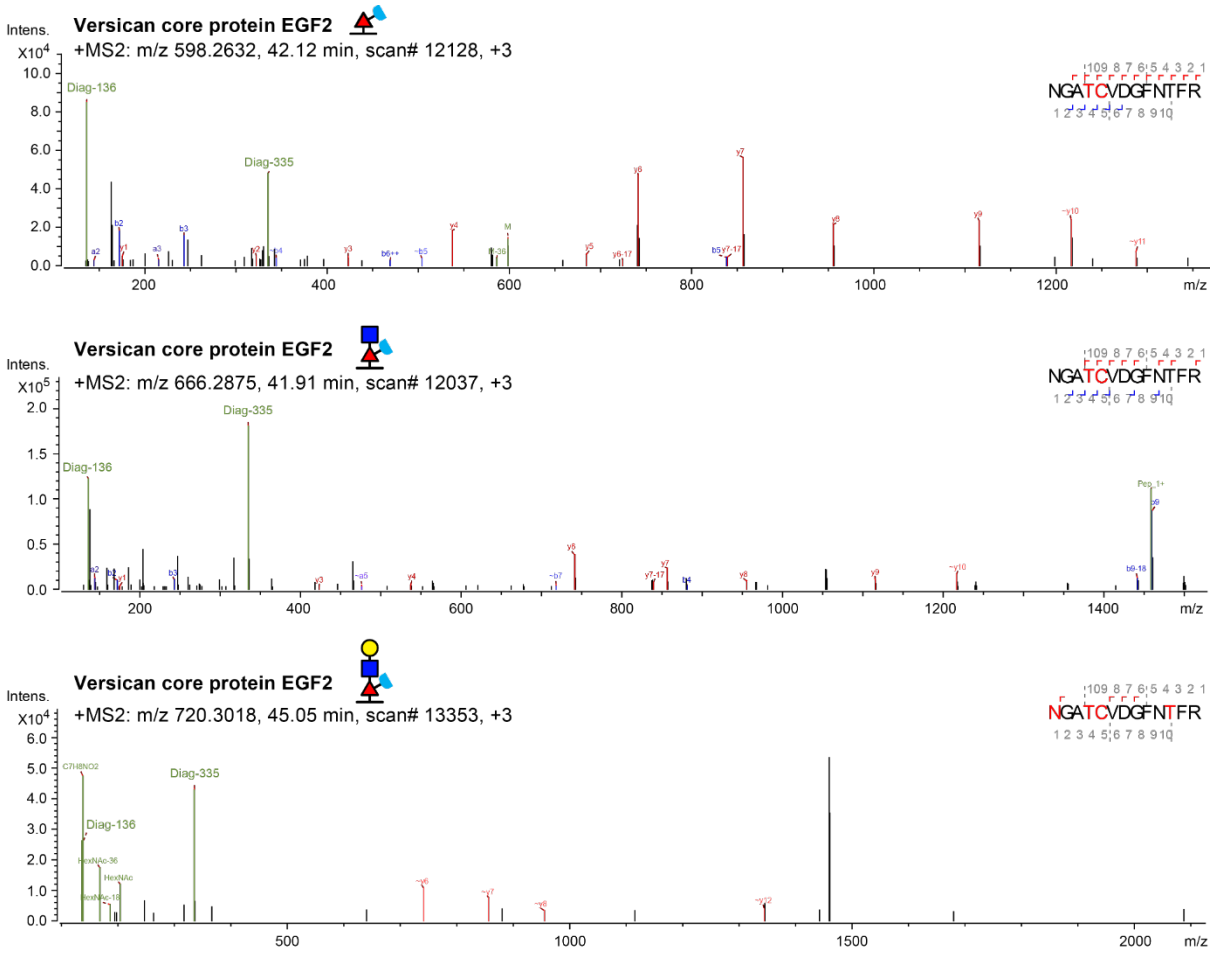


Figure 3S3B. Spectra from Versican.

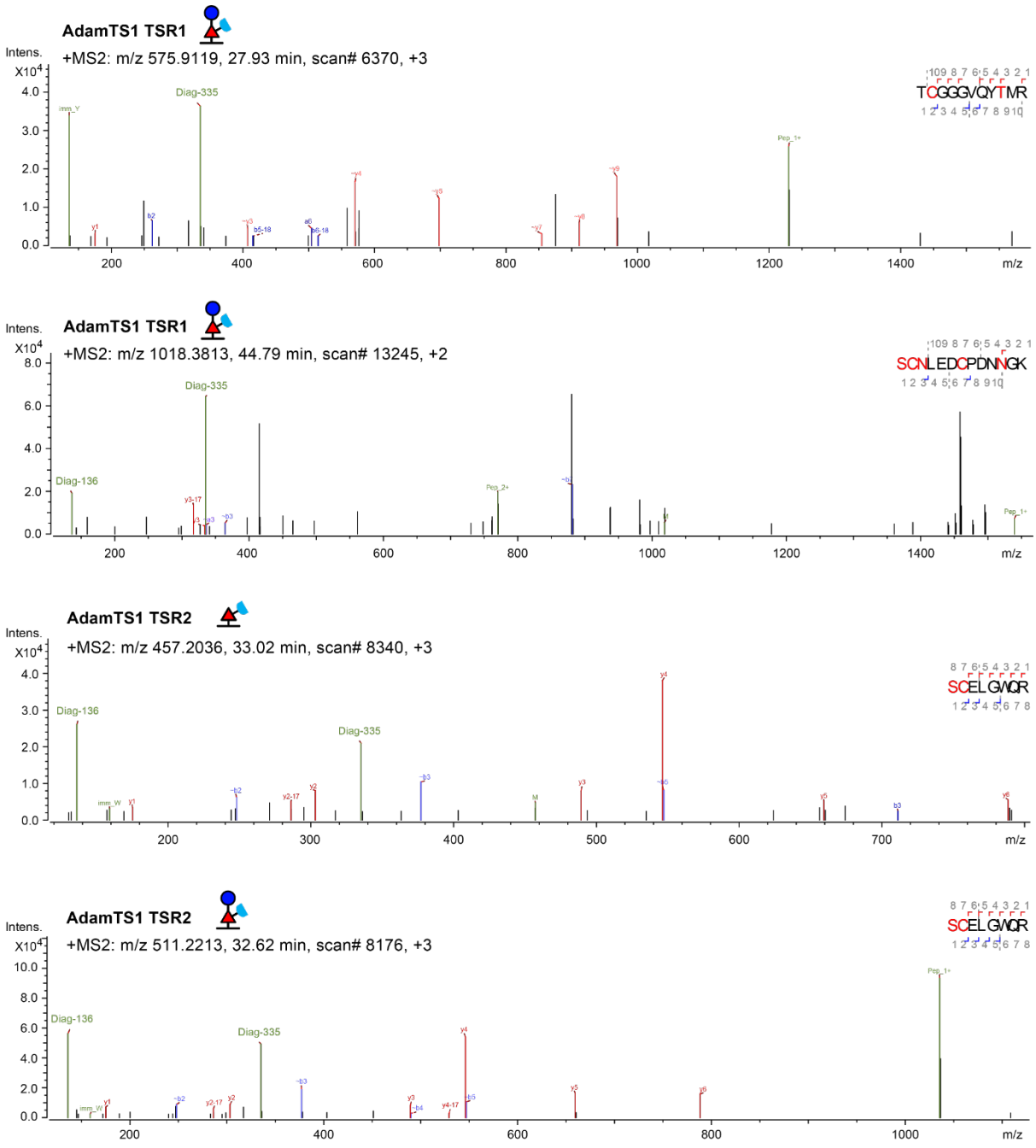


Figure 3S3C. Spectra from ADAMTS1.

