
A chemical biology approach to characterise the substrates of the glycosyltransferase GalNAc-T4 in breast cancer

Beatriz Calle

A thesis submitted for the degree of Doctor of Philosophy
Department of Chemistry
Imperial College London
September 2024

ABSTRACT

O-GalNAc glycosylation is one of the most abundant post-translational modifications. Aberrant O-GalNAc glycosylation is a hallmark of breast cancer. The glycosyltransferase GalNAc-T4 is thought to be heavily implicated in breast cancer, however its exact function in breast cancer and its protein substrates remain unclear. The aim of this project was therefore to characterise the protein substrates and glycosylation sites of GalNAc-T4 in the context of breast cancer and gain insight into the role of GalNAc-T4-mediated glycosylation in breast cancer.

We first developed a bioorthogonal reporter system for GalNAc-T4 using the "bump-and-hole" (BH) approach. We demonstrated that BH GalNAc-T4 recapitulates the *in vitro* activity of the wild-type enzyme whilst displaying a switch from the native substrate to the bumped analogue. We then equipped breast cancer cells with the ability to tag the substrates of GalNAc-T4 with a chemical handle. Analysing these chemically-tagged glycoproteins by mass spectrometry (MS) allowed us to determine the substrates of GalNAc-T4 in a panel of breast cancer cell lines. We also characterised the GalNAc-T4-mediated glycosylation sites in one of the newly identified substrates by MS-glycoproteomics and postulated a possible role of GalNAc-T4-mediated glycosylation on the function of the protein. To investigate the role of GalNAc-T4 and its substrates in a biologically-relevant model system, preliminary work was also performed to extend the BH approach to an *in vivo* setting. Deciphering the identity and the function of the breast cancer relevant-GalNAc-T4 substrates and glycosylation sites may help yield new diagnostic and therapeutic options towards breast cancer.

TABLE OF CONTENTS

ABSTRACT	2
TABLE OF CONTENTS	3
STATEMENT OF ORIGINALITY	7
COPYRIGHT DECLARATION	7
ACKNOWLEDGEMENTS	8
LIST OF FIGURES	10
LIST OF TABLES	12
ABBREVIATIONS	13
INTRODUCTION	16
Introduction to breast cancer	16
Breast Anatomy	17
Biology of Breast Cancer	18
Breast Cancer Subtypes	21
Diagnostic and therapeutic options for breast cancer	23
Introduction to glycobiology	24
O-GalNAc glycosylation and the GalNAc-T family	27
Breast cancer-associated changes in glycosylation	34
GalNAc-Ts implicated in breast cancer	39
Substrate identification of individual GalNAc-Ts	49
GalNAc-T Bump-and-Hole Engineering	51
PROJECT AIMS AND OBJECTIVES	60
METHODS	61
1. Generation of GalNAc-T4 KO BCCs	61
1.1. Cloning of plasmids for CRISPR-Cas9 KO of GalNAc-T4	61
1.2. Generation of GalNAc-T4 KO BCCs by CRISPR-Cas9 genome editing	61
2. <i>In vitro</i> growth assays with WT and GalNAc-T4 KO BCCs	62
3. <i>In vivo</i> growth assays with WT and GalNAc-T4 KO BCCs	62
4. <i>In vitro</i> glycosylation experiments with WT/BH GalNAc-T4 and UDP-sugars 1-1662	62
4.1. <i>In vitro</i> glycosylation reaction with WT/BH GalNAc-T4 and UDP-sugars 1-16	62
4.2. LC-MS analysis of synthetic glycopeptides	63
5. Michaelis-Menten Kinetics	63
6. Hygromycin kill curve of BCCs	63
7. Stable transfection of BH GalNAc-T4 into BCCs	64
8. Cell surface tagging of glycoproteins in transfected BCCs and in-gel fluorescence analysis	64

9. Tagging of glycoproteins in lysates of transfected T47D cells and streptavidin blot analysis.....	65
10. Tagging of glycoproteins in the secretome of transfected T47D cells and streptavidin blot analysis.....	66
11. Label-free MS-proteomics experiment on lysates and secretome of transfected T47D cells	66
11.1. Sample preparation for label-free MS-proteomics	66
11.2. Peptide desalting using UltraMicroSpin™ columns.....	67
11.3. LC-MS/MS analysis for label-free MS-proteomics	67
11.4. Analysis of label-free MS-proteomics peptide data using MaxQuant and Perseus	68
12. SILAC-based MS-(glyco)proteomics experiments on transfected BCCs	68
12.1. SILAC metabolic labelling and compound administration	68
12.2. Sample preparation to check for SILAC heavy label incorporation	69
12.3. LC-MS/MS analysis to check for SILAC heavy label incorporation	70
12.4. Checking SILAC heavy label incorporation using MaxQuant and Perseus	70
12.5. Sample preparation for SILAC-based MS-(glyco)proteomics.....	71
12.6. Peptide fractionation using UltraMicroSpin™ columns	72
12.7. LC-MS/MS analysis for SILAC-based MS-proteomics	72
12.8. Analysis of SILAC-based MS-proteomics peptide data using MaxQuant and Perseus	73
12.9. LC-MS/MS analysis for SILAC-based MS-glycoproteomics.....	73
13. Validation of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown	74
14. Testing different biorthogonal reporter reagents to facilitate MS-glycoproteomic analyses	75
14.1. <i>In vitro</i> glycosylation reaction with WT GalNAc-T2 and UDP-sugars 1 , 3 and 11 . 75	
14.2. CuAAC reaction with reporter reagents A , D , E (by Ganka Bineva-Todd)	75
14.3. CuAAC reaction with reporter reagents B , C , F	76
14.4. Solid Phase Extraction using Strata-X columns.....	76
14.5. LC-MS/MS analysis of synthetic glycopeptides tagged with the reporter reagents A-F	76
15. Puromycin kill curve of transfected T47D cells.....	77
16. Lentiviral transduction of transfected T47D cells with tagged CD46 and GOLM1	77
17. Protein expression and Ni-NTA/GFP-affinity purification of recombinant CD46. 77	
18. MS-glycoproteomics experiment with recombinant CD46 BC and CD46 C.....	78
18.1. Sample preparation for MS-glycoproteomics with recombinant CD46 using iTag-azide for the CuAAC reaction.....	78
18.2. LC-MS/MS analysis for MS-glycoproteomics.....	79

