Identification of Modulators of O-Glycosylation Using Isoform-Specific Sensors for GalNAc Transferase Activity

Lina Song

2016

Identification of Modulators of O-Glycosylation Using Isoform-Specific Sensors for GalNAc Transferase Activity

A Dissertation Presented By

Lina Song

to

The Department of Biological Sciences

Carnegie Mellon University

Pittsburgh, Pennsylvania

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in the field of

Biological Sciences

November 15th, 2016

Thesis Advisor: Dr. Adam D. Linstedt

Abstract

Humans express at least twenty isoforms of GalNAc transferase (termed GalNAc-T1 to T20) that localize to the Golgi apparatus and initiate mucin-type O-glycosylation. Regulation of this family of enzymes affects the function of a vast array of cellular proteins transiting the secretory pathway. This has an impact on many disease states including cancer, which frequently involves profoundly altered cell surface glycan profiles. As a consequence, drug-like modulators of GalNAc-Ts hold promise as entirely new therapeutics for major diseases such as osteoporosis, dyslipidemia, heart disease, and cancer. Additionally, specific modulators will be instrumental in elucidating the basic biology of this complex family of 20 isozymes carrying out a single reaction yet showing developmental and tissue specific expression and partially overlapping substrate specificities. However, there are currently no known inhibitors or activators of these initiating enzymes. To address this shortcoming, we developed a set of cell-based fluorescent sensors to be used for high-throughput screening to identify isoform specific modulators of O-glycosylation. The sensors traffic in the secretory pathway becoming activated if an added inhibitor blocks glycosylation. Simultaneous screening of compounds with two or more isoform specific sensors was used to minimize off-target effects allowing identification of candidates that directly target a particular GalNAc-T. Indeed, one of the lead compounds identified, named T3Inh-1, specifically inhibited purified GalNAc-T3 *in vitro*. T3Inh-1 also blocked glycosylation of FGF23, a specific substrate of GalNAc-T3, in cultured cells and in mice. Further, T3Inh-1 opposed the effect of upregulated GalNAc-T3 in driving metastatic-like cell invasion without affecting cell proliferation. Our hope is that this inhibitor, as well as others from our ongoing screening, can be used for new discoveries and therapeutics involving mucin-type O-glycosylation.

Contents

Chapter 1：Introduction 1

Glycosylation. 1

Generic functions of glycosylation. 1

N-glycosylation versus O-glycosylation. 1

Mucin-type O-glycosylation. 2

GalNAc-Ts. 2

GalNAc-T structure. 3

GalNAc-T related human diseases. 3

Two goals: isoform-specific sensors and drug-like modulators. 7

Chapter 2：Development of isoform-specific sensors of polypeptide GalNAc-T activity 8

2.1 Abstract 8

2.2 Introduction 8

2.3 Results 10

T2 sensor based on ANGPTL3. 10

T3 sensor based on FGF23. 15

T3 sensor based on artificial sequence. 20

Sensor tests in 96-well format. 21

2.4 Discussion 23

2.5 Experimental Procedures 27

Chapter 3: Activity Detection of GalNAc Transferases by Protein-Based Fluorescence Sensors *in vivo* 29

3.1 Abstract 29

3.2 Introduction 29

3.3 Materials 31

3.3.1 Confocal Assay Components 31

3.3.2 Flow Cytometry Assay Components 32

3.4 Methods 32

3.4.1 Assay GalNAc-T activity using biosensor and confocal microscopy 33

3.4.2 Assay GalNAc-T activity using biosensor and multi-well format flow-cytometry 34
3.5 Notes 35

Chapter 4：Inhibitor of GalNAc-T3-mediated O-glycosylation identified by cell-based screening 37
4.1 Abstract 37
4.2 Introduction 37
4.3 Results 39
The primary screening against the two sensors. 39
The kinetics study of T3Inh-1. 40
T3Inh-1 does not show off-target effects. 42
T3Inh-1 blocks invasion of breast cancer cells. 44
T3Inh-1 specifically enhances the cleavage of FGF23. 46
4.4 Discussion 47
4.5 Materials and Methods 49

Chapter 5：Progress towards Identifying additional GalNAc-Ts Modulators by Cell-Based Screening 53
5.1 Abstract 53
5.2 Introduction 53
5.3 Results 54

Chapter 6： Future Directions 58
6.1 T3Inh-1 inhibition mechanism. 58
6.2 Rational design of T3Inh-1 derivatives. 58
6.3 To assay anti-metastatic activity of T3Inh-1 *in vivo* using a mouse model. 59
6.4 Use T3Inh-1 to identify the T3 substrates that mediate the metastatic phenotype. 59
6.5 Determine the range of cancer types that are T3Inh-1 sensitive using the cultured cell invasion assay. 59
6.6 To identify GalNAc-T2-specific modulators and pan-modulators 60

Reference 61

Chapter 1：Introduction

Glycosylation. Glycosylation is the most abundant post-translational modification and it dramatically increases the diversity of the proteome. It involves the linking of glycans to the side chains of amino acids [1,2]. Unlike synthesis of DNA and protein, glycan synthesis is not template dependent. It has been estimated that in mammalian cells there are approximately 700 proteins including ~200 glycosyltransferases involved in the glycosylation processes, creating more than 7,000 different glycan structures [3,4]. Despite this complexity, in mammalian cells, glycans are assembled from only ten monosaccharide units: glucose (Glc), N-acetylglucosamine (GlcNAc), galactose (Gal), N-acetylgalatosamine (GalNAc), mannose (Man), fucose (Fuc), xylose (Xyl), sialic acid (SA), glucuronic acid (GlcA) and iduronic acid (IdoA). In large part physiological conditions regulate enzyme expression, sugar availability and enzyme activity to determine the ultimate glycan composition and abundance.

Generic functions of glycosylation. The initial focus of glycosylation research was on the "glycocalyx", which refers to glycans coating the cell surface to provide a protective barrier. As work progressed many more glycosylation functions were discovered. These include contributions to protein folding, quality control, proteolytic activation, protein localization and protein degradation. As a consequence, glycosylation is involved in a myriad of cellular processes including cell adhesion, cell movement, cell recognition, and many signaling pathways[1,7,8]. As a specific example, consider the sorting of soluble lysosomal proteins in the Golgi apparatus. Due to creation of a mannose-6 phopshate (M6P) signal, soluble lysosomal proteins are differentiated from other proteins and specifically transported to the lysosome by the M6P receptor. The M6P signal is created on mannose in the attached glycans of lysosomal proteins by addition of a phosphate at the 6^{th} position[5]. Another important example is glycan control over ectodomain shedding. In one case, glycosylation near a cleavage site in Notch receptors prevents cleavage of the receptors and this is critical to determination of laterality in *Xenopus tropicalis*[6].

N-glycosylation versus O-glycosylation. In the secretory pathway there are two major types of glycosylation: N-glycosylation and O-glycosylation[1]. They are different in many ways. N-glycosylation is initiated in the ER and the glycan is added *en bloc* to asparagine in the

consensus motif-N-X-S/T (X is any amino acids except proline). It is further processed in the Golgi apparatus and the resulting final N-glycan falls into three categories: high-mannose, hybrid and complex[9–11]. O-glycosylation is typically processed via stepwise attachment of carbohydrates beginning with covalent linkage to the hydroxyl group of a threonine and serine. It can be classified into subgroups based on the first monosaccharide added: O-mannosylation, O-GlcNAcsylation and O-GalNAcsylation[12–15]. O-GalNAcsylation, also called mucin-type O-glycosylation, is the focus of our study. It is described in more detail below. Note that whereas there have been comprehensive functional analyses of N-glycosylation due in large part to the discovery of tunicamycin, an N-glycosylation inhibitor, there is no equivalent inhibitor of mucin-type O-glycosylation and many of its functions remain elusive.

Mucin-type O-glycosylation. Mucin-type O-glycosylation, characterized by the initial addition of N-acetylgalactosamine to the hydroxyl group of serine or threonine residues (and possibly tyrosine), is a large and important subgroup of O-glycosylation[16–18]. This type of O-glycosylation was firstly discovered on mucin proteins, which mainly locate to the surface and extracellular matrix. Mucins are expressed by various epithelial cell types that exist in relatively harsh environments, such as the acidic environment of the stomach, or the complex environment of the intestinal tract. Mucin proteins normally contain a specific domain called mucin-like domain, which is mostly comprised of T-S-P repeats[7,19]. Many of the threonines and serines within this domain are conjugated with branched O-GalNAc glycans. The appended carbohydrate chains protect the cells from those harsh environments. Subsequently, the mucin-type O-glycosylation also has been found to be scattered located on non-mucin proteins and involved in more cellular processes, such as facilitating recognition and bind of ligand with receptor[19,20], affecting receptor's dimerization and autophosphorylation (EGFR)[21,22], directing lymphocyte and leukocyte homing[23,24], and modulating regulated proteolytic processing events to control protein activity, ectodomain shedding and cell signaling[8,25].

GalNAc-Ts. The mucin-type O-glycan chain is built in the Golgi apparatus by the stepwise addition of individual monosaccharides[1]. While one or two isoforms exist for each of the enzymes mediating chain extension, there are up to 20 distinct isoforms of the initiating enzyme UDP-N-acetyl-α-D-galactosamine polypeptide N-acetylgalactosaminyltransferase (termed GalNAc-Ts, T1-T20) in humans[26]. These isoforms have distinct yet overlapping

substrate specificities and show distinct temporal and spatial expression patterns[26]. Much remains to be determined regarding the purpose of this large number of isozymes covering a single reaction.

GalNAc-T structure. GalNAc-Ts are type II transmembrane proteins. All share a domain organization in which a short cytoplasmic tail and single transmembrane domain is followed by a variable length lumenal stem region and then catalytic and lectin domains (except for T20, which lacks the lectin domain)[26,27]. Binding of the catalytic domain to the substrate largely determines the substrate specificity of each isoform[28]. It has been shown *in vitro* that this depends mostly on the three residues flanking the glycan attachment site. However, the lectin domain influences this activity by binding GalNAc already linked to the substrate at a nearby site[29–31]. Binding by the lectin domain presumably increases the enzyme/substrate interaction and it also directly activates the catalytic domain[32].

GalNAc-T related human diseases. Because of the large number of proteins modified by O-glycosylation in the Golgi and the potential of O-glycosylation to profoundly affect functionality of the modified proteins, it should not be surprising that abnormal GalNAc-T activity perturbs cell function and causes disease[32]. Examples of normal cell functions that are perturbed by altered O-glycosylation include cell adhesion, migration and tissue lubrication. There are already at least 25 well-defined medical syndromes linked to defects in O-glycosylation and this number is likely to grow[8]. There is abundant evidence that altered expression of GalNAc-T isozymes is a common feature of many types of cancer and, importantly, that the resulting aberrant glycosylation has functional effects in the complex cascade of cancer metastasis. Table 1.1 summarizes how the misregulation of GalNAc-T expression correlates with many cancer types. Another example is a loss of function mutation in T3 that results in increased processing of fibroblast growth factor 23 (FGF23) leading to familial tumoral calcinosis and hyperphosphatemia-hyperstosis[33–35]. Similarly, T2 and T11 deficiencies are connected to high serum levels of triglyceride and high-density lipoprotein cholesterol[36] and heterotaxy disorder, respectively[6]. Also, T1 is involved in heart valve development and cardiac function[37].

Table 1.1. Many cancer types are correlated with misregulation of GalNAc-Ts.

Year	Isoform	Cancer Type	Authors	Substrates	Notes
2015	T1	Hepatocellular carcinoma	Miao-Juei Huang Min-Chuan Huang	EGFR	High expression of T1 is associated with poor patient survival. Knockdown decreased EGFR activation and increased its degradation by decreasing EGFR O-glycosylation[21].
2011	T2	Hepatocellular carcinoma	Yao-Ming Wu Min-Chuan Huang	EGFR	Restoring GalNAc-T2 expression in HCC cells suppressed EGF-induced cell growth, migration and invasion *in vitro* and *in vivo*[38].
2012	T2	Gastric cancer	Dong Hua Shiliang Wu	TGF-β1? MMP2?	GalNAc-T2 may exert anti-proliferative and anti-metastatic activity through the decrease of MMP-2 and TGF-β1[39].
2014	T2	Oral squamous cell carcinoma	Mei-Chun Lin Min-Chuan Huang	EGFR	GalNAc-T2 was overexpressed in OSCC tissue[22].
2014	T2	Neuroblastoma tumor	Wan-Ling Ho Min-Chuan Huang	IGF-1R	GalNAc-T2 expression is an independent prognostic factor for better survival for NB patients. GalNAc-T2 overexpression suppressed IGF-1-induced cell growth, migration, and invasion in NB cells[40].
2011	T3/T6	Pancreatic cancer	Z Li Y Sasaguri		Both GalNAc-T6 and T3 expressions correlated significantly with tumour differentiation, whereas only GalNAc-T6 expression predicted prognosis in pancreatic cancer[41].
2011	T3	Pancreatic cancer	K Taniuchi MA hollingsworth	GNAT1	GalNAc-T3 is overexpressed in human pancreatic cancer tissues and suppression of it significantly attenuates the growth of pancreatic cancers cells *in vitro* and *in vivo*[42].
2013	T3	Thyroid carcinoma	Yasuhiro Mochizuki Jun Amano		GalNAc-T3 highly expressed in papillary carcinomas that had invaded beyond the thyroid capsule[43].
2013	T3	Renal cell carcinoma	S Kitada Y Sasaguri		High expression of GalNAc-T3 showed a close

					relationship to poor clinical performance and large tumor size, or pathologically high Fuhrman's grading, and presence of vascular and necrosis[44].
2014	T3	Ovarian Cancer	Zhi-Qiang Wang Dimcho Bachvarov	MUC1	GalNAc-T3 suppression inhibited EOC cell migration and invasion.[45]
2015	T3	Early stage oral squamous cell carcinoma	Yoshikazu Harada Sohsuke Yamada		Patients with high expression of T3 have shorter disease-free survival[46].
2014	T5	Gastric cancer	H He Y Sun		GalNAc-T5 expression was markedly reduced in gastric cancer tissues[47].
2011	T6	Breast cancer	Jae-Hyun Park Yusuke Nakamura	Fibronectin	Overexpression of GalNAc-T6 disrupted acinar structure formation and EMT-like cellular alterations [48].
2016	T6	Pancreatic cancer	Yunus Emre Tarhan Jae-Hyun Park	Mucin4	Knockdown of GalNAc-T6 decreases glycosylation and transcription of Mucin 4. It also make pancreatic cells switch from P-cadherin to E-cadherin[49].
2016	T6	Lung adenocarcinoma	Zhi Li Yasuyuki Sasaguri		GalNAc-T6 expression independently predict poor overall survival[50].
2011	T7	Human melanoma	Avital Gaziel-Sovran Eva Hernando		Ectopic expression of miR-30b/30d promoted the metastatic behavior of melanoma cells by directly targeting the GalNAc-T7[51].
2012	T7	Cervical cancer	Rui-Qing Peng Hua Tang		The knockdown of GalNAc-T7 is markedly inhibits cervical cancer cell proliferation, migration and invasion, whereas ectopic expression of GalNAc T7 significantly enhances these properties[52].
2013	T7	Hepatocellular carcinoma	Sze Wan Shan Burton B. Yang		Silencing GalNAc-T7 by miR enhanced proliferation and migration[53].

2014	T7	Hep-2 laryngeal carcinoma	Wei Li Ji Sun		miR-34a and miR-34c may function as tumor suppressors in LSCC through down-regulation of GALNT7[54].
2015	T7	Nasopharyngeal carcinoma	Hong-Fang Duan Guo-Hui Nie		High-expression of GalNAc-T7 is associated with increased proliferation, colony formation, migration and invasion of NPC-derived cells. T7 is regulated by miR-494[55].
2013	T9	Neuroblastoma	Nora Berois Eduardo Osinaga		GalNAc-T9 expression was associated with high overall survival[56].
2012	T10	Gastric cancer	Yuan Gao Wenli Ma		GalNAc-T10 is an indicator of tumor differentiation in gastric cancer[57].
2014	T10	Clear-cell renal cell carcinoma	Qian Wu Jiejie Xu		High GalNAc-T10 expression indicated poor survial and early recurrence of patients with ccRCC[58].
2014	T10	HBV-associated hepatocellular carcinoma	Qian Wu Jian-Xin Gu	EGFR	GalNAc-T10 promotes proliferation and apoptosis resistance of Hepatoma cells. Its expression is regulated by miR-122[59].
2013	T11	Chronic lymphocytic leukemia	M. G. Libisch C. Robello		GalNAc-T11 is significantly overexpressed in CLL cells[60].
2012	T13	Murine levis lung cancer	Yasuyuki Matsumoto Koichi Furukawa		Stable transfection of GalNAc-T13 into C4 (T13-TF) resulted in increased invasion and motility[61].
2013	T13	Lung cancer	Yasuyuku Matsumoto Koichi Furukawa	Syndecan 1	GalNAc-T13 enhanceds metastasis[62].
2013	T14	Ovarian cancer	Ranran Wang Zhu Yang	Muc 13	GalNAc-T14 may contribute to ovarian carcinogenesis through aberrant glycosylation of MUC13, but not through the IL-8 pathway[63].
2015	T14	Breast cancer cell MCF-7	Tian Huanna Wu Chen	MUC1 MMP2	Knockdown of T14 decreased expression of N-cadherin, Vimentin, MMP2, VEGF, TGF-β, and increased expression of E-cadherin. Knockdown of T14 inhibited

					TGF-β1-induced cell morphological change[64].

Two goals: isoform-specific sensors and drug-like modulators. Given the biological and medical importance of GalNAc-T family, it is surprising that there are no tools to monitor their activity and no drug-like modulators to regulate it. To overcome this, we successfully developed GalNAc-T2 and T3 specific fluorescence sensors to monitor the corresponding enzymes[65,66]. The sensors traffic in the secretory pathway. Each sensor yielded little signal when glycosylated but was strongly activated in the absence of its glycosylation. Specificity of each sensor was confirmed in HEK cells with either the T2 or T3 enzyme deleted. These are the first reported tools to monitor the activity of single GalNAc-T isoforms and the strategy can now be used to develop sensors for the remaining isoforms. Using the isoform-specific sensors, we then designed a high-throughput screening strategy to identify drug-like inhibitors or activators of specific GalNAc-T isoforms. In our initial screening the best hit, named T3Inh-1, showed highly selective activation of T3 sensor and, crucially, it directly inhibited T3 *in vitro*. Further, T3Inh-1 blocked glycosylation of FGF23, specific substrate of GalNAc-T3, both *in vitro* and *in vivo*. Remarkably, it also strongly impaired cancer cell invasion. T3Inh-1 is the first validated inhibitor of any GalNAc-T. It will be useful to in vestigate the functions of T3 and, possibly as a therapeutic agent in certain human diseases arising from desregulation of T3.

Chapter 2：Development of isoform-specific sensors of polypeptide GalNAc-T activity

2.1 Abstract

Humans express up to twenty isoforms of GalNAc-T that localize to the Golgi apparatus and initiate the mucin-type O-glycosylation. Regulation of this enzyme family affects a vast array of proteins transiting the secretory pathway and diseases arise upon misregulation of specific isoforms. Surprisingly, molecular probes to monitor GalNAc-T activity are lacking and there exist no effective global or isoform-specific inhibitors. Here we describe the development of GalNAc-T2- and T3-isoform specific fluorescence sensors that traffic in the secretory pathway. Each sensor yielded little signal when glycosylated but was strongly activated in the absence of its glycosylation. Specificity of each sensor was assessed in HEK cells with either the GalNAc-T2 or T3 enzyme deleted. Although the sensors are based on specific substrates of the GalNAc-T2 and T3 enzymes, elements in or near the enzyme recognition sequence influenced their activity and required modification, which we carried out based on previous *in vitro* work. Significantly, the modified T2 and T3 sensors were activated only in cells lacking their corresponding isozymes. Thus, we have developed GalNAc-T2 and T3 specific sensors that will be valuable in both the study of GalNAc-T regulation and in high throughput screening for potential therapeutic regulators of specific GalNAc-Ts.

2.2 Introduction

Despite the biological and medical significance of the GalNAc-T family, GalNAc-T isoform-specific assays are lacking that monitor their activity in living cells and there are no known inhibitors. To address these shortcomings, we initiated the development of cell-based fluorescence sensors whose fluorescence is a readout of GalNAc-T activity (Figure 2.1A)[65]. The sensors are designed to traffic in the secretory pathway becoming activated if their glycosylation is inhibited. Fluorescence is dependent on inhibition because glycosylation prevents furin protease from removing a blocking domain, which occludes binding of the dye malachite green (MG) to a fluorescence activating protein domain (Figure 2.1B). Thus, when

glycosylation is inactive (e.g. in the presence of an active drug) the peptide sequence is cleaved by furin-like proprotein convertase processing and the sensor is activated. The biosensor operates in a "turn-on" mode and it is ratiometric because it also contains a fluorescent protein domain, GFP.

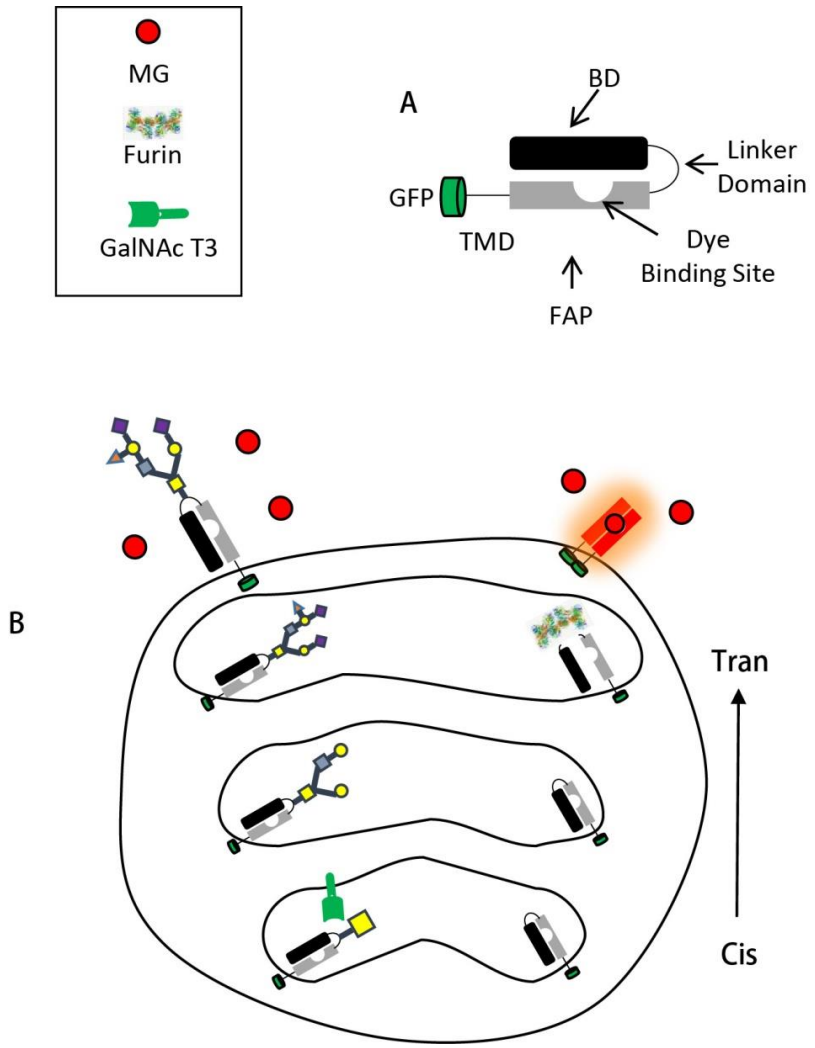

FIGURE 2.1. Sensor design and its traffic pathway. (A) The schematic shows an intact sensor with its blocking domain (BD), linker domain with a threonine site for GalNAc-T and an arginine site for furin, fluorescent activating protein domain (FAP), transmembrane domain (TMD), and green fluorescent protein domain (GFP). (B) The schematic shows the traffic pathway of sensor under different conditions. At the normal condition, GalNAc-Ts initiate the mucin-type O-glycan synthesis on threonine of linker peptide (left pathway). Glycan attachment masks the furin site but if an inhibitor blocks glycan attachment then furin cleaves the linker releasing the BD. As shown on the right pathway, the FAP then dimerizes and binds and activates the dye (malachite green).