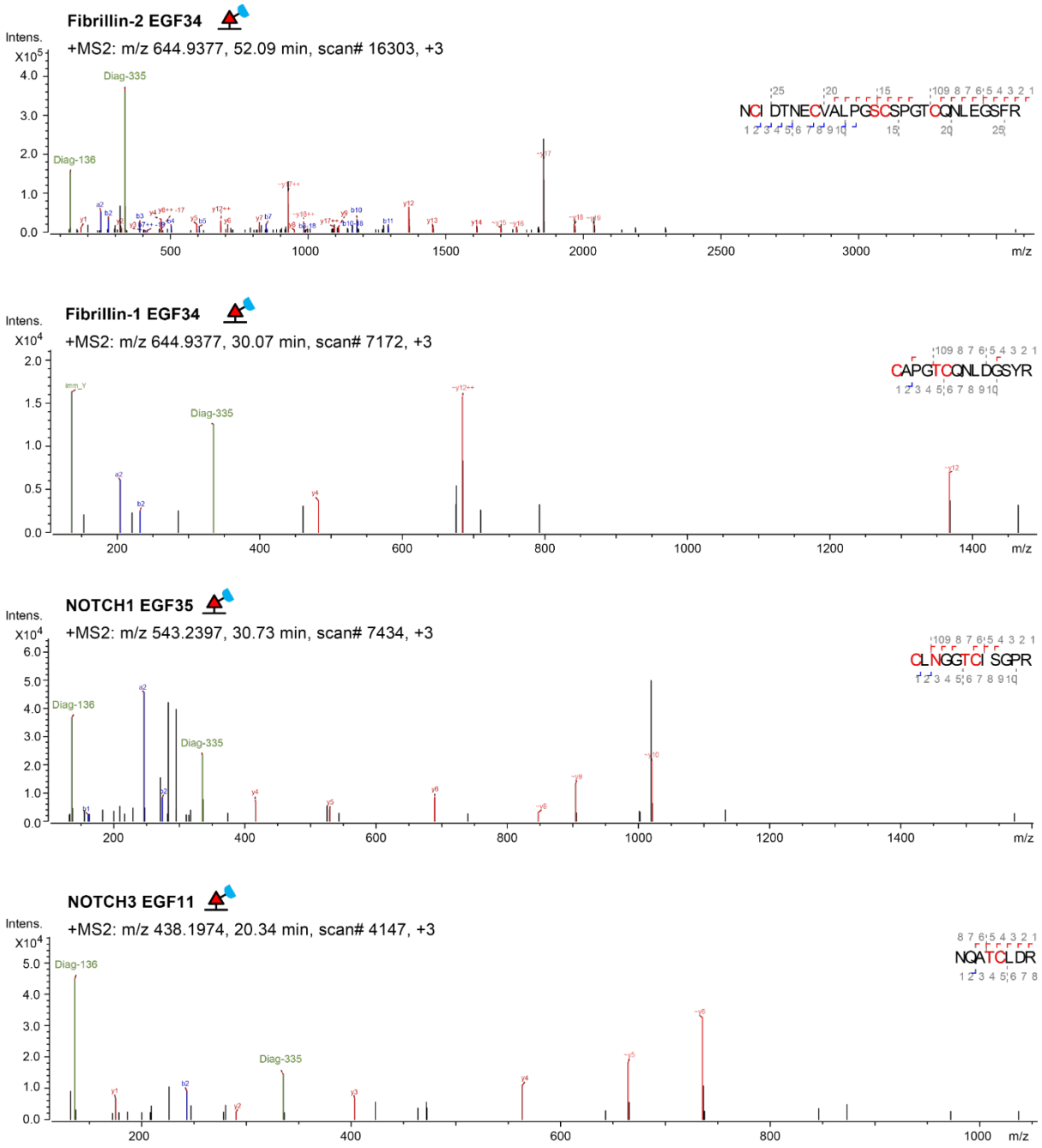


Figure 3S3D. Spectra from Fibrillin-2, Fibrillin-2, NOTCH1, and NOTCH3.

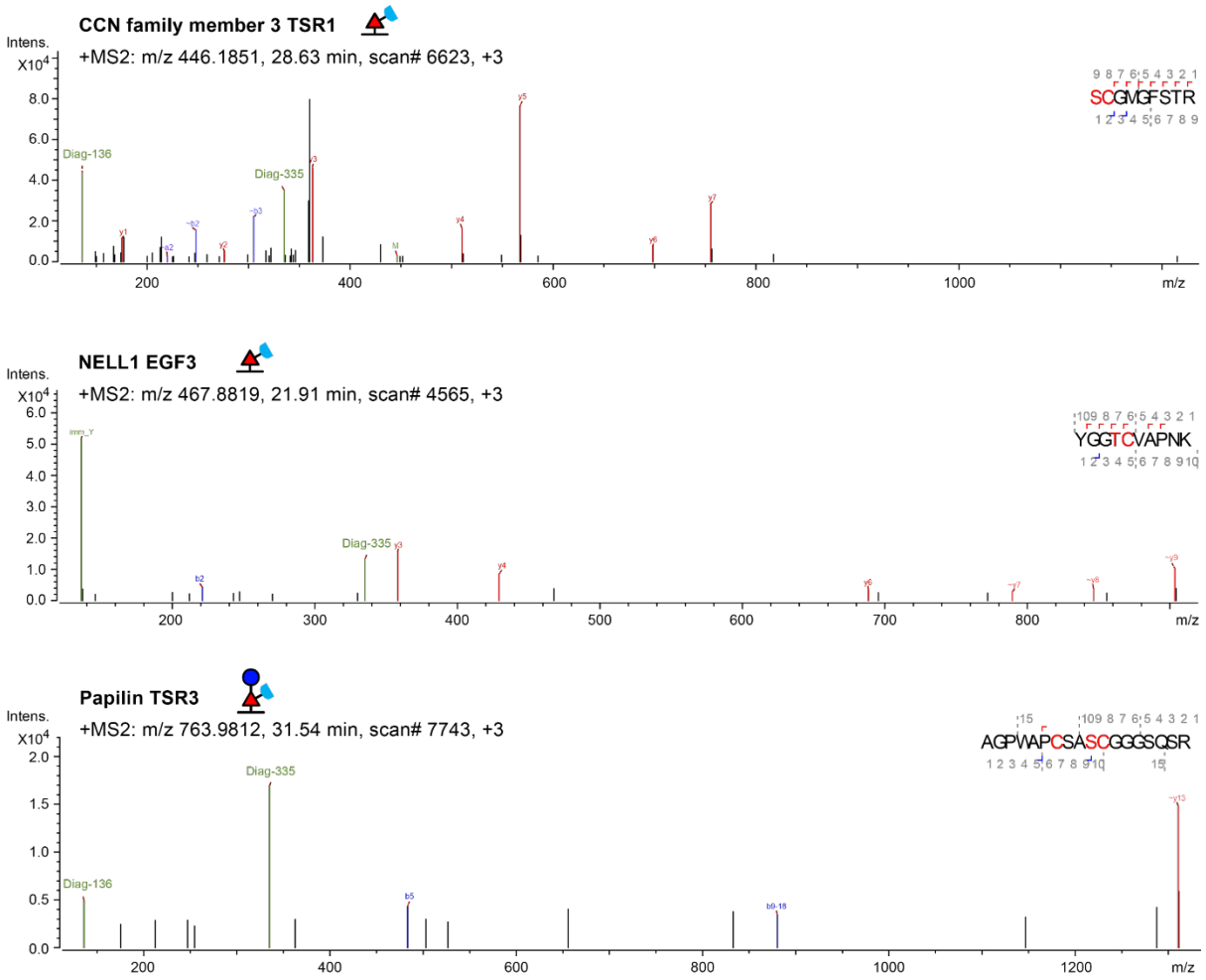


Figure 3S3E. Spectra from CCN3, NELL1, and Papilin.