19. Protein expression and Ni-NTA purification of recombinant CD46 BC and GOLM1	79
20. MS-glycoproteomics experiment with recombinant CD46 BC and GOLM1.....	80
20.1. Sample preparation for MS-glycoproteomics with recombinant CD46 and GOLM1 using biotin-DADPS-picolyl azide for the CuAAC reaction.....	80
20.2. Glycopeptide desalting using UltraMicroSpin™ columns	81
21. <i>In vivo</i> glycoprotein tagging with BH GalNAc-T2.....	81
21.1. Mammary fat pad tumour induction and compound administration	81
21.2. Tumour dissociation for cell isolation and protein expression control by Western blot.....	82
21.3. Tumour homogenization and glycoprotein tagging analysis by streptavidin blot ..	82
22. <i>In vivo</i> glycoprotein tagging with BH GalNAc-T4.....	83
RESULTS	84
CHAPTER 1: Investigating the role of GalNAc-T4 in breast cancer cells	84
1.1. Establishing GalNAc-T4 KO BCCs.....	84
1.2. <i>In vitro</i> growth assays of WT and GalNAc-T4 KO BCCs.....	86
1.3. <i>In vivo</i> growth assay of WT and GalNAc-T4 KO BCCs.....	88
CHAPTER 2: Developing BH GalNAc-T4 systems in breast cancer cells.....	90
2.1. <i>In vitro</i> validation of BH GalNAc-T4.....	90
2.1.1. In vitro glycosylation experiments with a panel of UDP-GalNAc analogues	90
2.1.2. Michaelis-Menten Kinetics	92
2.1.3. Competition experiment with the native substrate (UDP-GalNAc) and the bumped analogue (UDP-GalN6yne).....	96
2.2. Establishing a BH GalNAc-T4 system in BCCs	98
2.2.1. Cell surface tagging of glycoproteins in T47D ^{WT} and T47D ^{T4-KO} cells.....	98
2.2.2. Tagging of glycoproteins in lysates and secretome of transfected T47D cells	101
2.2.3. Cell surface tagging of glycoproteins in transfected MCF7 and MDA-MB-231 cells.....	102
CHAPTER 3: Characterising the substrates of GalNAc-T4 in breast cancer cells by MS-proteomics	104
3.1. Label-free MS-proteomics on lysates & secretome of transfected T47D ^{WT} cells..	104
3.2. Optimisation of MS-proteomics workflow.....	108
3.3. SILAC-based MS-(glyco)proteomics on lysates of transfected T47D ^{WT} cells.....	109
3.4. SILAC-based MS-proteomics on the secretome of transfected T47D ^{WT} cells.....	113
3.5. SILAC-based MS-proteomics on lysates of transfected MCF7 ^{WT} and MDA-MB-231 ^{WT} cells	116
3.6. Validation of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown	119
3.7. Analysing the expression of CD46, MUC1 and GOLM1 in WT and GalNAc-T4 KO BCCs	124

CHAPTER 4: Facilitating MS-glycoproteomic analyses on O-GalNAc glycoproteins using bioorthogonal reporter reagents	125
CHAPTER 5: Determining the GalNAc-T4-mediated glycosylation site(s) in CD46 and GOLM1 by MS-glycoproteomics	133
5.1. Identifying the GalNAc-T4-mediated glycosylation site(s) in CD46	135
5.1.1. Protein expression and Ni-NTA/GFP-affinity purification of recombinant CD46	137
5.1.2. Optimisation of the MS-glycoproteomics workflow for CD46	143
5.1.3. Protein expression and Ni-NTA purification of recombinant CD46 BC	146
5.1.4. MS-glycoproteomics experiment on recombinant CD46 BC	148
5.2. Identifying the GalNAc-T4-mediated glycosylation site(s) in GOLM1	150
5.2.1. Protein expression and Ni-NTA purification of recombinant GOLM1	152
5.2.2. MS-glycoproteomics experiment on recombinant GOLM1	153
5.2.3. In silico evaluation of GalNAc-T4-mediated glycosylation of GOLM1	155
CHAPTER 6: Developing BH GalNAc-T systems <i>in vivo</i>	157
6.1. Establishing a BH GalNAc-T2 system <i>in vivo</i>	157
6.2. Extending the BH GalNAc-T <i>in vivo</i> system to GalNAc-T4	160
DISCUSSION	164
CONCLUSION	177
BILBIOGRAPHY	178
SUPPLEMENTARY INFORMATION	201
List of supplementary figures	201
List of supplementary tables	202
Supplementary Figures	203
Supplementary Tables	215
Supplementary Bibliography	243

STATEMENT OF ORIGINALITY

I, Beatriz Calle, confirm that I have composed this thesis and that all the work presented in it is my own. Any published work or contributions from collaborators has been appropriately acknowledged and referenced. Any work presented in this thesis that was previously submitted for another degree at Imperial College London has been clearly stated. This work was undertaken at the Francis Crick Institute under the supervision of Dr Benjamin Schumann and Dr Ilaria Malanchi. This project was supported by a Crick-HEI studentship funded by the Department of Chemistry at Imperial College London and the Francis Crick Institute.

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

ACKNOWLEDGEMENTS

First and foremost, I wanted to thank Dr Ben Schumann and Dr Ilaria Malanchi for giving me the opportunity of doing this PhD project and allowing me to join their teams, I have learnt so much from both and have really appreciated the professional and personal support provided throughout my PhD.