Despite the biological and medical significance of the GalNAc-T family, GalNAc-T isoform-specific assays are lacking that monitor their activity in living cells and there are no

known inhibitors. To address these shortcomings, we initiated the development of cell-based fluorescence sensors whose fluorescence is a readout of GalNAc-T activity (Figure 2.1A)[65]. The sensors are designed to traffic in the secretory pathway becoming activated if their glycosylation is inhibited. Fluorescence is dependent on inhibition because glycosylation prevents furin protease from removing a blocking domain, which occludes binding of the dye malachite green (MG) to a fluorescence activating protein domain (Figure 2.1B). Thus, when glycosylation is inactive (e.g. in the presence of an active drug) the peptide sequence is cleaved by furin-like proprotein convertase processing and the sensor is activated. The biosensor operates in a "turn-on" mode and it is ratiometric because it also contains a fluorescent protein domain, GFP.

Our previously characterized the first generation sensor[65] contained a target substrate sequence with a non-restrictive GalNAc-T isoform consensus sequence for O-glycosylation. Because important diseases and regulatory changes involve single GalNAc-T isoforms and therapeutics should ideally target specific isoforms, our intention in the present study was to convert the biosensor to an isoform-specific modality, thereby addressing whether distinct isoforms exhibit true substrate specificity. Focusing on GalNAc-T2 and T3 we were able to design and generate isoform-specific detection but it required optimization of known sequences targeted by these enzymes. The strategy we used to develop GalNAc-T2 and T3 specific sensors can now be used to develop other isoform specific sensors. Significantly, with two or more isoform-specific sensors at hand, parallel screening can identify compounds selectively activating a single sensor thereby greatly reducing the likelihood of off-target effects.

2.3 Results

T2 sensor based on ANGPTL3. To develop a sensor specifically sensitive to the activity of GalNAc-T2 we used a sequence from angiopoietin-like 3 (ANGPTL3) as the linker in our previously described fluorescent glycosylation sensor[65]. All linker sequences used in this chapter are listed in Table 2.1. The chosen $_{219}$KPRAPRTTPF$_{228}$ sequence of ANGPTL3 is an *in vitro* target of GalNAc-T2 and to a lesser extent GalNAc-T3[67]. In liver cells, which naturally express ANGPTL3 and lack GalNAc-T3, the sequence is a specific target of GalNAc-T2[68]

indicating that other family members expressed in these cells do not recognize it. Glycosylation of the sequence on T_{226} blocks furin-like proprotein processing cleavage at R_{224}[1,67]. If this glycan masking were recapitulated in the context of the sensor, the sensor should be dark under normal conditions and activated upon inhibition of glycosylation at T_{226}. That is, absence of glycosylation should allow furin to cleave the linker, removing the inhibitory domain thereby allowing activation of the MG dye (Figure 2.1B). Because there are no available inhibitors against GalNAc-Ts, we used mutation of the glycan acceptor site in the sequence to test sensor functioning. Human embryonic kidney (HEK) cells were used in our experiments and they express all isoforms except GalNAc-T5, T9, T15, T17, T19 and T20 (www.proteinatlas.org). In anticipation of the likelihood that GalNAc-T3 would also glycosylate T_{226} we simultaneously tested a glycine substituted at position -1 ($T_{225}G$) relative to the glycosylation site because glycine substitution at this position interferes with GalNAc-T3 but not GalNAc-T2-mediated glycosylation *in vitro*[69].

Table 2.1 Sensor linker domain sequences

Sensor	Version	Sequence
ANGPTL3-based T2 sensor	WT	219-KP**RAPR**TTPF-228
	$T_{225}G$	-KP**RAPR**GTPF-
	$T_{226}G$ (acceptor site)	-KP**RAPR**TGPF-
	$T_{225}G/T_{226}G$	-KP**RAPR**GGPF-
FGF23-based T3 sensor	WT	168-HFNTPIPR**RHTR**SAEDDG-185
	$T_{171}G$ (upstream site)	-HFNGPIPR**RHTR**SAEDDG-
	$T_{178}G$ (acceptor site)	-HFNTPIPR**RHGR**SAEDDG-
	$H_{177}V$	-HFNTPIPR**RVTR**SAEDDG-
	$H_{177}V/T_{178}G$	-HFNTPIPR**RVGR**SAEDDG-
	$H_{177}V/S_{180}G/A_{181}P$	-HFNTPIPR**RVTRGP**EDDG-
	$H_{177}V/T_{178}G/S_{180}G/A_{181}P$	-HFNTPIPR**RVGRGP**EDDG-
Artificial sequence sensor	WT	1-R**RAYR**VTPGP-10
	$T_{7}G$ (acceptor site)	-R**RAYR**VGPGP-

Note: Residues comprising furin recognition are in bold and underlined and residues that are mutated from wildtype (WT) are in red.

HEK cells were generated that stably expressed one of four versions of the ANGPTL3-based sensor: wildtype (WT), $T_{225}G$, $T_{226}G$, or $T_{225}G/T_{226}G$. GFP fluorescence present in the sensor was used to verify expression and to quantify sensor activation as a ratio of MG fluorescence per unit GFP fluorescence. As expected, all cells expressed the sensors on the cell

surface as indicated by GFP fluorescence (Figure 2.2, left panel). Significantly, MG fluorescence was not apparent for both the WT version (Figure 2.2B) and the version with a glycine at position -1 (Figure 2.2E), whereas there was strong activation apparent for both versions lacking the T_{226} glycan acceptor site (Figure. 2.2H and K). Compared to WT, fluorescence was 267 and 558 fold greater for $T_{226}G$ and $T_{225}G/T_{226}G$, respectively (Figure 2.2M). Each increase was well over 500 standard deviations (SD) indicating a remarkable signal to noise ratio. Interestingly, the $T_{225}G/T_{226}G$ signal was about twice that of $T_{226}G$, indicating that glycine at -1 unexpectedly enhanced cleavage of the sensor. Note that, were glycosylation to take place at T_{225} it could account for this result but an observation described in the next paragraph counters this possibility.

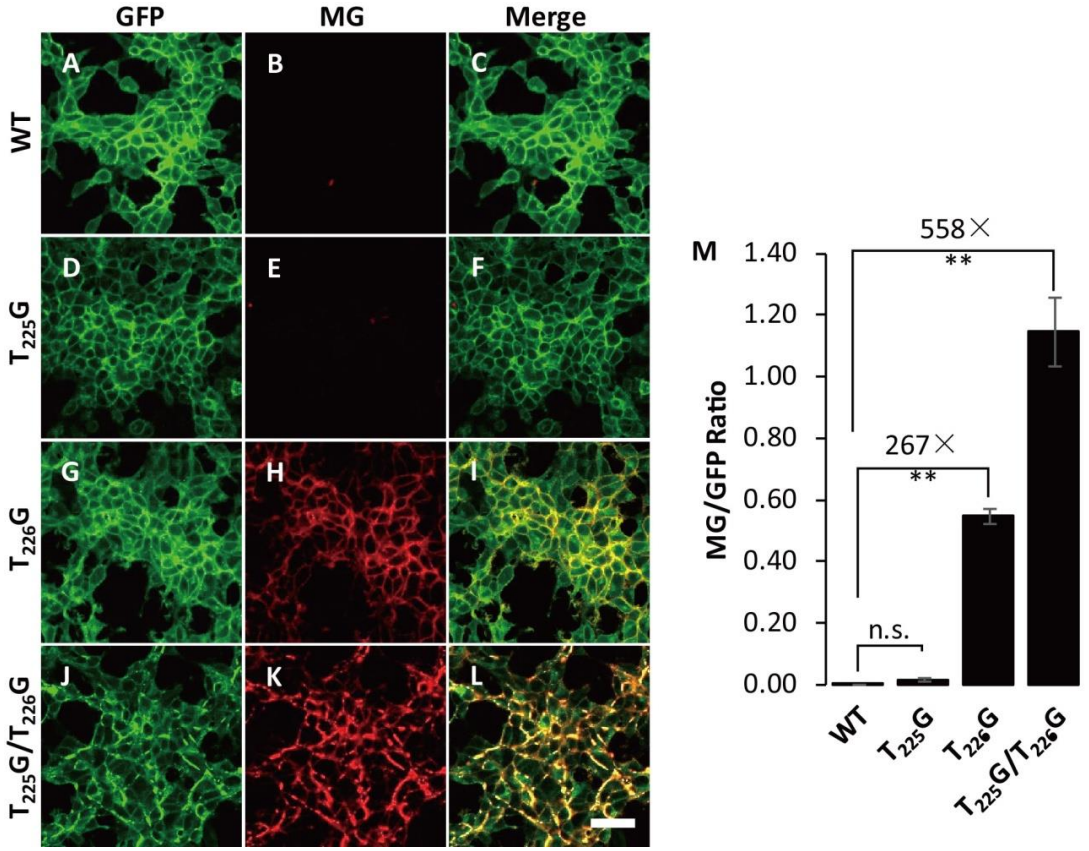

Figure 2.2. Evaluation of the ANGPTL3-based sensor for GalNAc-T2. (A-L) Representative images are shown for HEK cell lines expressing wildtype (WT, A-C), $T_{225}G$ (D-F), $T_{226}G$ (G-I) or $T_{225}G/T_{226}G$ (J-L) versions of the ANGPTL3-based T2 sensor. Images are GFP, malachite green (MG) and a merge of GFP (green) and MG (red). Image acquisition was identical within each channel and all but J-L were acquired on the same day. Bar=50μm. (M) The ratio of the average MG fluorescence per image field to the average GFP fluorescence for the same fields is plotted for the indicated constructs with fold differences shown (** = $p<0.001$, n.s. = not significant, 8 fields per condition per experiment, n = 4 ± SEM).

To test the GalNAc-T isoform specific control of sensor activation, we used zinc finger-nuclease gene engineered HEK cells lacking either the GalNAc-T2 or T3 enzymes (double knockout was not available). If the ANGPTL3- based sensor was GalNAc-T2 isoform specific in HEK cells, we expected activation of the sensor in ΔT2 cells. However, while GFP fluorescence indicated strong expression of the WT sensor, no activation of MG fluorescence was observed (Figure 2.3A-C). Importantly though, the T_{225}G version of the sensor was strongly activated in ΔT2 cells (Figure 2.3D-F). The specific and strong activation of the T_{225}G version of the sensor in ΔT2 cells (≈170 SD over control HEK cells expressing the same construct) meets the key criterion for isoform specific sensing. We speculate that glycine at position -1 of the acceptor site selectively interferes with recognition of T_{226} by the GalNAc-T3 isoform leaving GalNAc-T2 as the sole enzyme targeting T_{226}. The reason we favor this explanation rather than GalNAc-T3-mediated glycosylation of T_{225} is that glycine substitution of the T_{226} glycan acceptor site led to robust activation (Figure 2.3G-I). This activation argues that T_{225}, which remains present in this construct, is not glycosylated.

We also tested the WT and T_{225}G sensors in ΔT3 cells. Although well expressed, neither WT (Figure 2.3J-L) nor T_{225}G (Figure 2.3M-O) was activated, consistent with GalNAc-T2 being sufficient to carryout effective glycan masking of either sensor version. Quantification of all the results for ΔT2 and ΔT3 cells confirmed these conclusions (Figure 2.3P). Due to the lack of availability of other cell lines, we were limited to the ΔT3 cell line in controlling for participation of other isoforms. Nevertheless, such tests were arguably unimportant (even for the ΔT3 cell line) because the level of activation of the T_{225}G in ΔT2 cells was near maximal. This implies that GalNAc-T2 was the sole GalNAc-T in HEK capable of using the modified site (i.e. glycine at position -1) as a substrate. In summary, the T_{225}G modified ANGPTL3-based sensor provides a specific readout of GalNAc-T2 activity in HEK cells.

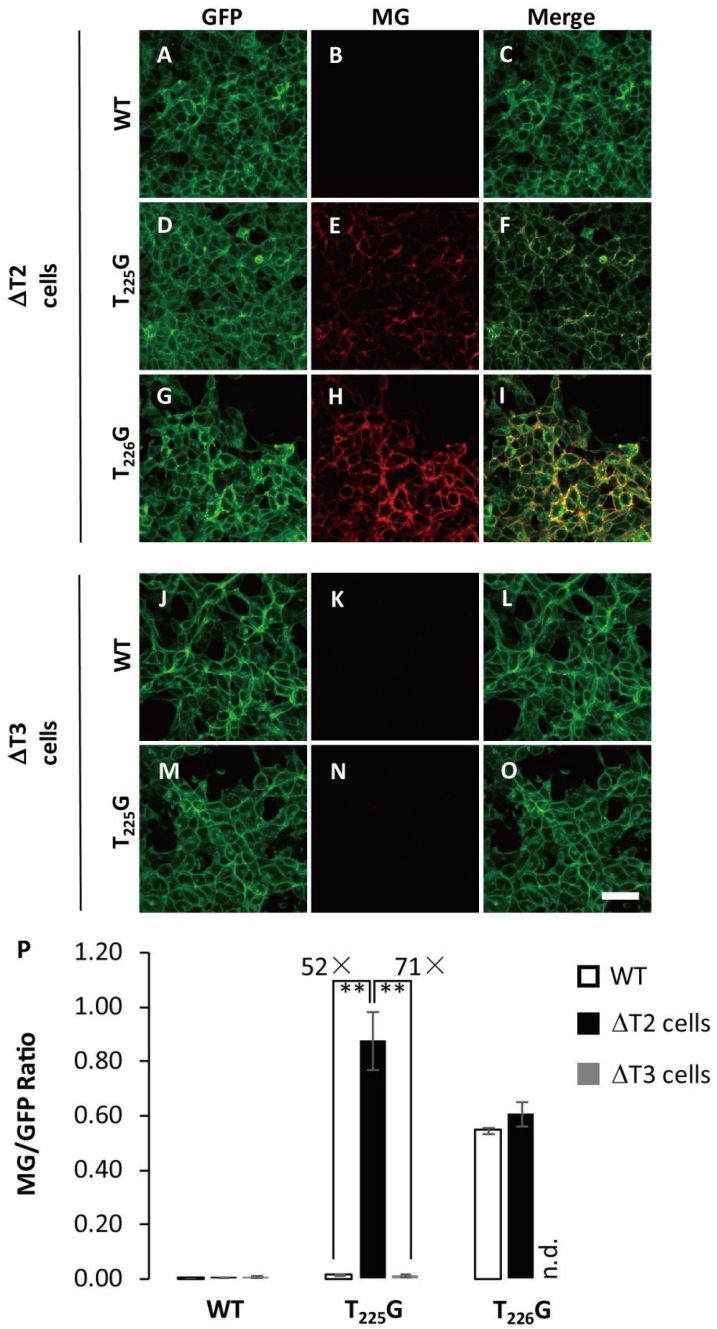

FIGURE 2.3. Activation of the ANGPTL3-based T2 sensor in ΔT2 and ΔT3 HEK cells. (A-O) Representative images are shown for HEK cell lines deleted for either GalNAc-T2 (ΔT2, A-I) or T3 (ΔT3, J-O) and expressing wildtype (WT, A-C, J-L), T225G (D-F, M-O), or T226G (G-I) versions of the ANGPTL3-based T2 sensor. Images are GFP, malachite green (MG) and a merge of GFP (green) and MG (red). Image acquisition was identical within each channel with J-O acquired on a separate day. Bar=50μm. (P) The ratio of the average MG fluorescence per image field to the average GFP fluorescence for the same fields is plotted for the indicated constructs with fold differences shown (** = $p<0.001$, n.d. = not determined, n = 4 ± SEM). Control HEK cell data is re-plotted from Figure 2.2.

T3 sensor based on FGF23. A similar strategy was adopted to develop a sensor specific to the activity of GalNAc-T3. The fibroblast growth factor 23 (FGF23) sequence $_{168}$HFNTPIPRRHTRSAEDDG$_{185}$ is a GalNAc-T3 specific target and its glycosylation on T_{178} blocks furin cleavage at R_{179}[35]. This sequence was used as the linker in the sensor and mutation of the glycan acceptor site was used to test sensor functioning. In this case, there was an additional site (T_{171}) to consider, which is a known target of glycosylation[35].

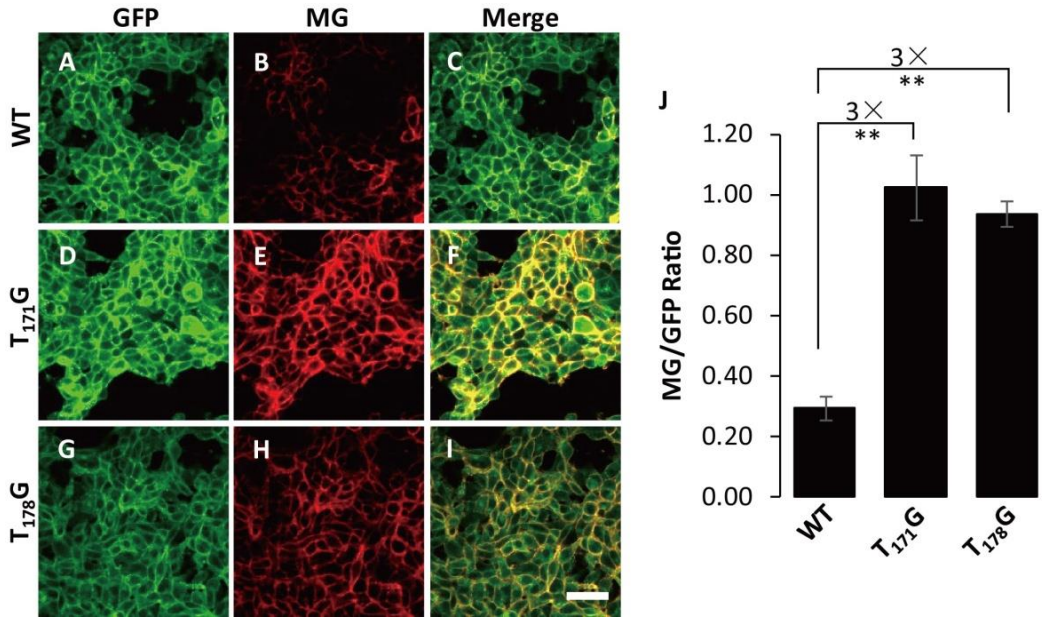

FIGURE 2.4. Evaluation of the T3 sensor based on FGF23. A–I, GFP, MG, and merged channels are shown for HEK cell lines expressing the WT (A–C), T_{171}G (D–F), and T_{178}G (G–I) versions of the FGF23-based T3 sensor. Bar=50μm. J, the MG/GFP ratio is shown for each (** = $p<0.001$, n = 4 ± S.E.).

For the FGF23-based sensor of GalNAc-T3 activity, HEK cells were generated stably expressing wildtype (WT), T_{171}G, or T_{178}G versions of the sensor. As before, GFP fluorescence confirmed sensor expression. Whereas relatively low levels of MG fluorescence were observed for WT (Figure 2.4A-C), activation was evident for both T_{171}G (Figure 2.4D-F) and T_{178}G (Figure 2.4G-I). While the effect of the T_{171}G mutation may seem surprising, in other work we observed that GalNAc-T3 requires glycosylation at T_{171} to glycosylate T_{178} in a process dependent on the GalNAc-T3 lectin domain (unpublished and ([31])), similar to the situation for several other GalNAc-T isoforms and their substrates[70,71]. More relevant to the present work, the fold increase for these constructs (Figure 2.4J) was markedly less than that above and yielded little more than 8 SD over background. Interestingly, this was not due to low activation.

The activated constructs yielded MG/GFP fluorescence ratios that were comparable to those of the ANGPTL3- based sensor (1.0 for FGF23-based $T_{171}G$ vs. 1.1 for ANGPTL3-based $T_{225}G/T_{226}G$). Rather, it was due to a higher level of background activation evident in the WT sensor. This is consistent with previous work showing that furin cleavage of FGF23 is highly efficient relative to its cleavage of ANGPTL3[72] but it could also be that the FGF23-based sensor was less efficiently glycosylated.

Before attempting to improve the FGF23- based sensor we tested whether its activation depended on GalNAc-T3. Indeed, its expression in ΔT3 cells led to greater activation of MG fluorescence (Figure 2.5D-F). The increase was small but significant (Figure 2.5J). The sensor was also tested in the ΔT2 cells. For an unknown reason, we were only able to recover cells stably expressing low levels, but, as expected, absence of GalNAc-T2 did not activate the sensor (Figure 2.5G-I). Thus, while the sensor needed improvement, it showed promise in being isoform specific.

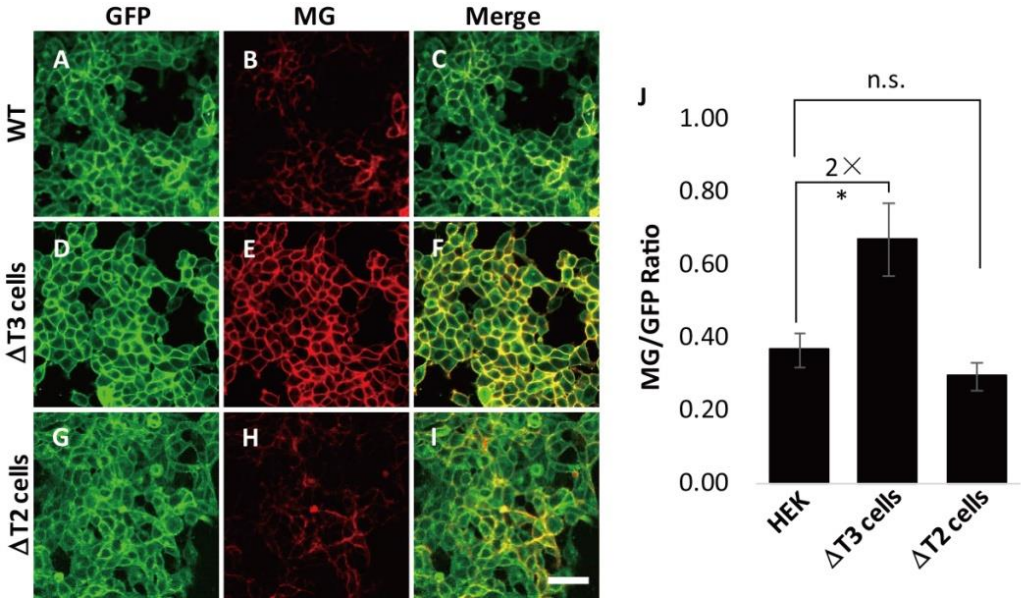

FIGURE 2.5. Specificity test of FGF23-based T3 sensor. A–I, HEK cell lines deleted for either GalNAc-T3 (D–F) or T2 (G–I) and expressing the wild type FGF23-based T3 sensor are shown. Bar=50μm. J, the MG/GFP ratio is shown for the wild type T3 sensor expressed in control HEK cells or HEK cells deleted for GalNAc-T2 or T3 (* = $p<0.05$, n.s. = not significant, n = 4 ± S.E.).