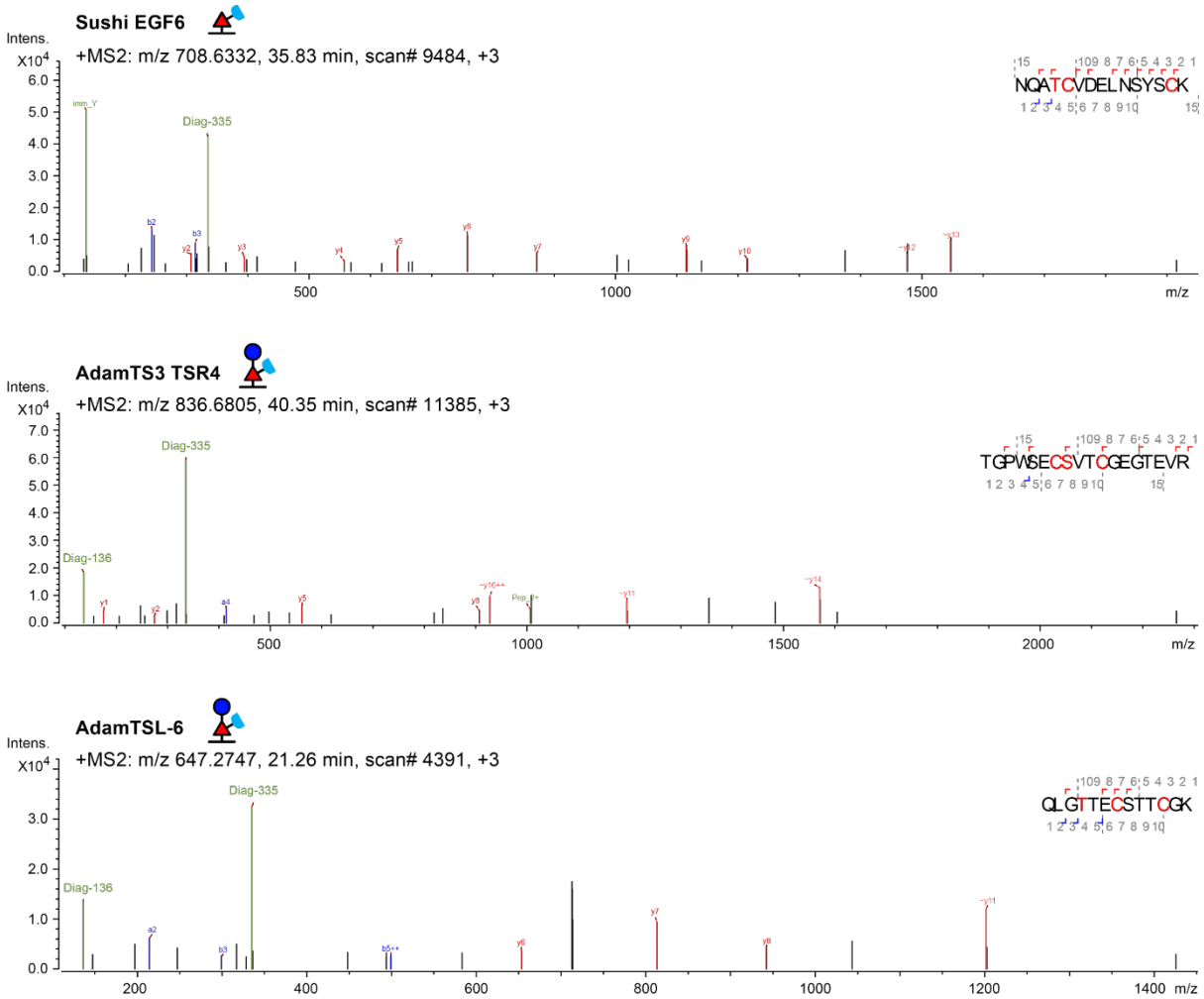


Figure 3S3F. Spectra from Sushi, ADAMTS3, and ADAMTSL-6.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

While domain-specific *O*-fucosylation is confined to a small subset of proteins containing specific protein domains, most of the targets identified to date play crucial roles in biological events and hold therapeutic values. Notable examples include Notch and ADAMTS proteins. In this dissertation, we expand the *O*-fucosylation repertoire by introducing a new type of domain-specific *O*-fucosylation mediated by two novel POFUTs and developing tools for identifying protein targets for POFUT1 and POFUT2.

In Chapter 2, to identify the enzyme responsible for the newly discovered *O*-fucosylation of EMI domains, we performed a series of assays to test whether this process is mediated by the existing POFUTs. Interestingly, knockout of either *POFUT1* or *POFUT2* had no effect on EMI *O*-fucosylation, indicating the existence of an unidentified POFUT for this modification. Of all the FUTs identified in humans, only FUT10 and FUT11 have undefined acceptor substrates. Furthermore, Alphafold2 screens of binding structures with the EMI domain have revealed significant EMI binding by FUT10 and FUT11 which is not seen in other FUTs (data from Dr. Mark Larance, not included in this dissertation). These clues prompted us to investigate the potentials of FUT10 and FUT11 for EMI *O*-fucosylation. In both *in vitro* enzymatic assays using purified proteins and in cellular experiments, we observed robust EMI *O*-fucosylation activity for both FUT10 and FUT11. Thus, we renamed them as POFUT3 and POFUT4. Interestingly, POFUT3 and POFUT4 showed similar behaviors as POFUT1 and POFUT2. These similarities include the requirement of folded substrate structures for efficient modification, functioning in

the ER rather than Golgi, and the participation in a non-canonical ER quality control pathway for the folding and secretion of their respective protein substrates.

In addition to the similarities, the EMI *O*-fucosylation mediated by POFUT3 and POFUT4 displays several unique characteristics that raise many new questions. Unlike EGF and TSR, which have a single fucose modification site, the EMI domain features two spatially adjacent fucose residues. These fucose residues may be oriented in a specific manner to prevent steric clashes. Obtaining crystal structures of the EMI domain with *O*-fucose modification would provide insights into how EMI accommodates two fucose residues simultaneously. As the two fucose sites are situated in separate but anti-parallel β-strands, intramolecular interactions may occur between the fucose and spatially nearby amino acid residues on the other β-strand. This interaction could bring the two beta-strands closer together, thereby restricting their flexibility and compacting the EMI structure. Supporting this hypothesis is the observation that MMRN1 T216A mutant displayed a reduced fucosylation level at the T265 site, indicating that the presence of the T216 fucose facilitates the addition of the T265 fucose (Figure 2S15A). This may be attributed to the T216 fucose generating interactions with residues on the other β-strand, compacting the EMI structure and make it more favorable for the T265 site modification compared to unmodified EMI structures. To test these hypotheses, we can generate co-crystal structures of POFUT3/4 with either wild-type EMI or EMI mutants (T216A, T265A, and T216A/T265A double mutants). Alternatively, we can employ NMR to monitor real-time conformational changes in EMI structures during fucose addition.

We observed that EMI *O*-fucosylation contributes to the secretion of MMRN1, indicating its involvement in a non-canonical ER quality control pathway for protein secretion. However, the detailed mechanisms remain unclear. Unlike *O*-fucose on TSR, which covers and stabilizes

the nearby disulfide bond, the two fucose residues of EMI are located in the middle of the two long anti-parallel β-strands and are not in close proximity to any disulfide bonds. *O*-fucosylation may potentially stabilize EMI structures by generating intramolecular interactions that bring the two β-strands closer together, as mentioned earlier. Conformational changes in EMI structure resulting from fucose addition could be monitored using HPLC, NMR, or Circular Dichroism (CD) spectroscopy. In addition, we can use *in vitro* unfolding assays to test whether the addition of fucose stabilizes the EMI structure and slows down its unfolding (131). To test whether the addition of fucose accelerates the folding of EMI structures, *in vitro* refolding assays can be used, as described in the study by Vasudevan et al (119). Refolding assays are also valuable in investigating whether the fucose residues are added after disulfide bond formation. Samples can be collected at various time points during the refolding process, their fucosylation and disulfide bond formation can be analyzed by mass spectrometry. Furthermore, EMI domains were observed to self-interact, presumably forming a trimer (150). The fucose residues on the EMI domain may play a role in its trimerization, thereby facilitating protein secretion. To examine this, we could mutate the *O*-fucose sites on the EMI domain and assess whether it leads to defects in trimerization. Fast performance liquid chromatography (FPLC) size-exclusion chromatography could be employed to monitor the trimerization.

We did not observe the incorporation of 6-AF into EMI domains when treating cells with 6-AF, suggesting that POFUT3 and POFUT4 may not utilize 6-AF in cells. However, the exact mechanism remains unknown. Predicting the binding structures of POFUT3/4 with 6-AF or GDP-6-AF using Alphafold2, combined with *in vitro* enzymatic assays, can be employed to determine whether GDP-6-AF acts as a competitive inhibitor or un- or non-competitive inhibitor, forming a stable complex with the enzyme that blocks enzymatic activity, or just fails to fit into

the enzyme pocket. Obtaining co-crystal structures of POFUT3/4 with GDP-fucose and EMI domains could provide more detailed information about the essential residues for GDP-fucose binding and the catalytic mechanisms involved in fucose transfer. This information could be utilized to engineer POFUT3/4 to accommodate 6-AF or other fucose analogs, generating a "bump-and-hole" pair to profile the glycoproteome of POFUT3/4, similar to the approach described in Schumann et al (178).

Screening additional EMI-containing proteins for the presence of *O*-fucose at sites equivalent to those observed in MMRN1 and utilizing our POFUT3/4 knockout cells to identify the modifying enzymes, would help refine the precision of the POFUT3/4 consensus sequence. In addition, while we only observed *O*-fucose as a monosaccharide modification on MMRN1 EMI domains, it could be elongated in other EMI-containing proteins. Furthermore, other types of modification could be identified on EMI domains in these proteins. Given the ongoing debate regarding the definition of EMI domains, proteins lacking an annotated EMI domain could still be modified by POFUT3/4 (150, 152, 179). To this end, we can utilize the Alphafold2 predicted folding structures of EMI domains to initiate a virtual screen in the human proteome, searching for candidates with similar folding structures.

In addition to the questions discussed above, numerous questions on the biological functions of POFUT3/4 remain unexplored. Do POFUT3 and POFUT4 display distinct modifying profiles in different tissue contexts or during different developmental stages? What is the biological significance of EMI *O*-fucosylation in physiological or disease contexts? Generating animal models with tissue-specific knockout of *POFUT3* or *POFUT4* would provide insights into these questions and help explore the therapeutic potential of this novel type of domain-specific *O*-fucosylation.