I wanted to thank Dr Omur Tastan (Schumann lab) for teaching me all the skills and knowledge to help me transition from a chemist to a biologist when I first started my project. I wanted to thank Dr Anna Cioce (Schumann lab) from whom I have learnt so much throughout my project. I wanted to thank Lucia Di Vagno (Schumann lab) for all her help and input in the MS-glycoproteomics experiments. I wanted to thank Ganka Bineva-Todd (Schumann lab) for synthesising all the modified sugars required for my experiments. I wanted to thank Dr Victoria Bridgeman and Dr Tatiana Rizou from the Malanchi lab for performing the *in vivo* experiments on my behalf, as well as the Biological Research Facility at the Francis Crick Institute for their support in these experiments. I wanted to thank Dr Jane Walker (Garcia Manyes lab) and Dr Maria Benito-Jardon (Treisman lab) for their input when designing the experiments to generate the GalNAc-T4 KO cells. I wanted to thank Sandhya Sridhar (Schumann lab) for her help when generating the GalNAc-T4 KO MCF7 cells. I wanted to thank Lotte Carr and Debipriya Das from the Flow Cytometry STP at the Francis Crick Institute for sorting the GFP+ cells on my behalf when generating the GalNAc-T4 KO cells. I wanted to thank the Advanced Sequencing Facility at the Francis Crick Institute for running my samples when validating the GalNAc-T4 KO cells by amplicon sequencing. I wanted to thank Jessica Olsen from the GeMS STP at the Francis Crick Institute for showing me how to analyse the output from the amplicon sequencing with CRISPResso2. I wanted to thank Dr Ioannis Loukas, Dr Thomas Wilson, and Dr Cristina Morales from the Scaffidi lab and Dr Michael Howell from the High Throughput Screening STP at the Francis Crick Institute for their help with the *in vitro* growth assays using WT and GalNAc-T4 KO cells. I wanted to thank Chloe Roustan from the Structural Biology STP for helping me with the protein expression and purification experiments. I wanted to thank the Chemical Biology STP at the Francis Crick Institute for synthesising the peptides and glycopeptides used in my experiments. I wanted to thank Cell Services at the Francis Crick Institute for all their support throughout my project. I wanted to thank Dr Keira Mahoney and Dr Stacy Malaker from the Malaker lab, as well as Dr Tania Auchynnikava, Helen Flynn, and Mark Skehel from the Proteomics STP at the Francis Crick Institute for their help and input in the MS-glycoproteomics experiments.

I wanted to thank Dr Deborah Schneider-Luftman from the Biostatistics STP at the Francis Crick Institute for her input on how to perform the statistical analyses of the SILAC-based MS-proteomics experiments. I wanted to thank the Viral Vector Core at the Francis Crick Institute for generating all the viruses required for my experiments.

I wanted to thank every member of the Schumann lab, past and present, for being such a supportive and amazing team, particularly Sophie Schmidt, Dr William Browne and Dr Edgar Gonzalez-Rodriguez who have been there for me time and time again over the past four years.

I also wanted to thank my partner Hamish Miller, who is my unconditional support and is there for me every single day through all the highs and lows. Last but not least, I wanted to thank my parents for always believing in me – without them I wouldn't be where and who I am today.

LIST OF FIGURES

Fig. 1. Breast Anatomy.	17
Fig. 2. N- and O-GalNAc glycans.	26
Fig. 3. Glycosylation modes and substrate preferences of the GalNAc-T family	28
Fig. 4. The GalNAc-T family initiates O-GalNAc glycosylation.	29
Fig. 5. Catalytic mechanisms of the GalNAc-T family.....	31
Fig. 6. BC-associated changes in glycosylation.	35
Fig. 7. GalNAc-T BH Engineering.	52
Fig. 8. Extending the GalNAc-T BH approach to living cells.....	55
Fig. 9. UDP-GalN6yne biosynthesis using peracetylated GalNAc analogues.....	56
Fig. 10. Extending the GalNAc-T BH system to GalNAc-T4.	59
Fig. 11. Establishing GalNAc-T4 KO BCCs.....	86
Fig. 12. In vitro growth assays of WT and GalNAc-T4 KO BCCs.....	87
Fig. 13. In vivo growth assay of WT and GalNAc-T4 KO T47D cells.	88
Fig. 14. In vivo growth assay of WT and GalNAc-T4 KO MCF7 cells.....	89
Fig. 15. Endpoint <i>in vitro</i> glycosylation experiments with WT/BH GalNAc-T4 and a panel of UDP-GalNAc analogues.	92
Fig. 16. Michaelis-Menten kinetics of WT GalNAc-T4/UDP-GalNAc and BH GalNAc-T4/UDP-GalN6yne.....	95
Fig. 17. Competition experiment with WT/BH GalNAc-T4 and UDP-GalNAc/UDP-GalN6yne.	97
Fig. 18. Cell surface tagging of glycoproteins in transfected T47D ^{WT} and T47D ^{T4-KO} cells using BH GalNAc-T4	99
Fig. 19. Cell surface, lysate and secretome tagging of glycoproteins in transfected T47D ^{WT} cells using BH GalNAc-T4.....	101
Fig. 20. Cell surface tagging of glycoproteins in transfected MCF7 ^{WT} and MDA-MB-231 ^{WT} cells using BH GalNAc-T4.....	103
Fig. 21. Tracking tagged glycoproteins in lysates and secretome of transfected T47D ^{WT} cells during neutravidin bead enrichment by streptavidin blot.....	105
Fig. 22. Statistical analysis of the label-free MS-proteomics experiments on lysates and secretome of transfected T47D ^{WT} cells.....	107
Fig. 23. Testing Amicon filters as an alternative approach to protein precipitation and resuspension for sample clean-up and buffer exchange after the CuAAC reaction	108
Fig. 24. Characterising the substrates of GalNAc-T4 in lysates of transfected T47D ^{WT} cells by SILAC-based MS-proteomicst.....	111
Fig. 25. Characterising the substrates of GalNAc-T4 in the secretome of transfected T47D ^{WT} cells by SILAC-based MS-proteomics.	115
Fig. 26. Characterising the substrates of GalNAc-T4 in lysates of transfected MCF7 ^{WT} and MDA-MB-231 ^{WT} cells by SILAC-based MS-proteomics.	117
Fig. 27. Validating the results of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown.....	121
Fig. 28. Comparing the expression levels of CD46, MUC1 and GOLM1 in WT and transfected BCCs.	123
Fig. 29. Comparing the expression levels of CD46, MUC1 and GOLM1 in WT and GalNAc-T4 KO BCCs	124
Fig. 30. Nomenclature for peptide fragmentation	126
Fig. 31. Unique MS signatures of GalN6yne-modified glycopeptides derivatised with biotin-DADPS-picolyl azide.	127