Our intention was to maximize signal to noise where feasible and ideally achieve a maximum signal at least 20 SD above background so that, in the context of a screen, a signal that is 10% of the maximum will be statistically meaningful (i.e. ≥2 SD above background).

To improve the signal to noise ratio we decided to modify the FGF23 insert sequence with the goal of making it a better substrate of GalNAc-T3 glycosylation. Fortunately, extensive *in vitro* assays were previously used to assess glycosylation of various peptide sequences[69] and recent work identified a phosphorylation site (S_{180}) that negatively regulates glycosylation[73]. Based on these results, we tested the single mutation $H_{177}V$ and the triple mutation $H_{177}V/S_{180}G/A_{181}P$. Compared to the level of "background" activation of WT (Figure 2.6A-C), both $H_{177}V$ (Figure 2.6G-I) and $H_{177}V/S_{180}G/A_{181}P$ (Figure 2.6M-O) substantially reduced the amount of activation. To test whether the constructs could still be activated by the absence of their glycosylation, we generated corresponding versions lacking the GalNAc-T3 acceptor site. Similar to mutation of the acceptor site in the WT to generate $T_{178}G$ (Figure 2.6D-F), MG fluorescence was activated in $H_{177}V/T_{178}G$ (Figure 2.6J-L). For an unknown reason, $H_{177}V/T_{178}G/S_{180}G/A_{181}P$ proved less useful as it did not yield comparable levels of MG fluorescence (Figure 2.6P-R). Quantification confirmed the significant improvement using the $H_{177}V$ version of the sensor (Figure 2.6S). Further because of the reduced noise, activation was >1000 SD over background.

Based on the improvement evident for the $H_{177}V$ version, we tested its isoform specificity using ΔT2 and ΔT3 cells. For comparison, the WT and $H_{177}V/S_{180}G/A_{181}P$ constructs were also tested. Both WT (Figure 2.7A-C) and $H_{177}V$ (Figure 2.7D-F) were active in ΔT3 cells indicating that these sensors depend on glycosylation by T3. Neither was activated in ΔT2 cells showing the specificity of this effect (Figure 2.7J-O). The $H_{177}V/S_{180}G/A_{181}P$ failed to show any activation (Figure 2.7G-I) and quantification confirmed that the $H_{177}V$ modified sensor was significantly improved in both signal to noise (≈900 SD) and isoform specificity (Figure 2.7P). These results support the utility of $H_{177}V$ modified FG23-based sensor for specific determination of GalNAc-T3 activity.

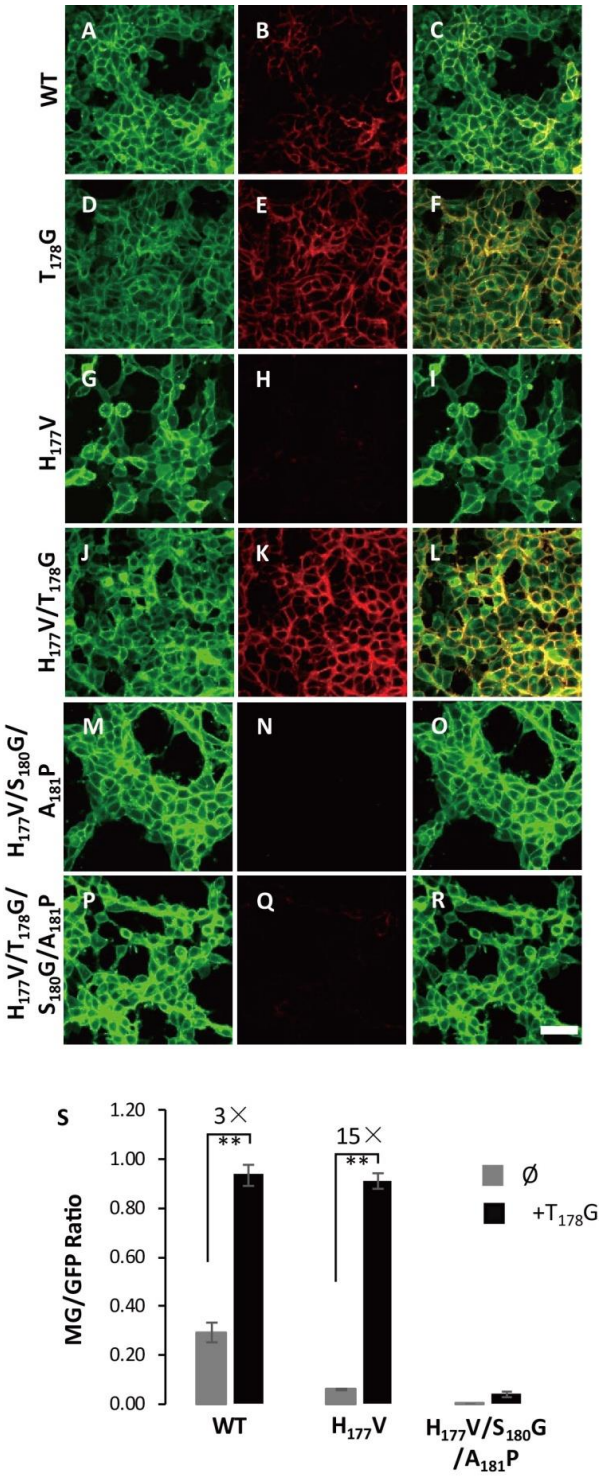

FIGURE 2.6. Optimization of the FGF23-based T3 sensor. (A-R) GFP, MG and merged channels are shown for HEK cell lines expressing WT (A-C), $T_{178}G$ (D-F), $H_{177}V$ (G-I), $H_{177}V/T_{178}G$ (J-L), $H_{177}V/S_{180}G/A_{181}P$ (M-O), and $H_{177}V/T_{178}G/S_{180}G/A_{181}P$ (P-R) versions of the FGF23-based T3 sensor. Note that those containing the $T_{178}G$ substitution (D-F, J-L, P-R) lack the glycan acceptor site. Bar=50μm. (S) Quantified MG/GFP ratio for the indicated constructs without (Ø) or with ($+T_{178}G$) the glycine substitution at the glycan acceptor site. (** = $p<0.001$, n = 4 ± SEM).

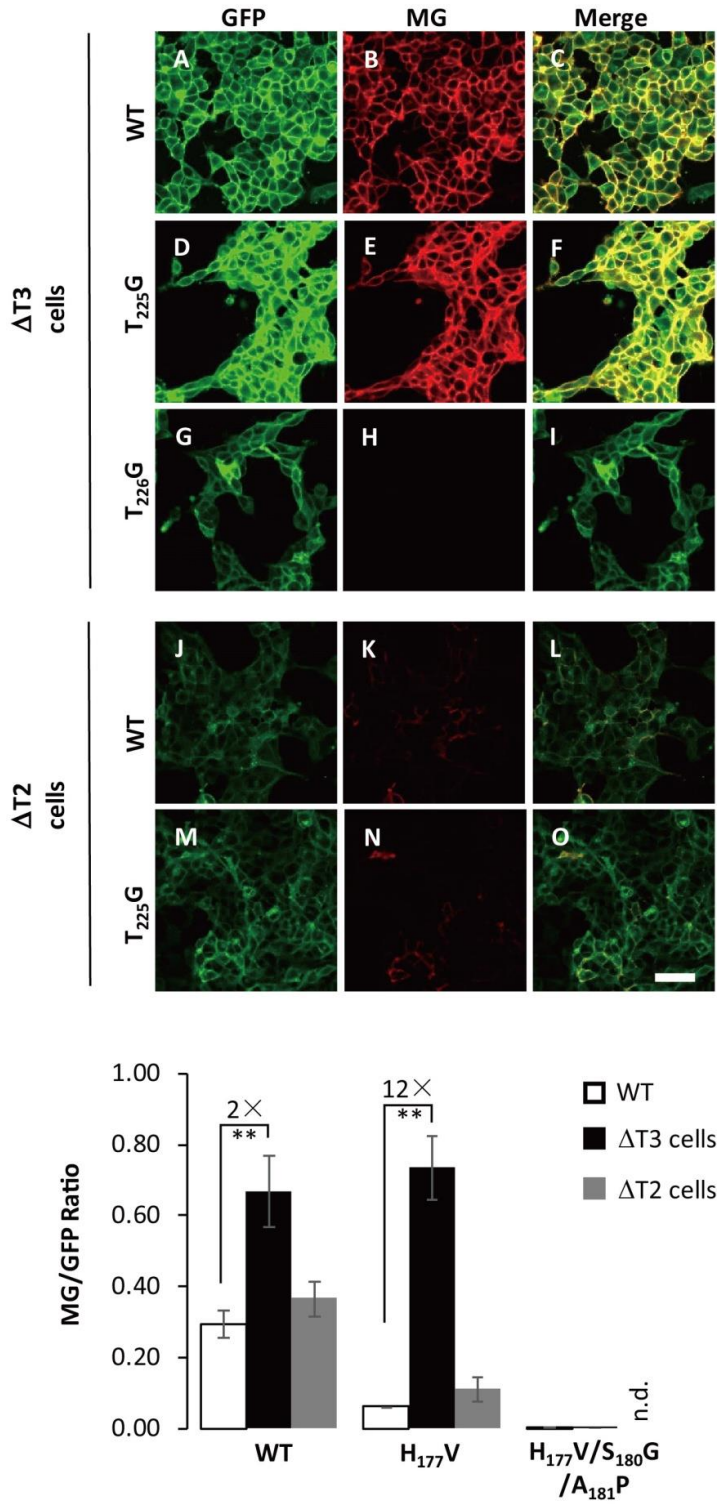

FIGURE 2.7. Activation of the FGF23-based T3 sensor in ΔT3 and ΔT2 HEK cells. (A-O) Representative images are shown for HEK cell lines deleted for either GalNAc-T3 (ΔT3, A-I) or T2 (ΔT2, J-O) and expressing wildtype (WT, A-C, J-L), $H_{177}V$ (D-F, M-O), or $H_{177}V/S_{180}G/A_{181}P$ (G-I) versions of the FGF23-based T3 sensor. Bar=50μm. (P) The MG/GFP ratio is shown for the indicated constructs expressed in control HEK cells or HEK cells deleted for GalNAc-T3 or T2 (** = $p<0.001$, n.d. = not determined, n = 4 ± SEM).

T3 sensor based on artificial sequence. The previous sections demonstrate how natural target sites for O-glycosylation and proprotein convertase processing can be tuned to improve signal to noise levels and isoform specificity of the sensors. Therefore, we wanted to test a linker sequence that was entirely designed *de novo*. The design was based on previous *in vitro* analysis of the substrate specificity of GalNAc-T3[69]. The sequence, $_{1}$RRAYRVTPGP$_{10}$, contained a single potential glycan acceptor, T_7, placed two residues C-terminal of an R_5 furin cleavage site in an attempt to maximize masking and thus decrease background sensor fluorescence. The sequence surrounding T_7 was chosen to optimize it as a substrate of GalNAc-T3. In HEK cells stably expressing the artificial sequence-based T3 sensor, GFP fluorescence indicated good expression and MG fluorescence indicated a reasonably low level of background activation (Figure 2.8A-C). To test its ability to function, we generated cells expressing the same sensor harboring a T_7G substitution. This version showed dramatically increased MG fluorescence (Figure 2.8D-F). Overall, the level of activation was 9 fold and 30 SD over background (Figure 2.8G).

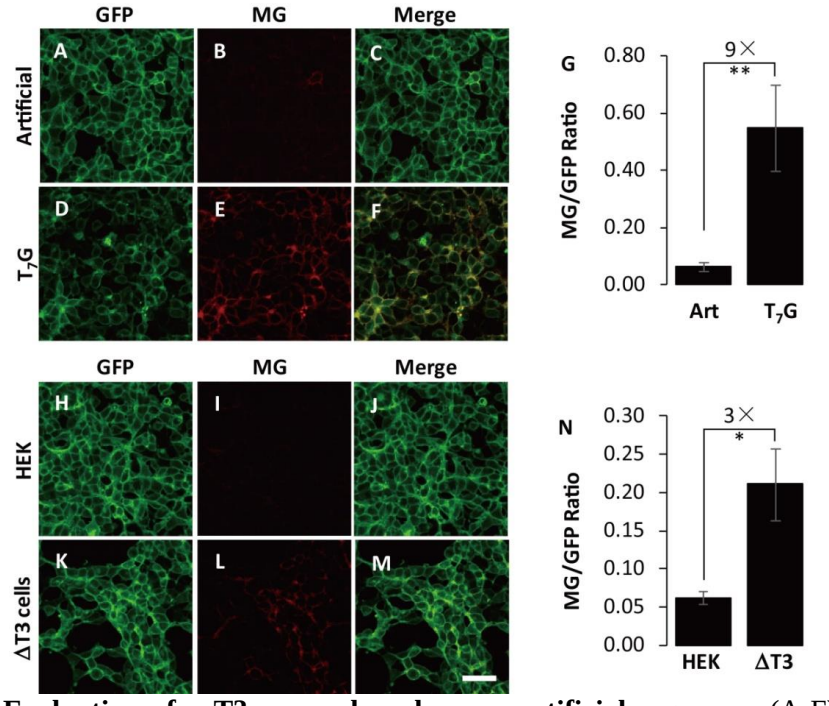

FIGURE 2.8. Evaluation of a T3 sensor based on an artificial sequence. (A-F) GFP, MG and merged channels are shown for HEK cell lines expressing either an unmodified version of the T3 sensor containing an artificial sequence (A-C) or a version with a glycine substitution of the glycan acceptor site (T_7G, D-F). Bar=50μm. (G) The MG/GFP ratio is shown for each (** = $p<0.001$, n = 4 ± SEM). (H-M) Images for the artificial sequence-based T3 sensor expressed in control (H-J) or ΔT3 (K-M) HEK cells. Bar=50μm. (N) The MG/GFP ratio is shown for the artificial sequence-based T3 sensor expressed in control HEK cells or HEK cells deleted for T3 (*= $p<0.05$, n = 4 ± SEM).

To test whether the artificial sequence-based sensor activity was controlled by GalNAc-T3, the sensor was next expressed in the ΔT3 cell line. Significantly, compared to control HEK cells (Figure 2.8H-J), the sensor was activated in ΔT3 cells (Figure 2.8K-M) and this was confirmed by quantification (Figure 2.8N). Although future modification of this sequence may lead to further optimization, these findings indicate that we achieved *de novo* design of a T3 sensor, which, unlike the FGF23-based version, functioned in the absence of an upstream glycosylation site.

Sensor tests in 96-well format. As a test of the utility of the sensors in high throughput screening for glycosylation inhibitors we adopted our assays to flow cytometry in a 96-well format. Cells expressing either the ANGPTL3-based T2 sensor or the FG23-based T3 sensor were released from 96-well dishes and collected in the presence of the MG dye. The GFP and MG fluorescence per cell for 10,000 cells per well was then determined. To assess whether sensor activation could be detected, the fluorescence distribution resulting from the wildtype sensors was compared to that of their corresponding glycine-substituted versions. These experiments were performed prior to our optimization of the FGF23-based T3 sensor, which served the secondary purpose of determining whether significant results were achievable with a low level of fold-activation. Indeed, both the ANGPTL3-based T2 sensor (Figure 2.9A) and the FGF23-based T3 sensor (Figure 2.9B) showed clear activation using this methodology. Quantification indicated an activation of over 600-fold (7816 SD) for the T2 sensor and 2-fold (38 SD) for the T3 sensor (Figure 2.9C) essentially matching the values determined for these versions using microscopy (Figure. 2.2M & Figure 2.4J). Indeed, both the fluorescence histograms and the calculated averages confirm that the difference in the fold-activation between the two sensors is primarily in the high background of the T3 sensor supporting the need for optimization of this sensor to improve its glycan masking.

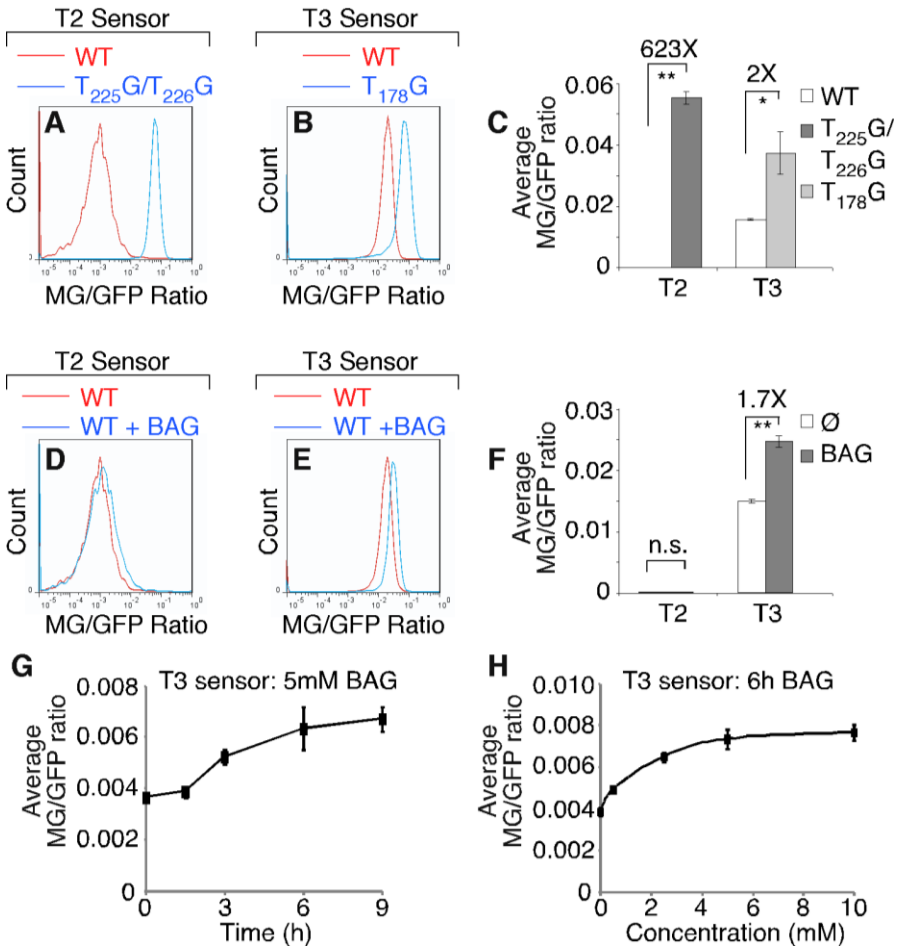

FIGURE 2.9. Sensor tests in 96-well format using inhibitor of glycan extension. (A-B) Flow cytometry-based histogram distribution of MG/GFP ratios is shown for HEK cell lines expressing the indicated versions of either the ANGPTL3-based T2 sensor (A) or the FGF23-based T3 sensor (B). Note that lines in red correspond to wildtype (WT) and lines in blue are for sensors lacking glycan acceptor sites. (C) Averages of the MG/GFP ratios are compared for each sensor version (*= $p<0.05$, ** = $p<0.001$, n = 3 ± SEM). (D-E) Histogram distribution of MG/GFP ratios is shown for the wildtype T2 (D) and T3 (E) sensors either without or with incubation (6 h) in 5 mM benzyl-N-acetyl-α-galactosaminide (BAG) to inhibit glycan extension. (F) Averages of the MG/GFP ratios are compared for each sensor with or without BAG treatment (** = $p<0.001$, n.s. = not significant, n = 3 ± SEM). (G-H) Averages of the MG/GFP ratios are compared for cells expressing the T3 sensor and treated with 5 mM BAG for the times indicated or treated with BAG for 6 h at the concentrations indicated (n = 3 ± SEM).

Next, to mimic the addition of an unknown drug that might score in a high throughput screen, we repeated the flow cytometry assay on cells expressing either of the wildtype sensors after their treatment with benzyl-N-acetyl-α- galactosaminide (BAG), which is an inhibitor of chain extension. If glycan masking in either sensor required a fully extended glycan chain, we expected BAG to activate the sensor. Interestingly, this was the case for only one of the sensors. Whereas the ANGPTL3-based T2 sensor was slightly, if at all, affected by BAG (Figure 2.9D),

the FGF23-based T3 sensor showed activation (Figure 2.9E). Quantification of the effect indicated a statistically significant activation of 1.7-fold and 17 SD over untreated (Figure 2.9F). This compares well with its 2-fold maximum activation suggesting that this sensor strongly depends on chain extension for its masking. On one hand, the differential effect is surprising in that the glycan site flanks the furin site in ANGPTL3 but resides directly in the FGF23 site (Table 2.1). Also, unlike ANGPTL3, our minimal first-generation sensor was activated by BAG[65] even though the relative positions of its glycan and furin sites are the same as ANGPTL3. On the other hand, the differential effect observed here is in agreement with previous work showing inhibition of processing by a single GalNAc in ANGPTL3[74] but not FGF23[35]. As confirmation of the effect, we determined its time course (Figure 2.9G) and concentration dependence (Figure 2.9H). Inhibition was detected within 3 h and with as little as 0.5 mM BAG. In summary, the experiments in this section provide a "proof of principle" for the utility of the sensors in high-throughput screening and also indicate that the importance of glycan chain length for glycan masking can vary depending the target sequence and enzyme involved.

2.4 Discussion

Despite the biological and medical importance of the GalNAc-Ts that initiate mucin type O-glycosylation in the Golgi, effective inhibitors are lacking. Central to the study of the GalNAc-Ts is their presence as a large family of isoforms that show distinct substrate specificities and expression patterns. Diseases arising from GalNAc-T misregulation typically involve single isoforms. Also, pan-specific or general inhibitors, while desirable as experimental tools, will block all mucin-type O-glycosylation, which is a critically important posttranslational modification for normal cell function. Thus, potential therapies will ideally target specific isoforms. With this consideration in mind, we sought to create an isoform-specific, cell-based, fluorescence biosensor to be used for highthroughput screening of potential inhibitors. Towards this goal, isoform-specific GalNAc-T target sequences for the GalNAc-T2 and T3 enzymes were used to replace a general consensus sequence in our previously characterized sensor, which works by recapitulating glycan masking[65]. The new sensors were activated in the absence of glycosylation and, after modification of the inserted sequences, yielded specific activation in response to ablation of their corresponding enzyme.

The T2 sensor used a target sequence in ANGPTL3 and revealed the potential of a glycine substitution flanking the glycan addition site to restrict isoform specificity of the sensor. The T3 sensor was based on a target sequence in FGF23 and its signal to noise ratio was dramatically improved by substitution of a valine next to its glycan acceptor site. Thus, we have developed GalNAc-T2 and T3 specific sensors that will be valuable in both the study of GalNAc-T regulation and in high-throughput screening for potential therapeutic regulators of specific GalNAc-Ts.