In Chapter 3, we developed the Fuc6nyl* tag for specific labeling of POFUT1/2 targets. Interestingly, we observed that the conjugation of Fuc6nyl* provided additional benefits for glycoproteomic analysis, such as increasing precursor ion charge intensity by introducing an extra proton and generating high-intensity, *O*-fucose-specific diagnostic ions for confident glycopeptide assignments. Leveraging Fuc6nyl*, we developed a glycoproteomic workflow, DidFAT, dedicated for the in-depth profiling of the *O*-fucose proteome of POFUT1/2. We showed the efficient enrichment and identification of *O*-fucosylated peptides from a transfected positive control, validating the effectiveness of this workflow. Initial success was achieved by employing DidFAT to identify endogenous targets of POFUT1/2 from cell secretome. Thirteen endogenous proteins were identified with fucose on their predicted sites, along with the identification of two unexpected *O*-fucose sites on Fibrillin-2 and Agrin. However, as discussed in Chapter 3, the limited number of hits we obtained is contrary to our expectations of a higher yield. This could be due to overexpression of ADAMTS9 TSR2-8, a highly abundant positive control that occupied the majority of secreted POFUT1/2 targets. Combined with the low abundance of *O*-fucosylated species in cell secretome and the inherent dynamic range limitation of the data-dependent acquisition (DDA) used for mass spectrometric data collection, it significantly reduced the available spots for endogenous *O*-fucose peptides to generate MS2 spectra. Note that the data presented here were obtained from a 10 mL input of secreted media from cells transfected with ADAMTS9 TSR2-8. We expect to identify more hits with a larger input of secreted media from untransfected cells. Furthermore, HEK293F cells might not be the ideal cell line for the in-depth profiling of POFUT1/2 target proteins. Applying DidFAT in cell lines with higher secretion levels, broader coverage of the POFUT1/2 modifying proteome, or

upregulated POFUT1/2 activity would be beneficial, such as pre-T cells, cortical epithelial cells, Caco2 cells, and Molt4 cells (140).

Additional DidFAT replicates will be obtained from the secretome of HEK293F cells and other above mentioned cell lines to validate and expand the list of identified *O*-fucosylated peptides after this dissertation. Furthermore, many POFUT1 predicted target proteins are membrane proteins. Applying DidFAT in cell lysates could potentially help us to achieve a deeper coverage of the POFUT1/2 modifying proteome.

References

1. Ginsburg V. Formation of guanosine diphosphate L-fucose from guanosine diphosphate D-mannose. J Biol Chem. 1960;235:2196-201.
2. Tonetti M, Sturla L, Bisso A, Benatti U, De Flora A. Synthesis of GDP-L-fucose by the human FX protein. J Biol Chem. 1996;271(44):27274-9.
3. Coffey JW, Miller ON, Sellinger OZ. THE METABOLISM OF L-FUCOSE IN THE RAT. J Biol Chem. 1964;239:4011-7.
4. Kaufman RL, Ginsburg V. The metabolism of L-fucose by HeLa cells. Exp Cell Res. 1968;50(1):127-32.
5. Ishihara H, Heath EC. The metabolism of L-fucose. IV. The biosynthesis of guanosine diphosphate L-fucose in porcine liver. J Biol Chem. 1968;243(6):1110-5.
6. Ishihara H, Massaro DJ, Heath EC. The metabolism of L-fucose. 3. The enzymatic synthesis of beta-L-fucose 1-phosphate. J Biol Chem. 1968;243(6):1103-9.
7. Wiese TJ, Dunlap JA, Yorek MA. L-fucose is accumulated via a specific transport system in eukaryotic cells. J Biol Chem. 1994;269(36):22705-11.
8. Ng BG, Sosicka P, Xia Z, Freeze HH. GLUT1 is a highly efficient L-fucose transporter. J Biol Chem. 2023;299(1):102738.
9. Lühn K, Laskowska A, Pielage J, Klämbt C, Ipe U, Vestweber D, et al. Identification and molecular cloning of a functional GDP-fucose transporter in Drosophila melanogaster. Exp Cell Res. 2004;301(2):242-50.
10. Lühn K, Wild MK, Eckhardt M, Gerardy-Schahn R, Vestweber D. The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. Nature Genetics. 2001;28(1):69-72.
11. Lu L, Varshney S, Yuan Y, Wei H, Tanwar A, Sundaram S, et al. In vivo evidence for GDP-fucose transport in the absence of transporter SLC35C1 and putative transporter SLC35C2. J Biol Chem. 2023;In Press.
12. Ashikov A, Routier F, Fuhlrott J, Helmus Y, Wild M, Gerardy-Schahn R, et al. The human solute carrier gene SLC35B4 encodes a bifunctional nucleotide sugar transporter with specificity for UDP-xylose and UDP-N-acetylglucosamine. J Biol Chem. 2005;280(29):27230-5.
13. Yurchenco PD, Atkinson PH. Fucosyl-glycoprotein and precursor polls in HeLa cells. Biochemistry. 1975;14(14):3107-14.
14. Yurchenco PD, Atkinson PH. Equilibration of fucosyl glycoprotein pools in HeLa cells. Biochemistry. 1977;16(5):944-53.
15. Feichtinger RG, Hullen A, Koller A, Kotzot D, Grote V, Rapp E, et al. A spoonful of L-fucose-an efficient therapy for GFUS-CDG, a new glycosylation disorder. EMBO Mol Med. 2021;13(9):e14332.
16. Sosicka P, Ng BG, Pepi LE, Shajahan A, Wong M, Scott DA, et al. Origin of cytoplasmic GDP-fucose determines its contribution to glycosylation reactions. J Cell Biol. 2022;221(10).
17. Skurska E, Szulc B, Maszczak-Seneczko D, Wiktor M, Wiertelak W, Makowiecka A, et al. Incorporation of fucose into glycans independent of the GDP-fucose transporter SLC35C1 preferentially utilizes salvaged over de novo GDP-fucose. J Biol Chem. 2022;298(8):102206.
18. Bisso A, Sturla L, Zanardi D, De Flora A, Tonetti M. Structural and enzymatic characterization of human recombinant GDP-D-mannose-4,6-dehydratase. FEBS Lett. 1999;456(3):370-4.

19. Sullivan FX, Kumar R, Kriz R, Stahl M, Xu GY, Rouse J, et al. Molecular cloning of human GDP-mannose 4,6-dehydratase and reconstitution of GDP-fucose biosynthesis in vitro. J Biol Chem. 1998;273(14):8193-202.
20. Richards WL, Kilker RD, Serif GS. Metabolite control of L-fucose utilization. J Biol Chem. 1978;253(23):8359-61.
21. Becker DJ, Lowe JB. Fucose: biosynthesis and biological function in mammals. Glycobiology. 2003;13(7):41R-53R.
22. Ma B, Simala-Grant JL, Taylor DE. Fucosylation in prokaryotes and eukaryotes. Glycobiology. 2006;16(12):158R-84R.
23. Schneider M, Al-Shareffi E, Haltiwanger RS. Biological functions of fucose in mammals. Glycobiology. 2017;27(7):601-18.
24. Flynn RA, Pedram K, Malaker SA, Batista PJ, Smith BAH, Johnson AG, et al. Small RNAs are modified with N-glycans and displayed on the surface of living cells. Cell. 2021;184(12):3109-24 e22.
25. Oriol R, Mollicone R. Fucosyltransferases 1, 2. GDP-Fucose Galactoside α2-Fucosyltransferases. FUT1 or H Blood Group, FUT2 or ABH Secretor Status and Sec1 (FUT1, FUT2, Sec1). Handbook of Glycosyltransferases and Related Genes2014. p. 515-30.
26. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42(Database issue):D490-5.
27. Lowe JB. The blood group-specific human glycosyltransferases. Baillieres Clin Haematol. 1993;6(2):465-92.
28. Scharberg EA, Olsen C, Bugert P. The H blood group system. Immunohematology. 2016;32(3):112-8.
29. Kannagi R. Fucosyltransferase 5. GDP-Fucose Lactosamine α3/4-Fucosyltransferase (FUT5). Handbook of Glycosyltransferases and Related Genes2014. p. 549-58.
30. Kannagi R. Fucosyltransferase 6. GDP-Fucose Lactosamine α3-Fucosyltransferase (FUT6). Handbook of Glycosyltransferases and Related Genes2014. p. 559-71.
31. Kudo T, Narimatsu H. Fucosyltransferase 3. GDP-Fucose Lactosamine α1,3/4-Fucosyltransferase. Lea and Leb Histo-Blood Groups (FUT3, Lewis Enzyme). Handbook of Glycosyltransferases and Related Genes2014. p. 531-9.
32. Kudo T, Narimatsu H. Fucosyltransferase 4. GDP-Fucose Lactosamine α1,3-Fucosyltransferase. Myeloid Specific (FUT4). Handbook of Glycosyltransferases and Related Genes2014. p. 541-7.
33. Kudo T, Narimatsu H. Fucosyltransferase 7. GDP-Fucose Lactosamine α1,3-Fucosyltransferase. Sialyl-Lex Specific (FUT7). Handbook of Glycosyltransferases and Related Genes2014. p. 573-80.
34. Kudo T, Narimatsu H. Fucosyltransferase 9. GDP-Fucose Lactosamine α1,3-Fucosyltransferase. Lex Specific (FUT9). Handbook of Glycosyltransferases and Related Genes2014. p. 597-603.
35. Stowell CP, Stowell SR. Biologic roles of the ABH and Lewis histo-blood group antigens Part I: infection and immunity. Vox Sang. 2019;114(5):426-42.
36. Ihara H, Tsukamoto H, Gu J, Miyoshi E, Taniguchi N, Ikeda Y. Fucosyltransferase 8. GDP-Fucose N-Glycan Core α6-Fucosyltransferase (FUT8). Handbook of Glycosyltransferases and Related Genes2014. p. 581-96.