Fig. 32. Testing different bioorthogonal reporter reagents to facilitate MS-glycoproteomic analyses.....	130
Fig. 33. HCD spectra of the synthetic glycopeptides derivatised with reporter reagents A-F	131
Fig. 34. ETD spectrum of the synthetic glycopeptide modified with sugar 3 and derivatised with the iTag-containing reagent B	132
Fig. 35. Function of CD46 in the innate and adaptive immune system.....	134
Fig. 36. Determining the GalNAc-T4-mediated glycosylation site(s) in CD46 and GOLM1 by MS-glycoproteomics.....	135
Fig. 37. Structure and isoforms of CD46.....	136
Fig. 38. Protein expression and Ni-NTA/GFP-affinity purification of recombinant CD46.....	138
Fig. 39. Tagging and protein expression control of recombinant CD46.....	140
Fig. 40. Glycosidase treatment of recombinant CD46	142
Fig. 41. Protein expression and Ni-NTA purification of recombinant CD46 BC.....	147
Fig. 42. Determining the GalNAc-T4-mediated glycosylation site(s) on recombinant CD46 BC by MS-glycoproteomics.....	149
Fig. 43. Structure of GOLM1	151
Fig. 44. Protein expression and Ni-NTA purification of recombinant GOLM1.	152
Fig. 45. Determining the GalNAc-T4-mediated glycosylation site(s) on recombinant GOLM1 by MS-glycoproteomics.....	154
Fig. 46. In silico evaluation of GalNAc-T4-mediated glycosylation of GOLM1	156
Fig. 47. Establishing a BH GalNAc-T2 system in vivo	158
Fig. 48. Optimisation of ex vivo tagging analysis of BH GalNAc-T2 tumours by streptavidin blot.....	160
Fig. 49. Extending the BH GalNAc-T in vivo system to GalNAc-T4	161
Fig. 50. Two possible extreme conformations of CD46 on the cell surface.....	169
Fig. 51. Oligomerisation of GOLM1.....	172
Fig. 52. Alternative splicing of the POC1B-GALNT4 gene.....	175

LIST OF TABLES

Table 1. Overview of the different BC subtypes and their molecular features.....	22
Table 2. Determining the incorporation efficiency of the heavy SILAC label on lysates of transfected T47D ^{WT} cells.....	109
Table 3. Determining the incorporation efficiency of the heavy SILAC label on the secretome of transfected T47D ^{WT} cells.....	114
Table 4. GalNAc-T4 substrates conserved between BCCs based on the SILAC-based MS-proteomics experiments.....	119
Table 5. Protein concentration and amount in each sample (50 µL) of the recombinant CD46 samples after Ni-NTA and GFP-affinity purification.....	139
Table 6. Conditions tested to maximise the coverage of CD46 for MS-glycoproteomics.....	144
Table 7. Conditions tested to maximise the coverage of CD46 for MS-glycoproteomics using the control CD46 BC construct and the CD46 BC1/BC2 FASTA files for the Byonic search.....	145

ABBREVIATIONS

Abbreviation	Definition
AmBic	Ammonium bicarbonate
BC	Breast cancer
BCC	Breast cancer cell
BH	Bump-and-hole
BH T4	Bump-and-hole GalNAc-T4
CA	Cancer antigen
CaCl ₂	Calcium chloride
Cas	CRISPR-associated endonuclease
CHX	Cycloheximide
CMP	Cytidine monophosphate
CuAAC	Copper-catalysed alkyne-azide cycloaddition
CuSO ₄	Copper sulfate
CV	Column volume
DADPS	Dialkoxydiphenylsilane
DC	Dendritic cell
DMSO	Dimethyl sulfoxide
DSB	Double strand break
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMT	Epithelial-to-mesenchymal transition
ER	Estrogen receptor
ESI	Electron spray ionisation
ETD	Electron transfer dissociation
FBS	Foetal bovine serum
FDR	False Discovery Rate
GAG	Glycosaminoglycan
Gal	Galactose
GALE	UDP-glucose 4-epimerase
GalNAc	<i>N</i> -acetylgalactosamine
GalNAc-T	<i>N</i> -acetylgalactosamine transferase
GFP	Green fluorescent protein
GFR	Growth-factor-reduced

GlcA	Glucuronic acid
GlcNAc	<i>N</i> -acetylglucosamine
gRNA	Guide RNA
HA	Hemagglutinin
HCD	Higher energy collisional dissociation
HCD-pd-ETD	Higher energy collisional dissociation product-dependent electron transfer dissociation
HER2	Human epidermal growth factor receptor-2
HS	Heparan sulfate
IAA	Iodoacetamide
IEDDA	Inverse electron demand Diels–Alder
IdoA	Iduronic acid
Indel	Insertion or deletion of nucleotides
i.p.	Intraperitoneally
i.t.	Intratumourally
iTag	Imidazolium tag
KD	Knock-down
KO	Knock-out
LC-MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
MACS	Magnetic activated cell sorting
Man	Mannose
MgCl ₂	Magnesium chloride
MGL	Macrophage galactose-specific lectin
MMP	Matrix metalloproteinase
MnCl ₂	Manganese chloride
MUC1	Mucin-1
MUC1-C	C-terminal portion of MUC1
MUC1-N	N-terminal portion of MUC1
MW	Molecular weight
m/z	Mass/charge ratio
NaCl	Sodium chloride
NahK	<i>N</i> -acetylhexosamine 1-kinase
Neu5Ac	5- <i>N</i> -acetylneuraminic acid
Neu5Gc	5- <i>N</i> -glycolylneuraminic acid
NiSO ₄	Nickel sulfate