Optimizing the sensors focuses on the signal to noise ratio and their isoform specificity. In relation to the insert sequences of the linker domain, the signal is largely a function of the efficiency of glycosylation, masking, and cleavage. GalNAc-T isoform specificity relates to differential substrate recognition. Ideally, one would modify the linker sequence to achieve efficient and isoform-specific GalNAc-T glycosylation with the site-specific glycosylation placed in or adjacent to the protease recognition site for maximal effect, and a sequence design that contains efficient protease cleavage in the absence of glycosylation. The degree to which these goals are mutually compatible for a particular GalNAc-T isoform is critical to successful sensor development. Presumably, these same criteria impacted the evolution of glycan masking sites. In that vein, we were fortunate to have the ANGPTL3 and FGF23 glycan masking sites as starting points for the GalNAc-T2- and T3-specific sensors, particularly as they were already well characterized. We envision extending this to other GalNAc-T isoforms and proprotein convertases using known substrate specificities and iterative modifications of the sensors.

Although we did not directly quantify the extent of glycosylation, masking and cleavage for these sensors, we can conclude that all three aspects were efficient for the ANGPTL3-based T2 sensor given its >500-fold activation. Further improvement seems unnecessary. With one caveat this was also true for this sensor's isoform specificity. The caveat is that the ANGPTL3 sequence needed to be modified so that it would be specifically activated in HEK ΔT2 cells. Thus, for use in isoform-specific screening, the T225G substitution is imperative. Interestingly, the glycine substitution at this position detectably activated the sensor indicated by an increase in the GFP/MG fluorescence ratio from 0.002 to 0.02. This implies that it was now less efficiently glycosylated and that further optimization might be possible, perhaps using alternative substitutions. Nevertheless, even with its higher level of background activation, the

modified sensor was activated greater than 50-fold in ΔT2 cells arguing that further optimization is not really needed.

In contrast, the wildtype FGF23-based T3 sensor was marginal, yielding only 3- and 2-fold activation in control and ΔT3 cells, respectively. This was due to background activation stemming from any of a number of possibilities including inefficient O-glycosylation, highly efficient furin cleavage or regulation by phosphorylation. Generally, peptide preferences of the GalNAc-Ts are sensitive to residues from -3 to +3 relative to the glycosylation site with a preference for proline at positions +1 and +3[69]. Further, as mentioned above, the residue at +2 (S180) impedes glycosylation when phosphorylated[73]. Consistent with these considerations, substitution of glycine at +2 and proline at +3 caused a significant drop in background activation of the T3 sensor (Figure 2.6S). However, this version of the sensor activated poorly suggesting the changes caused a problem with furin recognition. We had better luck with valine substitution of a histidine at position -1, which also potently reduced background but did not interfere with activation (Figure 2.6S). The presumed increase in glycosylation is consistent with the *in vitro* preference for a hydrophobic residue at this position[69]. We were also fortunate that the FGF23-based T3 sensors showed strong isoform specificity and recapitulated the lectin-domain-mediated mechanism by which T3 glycosylates T178 in FGF23 (unpublished) thereby raising the possibility of screening for drugs targeting either the catalytic unit or the lectin domain of T3.

In terms of applying the sensors to screening, it is important to mention that just as with sensors of N-glycosylation in the ER that are based on protein folding[75,76], anything contributing to successful glycosylation of the sensor is a potential drug target in their activation. This includes the extending enzymes, nucleotide sugar transporters, nucleotide sugar biosynthesis, trafficking components (critical for proper delivery of the sensor and localization of both the GalNAc-Ts and furin) and even factors controlling Golgi structure as Golgi structure can influence glycosylation efficiency[77]. Interestingly, glycosylation initiated by the addition of N-acetylglucosamine (GlcNAc) is strongly influenced by the metabolic flux of its precursor UDP-GlcNAc[13] but whether availability of UDP-GalNAc features as prominently in regulating mucin-type O-glycosylation is unknown.

In any case, relatively simple controls can be used to identify these types of "off target" hits. One relies on the fact that the sensors are ratiometric and traffic in the secretory pathway. By simply verifying a strong GFP signal at the cell surface, any compound that affects synthesis and trafficking of the sensor can be excluded. More importantly, one can now take advantage of the availability of both T2- and T3-specific sensors. Any compound that activates both sensors is either inhibiting both GalNAc-Ts directly (pan-specific) or it is acting indirectly, perhaps by affecting one of the processes just mentioned. Although such hits could be quite significant, highest priority should be given to compounds that specifically block a single sensor because they are likely to be acting directly on the corresponding GalNAc-T isoform. As we expand the sensor development to other isoforms this aspect will be further enhanced.

An exciting future application of the sensors is in determining conditions that alter a particular isoform's glycosylation activity. This might be a consequence of a change in signaling state in a differentiated cell type or a change occurring during differentiation. Once a change is identified, the responsible signaling pathway could be dissected using the sensor as an assay. Again, the availability of multiple isoform specific sensors will allow general (and possibly indirect) effects to be separated from those that are likely direct. The "turn-on" aspect of the sensors we describe here is most suited to signaling that inhibits an active enzyme. One simple way to generate a behavior capable of detecting both up and down regulation of a particular enzyme is to weaken the glycosylation site with respect to activity but not isoform specificity. We already saw this with the unmodified FGF23 insert. The "background" level of activation is significantly higher than an inactive sensor thus conditions that enhance T3 activity will detectably lower the signal (as was the case when we modified the sensor). Thus, once again the extensive *in vitro* work dissecting the roles of particular residues surrounding the glycan acceptor site in terms of both enzyme binding and specificity of binding will be enormously useful for de-tuning the sensors.

As a criterion for isoform specificity of the sensors, we used specific activation upon deletion of the corresponding isozyme in HEK cells. As mentioned above, all but 6 isoforms have been detected in HEK, so it remains possible that the missing isozymes could also target these sensors, possibly diminishing their utility outside HEK. Nevertheless, the T2- and T3-specific sensors were activated in ΔT2 and ΔT3 cells, respectively, even though these cells likely continued to express isozymes that are close paralogs with highly similar properties[26].

Further mitigating this concern is our ability to engineer cell systems with and without a given GalNAc-T expressed, as well as, with a repertoire of GalNAc-Ts[68,72]. This should enable sensor activity to be isoform-specific even if one or more isoforms may glycosylate the same substrate.

In summary, this report describes cell-based, fluorescent sensors for two isoforms of GalNAc-T. These sensors are promising for highthroughput screening for novel inhibitors. They can also be used to monitor isoform specific O-glycan masking in living cells under a variety of conditions. Finally, the approach described here can readily be extended to the development of sensors for many of the remaining GalNAc-T isoforms. It continues to be our hope that this technology will contribute to therapy development for disorders arising from GalNAc-T misregulation.

2.5 Experimental Procedures

Constructs. To improve recovery and expression of stable cell lines, the sequence encoding the previously described sensor[65] was excised using SpeI and NotI-HF and cloned into NheI and NotI sites of the pIRES vector. Primers were then designed encoding the target isoformspecific substrate sequences of ANGPTL3, FGF23 and an artificial sequence (see Table 1) and these were either cloned into the linker region by a loopin procedure using QuikChange (Stratagene, La Jolla, CA) or by direct insertion into XhoI and BamHI sites. Subsequent point mutation of the base linker sequences was by QuikChange. All constructs were confirmed by sequencing.

Cell culture. HEK cells, including HEK293 controls and HEK cells deleted for GalNAc-T2 or T3 (unpublished) produced by zinc finger nuclease mediated knockout similarly to previously described[68,78], were cultured in MEM medium (Thermo Scientific, Waltham, MA) with 10% fetal bovine serum (Atlanta Biologicals, Flowery Branch, GA) and 100 IU/ml penicillin–streptomycin (Sigma-Aldrich, St. Louis, MO). The cells were transfected with the constructs using the JetPEI transfection reagent (VWR International, Radnor, PA) exactly as described in the manufacture instructions. Selection with 2 μM puromycin (Sigma) was initiated 24-48 h post transfection. After about 2 weeks of culturing, the cells were collected using EDTA and sorted using a FacsVantage fluorescence cell sorter (BD Biosciences, San

Jose, CA). Cells (5,000 per construct) expressing GFP at 80-90% of the maximum observed were collected and propagated for use.

Imaging. The cells were passed onto coverslips at 50% confluence 24 h before imaging. After transfer to an imaging chamber containing 110 nM of the dye MG11p (MG, a generous gift from the Molecular Biosensor and Imaging Center, Carnegie Mellon University) they were imaged using an LSM 510 Meta DuoScan Spectral Confocal Microscope equipped with a 40X objective (Zeiss, Thornwood, NY). Single optical sections were acquired and quantified using ImageJ (National Institutes of Health, Bethesda, MD) as described previously[65]. Briefly, background was taken as the highest non-cell pixel value and was uniformly subtracted. The GFP channel was then used as a mask to select an area of measurement. The average pixel intensity in this area was then determined for the GFP and MG channels and expressed as a ratio. For every condition, 8 fields were analyzed per experiment and each experiment was repeated 4 times. For each channel (GFP or MG), acquisition parameters were identical for every experiment. Where necessary for presentation purposes, grayscale projection of the MG channel was adjusted equally for all images in an experiment.

Flow Cytometry. Control cells or cells stably expressing the sensors were passed into 96-well plates (50,000 cells/well) and returned to the incubator for 24 h. Where indicated, a further incubation for 0-9 h was carried out in media containing 0-10 mM benzyl-N-acetyl-α-galactosaminide (BAG, Sigma) added from a stock dissolved in DMSO. All samples contained equivalent amounts of DMSO (1%). Just prior to analysis the media was replaced with 0.1 ml of PBS solution containing 110 nM MG dye and 5 mM EDTA. After 5 min at 37°C the released cells were collected and GFP and MG fluorescence were measured by Accuri™ C6 flow cytometry (BD Biosciences) at 488 nm and 640 nm, respectively. Data presentation and quantification were carried out using FlowJo software (www.flowjo.com). For histograms, the background subtracted MG and GFP fluorescence measurements for each cell were expressed as a ratio and binned. Background was subtracted separately in each channel using the geometric mean of all values determined for the MG and GFP channels using untransfected HEK cells. To compare average values among multiple experiments, the individual ratios calculated as just indicated were collected and expressed as geometric means and these means were then compared using standard arithmetic mean calculations and statistics. Three independent experiments were performed for each.

Chapter 3: Activity Detection of GalNAc Transferases by Protein-Based Fluorescence Sensors *in vivo*

3.1 Abstract

Mucin-type O-glycosylation occurring in the Golgi apparatus is an important protein post-translational modification initiated by up to twenty GalNAc-transferase isozymes with largely distinct substrate specificities. Regulation of this enzyme family affects a vast array of proteins transiting the secretory pathway and misregulation causes human diseases. Here we describe the use of protein-based fluorescence sensors that traffic in the secretory pathway to monitor GalNAc-transferase activity in living cells. The sensors can either be "pan" or isozyme specific.

3.2 Introduction

Glycosylation enzymes change the surface characteristics of their protein substrates by appending or modifying carbohydrate chains termed glycans. Glycans impact protein solubility, stability, and interactions[1,7,8]. Mucin-type O-glycosylation is a large and important subgroup defined by the initial reaction in which N-acetylgalactosamine (GalNAc) is transferred to the hydroxyl group of serine or threonine (and possibly tyrosine) by UDP-N-acetyl-α-D-galactosamine polypeptide N-acetylgalactosaminyltransferases (GalNAc transferases). This occurs on secretory cargo as it passes through the Golgi apparatus and other Golgi-localized enzymes extend the glycan by further additions of individual monosaccharides. While one or two isoforms exist for each of the enzymes mediating chain extension, there are up to 20 distinct genes encoding GalNAc-transferase isozymes in humans, each with at least partial substrate selectivity[26]. Collectively, the isozymes modify a vast number of substrates and they are already linked to a significant number of medical syndromes[1,32,35].

Despite their biological and medical significance, assays have been lacking that monitor GalNAc-transferase activity in living cells and there are no known inhibitors. Therefore, we developed cell-based fluorescence sensors that are particularly sensitive to inhibition of GalNAc-transferase activity[65,66]. They are transfectable constructs encoding proteins that

traffic to the cell surface and become fluorescent if their glycosylation is inhibited. In brief, each sensor has a fluorescence activating protein domain (FAP) followed by a linker to a blocking domain that occludes dimerization necessary for binding of the dye malachite green to the FAP (Figure 3.1). The dye itself is non-fluorescent and only becomes fluorescent when bound to the FAP. Within the linker is a glycan acceptor site next to a furin protease site. Because GalNAc-transferases are localized to the cis/medial Golgi and furin is localized to the TGN, glycan addition occurs before, and therefore sterically blocks, processing by furin leaving the sensor inactive. Extending enzymes also act before furin and contribute to glycan-masking (*see* **Note 1**). However, if the GalNAc-transferase activity is inhibited, the entire glycan is absent and furin cleaves the linker releasing the blocking domain and allowing dye activation. The sensors are ratiometric because they also contain a fluorescent protein domain.

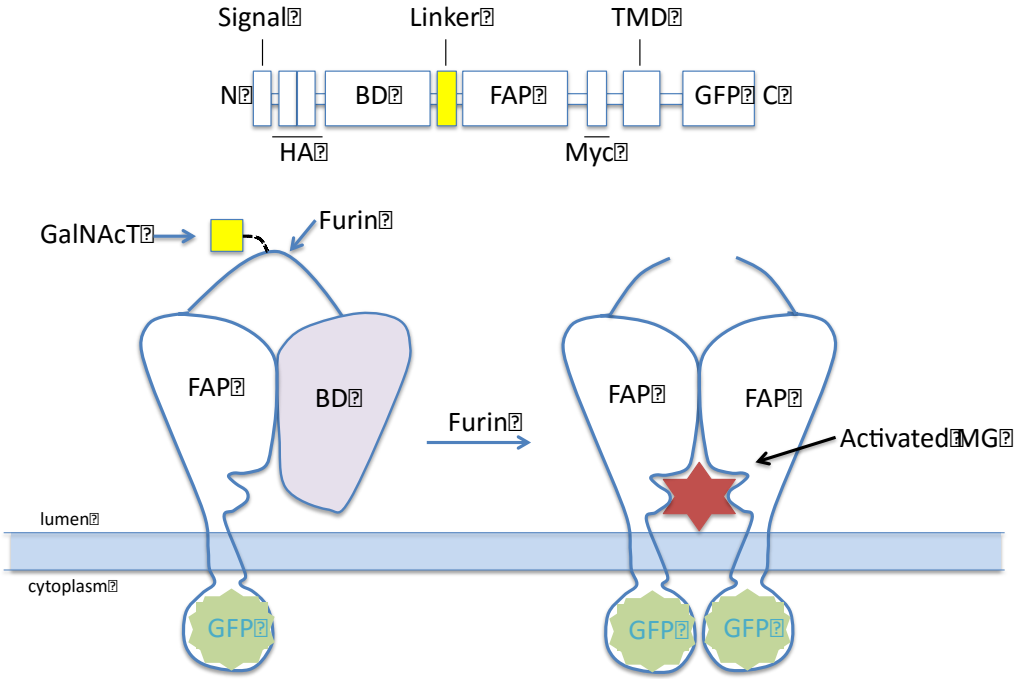

FIGURE 3.1. Sensor schematic. The linear diagram (not to scale) depicts the N-terminal cleaved signal sequence, tandem HA epitopes, blocking domain (BD), linker, fluorescence-activating protein (FAP), myc epitope, transmembrane domain (TMD), and fluorescent protein domain (GFP). The cartoon shows an intact sensor on the left with adjacent GalNAc-transferase and furin sites in the linker such that GalNAc addition masks the furin site. Reduced GalNAc-T activity allows furin cleavage releasing the BD. As shown on the right, the FAP then dimerizes and binds and activates the dye malachite green (MG). The signal is read as the MG/GFP ratio.

There are two critical aspects regarding sensor function. First, the sequence of the glycan acceptor site can either be "pan" or isozyme specific. The pan-specific version contains a minimal sequence recognized by GalNAc-transferases. Although not confirmed for every isozyme, it should be recognized as a substrate by most if not all[69,79]. For the isozyme-specific versions, the sequences are tailor-made for selective recognition by individual isozymes. At present only T2- and T3-specific versions are available. Their acceptor sequences were derived from known T2 or T3 substrates and then modified to improve selectivity[34,35,66,67,69]. The second critical aspect is that the starting or ground state signal can be significantly greater than true background (where true background is determined by mutating the furin site). This is advantageous if the desire is to assay enzyme activation. A less than optimal glycosylation site results in incomplete glycosylation and therefore an increased starting signal. Conditions up-regulating enzyme activity will then lower the signal. Thus, although the sensors were designed to assay for inhibitors, they can be further modified and used to monitor changes in GalNAc-transferase activity in either direction.

3.3 Materials

1. HEK293 growth medium: Minimal Essential Medium containing 10% fetal bovine serum and 100 IU/ml penicillin–streptomycin. Store at 4°C. Use in 5% humidified CO_2 incubator at 37°C.
2. MG11P (Sharp Edge Labs, Pittsburgh, PA) (*see* **Note 2)**: Reconstitute at 110 µM (1000x stock) in 95% methanol, 5% acetic acid. Store at 4°C in the dark.
3. MEM without phenol red: Adjust to 110 nM MG11P just before use.
4. EDTA/PBS: Reconstitute 8 g NaCl, 0.2 g KCl, 0.916 g Na_2HPO_4 and 0.2 g KH_2PO_4 in 1 L water at pH 7.3. Add 2.081g EDTA (5 mM final concentration) and adjust pH, if necessary. Adjust to 110 nM MG11P just before use.

3.3.1 Confocal Assay Components

1. Live cell imaging chamber (or glass bottom dishes) and inverted confocal microscope equipped with a 40X oil immersion objective and 488 nm and 633 nm filter sets (*see* **Note 3)**.

2. Image analysis software (e.g., ImageJ, http://imagej.nih.gov/ij/).

3.3.2 Flow Cytometry Assay Components

1. 96-well flat-bottom plastic dishes.
2. EDTA/PBS containing 110 nM MG11P.
3. Flow cytometer equipped to read at 488 nm and 640 nm (*see* **Note 3)** and accompanying software.

3.4 Methods

Below are two protocols for use of the GalNAc transferase biosensors. The first uses live-cell fluorescence microscopy and the second employs flow-cytometry. Flow cytometry is preferred because a large number of cells are readily quantified for each condition and multi-well dishes can be used allowing a large number of conditions to be tested (e.g., high throughput screening). The sensors can also be analyzed using immunoblotting if biochemical verification is desired (*see* **Note 4**). As a starting point, each protocol requires at least 3 cell lines: experimental, negative control and positive control. The experimental cell line(s) stably express the GalNAc-T biosensor(s) of interest. The linker sequence in these will contain either the pan-specific glycan acceptor site or one of the isozyme specific acceptor sites. The signal produced can be compared for differing conditions and normalized to the controls for comparison with other sensor signals. For the negative control, either untransfected cells or, optimally, cells stably expressing a matched version of the sensor with a mutated furin acceptor site (ΔFur) is used. This cell line will establish the true background, i.e. the signal in the absence of any cleavage, which theoretically corresponds to complete glycosylation of all sensor molecules. The positive control is a cell line expressing a matched sensor with a mutated glycan acceptor site (ΔGly). This cell line will establish the maximal possible signal given that it cannot be glycosylated. The protocols are described for use of the pre-existing HEK293 cell lines expressing pan-, T2- and T3-specific sensors and their matched controls. However, the linker sequence can be modified to produce new specificities (*see* **Note 5**). Also, the assay should work in most, if not all, cell lines but each cell type may require minor modifications in cell handling. After processing in the Golgi the sensor accumulates on the cell surface. It takes a

minimum of 3 h to develop a strong surface signal so, after a test condition, a period of at least this duration should be allowed before the cells are analyzed *(10)*.

3.4.1 Assay GalNAc-T activity using biosensor and confocal microscopy

1. Pass the cell lines (ΔFur, biosensor, ΔGly) to achieve 50% confluence on coverslips for imaging chamber or in glass bottom dishes and incubate 24 h (see Note 6).
2. Transfer cells to pre-warmed MEM without phenol red and containing 110 nM MG11P dye.
3. Mount on confocal microscope equipped with a 40X objective (e.g. LSM 510 Meta DuoScan Spectral Confocal Microscope). Use identical confocal settings for all data collection within the experiment (*see* **Note 7**). Focus using the GFP (488 nm) channel to yield sharp cell surface outlines of the near confluent cells (Figure 3.2). There should be about 100 cells per field. Acquire a single optical section using both the GFP and MG (633 nm) channels. Repeat for at least eight separate fields per coverslip.
4. To quantify, for each two-channel image define several regions of interest outside the cells. For each channel in each image, determine the average intensity value in these non-cell areas and use the highest value as the background. For each channel in each image, subtract the background value from all pixels (uniform subtraction). Next, in the GFP channel of each image, use auto threshold to define a threshold and select all above-threshold pixels. Record the average intensity value for these pixels for both the GFP and MG channels. The signal for that image is then expressed as the MG/GFP ratio. Repeat for all eight two-channel images per coverslip. The average of the eight ratios is then the value for a single condition in a single experiment.
5. Comparison of these values between ΔFur, the biosensor, and ΔGly will typically yield a low signal for the biosensor, an even lower signal for ΔFur and a much greater value for ΔGly. To normalize the values for comparison with other sensors subtract the background (ΔFur or untransfected) and then divide by the positive control (ΔGly).

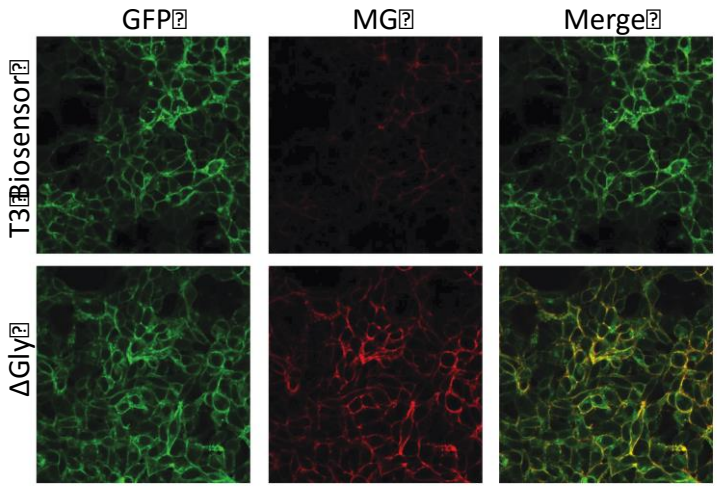

FIGURE 3.2. Fluorescence microscopy data. The figure compares cells expressing the T3 biosensor (upper row) to those expressing its activated ΔGly version (lower row) in the indicated channels.