37. Boruah BM, Kadirvelraj R, Liu L, Ramiah A, Li C, Zong G, et al. Characterizing human α-1,6-fucosyltransferase (FUT8) substrate specificity and structural similarities with related fucosyltransferases. J Biol Chem. 2020;295(50):17027-45.
38. Bastian K, Scott E, Elliott DJ, Munkley J. FUT8 Alpha-(1,6)-Fucosyltransferase in Cancer. Int J Mol Sci. 2021;22(1).
39. Ferrara C, Grau S, Jäger C, Sondermann P, Brünker P, Waldhauer I, et al. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcgammaRIII and antibodies lacking core fucose. Proc Natl Acad Sci U S A. 2011;108(31):12669-74.
40. Luo Y, Haltiwanger RS. O-fucosylation of notch occurs in the endoplasmic reticulum. J Biol Chem. 2005;280(12):11289-94.
41. Luo Y, Koles K, Vorndam W, Haltiwanger RS, Panin VM. Protein O-fucosyltransferase 2 adds O-fucose to thrombospondin type 1 repeats. J Biol Chem. 2006;281(14):9393-9.
42. Kakuda S, Haltiwanger RS. Fucosyltransferases 12, 13: Protein O-Fucosyltransferases 1 and 2 (POFUT1, POFUT2). Handbook of Glycosyltransferases and Related Genes2014. p. 623-33.
43. Holdener BC, Haltiwanger RS. Protein O-fucosylation: structure and function. Curr Opin Struct Biol. 2019;56:78-86.
44. Luo Y, Nita-Lazar A, Haltiwanger RS. Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. J Biol Chem. 2006;281(14):9385-92.
45. Kentzer EJ, Buko A, Menon G, Sarin VK. Carbohydrate composition and presence of a fucose-protein linkage in recombinant human pro-urokinase. Biochem Biophys Res Commun. 1990;171(1):401-6.
46. Buko AM, Kentzer EJ, Petros A, Menon G, Zuiderweg ER, Sarin VK. Characterization of a posttranslational fucosylation in the growth factor domain of urinary plasminogen activator. Proc Natl Acad Sci U S A. 1991;88(9):3992-6.
47. Harris RJ, Leonard CK, Guzzetta AW, Spellman MW. Tissue plasminogen activator has an O-linked fucose attached to threonine-61 in the epidermal growth factor domain. Biochemistry. 1991;30(9):2311-4.
48. Bjoern S, Foster DC, Thim L, Wiberg FC, Christensen M, Komiyama Y, et al. Human plasma and recombinant factor VII. Characterization of O-glycosylations at serine residues 52 and 60 and effects of site-directed mutagenesis of serine 52 to alanine. J Biol Chem. 1991;266(17):11051-7.
49. Harris RJ, Ling VT, Spellman MW. O-linked fucose is present in the first epidermal growth factor domain of factor XII but not protein C. J Biol Chem. 1992;267(8):5102-7.
50. Harris RJ, van Halbeek H, Glushka J, Basa LJ, Ling VT, Smith KJ, et al. Identification and structural analysis of the tetrasaccharide NeuAc alpha(2-->6)Gal beta(1-->4)GlcNAc beta(1-->3)Fuc alpha 1-->O-linked to serine 61 of human factor IX. Biochemistry. 1993;32(26):6539-47.
51. Wang Y, Lee GF, Kelley RF, Spellman MW. Identification of a GDP-L-fucose:polypeptide fucosyltransferase and enzymatic addition of O-linked fucose to EGF domains. Glycobiology. 1996;6(8):837-42.
52. Wang Y, Spellman MW. Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells. J Biol Chem. 1998;273(14):8112-8.

53. Wang Y, Shao L, Shi S, Harris RJ, Spellman MW, Stanley P, et al. Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase. J Biol Chem. 2001;276(43):40338-45.
54. Panin VM, Shao L, Lei L, Moloney DJ, Irvine KD, Haltiwanger RS. Notch ligands are substrates for protein O-fucosyltransferase-1 and Fringe. J Biol Chem. 2002;277(33):29945-52.
55. Shao L, Moloney DJ, Haltiwanger R. Fringe modifies O-fucose on mouse Notch1 at epidermal growth factor-like repeats within the ligand-binding site and the Abruptex region. J Biol Chem. 2003;278(10):7775-82.
56. Müller J, Rana NA, Serth K, Kakuda S, Haltiwanger RS, Gossler A. O-fucosylation of the notch ligand mDLL1 by POFUT1 is dispensable for ligand function. PLoS One. 2014;9(2):e88571.
57. Li M, Cheng R, Liang J, Yan H, Zhang H, Yang L, et al. Mutations in POFUT1, encoding protein O-fucosyltransferase 1, cause generalized Dowling-Degos disease. Am J Hum Genet. 2013;92(6):895-903.
58. Shi S, Stanley P. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. Proc Natl Acad Sci U S A. 2003;100(9):5234-9.
59. Okamura Y, Saga Y. Pofut1 is required for the proper localization of the Notch receptor during mouse development. Mech Dev. 2008;125(8):663-73.
60. Varshney S, Stanley P. Multiple roles for O-glycans in Notch signalling. FEBS Lett. 2018;592(23):3819-34.
61. Artavanis-Tsakonas S, Muskavitch MA. Notch: the past, the present, and the future. Curr Top Dev Biol. 2010;92:1-29.
62. Stanley P, Tanwar A. Regulation of myeloid and lymphoid cell development by O-glycans on Notch. Front Mol Biosci. 2022;9:979724.
63. Moloney DJ, Shair LH, Lu FM, Xia J, Locke R, Matta KL, et al. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. J Biol Chem. 2000;275(13):9604-11.
64. Moloney DJ, Panin VM, Johnston SH, Chen J, Shao L, Wilson R, et al. Fringe is a glycosyltransferase that modifies Notch. Nature. 2000;406(6794):369-75.
65. Brückner K, Perez L, Clausen H, Cohen S. Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. Nature. 2000;406(6794):411-5.
66. Chen J, Moloney DJ, Stanley P. Fringe modulation of Jagged1-induced Notch signaling requires the action of beta 4galactosyltransferase-1. Proc Natl Acad Sci U S A. 2001;98(24):13716-21.
67. Luca VC, Jude KM, Pierce NW, Nachury MV, Fischer S, Garcia KC. Structural biology. Structural basis for Notch1 engagement of Delta-like 4. Science. 2015;347(6224):847-53.
68. Luca VC, Kim BC, Ge C, Kakuda S, Wu D, Roein-Peikar M, et al. Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. Science. 2017;355(6331):1320-4.
69. Taylor P, Takeuchi H, Sheppard D, Chillakuri C, Lea SM, Haltiwanger RS, et al. Fringe-mediated extension of O-linked fucose in the ligand-binding region of Notch1 increases binding to mammalian Notch ligands. Proc Natl Acad Sci U S A. 2014;111(20):7290-5.
70. Kakuda S, Haltiwanger RS. Deciphering the Fringe-Mediated Notch Code: Identification of Activating and Inhibiting Sites Allowing Discrimination between Ligands. Dev Cell. 2017;40(2):193-201.