NK	Natural killer
NSG	NOD-SCID IL2R ^{gnull}
OE	Overexpressed
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PG	Proteoglycan
POI	Protein of interest
PR	Progesterone receptor
RT	Room temperature
SCR	Short consensus repeats
SD	Standard deviation
SDS	Sodium dodecyl sulfate
sgRNA	Single guide RNA
SILAC	Stable Isotope Labelling by Amino acids in Cell culture
SLe ^x	Sialyl Lewis X
S _N i	Substitution nucleophilic internal
SpCas9	<i>Streptococcus pyogenes</i> Cas9
TEV	Tobacco etch virus
TGFBR1/2	TGF- β receptor 1/2
T _H 1	T helper 1 effector cell
TNBC	Triple-negative breast cancer
TR	Tandem repeats
Tr1	Type 1 regulatory T cell
UDP	Uridine diphosphate
VSV-G	Vesicular Stomatitis Virus Glycoprotein
VVL	<i>Vicia Villosa</i> Lectin
WT	Wild-type

INTRODUCTION

Introduction to breast cancer

Breast cancer (BC) is the most common type of cancer in women worldwide, with an estimated 2.3 million women diagnosed in 2022 and 670 000 deaths globally¹. There are multiple factors that are associated with a higher risk of BC including but not limited to age, prior history of BC, genetic predisposition, and lifestyle factors^{2, 3}. Whilst men are also at risk, women are about 100 times more likely of developing BC³, possibly due to the greater fluctuations in sex hormone levels that women experience throughout their lifetime compared to men³. The risk of developing BC increases with age², where most BC cases are diagnosed in women older than 50 years³. There is also an association between the age at diagnosis and outcome prognosis, where women diagnosed with BC under the age of 50 have lower survival rates and present more aggressive BC phenotypes compared to older patients³. Earlier age at menarche and later age at menopause have also been associated with an increased risk of BC³, likely due to the greater lifetime exposure to endogenous estrogen³. Higher levels of endogenous sex hormone levels have also been linked to a higher risk of BC³. Furthermore, the risk of BC has been associated with the timing and duration of exogenous hormone exposure, where the risk is higher in women who start combined estrogen and progestin hormone replacement therapy closer to menopause and use it for a prolonged duration (>5 years)².

A personal history of BC is associated with a higher risk of developing a second BC in the same or opposite breast^{2, 3}. Furthermore, approximately 20-25% of BC patients have a positive family history² - the more first-degree relatives diagnosed with BC, the higher the risk of developing the disease^{2, 3}. The association between a positive family history and increased BC risk is linked to heritable genetic abnormalities³, with mutations in genes such as *BRCA1*, *BRCA2*, *TP53*, *PTEN*, and *CDH1*, among others, conferring a 40-85% lifetime risk of developing BC².

Lifestyle factors such as alcohol consumption and physical activity have also been associated with BC susceptibility³. Obesity (specifically in post-menopausal women), alcohol consumption and smoking have been associated with an increased risk of BC^{2, 3}. In contrast, consistent physical activity reduces the risk of BC^{2, 3}.

Breast Anatomy

The breast is primarily composed of three tissue types: glandular epithelium, fibrous stroma and adipose tissue (**Fig. 1**)^{4, 5}. The glandular epithelium comprises 10-15% of the breast and consists of 15-20 lobes which are connected to the nipple via mammary ducts^{4, 5}. Each lobe is composed of 20-40 lobules which are the source of milk production⁴. The epithelium of the duct and lobular system consists of two layers: an inner epithelial (luminal) cell layer and an outer myoepithelial (basal) cell layer (**Fig. 1**)⁶. The myoepithelial cells serve to propel milk forward through the ducts and are surrounded by a basement membrane which separates the epithelial and stromal compartments^{4, 5}. The fibrous stroma contains infiltrating immune cells, fibroblasts, and vascular endothelial cells (**Fig. 1**)^{5, 7}. The extracellular matrix of the stroma is composed of proteins secreted by epithelial and stromal cells such as collagens, fibronectins and proteoglycans⁵. The remainder of the breast consists of adipose tissue⁴. The breast is supplied by blood vessels, lymphatic vessels and nerves^{4, 5}. The majority of the lymphatics from the breast drain to the lymph nodes within the axilla^{4, 5}.

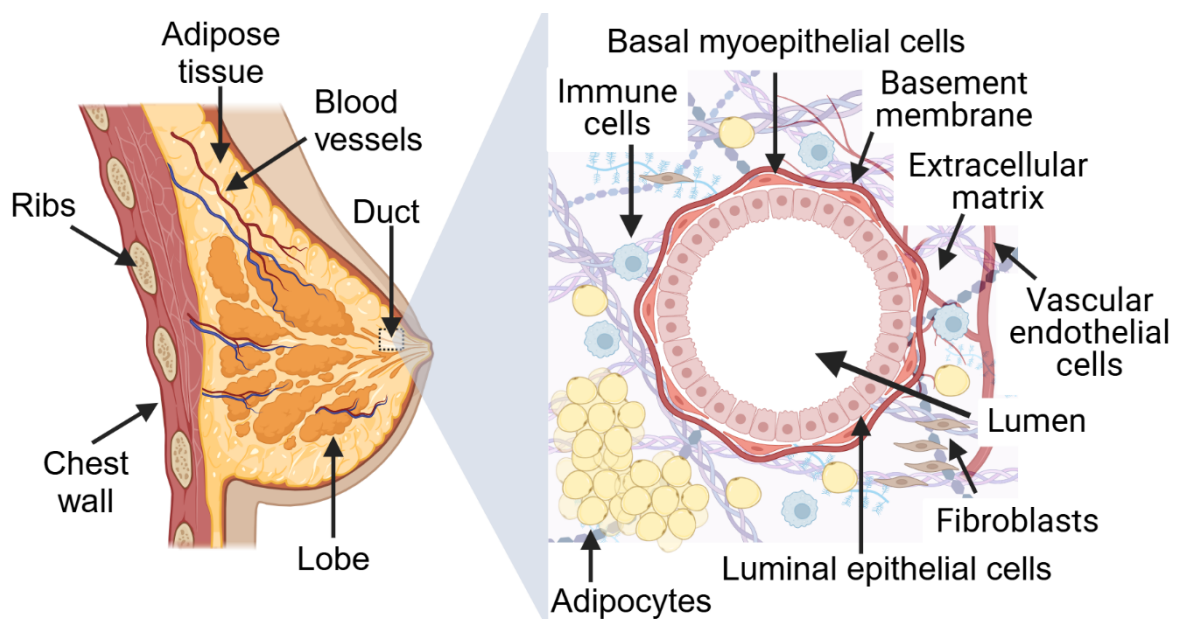


Fig. 1. Breast Anatomy. Figure adapted from De Silva et al. (2022)⁵. Figure created with BioRender.com.