3.4.2 Assay GalNAc-T activity using biosensor and multi-well format flow-cytometry

1. Pass the cell lines (ΔFur, biosensor, ΔGly) at 50,000 cells/well into 96-well plates (0.2 ml/well).
2. After 24 h replace the medium with 0.1 ml of the EDTA/PBS solution containing 110 nM MG dye and return to incubator for 5 min (*see* **Note 6**).
3. Suspend cells by pipetting each well up and down 2-3 times with a multi-channel pipette with 200 µl tips. Mount plate on flow cytometry instrument (e.g. BD Accuri™ C6 Flow Cytometer). Collect data for 10,000 cells per well using GFP (488 nm) and MG (640 nm) channels. (Figure 3.3)
4. To quantify, first determine the background. Calculate the geometric mean of values for each channel of each well of the untransfected or ΔFur cells (at least 3 wells). Use the average of these means (one for each channel) as the background values. Next, for each remaining well, determine the geometric mean of the values for each channel. Subtract the channel-specific background values and then calculate the MG/GFP ratio for each well. Averages of these values for multiple trials are then used for comparison.
5. Again, there should be a low signal for the biosensor, an even lower signal for ΔFur (equal to background) and a much greater value for ΔGly. To normalize the values for

comparison with other sensors subtract the background (untransfected or ΔFur) and then divide by the positive control (ΔGly).

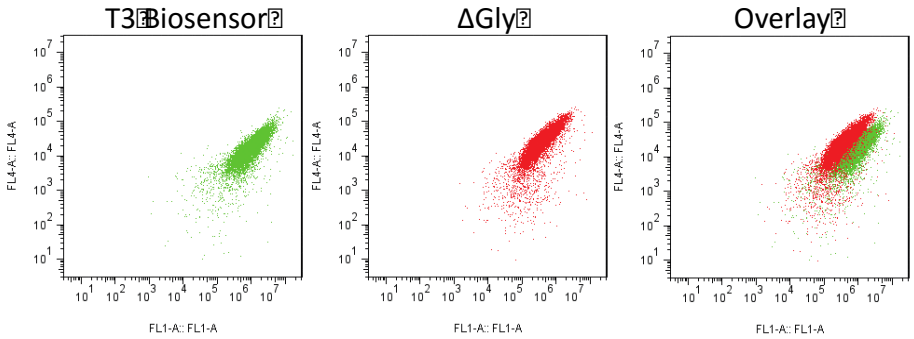

FIGURE 3.3. Flow cytometry data. Shown are scatter plots for the T3 biosensor (left), its activated ΔGly version (middle), and an overlay. Axes are 488 nm channel (X) and 633 nm channel (Y).

3.5 Notes

1. Glycan masking of adjacent protease-processing sites can occur upon addition of a single GalNAc *(13,15)* but the current sensors show variable requirements for chain extension. The pan- and T3-specific sensors appear to require extension whereas the T2-specific sensor does not *(9,10)*.
2. The protocols utilize a membrane impermeant version of the dye and therefore yield surface staining. A membrane permeant version can be used but also yields predominate surface staining because processing of the biosensor by furin occurs just before its exit from the Golgi. A Golgi signal can be detected if this exit is blocked (e.g. by cell incubation at 20°C).
3. Excitation and emission spectra for MG11P can be found at http://www.mbic.cmu.edu/images/datasheet/MG-11p-NH2-info_rev21.pdf.
4. Cleavage of the sensor is readily assayed by immunoblot *(9)*. Briefly, cell lysate proteins are separated by SDS-PAGE, transferred to nitrocellulose and blotted with anti-GFP antibodies to detect uncleaved (≈80 kD) and cleaved (≈60 kD) portions of the sensor. (The missing fragment is secreted into the cell media and detectable with anti-HA antibodies.) Quantify the cleaved fragment as a percent of total. The biosensor should give a value slightly higher than ΔFur and much lower than ΔGly.

5. **Optional linker modification**. The sensor plasmids *(9,10)* each contain in-frame fusions encoding a signal sequence, a blocking domain (HL4 heavy chain, (17)), a linker with furin and glycosylation sites, a FAP (L5*, *(11)*), the myc tag, a transmembrane domain from PDGFR, and either a Venus or GFP domain. The version with the pan-specific glycan acceptor site has a linker with the sequence 5'-c tcg aga aag aag aga tct acc ccc gct cca gct cca tcc ggt ggc ggt ggc agc-3' encoding NSRKKRSTPAPS where RKKR is the furin site and TPAP comprises the glycan site. To modify selectivity of the sensor, point mutations can be introduced or the entire linker can be substituted with another sequence encoding a glycan acceptor and protease-processing site. The optimal candidates are substrate sequence stretches known to be regulated by glycan masking by a particular GalNAc transferase. In either case, results from published in vitro assays establishing sequence preferences of particular GalNAc transferases *(13,15)* can be used to guide further mutagenesis to optimize sensor function *(10)*. Transient expression can be used for initial tests but ultimately stable cell lines will need to be generated to insure uniform and strong surface expression of the sensors. For any new version, comparison of its signal with and without an intact glycan acceptor site will test its effectiveness. Strong sensors yield maximal activation of 500-fold or more *(10)*. To confirm isoform-specificity of a new sensor, one needs to show that it is strongly activated in cells lacking the particular GalNAc transferase isoform *(10)*. Genomic editing or RNA interference can be used to suppress expression of the particular isoform.

6. Time after passage can be varied to achieve the desired density. Note that trypsin treatment to passage the cells degrades the surface population of sensor molecules. A minimum of 3 h and a recommended time of at least 6 h post-passage should be used to repopulate the cell surface. Cell release with EDTA/PBS (for passage or analysis) avoids this issue.

7. The intensity of the lasers should be adjusted to ensure maximum sensitivity without saturation. This will depend on expression level but matched control and experimental sensors should always be analyzed with identical settings.

Chapter 4：Inhibitor of GalNAc-T3-mediated O-glycosylation identified by cell-based screening

4.1 Abstract

Drug-like modulators of site-specific O-glycosylation by the GalNAc-T family hold promise as therapeutics but are currently unavailable. Employing simultaneous screening of compound libraries with cell-based sensors specific to individual GalNAc-T isoforms we identified a GalNAc-T3 sensor-specific hit that was subsequently found to directly bind and selectively inhibit purified GalNAc-T3. Further, the inhibitor opposed upregulated GalNAc-T3 in driving metastatic-like behavior of cancer cells and it blocked GalNAc-T3-mediated glycan-masking of FGF23 thereby reducing secretion of intact FGF23, a possible treatment of chronic kidney disease. These findings establish a pharmacological approach for the GalNAc-T family and suggest that targeting specific GalNAc-Ts will yield new therapeutics.

4.2 Introduction

Glycosylation, which fine-tunes the function of proteins, is the most abundant and diverse posttranslational modification[1,7,8]. Despite many documented roles in health and disease, the enzymes that mediate mucin-type O-glycosylation in the Golgi apparatus have yet to be discovered as druggable targets. The initiating enzymes, a family of 20 GalNAc-T isozymes, determine which substrates are modified and at which sites. Significant questions remain regarding their specificity, regulation, targets and functions and the lack of a pharmacological approach has been a critical limitation. The only confirmed inhibitor of mucin-type O-glycosylation, benzyl-N-acetyl-α-galactosaminide, blocks elongation rather than initiation and requires millimolar concentrations that can be toxic[80,81]. Not only are pan-specific modulators of the GalNAc Ts lacking, there is nothing isoform-specific. The latter will be essential to restrict effects to particular pathways and could possibly lead to a new basis for therapeutics conceptually related to the widespread use of drugs targeting individual protein kinases.

GalNAc-T3 serves as an important test case as it is implicated in at least two medically important pathways: stabilization of FGF23 and cancer metastasis[33,35,44,52,57,82–84]. GalNAc-T3

mediates glycan-masking of FGF23 in bone as part of a control mechanism determining the form of FGF23 that is secreted[35,73]. When present, the added O-glycan blocks FGF23 cleavage by the furin protease resulting in secretion of intact FGF23 that activates FGF23 receptor complexes at the kidney and intestine[85]. In contrast, non-glycosylated FGF23 is cleaved and the cleaved C-terminal product competitively blocks these same receptors[86]. Significantly, elevated intact FGF23 occurs in chronic kidney disease and upon kidney transplant where it is directly linked to poor prognosis due to its effects on renal phosphate reabsorption and 1,25-dihydroxyvitamin D biosynthesis[87,88]. Many studies have concluded that therapeutic control over FGF23 would be transformative in the clinic[89–91]. GalNAc-T3 is also overexpressed in cancerous tissue often correlating with shorter survival[43,44,46,83]. Further, knockdown of GalNAc-T3 expression inhibits invasive capacities arguing that GalNAc-T3 has potential as a therapeutic target[45].

As a step toward testing whether biologic discoveries and new therapeutics will result from developing drug-like modulators that control the activity of specific GalNAc-Ts we initiated high throughput cell-based screening of compound libraries. HEK cell lines were engineered to express fluorescent sensors that are specific to GalNAc-T2 or GalNAc-T3 activity[66]. For each sensor, glycosylation of its isozyme-specific target site prevents furin protease from removing a blocking domain (Figure 2.1). Thus, fluorescence increases upon GalNAc-T inhibition because removal of the blocking domain allows dimerization of a fluorogen activating protein domain so that it binds and activates the fluorescence of malachite green. They are ratiometric because the sensor backbone contains a green fluorescent protein as an internal control for expression. Each sensor showed clear activation after mutation of its glycan acceptor sites and these mutated constructs served as positive controls in the screen. The T3 sensor exhibited a background level of activation due to incomplete glycosylation[66] but this was considered advantageous for the possible identification of enzyme activators along with the desired inhibitors. As previously reported[66], sensor expression in HEK cell lines depleted of either GalNAc-T2 or T3 via zinc finger nuclease editing resulted in specific activation of the corresponding sensor confirming their isozyme selectivity.

4.3 Results

The primary screening against the two sensors. Our pilot screen used approved oncology drugs and compounds based on structural diversity (≈2000 compounds in total) with 6 h treatments at 10μM prior to flow cytometry to assay MG and GFP fluorescence on a cell-by-cell basis (Figure 4.1A). Each compound was tested in duplicate and against both sensors. Because each sensor requires essentially identical cellular reactions- the only difference being which GalNAc-T isoform modifies the sensor- most off-target hits (such as sugar nucleotide transporters, extending enzymes, or furin) will alter both sensors, whereas directly acting, isoform-specific candidates will be sensor-specific. Expressed as a fraction of the ΔGly positive controls, the fluorescence signals elicited by most compounds were considered background but one compound specifically and strongly activated the T3 sensor and became the focus of this phase of the study (Figure 4.1B). The compound is a quinoline of no known activity that we now refer to as GalNAc-T3 Inhibitor 1 or T3Inh-1 (Figure 4.1B, inset).

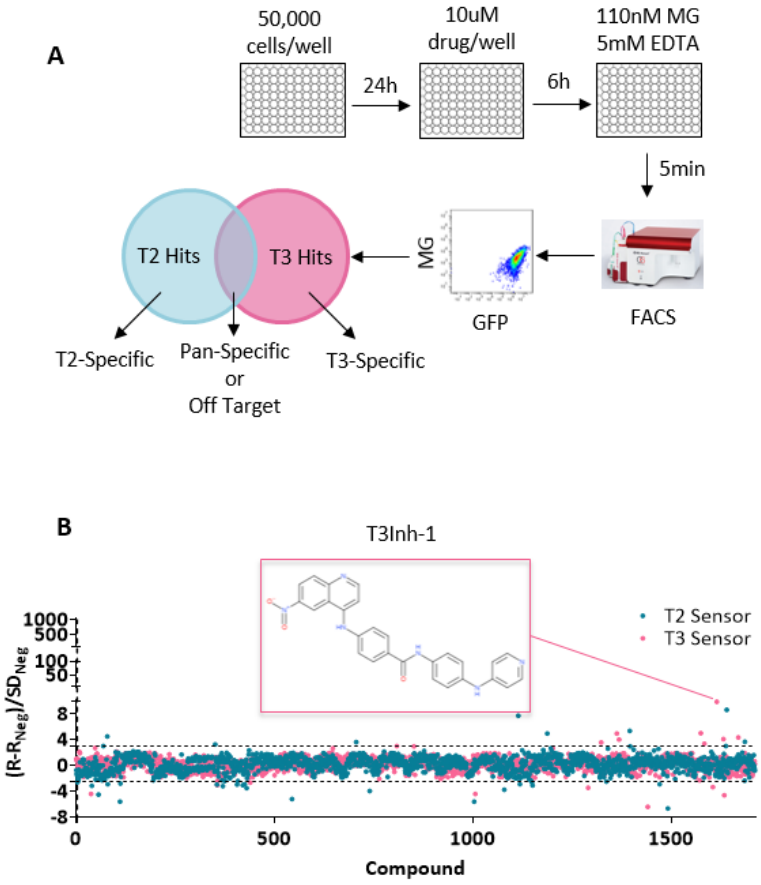

Figure 4.1. Screen for modulators of GalNAc-T2/T3. (A) Schematic showing cell plating, drug treatment, cell release, fluorescence measurement and parallel analysis using both T2 and T3 sensors.

Hits that activate both may be pan-specific or act on off-target pathways common to both sensors whereas sensor specific hits are likely acting directly on the corresponding GalNAc-T. (B) Q values ($Q=(R-R_{Neg})/SD_{Neg}$) were determined for each compound ($Q_{Compound}$) and the positive controls ($Q_{\Delta Gly}$) using the average of duplicate MG/GFP ratios for the compound (R), the vehicle-only control (R_{Neg}), and the standard deviation of the vehicle-only controls (SD_{Neg}). The plot shows the sensor signals for each compound as a fraction of the sensor positive controls and the structure of the indicated T3-specific hit is shown. Note that no data are covered by the inset.

The kinetics study of T3Inh-1. When retested at various doses against the T2 and T3 sensors, T3Inh-1 specifically activated the T3 sensor with an apparent IC_{50} of 12 μM (Figure 4.2A). Related compounds were not available and optimization of T3Inh-1 remains a future goal. Nevertheless, the strong isozyme specificity of the activity suggested that the compound acted directly on GalNAc-T3. As a test, we carried out *in vitro* glycosylation assays in which the purified lumenal domains of GalNAc-T2 or GalNAc-T3 (containing catalytic and lectin domains) were incubated with peptide and UDP-GalNAc substrates in the presence of increasing T3Inh-1 concentrations. A second stage reaction (UDP-Glo^{TM}), which showed no sensitivity to T3Inh-1, then converted the accumulated UDP product to ATP and then, via luciferase, to light. Corresponding to the sensor assay, inhibition in the *in vitro* assay was evident for GalNAc-T3 but not GalNAc-T2 and occurred with an IC_{50} of 7 μM (Figure 4.2B).

Towards characterizing the mechanism of inhibition we used the *in vitro* assay with purified GalNAc-T3 and individually varied both peptide and UDP-GalNAc substrate concentrations in the presence of 0, 7.5 or 15 μM T3Inh-1. The results were similar for both substrates (Figure 4.2C-D), where T3Inh-1 decreased the Vmax and increased the Km (Table 4.1) indicating a mixed-mode of inhibition in which the inhibitor most likely binds both free enzyme to reduce substrate binding and enzyme-substrate complexes to reduce turnover. Implied in this model of action is direct binding to the enzyme typically at an allosteric site. To test for direct binding, intrinsic tryptophan fluorescence of GalNAc-T3 was determined in the presence of increasing T3Inh-1 concentrations. At all concentrations, the drug itself yielded miniscule signals, whereas the drug had a profound and dose-dependent effect on the GalNAc-T3 emission spectrum (Figure 4.2E). These results confirmed direct binding with an apparent Kd of 17 μM (Figure 4.2F). The similarity in concentration dependence of sensor activation in cells (12μM), *in vitro* inhibition (7μM) and direct binding (17μM) argues that T3Inh-1 acts directly on cellular GalNAc-T3 and inhibits its activity.

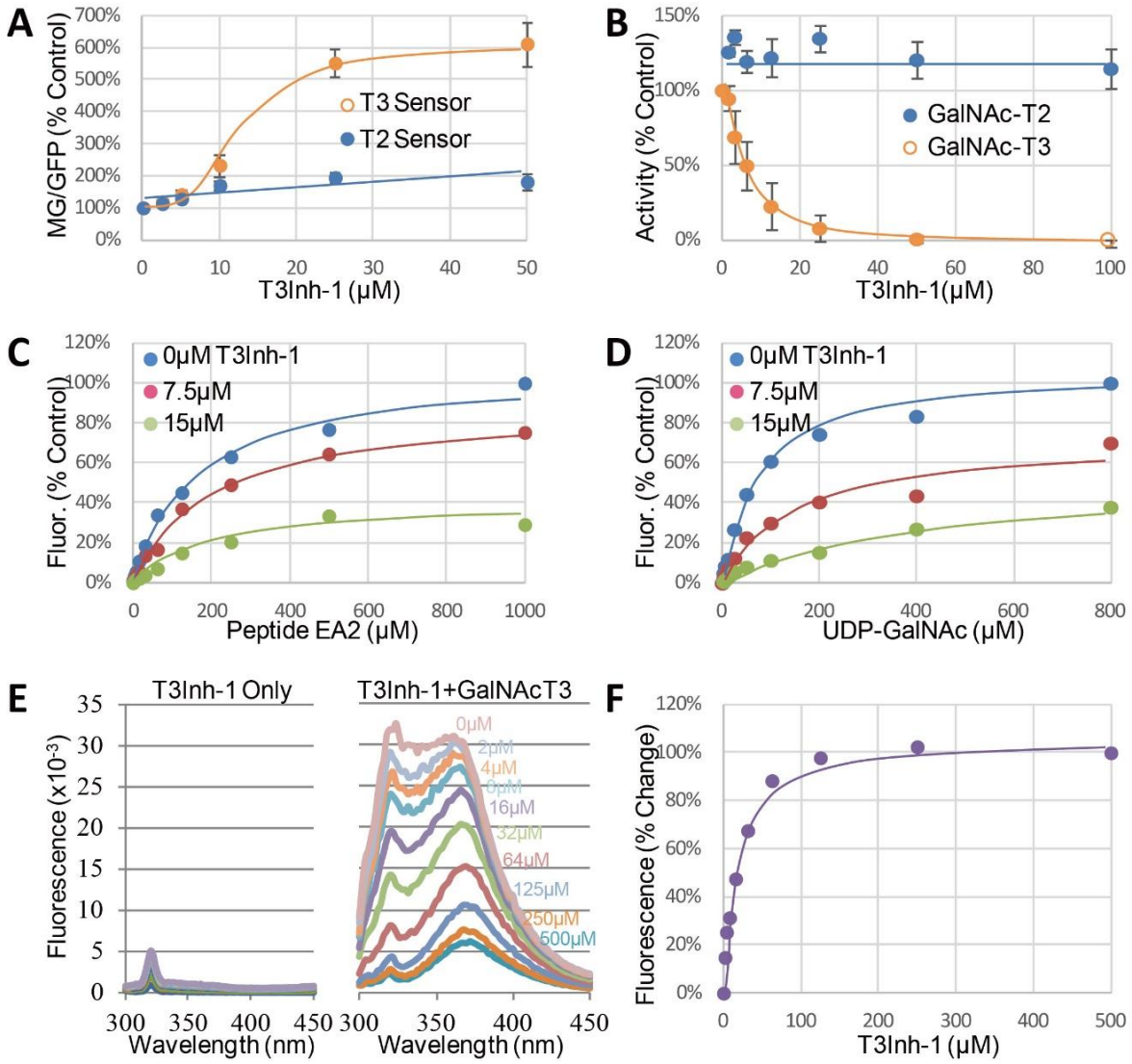

Figure 4.2. T3Inh-1 is a direct mixed-mode inhibitor of GalNAc-T3. (A) Comparison of T2 and T3 sensor activation at the indicated concentrations of T3Inh-1 (n=3±SEM). MG/GFP ratio was determined for 20,000 cells by FACS and average value is plotted as percent of the positive control (i.e. the ΔGly version of each sensor). (B) Comparison of effect of the indicated concentrations of T3Inh-1 on *in vitro* glycosylation mediated by purified GalNAc-T2 or GalNAc-T3. Values are averages expressed as percentage of the control "vehicle-only" reactions (n=3±SEM). (C-D) The *in vitro* assay was carried out in the presence of 0, 7.5, or 15 μM T3Inh-1 at the indicated concentrations of peptide or UDP-GalNAc substrate. Values are averages expressed as percent of the control reactions with no inhibitor and saturating substrates (n=3± SEM). (E) Representative fluorescence spectra are shown for T3Inh-1 alone or for purified GalNAc-T3 in the presence of the indicated concentrations of T3Inh-1. Note dose-dependent quenching of tryptophan fluorescence indicating direct binding. (F) Fluorescence quenching was quantified at each concentration using the peak value at 324 nm (n=3± SEM). Note that all graphs have error bars but some are too small to be apparent.

Table 4.2. Determination of inhibition type of T3Inh-1

Table 1A. Vmax and Km were determined in peptide (EA2) titration assay

146771	0 μM	7.5 μM	15 μM
Vmax	100%	82%	36%
Km (μM)	173.7	208.4	210.3
Ki (μM)		9.9	

Table 1B. Vmax and Km were determined in UDP-GalNAc titration assay

146771	0 μM	7.5 μM	15 μM
Vmax (luminescence unit)	100%	71%	56%
Km (μM)	74.9	153.4	448.3
Ki (μM)		2.9	

T3Inh-1 does not show off-target effects. Importantly, T3Inh-1 did not affect the staining patterns of the VVA, ConA or GSII lectins, which bind terminal GalNAc, branched alpha-mannose, or terminal GlcNAc, respectively (Figure 4.3). This implies that the major enzymes contributing to overall N- and O-glycosylation were unaffected and is consistent with GalNAc-T3 modifying a relatively limited number of substrates[92]. Also, there was no change in localization or expression level of GalNAc-T3, GalNAc-T2 or any other Golgi marker tested (Figure 4.4). Finally, T3Inh-1 had no effect on cell proliferation (Figure 4.5).

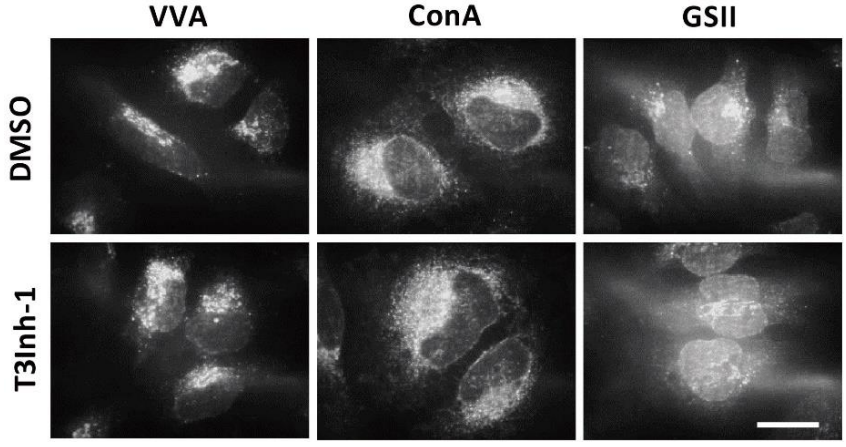

Figure 4.3. General N- and O-glycosylation are unaffected. Representative images of untreated or T3Inh-1 treated (6h, 10μM) HeLa cells after fixation and staining with the indicated lectin.