71. Hofsteenge J, Huwiler KG, Macek B, Hess D, Lawler J, Mosher DF, et al. C-mannosylation and O-fucosylation of the thrombospondin type 1 module. J Biol Chem. 2001;276(9):6485-98.
72. Valero-González J, Leonhard-Melief C, Lira-Navarrete E, Jiménez-Osés G, Hernández-Ruiz C, Pallarés MC, et al. A proactive role of water molecules in acceptor recognition by protein O-fucosyltransferase 2. Nat Chem Biol. 2016;12(4):240-6.
73. Ricketts LM, Dlugosz M, Luther KB, Haltiwanger RS, Majerus EM. O-fucosylation is required for ADAMTS13 secretion. J Biol Chem. 2007;282(23):17014-23.
74. Wang LW, Dlugosz M, Somerville RP, Raed M, Haltiwanger RS, Apte SS. O-fucosylation of thrombospondin type 1 repeats in ADAMTS-like-1/punctin-1 regulates secretion: implications for the ADAMTS superfamily. J Biol Chem. 2007;282(23):17024-31.
75. Adams JC, Tucker RP. The thrombospondin type 1 repeat (TSR) superfamily: diverse proteins with related roles in neuronal development. Dev Dyn. 2000;218(2):280-99.
76. Leonhard-Melief C, Haltiwanger RS. O-fucosylation of thrombospondin type 1 repeats. Methods Enzymol. 2010;480:401-16.
77. Shcherbakova A, Preller M, Taft MH, Pujols J, Ventura S, Tiemann B, et al. C-mannosylation supports folding and enhances stability of thrombospondin repeats. Elife. 2019;8.
78. Furmanek A, Hofsteenge J. Protein C-mannosylation: facts and questions. Acta Biochim Pol. 2000;47(3):781-9.
79. Du J, Takeuchi H, Leonhard-Melief C, Shroyer KR, Dlugosz M, Haltiwanger RS, et al. O-fucosylation of thrombospondin type 1 repeats restricts epithelial to mesenchymal transition (EMT) and maintains epiblast pluripotency during mouse gastrulation. Dev Biol. 2010;346(1):25-38.
80. Benz BA, Nandadasa S, Takeuchi M, Grady RC, Takeuchi H, LoPilato RK, et al. Genetic and biochemical evidence that gastrulation defects in Pofut2 mutants result from defects in ADAMTS9 secretion. Dev Biol. 2016;416(1):111-22.
81. Lesnik Oberstein SA, Kriek M, White SJ, Kalf ME, Szuhai K, den Dunnen JT, et al. Peters Plus syndrome is caused by mutations in B3GALTL, a putative glycosyltransferase. Am J Hum Genet. 2006;79(3):562-6.
82. Iwasaki H, Akashi K. Hematopoietic developmental pathways: on cellular basis. Oncogene. 2007;26(47):6687-96.
83. Stier S, Cheng T, Dombkowski D, Carlesso N, Scadden DT. Notch1 activation increases hematopoietic stem cell self-renewal in vivo and favors lymphoid over myeloid lineage outcome. Blood. 2002;99(7):2369-78.
84. Ohishi K, Varnum-Finney B, Bernstein ID. Delta-1 enhances marrow and thymus repopulating ability of human CD34(+)CD38(-) cord blood cells. J Clin Invest. 2002;110(8):1165-74.
85. Yao D, Huang Y, Huang X, Wang W, Yan Q, Wei L, et al. Protein O-fucosyltransferase 1 (Pofut1) regulates lymphoid and myeloid homeostasis through modulation of Notch receptor ligand interactions. Blood. 2011;117(21):5652-62.
86. Yu VW, Saez B, Cook C, Lotinun S, Pardo-Saganta A, Wang YH, et al. Specific bone cells produce DLL4 to generate thymus-seeding progenitors from bone marrow. J Exp Med. 2015;212(5):759-74.
87. Bhandoola A, Sambandam A. From stem cell to T cell: one route or many? Nature Reviews Immunology. 2006;6(2):117-26.

88. Maillard I, Fang T, Pear WS. Regulation of lymphoid development, differentiation, and function by the Notch pathway. Annu Rev Immunol. 2005;23:945-74.
89. Tanigaki K, Honjo T. Regulation of lymphocyte development by Notch signaling. Nat Immunol. 2007;8(5):451-6.
90. Koch U, Lacombe TA, Holland D, Bowman JL, Cohen BL, Egan SE, et al. Subversion of the T/B lineage decision in the thymus by lunatic fringe-mediated inhibition of Notch-1. Immunity. 2001;15(2):225-36.
91. Feyerabend TB, Terszowski G, Tietz A, Blum C, Luche H, Gossler A, et al. Deletion of Notch1 converts pro-T cells to dendritic cells and promotes thymic B cells by cell-extrinsic and cell-intrinsic mechanisms. Immunity. 2009;30(1):67-79.
92. Visan I, Tan JB, Yuan JS, Harper JA, Koch U, Guidos CJ. Regulation of T lymphopoiesis by Notch1 and Lunatic fringe-mediated competition for intrathymic niches. Nat Immunol. 2006;7(6):634-43.
93. Sheng Y, Yahata T, Negishi N, Nakano Y, Habu S, Hozumi K, et al. Expression of Delta-like 1 in the splenic non-hematopoietic cells is essential for marginal zone B cell development. Immunol Lett. 2008;121(1):33-7.
94. Saito T, Chiba S, Ichikawa M, Kunisato A, Asai T, Shimizu K, et al. Notch2 is preferentially expressed in mature B cells and indispensable for marginal zone B lineage development. Immunity. 2003;18(5):675-85.
95. Tan JB, Xu K, Cretegny K, Visan I, Yuan JS, Egan SE, et al. Lunatic and manic fringe cooperatively enhance marginal zone B cell precursor competition for delta-like 1 in splenic endothelial niches. Immunity. 2009;30(2):254-63.
96. Roy M, Pear WS, Aster JC. The multifaceted role of Notch in cancer. Curr Opin Genet Dev. 2007;17(1):52-9.
97. Postma C, Hermsen MA, Coffa J, Baak JP, Mueller JD, Mueller E, et al. Chromosomal instability in flat adenomas and carcinomas of the colon. J Pathol. 2005;205(4):514-21.
98. Chabanais J, Labrousse F, Chaunavel A, Germot A, Maftah A. POFUT1 as a Promising Novel Biomarker of Colorectal Cancer. Cancers (Basel). 2018;10(11).
99. Deschuyter M, Pennarubia F, Pinault E, Legardinier S, Maftah A. Functional Characterization of POFUT1 Variants Associated with Colorectal Cancer. Cancers (Basel). 2020;12(6).
100. Ma L, Dong P, Liu L, Gao Q, Duan M, Zhang S, et al. Overexpression of protein O-fucosyltransferase 1 accelerates hepatocellular carcinoma progression via the Notch signaling pathway. Biochem Biophys Res Commun. 2016;473(2):503-10.
101. Sawey ET, Chanrion M, Cai C, Wu G, Zhang J, Zender L, et al. Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. Cancer Cell. 2011;19(3):347-58.
102. Li Q, Wang J, Ma X, Wang M, Zhou L. POFUT1 acts as a tumor promoter in glioblastoma by enhancing the activation of Notch signaling. J Bioenerg Biomembr. 2021;53(5):621-32.
103. Kroes RA, Dawson G, Moskal JR. Focused microarray analysis of glyco-gene expression in human glioblastomas. J Neurochem. 2007;103 Suppl 1:14-24.
104. Dong S, Wang Z, Huang B, Zhang J, Ge Y, Fan Q, et al. Bioinformatics insight into glycosyltransferase gene expression in gastric cancer: POFUT1 is a potential biomarker. Biochem Biophys Res Commun. 2017;483(1):171-7.