Biology of Breast Cancer

Breast tumours develop from benign to malignant growths by the gradual accumulation of genetic and epigenetic changes in a population of cells alongside clonal expansion and selection⁸. The glandular epithelium can initially undergo benign proliferative changes, where the cells increase in number but retain the overall ability to assemble into epithelial structures – these growths are termed hyperplastic^{7, 9}. Hyperplastic growths can become dysplastic as the cells become morphologically abnormal⁹, and later progress into *in situ* carcinomas as the number of myoepithelial cells decreases, the number of stromal cells increases, and the basement membrane starts to be degraded⁷. If the cells breach the basement membrane and invade the surrounding stroma, the growth is reclassified as malignant^{9, 10}. These cancer cells may then intravasate lymphatic or blood vessels and gain access to the general circulation, allowing them to metastasise into distant tissues¹⁰.

Cancer cells develop driver mutations which are critical for cancer progression and provide a clonal selectivity advantage, in addition to passenger mutations which are accidental byproducts of their genetic instability^{8, 11}. Mutations in the genes *AKT1*, *AKT2*, *BRCA1*, *CCND1*, *CDH1*, *MAP3K1*, *MYC*, *PIK3CA*, *PTEN*, *RB1*, and *TP53* amongst others are known to be drivers of BC¹¹⁻¹³. Epigenetic changes occur in parallel with genetic alterations to promote tumorigenesis¹¹. In cancer cells tumour suppressor genes become packaged into heterochromatin and/or the CpG sequences in their promoters become methylated, ultimately leading to gene silencing, whilst the opposite mechanisms occur with oncogenes^{8, 11}.

Germ-line mutations in *BRCA1* and *BRCA2* account for up to 10% of all BC cases and approximately half of all hereditary BCs^{11, 14}. These genes are implicated in the repair of DNA double-strand breaks (DSBs) by homologous recombination and hence are important in the maintenance of genomic stability⁸. Cells lacking *BRCA1* or *BRCA2* function cannot repair DSBs by homologous recombination and instead exploit the error-prone non-homologous end joining mechanism to repair DSBs, which results in genomic instability¹¹. Mutations and deletions of the *TP53* tumour suppressor gene are also frequently observed in BC^{11, 15}. This gene encodes the p53 protein, a transcription regulator which causes cell cycle arrest or apoptosis in response to cellular stress or DNA damage^{8, 15}. Inactivating p53 therefore enables cancer cells to survive and proliferate in the presence of cellular stress and/or DNA damage^{8, 11, 15}. Amplification of the genes *CCND1* and *MYC* and mutations of the tumour suppressor gene *RB1* are also associated with BC¹¹⁻¹³. These genes encode the G1/S-specific cyclin D1, the transcription factor Myc, and the retinoblastoma-associated protein, respectively, all of which are involved in regulating cell cycle entry and cell growth^{16, 17}.

Deregulation of growth factor receptor activity can lead to oncogenic signalling and drive carcinogenesis^{11, 18}. In healthy cells, the binding of an extracellular ligand to its cognate receptor activates one or more intracellular signalling pathways which ultimately affects the behaviour of the cell^{18, 19}. In cancer cells, growth factor receptor pathways may be constitutively activated by mutations in the genes encoding the receptors, leading to ligand-independent signalling, or by overexpression of the receptors and/or the ligands^{11, 18}. The human epidermal growth factor receptor-2 (HER2) is a tyrosine kinase receptor that regulates epithelial cell growth, differentiation, and survival²⁰. HER2 is encoded by the gene *ERBB2* and it is overexpressed in 30% of BCs^{11, 18, 21}. HER2 has no known ligand^{18, 22}; nevertheless, it can heterodimerize with the other members of the HER family (HER1/EGFR, HER3, and HER4), leading to the activation of downstream signalling pathways such as the PI3K/Akt/mTOR and RAS/RAF/MEK/MAPK pathways^{11, 22}. Activating mutations in *AKT1*, *AKT2*, *MAP3K1*, and *PIK3CA*, as well as inactivation of the tumour suppressor gene *PTEN* by mutation or deletion, are commonly observed in BC and result in aberrant signalling¹¹.

Hyperactivation of the PI3K/Akt/mTOR signalling pathway can also cause changes in glucose metabolism in cancer cells⁸. In healthy nonproliferating cells with access to oxygen, glucose is imported into the cytoplasm and broken down into pyruvate through glycolysis, generating 2 ATP molecules in the process^{9, 23}. The pyruvate is then transported into the mitochondria and decarboxylated into acetyl-CoA by the pyruvate dehydrogenase complex before entering the citric acid cycle^{9, 23}. The citric acid cycle produces NADH, FADH₂ and CO₂ as a waste product²³. The NADH and FADH₂ are then used as electron donors to generate ATP by oxidative phosphorylation²³. The complete oxidation of a molecule of glucose through this process can generate up to 36 ATP molecules⁹. Cancer cells increase their glucose uptake, but only a small proportion of the imported glucose is used for ATP production via oxidative phosphorylation, even in the presence of abundant oxygen^{8, 9}. The glucose is instead processed through glycolysis, producing lactate which is secreted by the cells^{8, 9}. The remaining carbon atoms derived from glucose are used for the biosynthesis of proteins, nucleic acids, and lipids required for tumour growth^{8, 9}. This switch in oxidative phosphorylation to aerobic glycolysis is termed the Warburg effect^{8, 9, 24}.