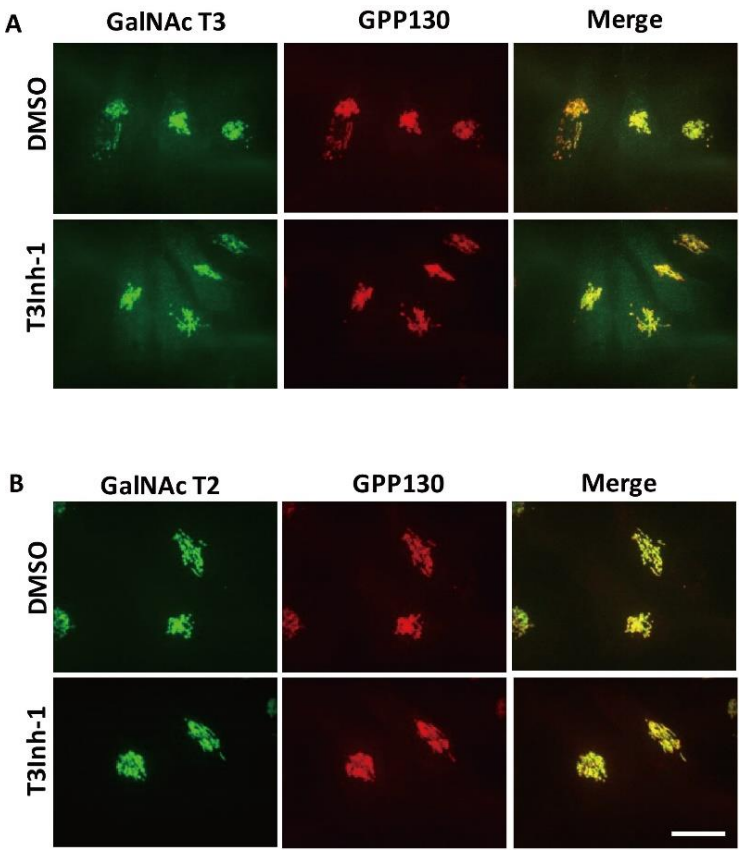

Figure 4.4. The localization and expression level of GalNAc-Ts are unaffected under T3Inh-1 treatment. Representative images of untreated or T3Inh-1 treated (6h, 10μM) HeLa cells after fixation and staining with GalNAc-T3 (A), GalNAc-T2 (B), and the Golgi marker GPP130. Bar=10μm.

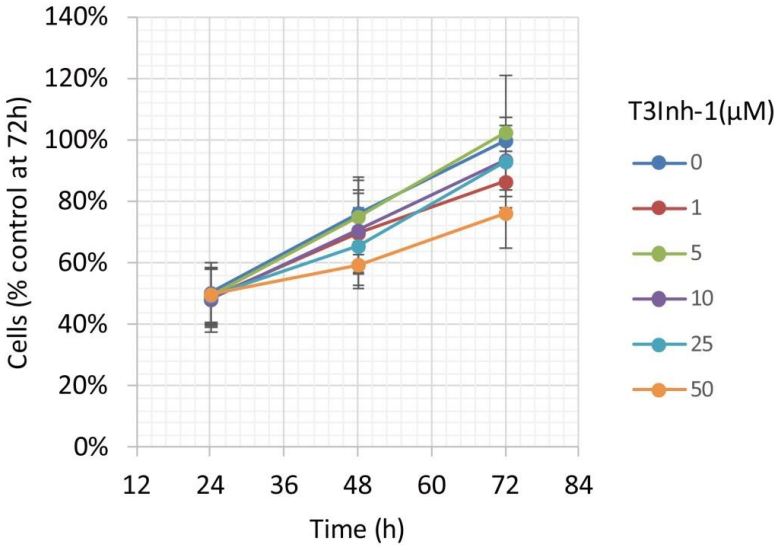

Figure 4.5. Cell growth at various T3Inh-1 exposures. Identical numbers of HeLa cells were plated and grown in the continuous presence of the indicated concentrations of T3Inh-1 and then at 24, 48, or

72 h they were released and counted. Averages are shown normalized using the untreated sample at 72 h (n=3±SEM).

T3Inh-1 blocks invasion of breast cancer cells. Given its validation as a direct inhibitor of GalNAc-T3 without obvious off-target effects we turned to biologically relevant tests of T3Inh-1. As mentioned, overexpression of GalNAc-T3 is linked to cancer cell invasiveness as well as poor outcomes in patients[43,44,46,83]. Indeed, our analysis of publically available data for 1117 breast cancer patients[93] using Kaplan-Meier survival plots shows that high expression of GalNAc-T3 correlates with poor patient overall survival (Figure 4.6A). Therefore, we carried out migration and invasion assays with the breast cancer cell line MDA-MB231, which expresses a relatively high level of GalNAc-T3 (Figure 4.6B) in the absence or presence of 5 μM T3Inh-1. Cells were cultured on uncoated (to assay migration) or Matrigel-coated (to assay invasiveness) Bioboat™ filters for 24 or 48h and those cells that moved to the underside of the filters were imaged and quantified. T3Inh-1 was strikingly effective, inhibiting migration by >80% (Figure 4.7A and B) and invasion by 98% (Figure 4.7C and D) while causing no discernable effect on cell proliferation (Figure 4.7E). To confirm that the effect was due to GalNAc-T3, the same experiment was carried out using MCF7 cells, which is a breast cancer cell line that expresses low levels of GalNAc-T3 (Figure 4.6B). Invasion by MCF7 cells was significantly increased by transfection with GalNAc-T3 and this increase was strongly blocked by T3Inh (Figure 4.8).

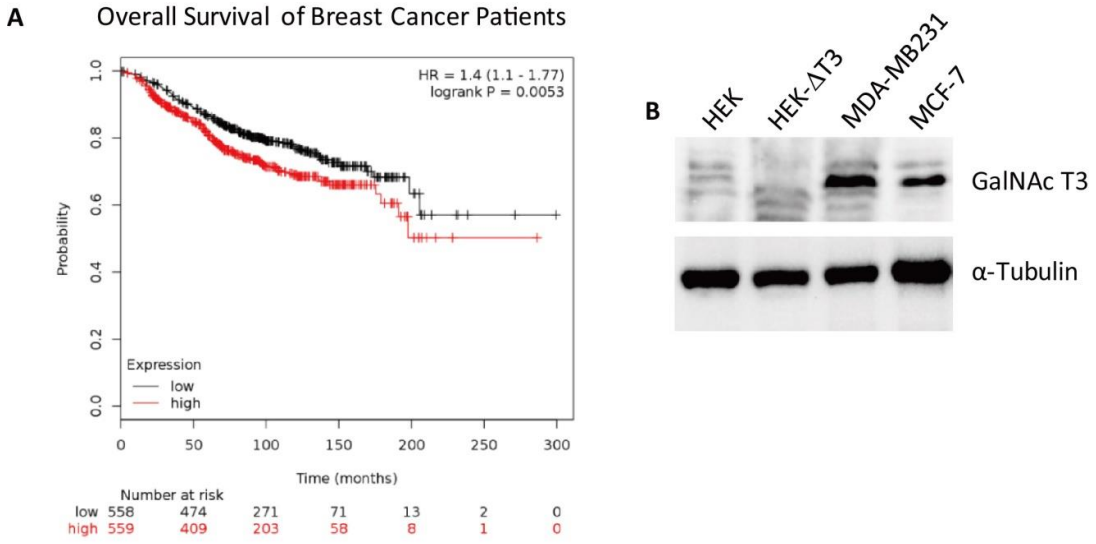

Figure 4.6. Breast cancer survival as a function of GalNAc-T3 expression and GalNAc-T3 expression in cultured breast cancer cell lines. (A) Kaplan-Meier curves compare overall survival in

patients with breast cancer between groups with high or low expression of GalNAc-T3. (B) Immunoblots of cell lysates from the indicated cell types [HEK, HEK-ΔT3 (edited to lack GalNAc-T3 expression), MDA-MB231, and MCF-7 using anti-GalNAc-T3 and anti-tubulin antibodies.

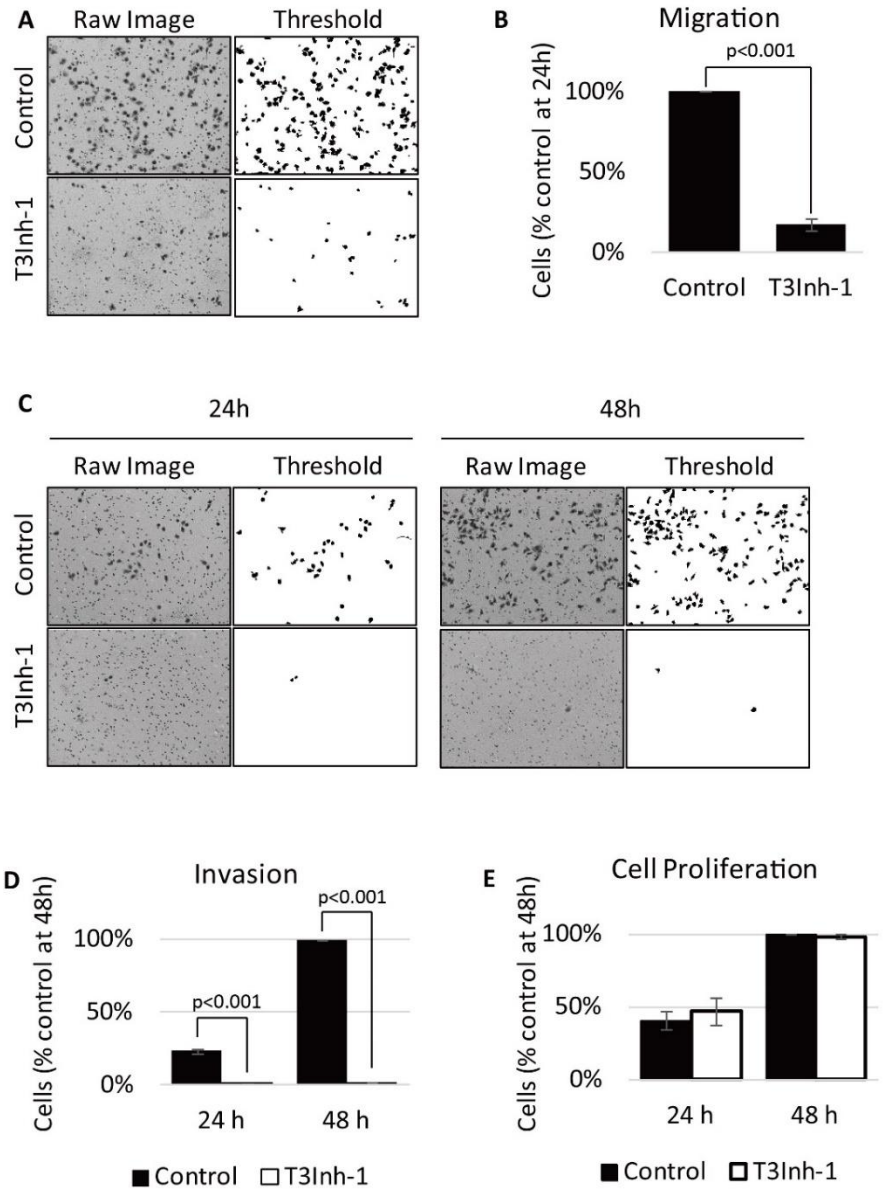

Figure 4.7. T3Inh-1 inhibits invasion and migration of MDA-MB231 breast cancer cells. (A) Cell migration through uncoated filters was determined for the MDA-MB231 breast cancer cell line grown in the absence or presence of 5μM T3Inh-1. The raw image of the filter shows both cells and the filter holes whereas a size-cut off was used in the thresholded image to specifically visualize the cells. (B) Cell migration results were quantified by counting cells that migrated to the underside of the filter and each experiment was normalized using the average determined for controls at 24 h (n=3±SEM). (C and D) Identical analysis except that the filters were pre-coated with Matrigel so that the assay measures invasion not just migration and the 48 h control was used for normalization (n=3±SEM). (E) MDA-MB231 proliferation was determined for cells grown in the presence or absence of 5 μM T3Inh-1 by

cell counting at 24 or 48 h. Quantification normalized by the value determined for untreated cells at 48 h (n=3±SEM).

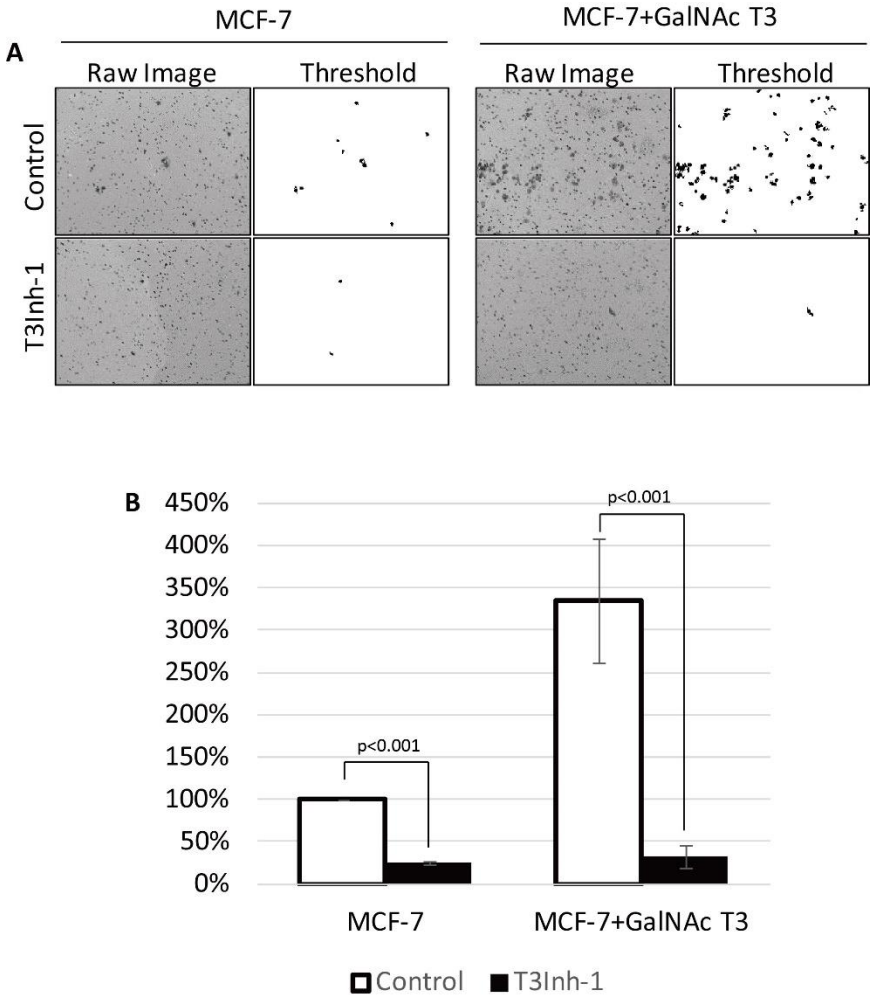

Figure 4.8. Overexpression of GalNAc-T3 increases invasion of MCF-7 breast cancer cells and T3Inh-1 can inhibits GalNAc-T3 induced invasion. (A) Mock and GalNAc-T3 transfected MCF7 cells were plated on Matrigel-coated filters in the absence or presence of 5μM T3Inh-1 for 24 h. Thresholded images show cells on underside of filters. (B) Cell counts are shown relative to untreated controls after normalization using the total number of cells (determined using parallel wells 24 h post-plating).

T3Inh-1 specifically enhances the cleavage of FGF23. As the relevant target(s) of GalNAc-T3 driving these metastatic-like cell behaviors have not yet been identified, we also tested if T3Inh-1 could inhibit glycan masking of FGF23, a known GalNAc-T3 target. If so, we expected reduced secretion of intact FGF23. HEK cells co-expressing transfected FGF23 and GalNAc-T3 were treated with increasing concentrations of T3Inh-1 and secreted FGF23 was assayed by immunoblot. There was a clear dose-dependent loss of intact FGF23, with a half-max of 14 μM for this effect (Figure 4.9A and B). FGF23 levels in the cell lysates were

unaffected arguing against any change in expression (not shown). For an unknown reason, perhaps related to its instability in culture, the cleaved fragment did not increase, but there was a clear increase in the ratio of cleaved/intact, which reflects the net signaling activity (Figure 4.9B). Importantly, we also tested the effect of T3Inh-1 on cleavage of ANGPTL3, which is controlled by GalNAc-T2-mediated glycan masking[67,68]. Secreted ANGPTL3 remained high at all concentrations, confirming the selectivity of T3Inh-1 towards GalNAc-T3 (Figure 4.9A-B). Reasoning that we might be able to see a similar effect on secreted FGF23 in an animal model, mice were injected intraperitoneally with T3Inh-1 and serum levels of cleaved FGF23 were determined. Three groups (0, 25 and 50 mg/kg T3Inh-1) of 8 animals per group received one or two injections separated by 24 h, followed by blood collection after another 24 h. An ELISA assay with antibodies against the N- and C-terminal portions of FGF23 was used to determine the amount of cleaved FGF23 in the blood[94,95]. Remarkably, T3Inh-1 caused a robust and statistically significant increase in the cleaved form of FGF23 at the tested 25 and 50 mg/kg concentrations (Figure 4.9C). While this is the hoped-for result for possibly mitigating the effects of elevated intact FGF23 in chronic kidney disease patients, further work with animal models can only be carried out once a sufficient supply of T3Inh-1 is available.

4.4 Discussion

This study identifies an isozyme-specific inhibitor targeting GalNAc-T3. The compound binds directly conferring a mixed-mode of inhibition and is equally active *in vitro* and in cells. Its discovery paves the way for structural studies that will contribute to our understanding of the enzyme reaction mechanism and undoubtedly reveal unexpected insights into how modulators such as T3Inh-1 can specifically alter the function a single member of a highly conserved family of enzymes. Structural studies are also anticipated to guide rational design of modified versions of T3Inh-1 to improve its binding affinity and efficacy.

The compound is already effective in both cellular and animal contexts providing the first chance for acute inhibition of GalNAc-T3 to study its role in diverse cellular processes. Further, forthcoming tests of disease models, possibly employing higher affinity versions, may strengthen the case for therapeutic uses of T3Inh-1. GalNAc-T3 and the other GalNAc Ts each have multiple (and sometimes overlapping) substrates. Thus, the full extent of unwanted side

effects may be difficult to predict and targeted drug delivery may be a necessary component to an effective therapy. However, in the case of GalNAc-T3, known effects of its knockout in the mouse model are all attributed to FGF23 processing[96]. Thus, use of T3Inh-1 to reduce intact FGF23 (and increase the inhibitory cleaved fragment) to treat chronic kidney disease may have limited side effects. In contrast, the use T3Inh-1 to oppose upregulated GalNAc-T3 activity in cancer in a patient or animal model would be expected to also alter FGF23 levels. Such effects may be mitigated by precise titration and time of exposure. Clearly, the issue of multiple substrates (and even functionally distinct sites within a given substrate) has been overcome in successful therapies targeting protein kinases.

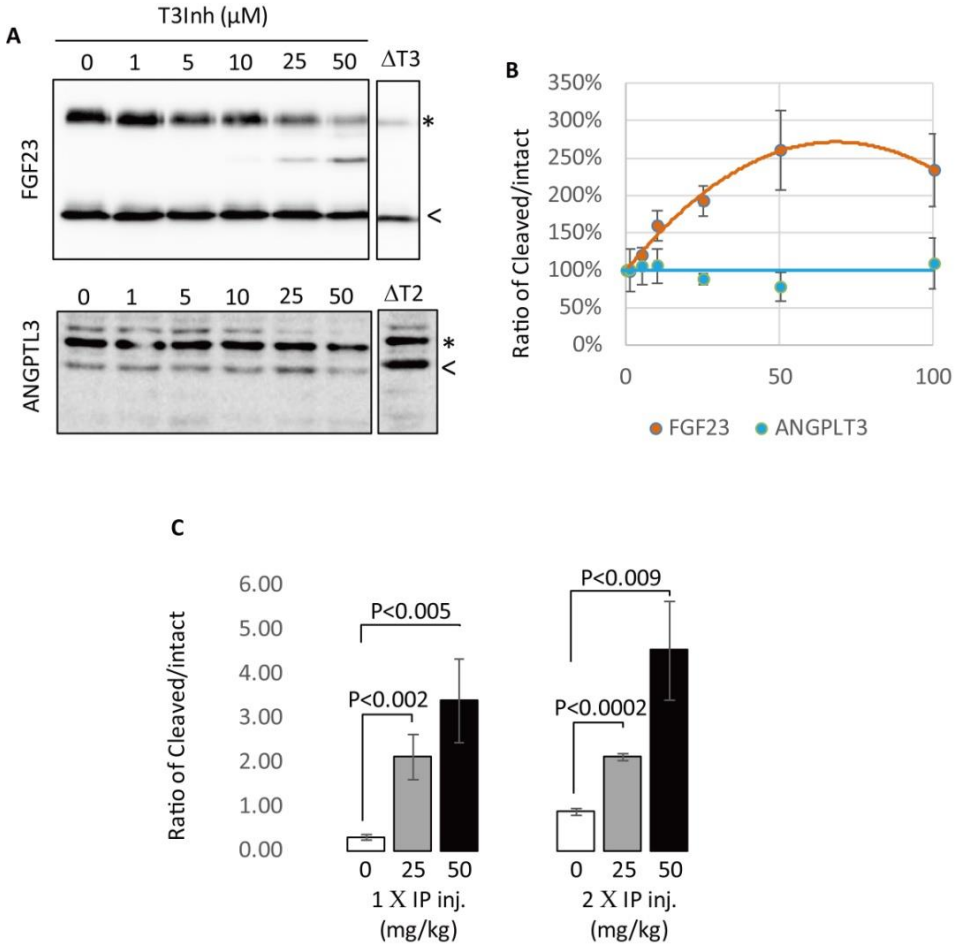

Figure 4.9. T3Inh-1 promotes FGF23 processing. (A) Immunoblot of media collected from cells after a 6 h period in the presence of the indicated concentrations of T3Inh-1. HEK cells were transfected with FLAG-FGF23 and GalNAc-T3 or Myc-ANGPTL3 and anti-FLAG and anti-Myc antibodies were used to assay intact (*) and cleaved (<) FGF23 and ANGPTL3, respectively. (B) Quantified results showing the percent ratio change of cleaved/intact FGF23 or ANGPTL3 normalized to the amount present in untreated controls (n≥3±SEM). (C) Serum ELISA assay results showing ratio of

cleaved/intact FGF23 in mouse sera collected 24 h after either 1 or 2 (consecutive day) intraperitoneal injections of the indicated amount of T3Inh-1 (averages of 4 animals ±SEM).

Further analysis of other hits in this and ongoing screens will likely identify additional inhibitors and/or activators of GalNAc-T2 and GalNAc-T3 and should also identify pan-specific modulators. We anticipate that discovery of the latter, as well as isozyme-specific modulators of other GalNAc Ts, will be aided by both development of new sensors (with new specificities built into the linker domain) and rational design approaches aided by T3Inh-1 identified here. Because many individual GalNAc Ts are associated with unique diseases[1,7,8], we may have the basis for a new class of therapeutics capable of treating an array of differing diseases once a panel of modulators is available targeting individual isozymes.