105. Yokota S, Ogawara K, Kimura R, Shimizu F, Baba T, Minakawa Y, et al. Protein O-fucosyltransferase 1: a potential diagnostic marker and therapeutic target for human oral cancer. Int J Oncol. 2013;43(6):1864-70.
106. Milde-Langosch K, Karn T, Schmidt M, zu Eulenburg C, Oliveira-Ferrer L, Wirtz RM, et al. Prognostic relevance of glycosylation-associated genes in breast cancer. Breast Cancer Res Treat. 2014;145(2):295-305.
107. Wahby S, Jarczyk J, Fierek A, Heinkele J, Weis CA, Eckstein M, et al. POFUT1 mRNA expression as an independent prognostic parameter in muscle-invasive bladder cancer. Transl Oncol. 2021;14(1):100900.
108. Rampias T, Vgenopoulou P, Avgeris M, Polyzos A, Stravodimos K, Valavanis C, et al. A new tumor suppressor role for the Notch pathway in bladder cancer. Nat Med. 2014;20(10):1199-205.
109. Larose H, Prokoph N, Matthews JD, Schlederer M, Högler S, Alsulami AF, et al. Whole Exome Sequencing reveals NOTCH1 mutations in anaplastic large cell lymphoma and points to Notch both as a key pathway and a potential therapeutic target. Haematologica. 2021;106(6):1693-704.
110. Pennarubia F, Ito A, Takeuchi M, Haltiwanger RS. Cancer-associated Notch receptor variants lead to O-fucosylation defects that deregulate Notch signaling. J Biol Chem. 2022;298(12):102616.
111. Weh E, Reis LM, Tyler RC, Bick D, Rhead WJ, Wallace S, et al. Novel B3GALTL mutations in classic Peters plus syndrome and lack of mutations in a large cohort of patients with similar phenotypes. Clin Genet. 2014;86(2):142-8.
112. Aliferis K, Marsal C, Pelletier V, Doray B, Weiss MM, Tops CM, et al. A novel nonsense B3GALTL mutation confirms Peters plus syndrome in a patient with multiple malformations and Peters anomaly. Ophthalmic Genet. 2010;31(4):205-8.
113. Zhang A, Venkat A, Taujale R, Mull JL, Ito A, Kannan N, et al. Peters plus syndrome mutations affect the function and stability of human beta1,3-glucosyltransferase. J Biol Chem. 2021;297(1):100843.
114. Mead TJ, Apte SS. ADAMTS proteins in human disorders. Matrix Biol. 2018;71-72:225-39.
115. Kelwick R, Desanlis I, Wheeler GN, Edwards DR. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. Genome Biol. 2015;16(1):113.
116. Zhang A, Berardinelli SJ, Leonhard-Melief C, Vasudevan D, Liu TW, Taibi A, et al. O-Fucosylation of ADAMTSL2 is required for secretion and is impacted by geleophysic dysplasia-causing mutations. J Biol Chem. 2020;295(46):15742-53.
117. Le Goff C, Morice-Picard F, Dagoneau N, Wang LW, Perrot C, Crow YJ, et al. ADAMTSL2 mutations in geleophysic dysplasia demonstrate a role for ADAMTS-like proteins in TGF-beta bioavailability regulation. Nat Genet. 2008;40(9):1119-23.
118. Dagoneau N, Benoist-Lasselin C, Huber C, Faivre L, Mégarbané A, Alswaid A, et al. ADAMTS10 mutations in autosomal recessive Weill-Marchesani syndrome. Am J Hum Genet. 2004;75(5):801-6.
119. Vasudevan D, Takeuchi H, Johar SS, Majerus E, Haltiwanger RS. Peters plus syndrome mutations disrupt a noncanonical ER quality-control mechanism. Curr Biol. 2015;25(3):286-95.
120. Neupane S, Berardinelli SJ, Cameron DC, Grady RC, Komatsu DE, Percival CJ, et al. O-fucosylation of thrombospondin type 1 repeats is essential for ECM remodeling and signaling during bone development. Matrix Biol. 2022;107:77-96.

121. Holdener BC, Percival CJ, Grady RC, Cameron DC, Berardinelli SJ, Zhang A, et al. ADAMTS9 and ADAMTS20 are differentially affected by loss of B3GLCT in mouse model of Peters plus syndrome. Hum Mol Genet. 2019;28(24):4053-66.
122. Dubail J, Vasudevan D, Wang LW, Earp SE, Jenkins MW, Haltiwanger RS, et al. Impaired ADAMTS9 secretion: A potential mechanism for eye defects in Peters Plus Syndrome. Sci Rep. 2016;6:33974.
123. Hubmacher D, Schneider M, Berardinelli SJ, Takeuchi H, Willard B, Reinhardt DP, et al. Unusual life cycle and impact on microfibril assembly of ADAMTS17, a secreted metalloprotease mutated in genetic eye disease. Sci Rep. 2017;7:41871.
124. Kim ML, Chandrasekharan K, Glass M, Shi S, Stahl MC, Kaspar B, et al. O-fucosylation of muscle agrin determines its ability to cluster acetylcholine receptors. Mol Cell Neurosci. 2008;39(3):452-64.
125. Wang J, Miao Y, Wicklein R, Sun Z, Wang J, Jude KM, et al. RTN4/NoGo-receptor binding to BAI adhesion-GPCRs regulates neuronal development. Cell. 2021;184(24):5869-85 e25.
126. Houlahan CB, Kong Y, Johnston B, Cielesh M, Chau TH, Fenwick J, et al. Ultra-sensitive platelet proteome maps the O-glycosylation landscape and identifies a new form of domain-specific O-fucosylation. Mol Cell Proteomics. 2023;In Revision.
127. Kakuda S, LoPilato RK, Ito A, Haltiwanger RS. Canonical Notch ligands and Fringes have distinct effects on NOTCH1 and NOTCH2. J Biol Chem. 2020;295(43):14710-22.
128. Harvey BM, Rana NA, Moss H, Leonardi J, Jafar-Nejad H, Haltiwanger RS. Mapping Sites of O-Glycosylation and Fringe Elongation on Drosophila Notch. J Biol Chem. 2016;291(31):16348-60.
129. Li Z, Han K, Pak JE, Satkunarajah M, Zhou D, Rini JM. Recognition of EGF-like domains by the Notch-modifying O-fucosyltransferase POFUT1. Nat Chem Biol. 2017;13(7):757-63.
130. Takeuchi H, Yu H, Hao H, Takeuchi M, Ito A, Li H, et al. O-Glycosylation modulates the stability of epidermal growth factor-like repeats and thereby regulates Notch trafficking. J Biol Chem. 2017;292(38):15964-73.
131. Berardinelli SJ, Eletsky A, Valero-Gonzalez J, Ito A, Manjunath R, Hurtado-Guerrero R, et al. O-fucosylation stabilizes the TSR3 motif in thrombospondin-1 by interacting with nearby amino acids and protecting a disulfide bond. J Biol Chem. 2022;298(6):102047.
132. Okajima T, Xu A, Lei L, Irvine KD. Chaperone activity of protein O-fucosyltransferase 1 promotes notch receptor folding. Science. 2005;307(5715):1599-603.
133. Ajima R, Suzuki E, Saga Y. Pofut1 point-mutations that disrupt O-fucosyltransferase activity destabilize the protein and abolish Notch1 signaling during mouse somitogenesis. PLoS One. 2017;12(11):e0187248.
134. McMillan BJ, Zimmerman B, Egan ED, Lofgren M, Xu X, Hesser A, et al. Structure of human POFUT1, its requirement in ligand-independent oncogenic Notch signaling, and functional effects of Dowling-Degos mutations. Glycobiology. 2017;27(8):777-86.
135. Stahl M, Uemura K, Ge C, Shi S, Tashima Y, Stanley P. Roles of Pofut1 and O-fucose in mammalian Notch signaling. J Biol Chem. 2008;283(20):13638-51.
136. Okeley NM, Alley SC, Anderson ME, Boursalian TE, Burke PJ, Emmerton KM, et al. Development of orally active inhibitors of protein and cellular fucosylation. Proc Natl Acad Sci U S A. 2013;110(14):5404-9.