Changes in hormone metabolism are also commonly observed in BC¹¹. The glandular epithelium is highly sensitive to fluctuations in hormone levels, in particular estrogen and progesterone which stimulate the growth and differentiation of mammary epithelial cells^{5, 22}. Both hormones mediate their effects by binding to their cognate receptors – the estrogen receptor (ER) and the progesterone receptor (PR), respectively²². Both receptors belong to the family of nuclear receptors which function as transcription factors^{18, 22}. These receptors may reside in the cytoplasm until ligand binding triggers their import into the nucleus, or they may be constitutively bound to DNA even in the absence of their ligand^{18, 19}. The hydrophobicity and small size of estrogen and progesterone allows them to readily diffuse across the plasma membrane into the cell, where they bind to their receptor¹⁹. Ligand binding induces a conformational change in the receptor, causing the dissociation of co-repressors and the binding of co-activators^{18, 19}. The ligand-bound receptor can then dimerise and bind to DNA sequences called hormone response elements to regulate the expression of target genes^{11, 18, 22}. Two isoforms of the ER exist that are products of two different genes: ER α and ER β ^{22, 25}. Estrogen binds to both isoforms, but its affinity for ER α is 10 times higher than for ER β ²⁵. The effect of estrogen on the growth of mammary epithelial cells is primarily mediated by ER α ²². ER β is thought to reduce the cell proliferation induced by ER α and inhibit ER α -mediated transcription^{22, 25}. The ratio of ER α /ER β is higher in BC cells (BCCs) compared to healthy breast tissue²⁵. Mutations in the *ESR1* gene, which encodes the ER α , have also been reported in BC patients, resulting in hypersensitivity to the growth-promoting effects of estrogen or ligand-independent ER signalling²².

Metastatic BC is the most common cause of BC-related mortality^{10, 26}. The initial step in metastasis, local invasiveness, involves significant changes in the phenotype of cancer cells within the primary tumour^{8, 10}. The epithelial cells undergo changes in their shape and gene expression profiles to become less adhesive and more migratory through the process termed epithelial-to-mesenchymal transition (EMT)^{8, 10}. During EMT, epithelial cells lose their characteristic morphology and the expression of epithelial genes such as E-cadherin and cytokeratins, acquiring instead a fibroblast-like shape and the expression of genes typical of mesenchymal cells, such as vimentin, fibronectin, matrix metalloproteinases (MMPs), and N-cadherin^{8, 10}. E-cadherin is normally expressed on the plasma membrane of epithelial cells, where it plays a key role in cell-cell adhesion by binding epithelial cells together through adherens junctions^{8, 10}. Mutations, deletions, and methylation of the *CDH1* gene, which encodes E-cadherin, are frequently observed in BC and result in the suppression of E-cadherin expression^{10, 27}. The switch from E-cadherin to N-cadherin expression in BCCs that have undergone EMT facilitates invasiveness through homophilic interactions with N-cadherin displayed on stromal cells¹⁰.

Breast Cancer Subtypes

BC can be classified into different subtypes depending on multiple factors including histopathologic analysis, receptor status and gene expression profiling¹³. Identifying the BC subtype can help establish the most suitable treatment as well as predicting the clinical course including the probability of developing distant metastasis, risk of relapse, and overall patient survival²⁸. Histopathologic analysis reveals the tumour type, stage, and grade of breast tumours²⁹. Depending on whether the tumour is restricted to the ductal-lobular system or not, BC can be histopathologically classified as either an *in situ* or invasive carcinoma³⁰. Most breast tumours originate in the lining of the ducts, with 70-80% of histological subtypes being invasive ductal carcinomas^{13, 31}. The second most common histological subtype is invasive lobular carcinomas (around 15%), whilst others such as medullary, mucinous, or tubular subtypes are less frequent^{13, 32}. The BC stage reflects the degree of tumour spread, where Stage 0 is pre-invasive BC and Stage 4 is metastatic BC³³. The overall survival rate is inversely correlated with the tumour stage²⁹. The BC grade reflects the degree of differentiation of cancer cells – the higher the grade, the lower the degree of differentiation²⁹. Classifying BC based on receptor status depends on the expression of three receptors: the ER, the PR, and HER2¹³. More than 75% of BC cases express the ER and/or the PR, while HER2 positivity is detected in 10-15%¹³. If none of these receptors are expressed, the tumour is classified as triple-negative BC (TNBC)^{11, 28}.

Gene expression profiling of BC samples initially revealed five subclusters termed the Intrinsic Subtypes¹³. Luminal A and luminal B tumours are ER+ and PR+ and are characterised by the expression of luminal genes such as *GATA3*, *FOXA1* and *XBP1*^{34, 35}. Luminal A tumours express higher levels of the hormone receptors (especially the PR) and luminal genes than Luminal B tumours³⁴. In contrast, Luminal B tumours are more proliferative (higher Ki-67 status), display a higher expression of genes involved in mitosis and proliferation, and are clinically more aggressive than Luminal A tumours^{13, 34}. HER2-enriched tumours are characterised by a high expression of HER2-related genes and are frequently higher-grade tumours³⁴. TNBC tumours lack the expression of the ER/PR/HER2^{11, 28}. These tumours often express genes which are typical of myoepithelial/basal cells and are therefore commonly referred to as basal-like tumours¹³. Normal-like tumours are characterised by a gene expression pattern similar to healthy breast tissue¹³.

A new classification of BC subtypes was later developed based on the finding that a significant portion of the gene expression landscape is dominated by recurrent copy number alterations that drive expression in *cis*, meaning on the same chromosome as the alteration itself^{12, 13}. These *cis*-driven genes were used to cluster breast tumours into 10 Integrative Clusters^{12, 13}. **Table 1** correlates the Intrinsic Subtypes and the Integrative Clusters classifications and summarises some of the key molecular features of each¹¹.

Table 1. Overview of the different BC subtypes and their molecular features. Table adapted from Gedik et al. (2019)¹¹.