4.5 Materials and Methods

Primary screen. HEK cells stably expressing the T2 sensor (containing ANGPTL3 linker sequence with $T_{225}G$ modification[66]) or the T3 sensor (containing FGF23-based linker[66]) were cultured in MEM (Corning, NY) with 10% fetal bovine serum (FBS, Atlanta Biologicals, Flowery Branch, GA) and 100 IU/ml penicillin-streptomycin (Sigma-Aldrich, St. Louis, MO) at 37 °C, 5% CO_2. Positive controls were cell lines expressing matched sensors with the glycosylation site mutated (ΔGly), specifically $T_{225}G/T_{226}G$ and $T_{178}G$ for the T2 and T3 sensors, respectively (*26*). Cells (50,000/well) were seeded in 96-well-plates and grown for 24 h. Compounds (the approved oncology drugs set V and the diversity set II from the National Cancer Institute Developmental Therapeutics Program) were then added to achieve a 10 μM final concentration. After 6 h, the medium was aspirated and the cells were released by adding 100 μl 5 mM EDTA/PBS containing 110 nM MG dye for 5 min at 37°C. The plates were then transferred to an Accuri™ C6 flow cytometer (BD Biosciences) where GFP and MG fluorescence was measured using 488 nm and 640 nm for 10,000 cells per well. Data analysis used FlowJo software (www.flowjo.com). For each well the geometric means of the MG and GFP signals were used to compute the MG/GFP ratio. Each compound was analyzed in two wells and the average of the two resulting ratios (R) was recorded. Each daily run included at least 16 wells of vehicle-only controls (sensor-expressing cells treated with a matching DMSO concentration) and a similar analysis was used to calculate their average MG/GFP ratio (R_{Neg})

and its standard deviation (SD_{Neg}). The Q value of each compound (as well as the untreated ΔGly positive controls) was calculated by using the following equation: $Q=(R-R_{Neg})/SD_{Neg.}$ Background fell within the range $-2.5 \leq Q \leq 3$ and the average $Q_{\Delta Gly}$ was 135 and 38 for the T2 and T3 sensors, respectively.

Secondary screen. Glycosylation assays using recombinant GalNAc-T2 and GalNAc-T3 were carried out using the UDP-Glo™ Glycosyltransferase assay kit (Promega, Madison, WI), according to the manufacturer's recommendation. The reaction (25 µL) included 2.5 ng/ µl purified enzyme, 25 µM UDP-GalNAc (Sigma-Aldrich), 12.5 µM EA2 peptide (AnaSpec, Mucin 10, AA153-165, PTTDSTTPAPTTK), 25 mM Tris-HCl (pH7.5), 5mM $MnCl_2$, 2.5 mM $CaCl_2$ and 50 µM compound. The negative control was vehicle only (same reaction mixture with a matched percentage of DMSO instead of compound), whereas background was from the reaction carried out without enzyme and DMSO instead of compound. All reactions were incubated at 37 °C in a water bath for 30 min and then cooled to room temperature. Aliquots (5 µL) were then added to 384-well plate (Thermo Scientific, Waltham, MA) to which 5 µl of UDP Detection Reagent was also added. Duplicate measures were made for all reactions. After 1 h at room temperature the luminescent signals were determined using a Tecan Infinite M1000 (Tecan Group Ltd., Männedorf, Switzerland) with integration time set to 1000 msec. The background-subtracted average for each compound was expressed as a percentage of the negative control (taken as 100%). Compounds with ≥50% effect were considered direct modulators.

Titration assays. For the sensor assay, about 200,000 cells expressing either sensor were seeded into 24-well plates. After 24 h, the cells were incubated for another 6 h in the presence of 0-50 µM compound. To release the cells the medium was replaced with 200µl 5mM EDTA/PBS containing 110nM MG dye. After 5 min at 37°C, fluorescence measurements (20,000 cells/well) were carried out as described above. For the biochemical assay, the assay conditions were identical except for variations in the compound or substrate concentrations as indicated in the figure legends. Data analysis was using Prism (Graph Pad Prism Inc.).

Tryptophan fluorescence quenching assay. The purified lumenal domain of GalNAc-T3 (30 ng/ µL) was incubated with 0-500 µM compound at room temperature for 10 min and 200 µl aliquots were transferred to a 96-well glass-bottom plate and the fluorescent emission was

scanned (300-450nm) using a Tecan Infinite M1000 with excitation at 290nm, gain set to 150, number of flashes at 50 and flash frequency at 400Hz. The value at the peak of emission at 324nm was used for the binding curve analysis by Prism.

Microscopy. For determination of sensor activation, spinning-disk confocal microscopy was used exactly as described[66]. To assess possible effects of compounds on Golgi markers, including GalNAc-T2 and GalNAc-T3, immunofluorescence was carried on HeLa cells treated with 10 μM compound for 6 h. Briefly, the cells were grown on 12mm diameter coverslips for 48 h, treated with the compounds, washed with PBS and fixed with 3% paraformaldehyde for 15 min. Blocking, Triton X-100 permeabilization, antibody incubations and image capture by spinning-disk confocal were as described[97]. Monoclonal antibodies against GalNAc-T2 and T3 were used undiluted and the polyclonal against GPP130 was used at 1:2000. Lectin staining was identical to immunofluorescence except that the fluorescent lectins (HPA, VVA, ConA and GSII from Vector Laboratories, INC. Burlingame, CA) were used at 1:200 in place of antibodies. All corresponding images were acquired and adjusted using identical parameters.

Proliferation assay. Equal numbers of HEK cells were plated in 24-well dishes in growth medium containing 0-50 μM T3Inh-1. After 24, 48 or 72 h, the cells were released using trypsin and counted twice using a hemocytometer.

Cell invasion assay. Breast cancer MDA-MB231 cells were grown in DMEM medium (VWR) with 10% FBS and 100 IU/ml penicillin-streptomycin at 37 °C and 5% CO_2 and then plated at a density of 1.32×10^4 in 0.3 ml of RPMI1640 medium without FBS into the upper chamber of a Bioboat™ insert fitted with a 8.0 μm PET membrane (Corning, NY). For migration assays, the filter was uncoated. For invasion assays, it was pre-coated with 100 μl Matrigel™ (BD Biosciences) at concentration of 272 μg /ml for 1-2 hours at room temperature. Medium containing 10% FBS (0.6 ml) was placed into the lower chamber as a chemoattractant. The compound (final concentration of 5 μM) or a matching amount of DMSO was added to both chambers. After 24h or 48h the cells were fixed with -20°C methanol for 15 min and then stained with Trypan blue for 5 min. Cells on the upper surface were removed using cotton swabs. Cells present on the underside of the membrane were photographed using an EVOS FL Cell Imaging System (Invitogen, CA) and the images were thresholded for presentation and

counting using Image J (National Institutes of Health, Bethesda, MD). Proliferation of MB231 cells was assayed as described above for HEK cells.

Immunoblotting. The FLAG-tagged FGF23, His-tagged ANGPTL3[67], and untagged GalNAc-T3 (from ([35]) and cloned into PCDNA 3.0 using BamH1 sites) were transfected into HEK cells using the JetPEI transfection reagent (VWR International, Radnor, PA) according to the manufacturer's instructions. After 24 h, the medium was replaced with serum-free MEM containing the compound for 6 h. The medium and cells were then collected and, after trichloroacetic acid precipitation of the medium, analyzed by immunoblot using anti-FLAG antibody at 1:1000 (Bethyl Labs Inc., TX) or anti-His antibody at 1:2000 (Bethyl Labs Inc.) and a peroxidase-coupled secondary antibody (BioRad Labs Inc., CA). Emission was captured and quantified using a ChemiDoc^{TM} Touch Imaging System with Image Lab Software (Biorad).

Animal analysis. Wild-type C57BL/6 six to eight week mice were purchased from Charles River Laboratories international Inc. (Wilmington, MA). Protocols, handling, and care of the mice conformed to protocols approved by the Institutional Animal Care and Use Committee of Carnegie Mellon University. The compound was dissolved in DMSO at 25 and 50 mg/ml then further diluted with PEG400 to create 5 and 10 mg/ml stocks for injection. Control (vehicle only: 20% DMSO, 80% PEG400) and experimental (25 or 50 mg/kg compound) animals received either single or double (separated by 24 h) intraperitoneal injections and, 24 h after the last injection, a cardiac blood draw was carried out. Cleaved FGF23 was determined using ELISA kits from Immunotopics (Carlsbad, VA, USA) by subtracting intact from total.

Chapter 5：Progress towards Identifying additional GalNAc-Ts Modulators by Cell-Based Screening

5.1 Abstract

In addition to the GalNAc-T3 inhibitor described in the previous chapter, there is biological and medical significance in identifying other drug-like modulators against GalNAc-Ts, including activators of T3, inhibitors and activators of T2, and "pan-specific" modulators that target all isoforms. To date, we carried out screening of 21,710 compounds using the cell-based sensors specific to T2 and T3. There were 75 above threshold hits. Of these, 53 compounds were specific to T2, 19 were specific to T2 and 3 acted on both sensors (including a compound that activated the T2 sensor but inhibited the T3 sensor. A fraction of the hits (51 compounds) have been tested at 50μM in the *in vitro* assay and 10 compounds were active *in vitro*. However, only one compound showed well-behaved, dose-dependent inhibition. Since it was active against both purified enzymes, it is a potentially a pan-specific inhibitors of all GalNAc isoenzymes. It is hoped that our ongoing analysis of the untested hits (24 compounds) will identify additional drug-like modulators targeting GalNAc-Ts.

5.2 Introduction

Similar to GalNAc-T3 introduced in the previous chapter, GalNAc-T2 is another well-studied isoform with crucial functions. GalNAc-T2 is ubiquitously expressed, which contrasts with the limited spatial and temporal expression pattern of many other GalNAc-T isoforms including GalNAc-T3[28]. GalNAc-T2 is involved in lipid metabolism by affecting ANGPTL3's cleavage[67,98,99]. ANGPTL3, which is expressed and secreted by the liver functions as an important inhibitor of the lipoprotein lipase and endothelial lipase. Its activation requires furin cleavage at the C-terminal side of R_{224} but glycosylation by GalNAc-T2 at T_{226} blocks this cleavage. Humans with loss-of-function mutations in *GALNT2* and mice with a *GALNT2* knockout show lower high-density lipoprotein cholesterol because of increased ANGPTL3 cleavage. In addition, expression of GalNAc-T2 is upregulated for extravillus trophoblast

invasion in placental development[100]. It has been shown that GalNAc-T2 regulates extravillus trophoblast cell adhesion, migration, and invasion through specifically modulating the glycosylation and functions of β1-integrin. Like GalNAc-T3, misregulation of GalNAc-T2 also correlates with carcinogenesis[22,38–40]. For example, GalNAc-T2 is a negative indicator in gastric cancer, neuorablastoma and human glioma metastasis. In addition, the glycosylation and activity of EGFR is regulated by GalNAc-T2 in hepatocellular carcinoma (HCC) and oral squamous cell carcinoma but in opposite directions. GalNAc-T2 is downregulated in HCC. Huang MC et al. reported that restoring GalNAc-T2 expression in HCC cells suppresses EGF-induced cell growth, migration and invasion *in vitro* and *in vivo*. In oral squamous cell carcinoma, overexpressed GalNAc-T2 enhanced invasiveness by regulating glycosylation and activity of EGFR. The apparent contradictory role of glycosylation in regulating EGFR function in different cancer cells remains unclear.

In summary, similar to GalNAc-T3, there are several disease-relevant functions of GalNAc-T2 that imply the therapeutic potential of GalNAc-T2 modulators. To identify drug-like compounds targeting GalNAc-T2, GalNAc-T3 or both, 21,710 compounds were screened using the isoform specific sensors described in Chapter 3. Of the 75 confirmed hits resulting from the primary screen only a fraction have yet been analyzed using the secondary *in vitro* assay screen. To date, no additional compounds with direct isoform-specific effects were identified but two potential pan-specific inhibitors were obtained.

5.3 Results

Primary screening results against two sensors for the large library. To identify compounds that act on GalNAc-T2/T3, a large compound library was screened using the strategy described in chapter 4 (Figure 4.2). Figure 5.1A show the results for the primary screening of 21,710 compounds against the two the sensors. By using the same threshold cut-offs of -2.5 and +3, 75 compounds were considered hits after the primary screening. Of these, 19 compounds selectively enhanced the T2 sensor and 34 selectively suppressed it corresponding to putative inhibitors and activators of the T2 enzymes, respectively (Figure 5.1B). We also identified 10 potential inhibitors and 9 potential activators of the T3 enzyme. Two compounds enhanced the signal of both the T2 sensor and the T3 sensor, suggesting they

are potentially pan-inhibitors. Strangely, one compound (30978-F07) activated the T2 sensor (Q value=5.8) while also inhibiting the T3 sensor (Q value=-3.6). Further testing will be needed to determine the reason of this behavior.

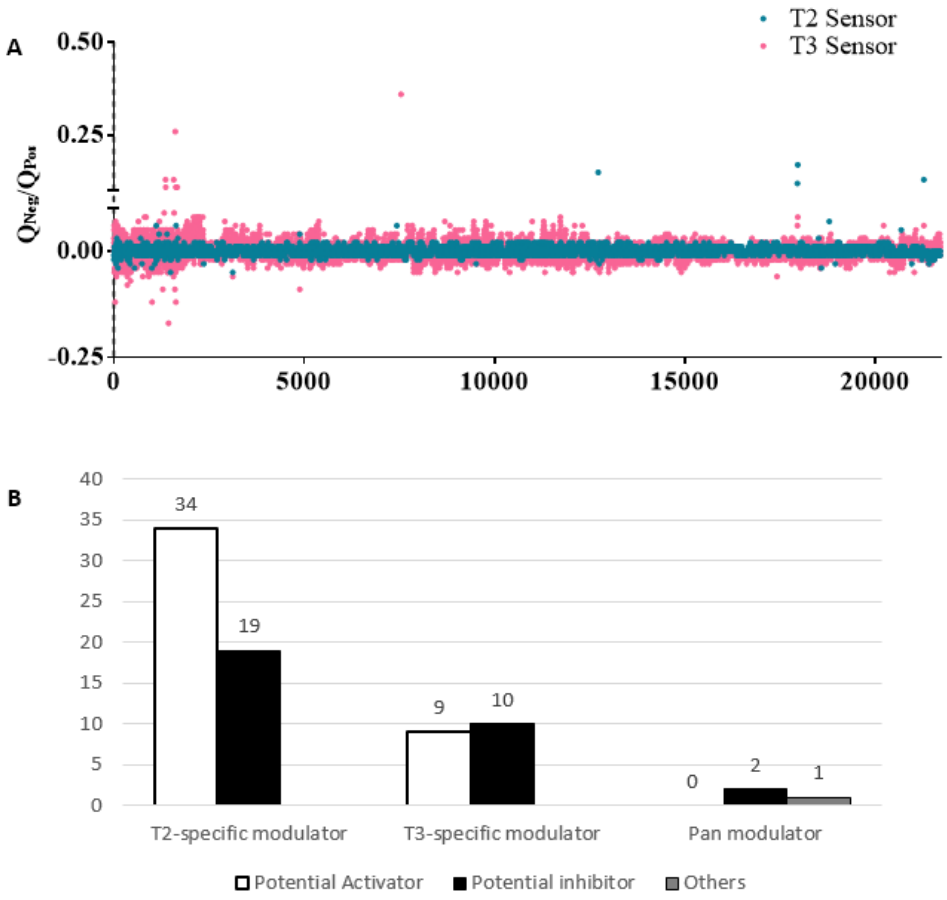

Figure 5.1. Primary screen results for modulators of GalNAc-T2/T3 from the large library. (A) Q values ($Q=(R-R_{Neg})/SD_{Neg}$) were determined for each compound ($Q_{Compound}$) using the average of duplicate MG/GFP ratios for the compound (R), the vehicle-only control (R_{Neg}), and the standard deviation of the vehicle-only controls (SD_{Neg}). The plot shows the sensor signals for each compound. (B) Bar graph summarizes the candidates have been tested *in vitro* assay. (B) Bar graph summarizes the candidates from the primary sensor screening

Secondary screening results using purified GalNAc-Ts *in vitro*. The 75 primary screen hits were tested for their activity *in vitro* at 50µM against the purified enzymes (GalNAc-T2 and T3). In this secondary screen, to be considered a hit a compound needed to either increase or decrease the luminescence signal by 50% compared to the vehicle control. Figure 4.2 shows the results of the *in vitro* assay. Compound 4774-A06 yielded a large standard deviation indicating a need for further testing. Compound 4788-F11 is the T3-specific inhibitor (T3Inh-

1) discussed in Chapter 3. We noted some discrepancies compared to those of the primary screen that may have been due to concentration differences since the primary screen was carried out at a lower concentration (10μM). In all, twenty candidates were considered hits in the secondary screen and only nine of these yield results were consistent with primary screen in terms of direction (activation or inhibition). Of those, 2 are potential activators (4773-B07 and 30974-D02) and 7 are potential inhibitors (1-G10, 4788-F11, 4789-A04, 31011-E05, 31012-H05, 60213-D08 and 60239-A10).

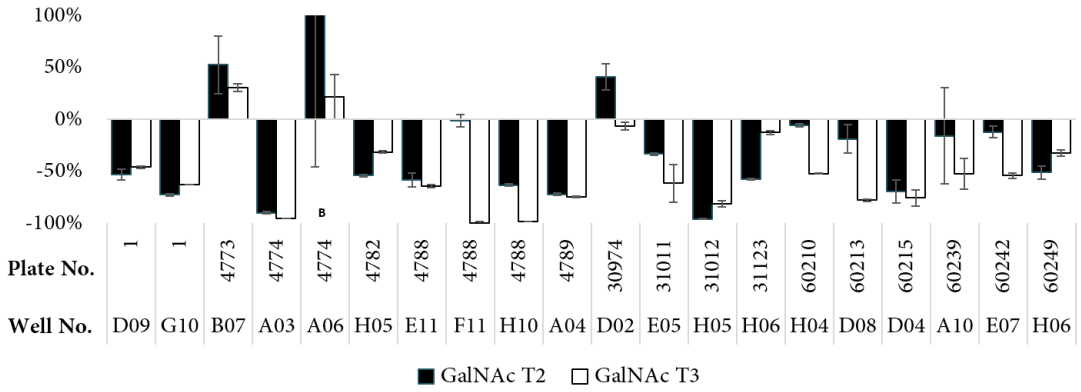

Figure 5.2 The secondary screen results using purified GalNAc-Ts. The purified GalNAc-T catalyzes glycosylation reaction and transfer GalNAc from UDP-GalNAc to the potential glycosylation sites (T/S) of peptide at proper condition. The UDP released from reaction will be transferred to ATP, which binds to luciferase and give detectable signal. The figure shows *in vitro* assay results for the twenty compounds above the threshold (>50% or <-50%). Among those compounds, six compounds are pan-inhibitors, one compound relatively inhibits GalNAc-T2 and one compound inhibits GalNAc-T3 only, and two activate GalNAc-T2.

Analysis of a putative pan-specific compound. The 9 candidates from the secondary screen will be next tested for dose-dependence in both the sensor assay (i.e. against the two sensors) and in the *in vitro* assay (i.e. against the purified GalNAc-Ts). This was initiated for 4789-A04. It activated only the T2 sensor in the primary screen. It inhibited both enzymes in the secondary screen. When titrated against sensors, it activated both. And, it inhibited both enzymes in the *in vitro* titration test (Figure 5.3). Note that 4789-A04 had no effect in the *in vitro* assay if only UDP was added. This shows that the compound does not act off-target in the assay. Nevertheless, it's IC50 of 79μM for GalNAc-T2 and 40μM for GalNAc-T3 suggests that the compound will need further optimization to be useful.

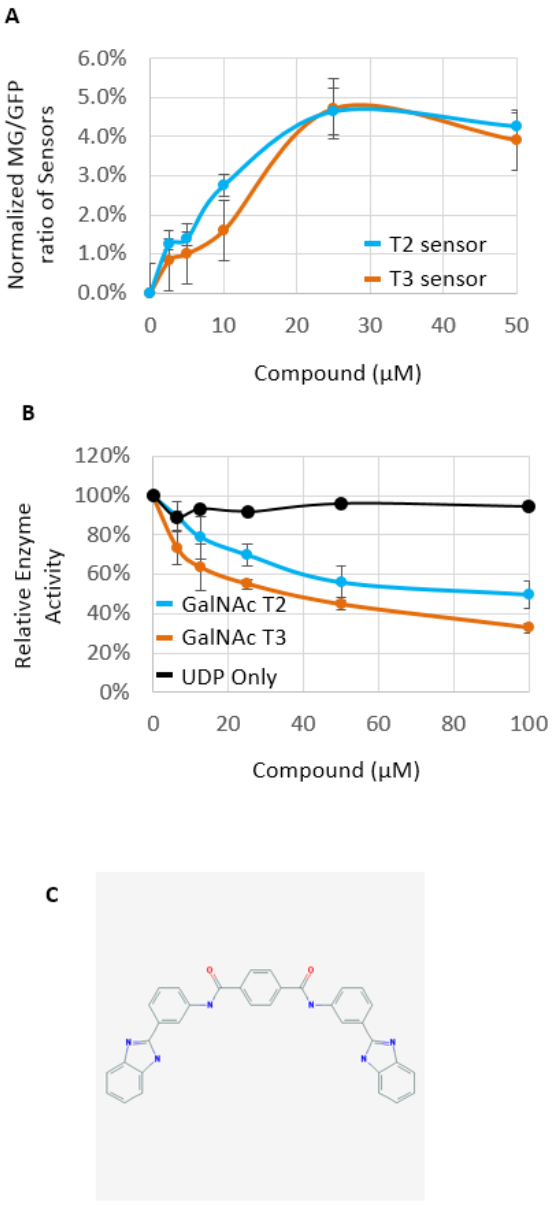

Figure 5.3. The titration results of 4789-A04. The results of sensor titrations (A) (n=2±SEM), *in vitro* titration (B) (n=2±SEM), and structures (C) for 4789-A04. The structures were downloaded from PubMed website.

In conclusion, in addition to T3Inh-1, we have identified eight promising candidates and our preliminary results on one of these indicates that we will identify at least one compound that inhibits both GalNAc-T2 and T3 (i.e. a potential pan-specific inhibitor).

Chapter 6： Future Directions

6.1 T3Inh-1 inhibition mechanism.

Using our GalNAc-T sensors, we discovered T3Inh-1, the first inhibitor of GalNAc-T3. Interestingly, T3Inh-1 does not inhibit the activity of GalNAc-T2 but it does bind GalNAc-T2 with a low affinity. The two isozymes share a high degree of sequence similarity in their catalytic domains and less similarity in their lectin domains. To determine the T3Inh-1 binding site, the catalytic and the lectin domains of GalNAc-T3 will be purified and direct binding will be carried out *in vitro*. Once binding is observed, mutation will be used to map the binding site. The focus will be on residues near known functional sites such as the sites for binding UDP-GalNAc or peptide of the catalytic domain, or the glycan binding site of the lectin domain. If enzyme activity can be retained this should result in a T3Inh-1-insensitive version, which will be useful for future experiments. In addition to these binding studies, we will carry out co-crystallization trials (in collaboration with Dr. Ramon Hurtado Guerrero) to determine the structure of the enzyme/inhibitor complex at high resolution.