137. Kizuka Y, Nakano M, Yamaguchi Y, Nakajima K, Oka R, Sato K, et al. An Alkynyl-Fucose Halts Hepatoma Cell Migration and Invasion by Inhibiting GDP-Fucose-Synthesizing Enzyme FX, TSTA3. Cell Chem Biol. 2017;24(12):1467-78 e5.
138. Al-Shareffi E, Chaubard JL, Leonhard-Melief C, Wang SK, Wong CH, Haltiwanger RS. 6-alkynyl fucose is a bioorthogonal analog for O-fucosylation of epidermal growth factor-like repeats and thrombospondin type-1 repeats by protein O-fucosyltransferases 1 and 2. Glycobiology. 2013;23(2):188-98.
139. Schneider M, Kumar V, Nordstrom LU, Feng L, Takeuchi H, Hao H, et al. Inhibition of Delta-induced Notch signaling using fucose analogs. Nat Chem Biol. 2018;14(1):65-71.
140. Ma C, Takeuchi H, Hao H, Yonekawa C, Nakajima K, Nagae M, et al. Differential Labeling of Glycoproteins with Alkynyl Fucose Analogs. Int J Mol Sci. 2020;21(17).
141. Li J, Hsu HC, Mountz JD, Allen JG. Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. Cell Chem Biol. 2018;25(5):499-512.
142. Miyoshi E, Moriwaki K, Nakagawa T. Biological function of fucosylation in cancer biology. J Biochem. 2008;143(6):725-9.
143. Sperandio M. Selectins and glycosyltransferases in leukocyte rolling in vivo. FEBS J. 2006;273(19):4377-89.
144. Kumar A, Torii T, Ishino Y, Muraoka D, Yoshimura T, Togayachi A, et al. The Lewis X-related alpha1,3-fucosyltransferase, Fut10, is required for the maintenance of stem cell populations. J Biol Chem. 2013;288(40):28859-68.
145. Mollicone R, Moore SE, Bovin N, Garcia-Rosasco M, Candelier JJ, Martinez-Duncker I, et al. Activity, splice variants, conserved peptide motifs, and phylogeny of two new alpha1,3-fucosyltransferase families (FUT10 and FUT11). J Biol Chem. 2009;284(7):4723-38.
146. Mollicone R, Oriol R. Fucosyltransferases 10, 11. GDP-Fucose N-Glycan Core α1,3-Fucosyltransferases (FUT10, FUT11). Handbook of Glycosyltransferases and Related Genes2014. p. 605-22.
147. Patnaik S. Characterization of Fut10 and Fut11, Putative Alpha-1-3/4 Fucosyltransferase Genes Important for Vertebrate Development. Nature Precedings. 2007.
148. Roos C, Kolmer M, Mattila P, Renkonen R. Composition of Drosophila melanogaster proteome involved in fucosylated glycan metabolism. J Biol Chem. 2002;277(5):3168-75.
149. Du Y, Li D, Li N, Su C, Yang C, Lin C, et al. POFUT1 promotes colorectal cancer development through the activation of Notch1 signaling. Cell Death Dis. 2018;9(10):995.
150. Doliana R, Bot S, Bonaldo P, Colombatti A. EMI, a novel cysteine-rich domain of EMILINs and other extracellular proteins, interacts with the gC1q domains and participates in multimerization. FEBS Lett. 2000;484(2):164-8.
151. Braghetta P, Ferrari A, De Gemmis P, Zanetti M, Volpin D, Bonaldo P, et al. Overlapping, complementary and site-specific expression pattern of genes of the EMILIN/Multimerin family. Matrix Biol. 2004;22(7):549-56.
152. Colombatti A, Spessotto P, Doliana R, Mongiat M, Bressan GM, Esposito G. The EMILIN/Multimerin family. Front Immunol. 2011;2:93.
153. Bot S, Andreuzzi E, Capuano A, Schiavinato A, Colombatti A, Doliana R. Multiple-interactions among EMILIN1 and EMILIN2 N- and C-terminal domains. Matrix Biol. 2015;41:44-55.
154. Marastoni S, Andreuzzi E, Paulitti A, Colladel R, Pellicani R, Todaro F, et al. EMILIN2 down-modulates the Wnt signalling pathway and suppresses breast cancer cell growth and migration. J Pathol. 2014;232(4):391-404.

155. Schiavinato A, Becker AK, Zanetti M, Corallo D, Milanetto M, Bizzotto D, et al. EMILIN-3, peculiar member of elastin microfibril interface-located protein (EMILIN) family, has distinct expression pattern, forms oligomeric assemblies, and serves as transforming growth factor β (TGF-β) antagonist. J Biol Chem. 2012;287(14):11498-515.
156. Zacchigna L, Vecchione C, Notte A, Cordenonsi M, Dupont S, Maretto S, et al. Emilin1 links TGF-beta maturation to blood pressure homeostasis. Cell. 2006;124(5):929-42.
157. Matsumoto K, Yuan Y, Luther KB, Matsuno K, Haltiwanger RS. Distinct Functions of O-fucose for Notch Trafficking in Mammalian Cells and Drosophila. 2023;In Preparation.
158. Kizuka Y, Nakano M, Yamaguchi Y, Nakajima K, Oka R, Sato K, et al. An Alkynyl-Fucose Halts Hepatoma Cell Migration and Invasion by Inhibiting GDP-Fucose-Synthesizing Enzyme FX, TSTA3. Cell Chem Biol. 2017;24(12):1467-78.e5.
159. Moremen KW, Ramiah A, Stuart M, Steel J, Meng L, Forouhar F, et al. Expression system for structural and functional studies of human glycosylation enzymes. Nat Chem Biol. 2018;14(2):156-62.
160. Takeuchi H, Kantharia J, Sethi MK, Bakker H, Haltiwanger RS. Site-specific O-glucosylation of the epidermal growth factor-like (EGF) repeats of notch: efficiency of glycosylation is affected by proper folding and amino acid sequence of individual EGF repeats. J Biol Chem. 2012;287(41):33934-44.
161. Takeuchi H, Schneider M, Williamson DB, Ito A, Takeuchi M, Handford PA, et al. Two novel protein O-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling. Proc Natl Acad Sci U S A. 2018;115(36):E8395-e402.
162. Sakaidani Y, Ichiyanagi N, Saito C, Nomura T, Ito M, Nishio Y, et al. O-linked-N-acetylglucosamine modification of mammalian Notch receptors by an atypical O-GlcNAc transferase Eogt1. Biochem Biophys Res Commun. 2012;419(1):14-9.
163. Jadot M, Boonen M, Thirion J, Wang N, Xing J, Zhao C, et al. Accounting for Protein Subcellular Localization: A Compartmental Map of the Rat Liver Proteome. Mol Cell Proteomics. 2017;16(2):194-212.
164. Geladaki A, Kočevar Britovšek N, Breckels LM, Smith TS, Vennard OL, Mulvey CM, et al. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. Nat Commun. 2019;10(1):331.
165. Fabini G, Freilinger A, Altmann F, Wilson IB. Identification of core alpha 1,3-fucosylated glycans and cloning of the requisite fucosyltransferase cDNA from Drosophila melanogaster. Potential basis of the neural anti-horseadish peroxidase epitope. J Biol Chem. 2001;276(30):28058-67.
166. Rendic D, Linder A, Paschinger K, Borth N, Wilson IB, Fabini G. Modulation of neural carbohydrate epitope expression in Drosophila melanogaster cells. J Biol Chem. 2006;281(6):3343-53.
167. Loriol C, Dupuy F, Rampal R, Dlugosz MA, Haltiwanger RS, Maftah A, et al. Molecular evolution of protein O-fucosyltransferase genes and splice variants. Glycobiology. 2006;16(8):736-47.
168. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005. p. 571-607.
169. Bagdonaite I, Malaker SA, Polasky DA, Riley NM, Schjoldager K, Vakhrushev SY, et al. Glycoproteomics. Nature Reviews Methods Primers. 2022;2(1).

170. de Haan N, Narimatsu Y, Koed Moller Aasted M, Larsen ISB, Marinova IN, Dabelsteen S, et al. In-Depth Profiling of O-Glycan Isomers in Human Cells Using C18 Nanoliquid Chromatography-Mass Spectrometry and Glycogenomics. Anal Chem. 2022;94(10):4343-51.
171. Wohlgemuth J, Karas M, Eichhorn T, Hendriks R, Andrecht S. Quantitative site-specific analysis of protein glycosylation by LC-MS using different glycopeptide-enrichment strategies. Anal Biochem. 2009;395(2):178-88.
172. Schiapparelli LM, McClatchy DB, Liu HH, Sharma P, Yates JR, 3rd, Cline HT. Direct detection of biotinylated proteins by mass spectrometry. J Proteome Res. 2014;13(9):3966-78.
173. Zhou Q, Xie Y, Lam M, Lebrilla CB. N-Glycomic Analysis of the Cell Shows Specific Effects of Glycosyl Transferase Inhibitors. Cells. 2021;10(9).
174. Williamson DB, Sohn CJ, Ito A, Haltiwanger RS. POGLUT2 and POGLUT3 O-glucosylate multiple EGF repeats in fibrillin-1, -2, and LTBP1 and promote secretion of fibrillin-1. J Biol Chem. 2021;297(3):101055.
175. Okajima T, Irvine KD. Regulation of notch signaling by o-linked fucose. Cell. 2002;111(6):893-904.
176. Rampal R, Li AS, Moloney DJ, Georgiou SA, Luther KB, Nita-Lazar A, et al. Lunatic fringe, manic fringe, and radical fringe recognize similar specificity determinants in O-fucosylated epidermal growth factor-like repeats. J Biol Chem. 2005;280(51):42454-63.
177. Roushan A, Wilson GM, Kletter D, Sen KI, Tang W, Kil YJ, et al. Peak Filtering, Peak Annotation, and Wildcard Search for Glycoproteomics. Mol Cell Proteomics. 2021;20:100011.
178. Schumann B, Malaker SA, Wisnovsky SP, Debets MF, Agbay AJ, Fernandez D, et al. Bump-and-Hole Engineering Identifies Specific Substrates of Glycosyltransferases in Living Cells. Mol Cell. 2020;78(5):824-34 e15.
179. Callebaut I, Mignotte V, Souchet M, Mornon JP. EMI domains are widespread and reveal the probable orthologs of the Caenorhabditis elegans CED-1 protein. Biochem Biophys Res Commun. 2003;300(3):619-23.