6.2 Rational design of T3Inh-1 derivatives.

As the first identified GalNAc-T3 inhibitor T3Inh-1 is a great tool to study the catalytic mechanism and the functions of GalNAc-T3. However, the limited solubility of T3Inh-1 in aqueous solution is a problem. Even in DMSO, the maximum achievable concentration is only 50mM. Using 80% PEG400/ 20% DMSO as vehicle allowed our animal studies, but solubility is still an issue hindering full-scale studies especially because we observed a certain degree of drug precipitation in this formulation after extended storage. Therefore, we will continue our work and optimize the structure of T3Inh-1 for optimal efficacy, solubility and other improved properties such as absorption, distribution, metabolism, excretion, and toxicity (ADMET). In collaboration with a medicinal chemist, we expect to develop a second generation of GalNAc-T3 inhibitors with improved potency and physicochemical properties through our further structure-activity relationship (SAR) studies.

6.3 To assay anti-metastatic activity of T3Inh-1 *in vivo* using a mouse model.

Based on our results, T3Inh-1 significantly inhibits migration and invasion of MDA-MB231 breast cancer cells *in vitro* without affecting cell proliferation. MDA-MB231 shows a higher expression level of GalNAc-T3 than MCF-7, which correlates with its significantly higher invasiveness *in vitro*. Overexpression of GalNAc-T3 in MCF-7 cells enhanced cell invasiveness and this was inhibited by 5μM T3Inh-1. Because these results suggest that T3Inh-1 might be a therapeutic agent for cancer metastasis, one of our future goals is to test its effectiveness against metastasis using the mouse model. Both spontaneous and passive metastasis models will be tested. If we observe T3Inh-1 treatment attenuates in the passive model, but not in the spontaneous metastasis model, it will suggest that T3 acts during tumor circulation after dissemination rather than during initial invasion. Further, to validate possible results implicating T3 in metastasis, we will use overexpression and CRISPR-Cas9 knockout of T3 in the metastasis assay.

6.4 Use T3Inh-1 to identify the T3 substrates that mediate the metastatic phenotype.

Our preliminary data suggests that substrates of GalNAc-T3 in MDA-MB231 cells drive metastatic-like behavior. To identify these substrates and their putative GalNAc-T3-specific glycosylation sites, MDA-MB231 will be cultured with or without T3Inh-1 while differentially labeling cellular protein with stable isotope labeling with amino acids in cell culture (SILAC). Mass spectrometry will be used to identify the substrates and their glycosylation sites based on quantitative differences between the two groups. To confirm biological relevance of identified potential substrates, they will be knocked out and rescued. The expectation is that knockout will be block invasiveness and rescue will restore it if the substrate contributes to the invasive phenotype.

6.5 Determine the range of cancer types that are T3Inh-1 sensitive using the cultured cell invasion assay.

High expression level of GalNAc-T3 correlates not only with invasiveness of breast cancer, but also with invasiveness of other cancers, including pancreatic, thyroid, renal and ovarian. This suggests that T3Inh-1 might block invasiveness of cell lines derived from these other cancers. A future goal is to first determine GalNAc-T3 expression levels across a wide-array of human cancer cell lines covering major types of cancer (selected from the NCI-60 panel and

then, for the same cells, determine their sensitivity to T3Inh-1 in the cell invasion assay used in Chapter 3. This will correlate expression level with invasiveness and provide a function test of the role of GalNAc-T3. Further, sensitivity will imply that the cancer is potentially treatable with T3Inh-1.

6.6 To identify GalNAc-T2-specific modulators and pan-modulators

The work in chapter 4 is incomplete and we have not yet found any GalNAc-T2-specific modulators. The secondary screening will be completed and any hits will be validated along the lines of what was accomplished for T3Inh-1. Our hope is to not only identify inhibitors of GalNAa-T2 but also compounds that act universally to inhibit all GalNAc-Ts. Our argument is that such reagents will be transformative in glycobiology research and represent a potentially represent a new class of therapeutic.

Reference

1. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation : diversity , synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13,** 448–462 (2012).
2. Stowell, S. R., Ju, T. & Cummings, R. D. Protein Glycosylation in Cancer. (2015). doi:10.1146/annurev-pathol-012414-040438
3. Cummings, R. D. & Wright, P. C. Molecular BioSystems. (2009). doi:10.1039/b907931a
4. Nairn, A. V *et al. Regulation of Glycan Structures in Animal Tissues*. **283,** (2008).
5. Canuel, M., Libin, Y. & Morales, C. R. The interactomics of sortilin : an ancient lysosomal receptor evolving new functions Histology and. 481–492 (2009). doi:10.14670/HH-24.481
6. Boskovski, M. T. *et al.* The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature* **504,** 456–9 (2013).
7. Post, S. Van Der *et al.* Site-specific O -Glycosylation on the MUC2 Mucin Protein Inhibits Cleavage by the Porphyromonas gingivalis Secreted Cysteine Protease (RgpB) * □. **288,** 14636–14646 (2013).
8. Schjoldager, K. T.-B. G. & Clausen, H. Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochim. Biophys. Acta* **1820,** 2079–94 (2012).
9. Planinc, A., Bones, J., Dejaegher, B. & Antwerpen, P. Van. Analytica Chimica Acta Glycan characterization of biopharmaceuticals : Updates and perspectives. **921,** (2016).
10. Zhang, X. & Wang, Y. Glycosylation Quality Control by the Golgi Structure. *J. Mol. Biol.* **428,** 3183–3193 (2016).
11. Hemming, F. W. & Centre, M. Assembly of the Oligosaccharide Chains. **221,** 203–221 (1982).
12. Neubert, P. & Strahl, S. ScienceDirect Protein O -mannosylation in the early secretory pathway. *Curr. Opin. Cell Biol.* **41,** 100–108 (2016).
13. Hart, G. W. Minireview Nutrient Regulation of Signaling , Transcription , and Cell Physiology by O -GlcNAcylation Minireview. (2014). doi:10.1016/j.cmet.2014.07.014
14. Ma, Z. & Vocadlo, D. J. Hyper- O -GlcNAcylation Is Anti-apoptotic and Maintains Constitutive NF- □ B Activity in Pancreatic Cancer Cells * □. (2013). doi:10.1074/jbc.M113.470047
15. Tan, E. P. *et al.* Altering O-linked β -N-acetylglucosamine cycling disrupts mitochondrial function. *J. Biol. Chem.* **289,** 14719–14730 (2014).
16. Bennun, S. V *et al.* Systems Glycobiology : Integrating Glycomics , and Other ‘ Omics Data Sets to Characterize Cellular Glycosylation Processes. *J. Mol. Biol.* **428,** 3337–3352 (2016).
17. Mills, I. G. Seminars in Cell & Developmental Biology Nuclear translocation and functions of growth factor receptors. *Semin. Cell Dev. Biol.* **23,** 165–171 (2012).
18. Ten Hagen, K. G., Fritz, T. a. & Tabak, L. a. All in the family: The UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* **13,** 1–16 (2003).
19. Nordén, R. *et al.* O -Linked Glycosylation of the Mucin Domain of the Herpes Simplex Virus Type 1-specific Glycoprotein gC-1 Is Temporally Regulated in a Seed-and-spread Manner * □. **290,** 5078–5091 (2015).

20. Serafini-Cessi, F., Dall'Olio, F., Malagolini, N. & Campadelli-Fiume, G. Temporal aspects of O-glycosylation of glycoprotein C from herpes simplex virus type-1. *Biochem J* **262,** 479–484 (1989).

21. Huang, M., Hu, R., Chou, C. & Hsu, C. Knockdown of GALNT1 suppresses malignant phenotype of hepatocellular carcinoma by suppressing EGFR signaling. **6,**

22. Lin, M., Huang, M., Liu, C., Yang, T. & Huang, M. GALNT2 enhances migration and invasion of oral squamous cell carcinoma by regulating EGFR glycosylation and activity. *Oral Oncol.* **50,** 478–484 (2014).

23. Yeh, J. *et al.* Novel Sulfated Lymphocyte Homing Receptors and Their Control by a Core1 Extension [illegible] 1 , 3- N -Acetylglucosaminyltransferase. **105,** 957–969 (2001).

24. Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the Molecular Basis of Leukocyte Tethering and Rolling Revealed by Structures of P- and E-Selectin Bound to SLe X and PSGL-1. **103,** 467–479 (2000).

25. Bennett, E. P. *et al.* Cloning and characterization of a close homologue of human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *J. Biol. Chem.* **274,** 25362–25370 (1999).

26. Bennett, E. P. *et al.* Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22,** 736–756 (2012).

27. J.C. Paulson, K. J. C. Minireview: Glycosyltransferases. *J. Biol. Chem.* **264,** 17615–17618 (1989).

28. 2011-R-Location, location, location-new insights into O-GalNAc protein glycosylation.pdf.

29. Fritz, T. a, Raman, J. & Tabak, L. a. Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferase-2. *J. Biol. Chem.* **281,** 8613–8619 (2006).

30. Yoshimura, Y. *et al.* Elucidation of the sugar recognition ability of the lectin domain of UDP-GalNAc : polypeptide N -acetylgalactosaminyltransferase 3 by using unnatural glycopeptide substrates. **22,** 429–438 (2012).

31. Gerken, T. a. *et al.* The lectin domain of the polypeptide GalNAc transferase family of glycosyltransferases (ppGalNAc Ts) acts as a switch directing glycopeptide substrate glycosylation in an N- or C-terminal direction, further controlling mucin type O-Glycosylation. *J. Biol. Chem.* **288,** 19900–19914 (2013).

32. Gill, D. J. *et al.* Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. (2013). doi:10.1073/pnas.1305269110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1305269110

33. Chefetz, I. & Sprecher, E. Familial tumoral calcinosis and the role of O-glycosylation in the maintenance of phosphate homeostasis. *Biochim. Biophys. Acta* **1792,** 847–52 (2009).

34. Schjoldager, K. T.-B. G. *et al.* A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing. *J. Biol. Chem.* **286,** 40122–40132 (2011).

35. Kato, K. *et al.* Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation. *J. Biol. Chem.* **281,** 18370–7 (2006).

36. Willer, C. J. *et al.* Newly identified loci that influence lipid concentrations and risk of

coronary artery disease. *Nat. Genet.* **40,** 161–169 (2008).

37. Tian, E., Stevens, S. R., Guan, Y., Springer, D. A. & Anderson, S. A. Galnt1 Is Required for Normal Heart Valve Development and Cardiac Function. 1–19 (2015). doi:10.1371/journal.pone.0115861

38. Wu, Y., Liu, C., Hu, R., Huang, M. & Lee, J. Mucin Glycosylating Enzyme GALNT2 Regulates the Malignant Character of Hepatocellular Carcinoma by Modifying the EGF Receptor. **71,** 7270–7279 (2011).

39. Hua, D. *et al.* Polypeptide N-acetylgalactosaminyltransferase 2 regulates cellular metastasis-associated behavior in gastric cancer. 1267–1274 (2012). doi:10.3892/ijmm.2012.1130

40. Ho, W. *et al.* GALNT2 suppresses malignant phenotypes through IGF-1 receptor and predicts favorable prognosis in neuroblastoma. **5,** (2014).

41. Li, Z. *et al.* Polypeptide N -acetylgalactosaminyltransferase 6 expression in pancreatic cancer is an independent prognostic factor indicating better overall survival. 1882–1889 (2011). doi:10.1038/bjc.2011.166

42. Taniuchi, K. *et al.* Overexpression of GalNAc-transferase GalNAc-T3 promotes pancreatic cancer cell growth. 4843–4854 (2011). doi:10.1038/onc.2011.194

43. Mochizuki, Y., Ito, K., Izumi, H., Kohno, K. & Amano, J. Expression of Polypeptide N-Acetylgalactosaminyl Transferase-3 and Its Association with Clinicopathological Factors in Thyroid Carcinomas. **23,** (2013).

44. Kitada, S. *et al.* Polypeptide N-acetylgalactosaminyl transferase 3 independently predicts high-grade tumours and poor prognosis in patients with renal cell carcinomas. *Br. J. Cancer* **109,** 472–81 (2013).

45. Wang, Z. *et al.* Role of the polypeptide N-acetylgalactosaminyltransferase 3 in ovarian cancer progression : possible implications in abnormal mucin O-glycosylation ABSTRACT : **5,** (2014).

46. Harada, Y., Izumi, H., Noguchi, H. & Kuma, A. Strong expression of polypeptide N -acetylgalactosaminyltransferase 3 independently predicts shortened disease-free survival in patients with early stage oral squamous cell carcinoma. 1357–1368 (2016). doi:10.1007/s13277-015-3928-7

47. He, H. *et al.* Clinical significance of polypeptide N -acetylgalactosaminyl transferase-5 (GalNAc-T5) expression in patients with gastric cancer. **5,** 2021–2029 (2014).

48. Fibronectin, O., Park, J. & Katagiri, T. Polypeptide N -acetylgalactosaminyltransferase 6 Disrupts Mammary Acinar Morphogenesis through. **13,** 320–326 (2011).

49. Tarhan, Y. E., Kato, T. & Jang, M. Morphological Changes , Cadherin Switching , and Growth Suppression in Pancreatic Cancer by GALNT6. *Neoplasia* **18,** 265–272 (2016).

50. Li, Z., Yamada, S., Wu, Y., Wang, K. & Liu, Y. Polypeptide N-acetylgalactosaminyltransferase-6 expression independently predicts poor overall survival in patients with lung adenocarcinoma after curative resection. (2016).

51. Gaziel-sovran, A. *et al.* Article miR-30b / 30d Regulation of GalNAc Transferases Enhances Invasion and Immunosuppression during Metastasis. *Cancer Cell* **20,** 104–118 (2011).

52. Peng, R. *et al.* MicroRNA-214 Suppresses Growth and Invasiveness of Cervical Cancer Cells by Targeting UDP- N -acetyl- ⎵ - D -galactosamine : Polypeptide N -Acetylgalactosaminyltransferase 7 * □. **287,** 14301–14309 (2012).

53. Shan, S. W., Fang, L., Shatseva, T., Rutnam, Z. J. & Yang, X. Mature miR-17-5p and passenger miR-17-3p induce hepatocellular carcinoma by targeting PTEN , GalNT7 and vimentin in different signal pathways. (2013). doi:10.1242/jcs.122895

54. Li, W. E. I., Ma, H. & Sun, J. I. microRNA‐34a / c function as tumor suppressors in Hep‐2 laryngeal carcinoma cells and may reduce GALNT7 expression. 1293–1298 (2014). doi:10.3892/mmr.2014.1929

55. Duan, H., Li, X., Hu, H. & Li, Y. Functional elucidation of miR-494 in the tumorigenesis of nasopharyngeal carcinoma. 6679–6689 (2015). doi:10.1007/s13277-015-3356-8

56. Berois, N. *et al.* GALNT9 Gene Expression Is a Prognostic Marker in Neuroblastoma Patients BACKGROUND : **233,** 225–233 (2013).

57. Gao, Y., Liu, Z., Feng, J., Sun, Q. & Zhang, B. A. O. Expression pattern of polypeptide N-acetylgalactosaminyltransferase-10 in gastric carcinoma. 113–116 (2013). doi:10.3892/ol.2012.980

58. Wu, Q. *et al.* Elevated Expression of N -Acetylgalactosaminyltransferase 10 Predicts Poor Survival and Early Recurrence of Patients with Clear-Cell Renal Cell Carcinoma. 2446–2453 (2015). doi:10.1245/s10434-014-4236-y

59. Wu, Q. *et al.* Decreased Expression of Hepatocyte Nuclear Factor 4 ␣ (Hnf4 ␣)/ MicroRNA-122 (miR-122) Axis in Hepatitis B Virus-associated Hepatocellular Carcinoma Enhances Potential Oncogenic GALNT10 Protein Activity *. **290,** 1170–1185 (2015).

60. Libisch, M. G. *et al.* GALNT11 as a new molecular marker in chronic lymphocytic leukemia. *Gene* **533,** 270–279 (2014).

61. Matsumoto, Y., Zhang, Q., Akita, K., Nakada, H. & Hamamura, K. Biochemical and Biophysical Research Communications pp-GalNAc-T13 induces high metastatic potential of murine Lewis lung cancer by generating trimeric Tn antigen. *Biochem. Biophys. Res. Commun.* **419,** 7–13 (2012).

62. Matsumoto, Y. *et al.* Trimeric Tn antigen on syndecan 1 produced by ppGalNAc-T13 enhances cancer metastasis via a complex formation with integrin α5β1 and matrix metalloproteinase 9. *J. Biol. Chem.* **288,** 24264–76 (2013).

63. Wang, R., Yu, C., Zhao, D., Wu, M. & Yang, Z. H. U. The mucin-type glycosylating enzyme polypeptide N-acetylgalactosaminyltransferase 14 promotes the migration of ovarian cancer by modifying mucin 13. 667–676 (2013). doi:10.3892/or.2013.2493

64. Huanna, T. *et al.* GALNT14 Mediates Tumor Invasion and Migration in Breast Cancer Cell MCF-7. **1171,** 1159–1171 (2015).

65. Bachert, C. & Linstedt, A. D. A Sensor of Protein O-Glycosylation Based on Sequential Processing in the Golgi Apparatus. **24,** (2012).

66. Song, L., Bachert, C., Schjoldager, K. T., Clausen, H. & Linstedt, A. D. Development of isoform-specific sensors of polypeptide GalNAc-transferase activity. *J. Biol. Chem.* **289,** 30556–30566 (2014).

67. Schjoldager, K. T.-B. G. *et al.* O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids. *J. Biol. Chem.* **285,** 36293–303 (2010).

68. Schjoldager, K. T.-B. G. *et al.* Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc. Natl. Acad. Sci. U.*

S. A. **109,** 9893–8 (2012).

69. Gerken, T. a. *et al.* Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J. Biol. Chem.* **286,** 14493–14507 (2011).

70. Hassan, H. *et al.* The lectin domain of UDP-N-acetyl-D-galactosamine: Polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.* **275,** 38197–38205 (2000).

71. Wandall, H. H. *et al.* The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: Lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology* **17,** 374–387 (2007).

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

73. Tagliabracci, V. S. *et al.* Dynamic regulation of FGF23 by Fam20C phosphorylation, GalNAc-T3 glycosylation, and furin proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 5520–5 (2014).

74. Egea, G. *et al.* cis -Golgi resident proteins and O -glycans are abnormally compartmentalized in the RER of colon cancer cells. **830,** 819–830 (1993).

75. Losfeld, M. E., Soncin, F., Ng, B. G., Singec, I. & Freeze, H. H. A sensitive green fluorescent protein biomarker of N-glycosylation site occupancy. *FASEB J.* **26,** 4210–4217 (2012).

76. Contessa, J. N. *et al.* Molecular imaging of N-linked glycosylation suggests glycan biosynthesis is a novel target for cancer therapy. *Clin. Cancer Res.* **16,** 3205–14 (2010).

77. Puthenveedu, M. A., Bachert, C., Puri, S., Lanni, F. & Linstedt, A. D. GM130 and GRASP65-dependent lateral cisternal fusion allows uniform Golgi-enzyme distribution. **8,** (2006).

78. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 5809–5814 (2008).

79. Kong, Y. *et al.* Probing polypeptide GalNAc-transferase isoform substrate speci fi cities by in vitro analysis. **25,** 55–65 (2015).

80. Kuan, S. F., Byrd, J. C., Basbaum, C. & Kim, Y. S. Inhibition of mucin glycosylation by aryl-N-acetyl-alpha-galactosaminides in human colon cancer cells. *J. Biol. Chem.* **264,** 19271–7 (1989).

81. Patsos, G. *et al.* O-Glycan inhibitors generate aryl-glycans, induce apoptosis and lead to growth inhibition in colorectal cancer cell lines. *Glycobiology* **19,** 382–398 (2009).

82. Kohsaki, T. *et al.* Expression of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase isozymes T1 and T2 in human colorectal cancer. *J. Gastroenterol.* **35,** 840–8 (2000).

83. Brockhausen, I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep.* **7,** 599–604 (2006).

84. Brooks, S. a., Carter, T. M., Bennett, E. P., Clausen, H. & Mandel, U. Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer. *Acta Histochem.* **109,** 273–284 (2007).

85. Rowe, P. S. N. HHS Public Access. **6020,** 64–71 (2016).

86. Goetz, R. *et al.* Isolated C-terminal tail of FGF23 alleviates hypophosphatemia by inhibiting FGF23-FGFR-Klotho complex formation. *Proc. Natl. Acad. Sci. U. S. A.* **23,** (2010).

87. Liu, S. *et al.* Pathogenic role of Fgf23 in Hyp mice. 38–49 (2006). doi:10.1152/ajpendo.00008.2006.

88. Larsson, T. *et al.* Fibroblast growth factor-23 mutants causing familial tumoral calcinosis are differentially processed. *Endocrinology* **146,** 3883–91 (2005).

89. Fukumoto, S. FGF23-FGF Receptor/Klotho Pathway as a New Drug Target for Disorders of Bone and Mineral Metabolism. *Calcif. Tissue Int.* **98,** 334–340 (2016).

90. Isakova, T. *et al.* Rationale and Approaches to Phosphate and Fibroblast Growth Factor 23 Reduction in CKD. *J. Am. Soc. Nephrol.* **26,** 2328–39 (2015).

91. Smith, E. R. The use of fibroblast growth factor 23 testing in patients with kidney disease. *Clin. J. Am. Soc. Nephrol.* **9,** 1283–1303 (2014).

92. Schjoldager, K. T. *et al.* Deconstruction of O-glycosylation — GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. **16,** 1713–1722 (2015).

93. Szász, A. M. *et al.* Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* **7,** 49322–49333 (2016).

94. Sun, N. *et al.* FGF23 neutralization improves bone quality and osseointegration of titanium implants in chronic kidney disease mice. *Sci Rep* **5,** 8304 (2015).

95. Bai, X. *et al.* CYP24 inhibition as a therapeutic target in FGF23- mediated renal phosphate wasting disorders. **126,** 667–680 (2016).

96. Ichikawa, S. *et al.* Genetic rescue of glycosylation-deficient FGF23 in the GALNT3 knockout mouse. *Endocrinology* **155,** 3891–3898 (2014).

97. Mukhopadhyay, S., Bachert, C., Smith, D. R. & Linstedt, A. D. Manganese-induced Trafficking and Turnover of the cis -Golgi Glycoprotein GPP130. **21,** 1282–1292 (2010).

98. Schjoldager, K. T., Christoffersen, C., Leguern, E., Clausen, H. & Rader, D. J. Clinical and Translational Report Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans , Nonhuman Primates , and Clinical and Translational Report Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans , Nonhuman Primates ,. 234–245 (2016).

99. Essalmani, R. *et al.* Furin is the primary in vivo convertase of angiopoietin-like 3 and endothelial lipase in hepatocytes. *J. Biol. Chem.* **288,** 26410–8 (2013).

100. Liao, W. *et al.* Expression of GALNT2 in human extravillous trophoblasts and its suppressive role in trophoblast invasion. *Placenta* **33,** 1005–1011 (2012).