Intrinsic subtype	ER status	HER2 status	Ki-67 status	Molecular features	Integrative Cluster(s)
Luminal A	ER+	HER2-	Low	Mutations in <i>GATA3</i> , <i>FOXA1</i> , <i>MAP3K1</i> and <i>PIK3CA</i> High <i>ESR1</i> expression Gain of chromosomes 1q and 8q Loss of chromosomes 8p and 16q	2, 3, 4, 7, 8
Luminal B	ER+	HER2±	High	Mutations in <i>PIK3CA</i> and <i>TP53</i> Amplification of <i>CCDN1</i> Enhanced genomic instability Focal amplifications (e.g., 8p12, 11q13)	1, 2, 6, 9
HER2	ER-	HER2+	High	Mutations in <i>PIK3CA</i> and <i>TP53</i> Amplification of <i>CCDN1</i> and <i>ERBB2</i> High EGFR expression High genomic instability	5
Basal-like/ TNBC	ER-	HER2-	High	Mutations in <i>TP53</i> Loss of <i>BRCA1</i> and <i>RB1</i> High expression of DNA repair proteins High genomic instability Focal amplifications (e.g., 8q24)	4, 10

BCC lines are commonly used for BC modelling *in vitro* and *in vivo*. Cell lines such as T47D, MCF7 and MDA-MB-231 cells account for two-thirds of BCCs used for this purpose³⁶. Different cell lines can be used to model different BC subtypes. For example, the T47D and MCF7 cell lines display a Luminal A phenotype, whilst MDA-MB-231 cells display a TNBC phenotype³⁶.

Diagnostic and therapeutic options for breast cancer

Early detection of BC is linked to a better prognosis and lower mortality rates³⁷. Breast screening and increased BC awareness facilitates early detection, leading to better outcomes³⁷. Mammography continues to be the standard screening method, and it plays a key role in detecting clinically unsuspected BC³⁸. Additional imaging methods, such as ultrasound and magnetic resonance imaging, can aid in cancer detection and are recommended for women at higher risk of BC³⁸. Breast lesions identified through physical examination or imaging are biopsied and examined by a pathologist, who then determines if further tests are necessary³⁸.

Liquid biopsies can also be used to detect cancer-related factors in the bloodstream, including circulating tumour cells, tumour DNA fragments, and tumour markers, which are released from both primary and metastatic tumour sites³⁹. Two serum tumour markers commonly used in the clinic to monitor BC progression and response to treatment are the cancer antigen 15-3 (CA 15-3) and the CA 27-29^{7, 40-43}. Both markers detect different epitopes within the Mucin-1 (MUC1) protein⁴⁰⁻⁴³. MUC1 is a member of the mucin family of glycoproteins which is normally expressed in epithelial cells of the mammary gland and is overexpressed (OE) and aberrantly glycosylated in more than 90% of BCs⁴⁴. The most widely used CA 15-3 assay uses two monoclonal antibodies: the DF3 antibody, which binds to the TRPAPGS epitope, and the 115D8 antibody, whose epitope is not fully mapped but is believed to target the glycan chains on MUC1, as its binding is disrupted by desialylation^{40, 43, 45-48}. The CA 27-29 assay uses the monoclonal antibody B27.29, which targets the PDTRPAP epitope^{40, 43, 46, 49}. Nevertheless, these biomarkers have limited sensitivity and specificity, as elevated levels can also be observed in other conditions^{7, 40}.

Surgery, either in the form of lumpectomy or mastectomy, is a crucial treatment for early-stage BC^{26, 50}. Adjuvant therapy, including chemotherapy and/or radiotherapy, is designed to eradicate any remaining BCCs, thereby reducing the risk of post-surgical recurrence^{26, 51}. If there is evidence that the cancer has spread to the lymph nodes, partial or complete lymph node removal may also be performed⁴. The choice of treatment depends on the receptor status, lymph node involvement and the size of the tumour². BRCA-deficient BCCs cannot repair DSBs in DNA and are therefore especially sensitive to DNA-damaging agents such as platinum-based chemotherapeutics^{2, 14}. Furthermore, tumours that have lost BRCA1 or BRCA2 function depend on the enzyme PARP to repair single-strand breaks in DNA and are hence highly sensitive to PARP inhibition^{2, 8, 14}.

Since over 75% of BC cases are hormone receptor-positive¹³, these tumours can be treated with hormonal therapies²². Selective estrogen receptor modulators (SERMs) are ER α antagonists that promote the recruitment of co-repressors, hence inhibiting ER signalling and the expression of target genes^{18, 52}. Selective estrogen receptor degraders (SERDs) induce polyubiquitination of the ER α and degradation by the ubiquitin-proteasome system⁵². Inhibition of the enzyme aromatase, which catalyses the conversion of androgens to estrogens, reduces estrogen production and thereby inhibits ER signalling^{2, 11, 52}. Tumours that are HER2 positive can be treated with anti-HER2 agents or tyrosine kinase inhibitors^{11, 22}. Cancer immunotherapy aims to leverage the immune system to target and eliminate cancer cells⁸. One approach involves using monoclonal antibodies that target BC-specific antigens to induce BCC apoptosis, promote macrophage-mediated phagocytosis, or suppress target activity^{8, 53}. An alternative strategy involves blocking the inhibitory receptors PD-1 and CTLA-4 on T cells, thereby enhancing T cell activity and anti-tumour response⁸. However, the development of resistance to initially effective therapies remains a significant clinical challenge⁵⁴.

Introduction to glycobiology

Glycobiology studies the biosynthesis, structure, and biology of glycans (also known as saccharides, sugar chains or carbohydrates)⁵⁵. Every single cell is covered by a dense glycan coating known as the glycocalyx.⁵⁵ Their presence on the cell surface allows glycans to mediate and modulate a variety of cell-cell and cell-matrix interactions, which are crucial in several processes such as cell adhesion, intracellular signalling and immune cell recognition⁵⁶. Monosaccharides are the simplest structural units of glycans⁵⁷. Glycans can exist as freestanding entities or as glycoconjugates, where they are covalently attached by glycosidic bonds to other biomolecules such as amino acids within proteins to form glycoproteins⁵⁵. Other major classes of glycoconjugates exist such as proteoglycans (PGs) and glycolipids²⁴. PGs consist of one or more glycosaminoglycan (GAG) chains attached to a core protein⁵⁸. GAGs are linear polysaccharide chains of disaccharide repeating units consisting of an amino sugar (such as *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine (GalNAc)) and galactose or a uronic acid (glucuronic acid (GlcA) or iduronic acid (IdoA))^{55, 58}. Glycolipids consist of a glycan linked to a lipid, and they can be neutral or anionic⁵⁵. A ganglioside is an anionic glycolipid carrying one or more sialic acid residues^{24, 55}.

Glycosidic bonds exist as α - or β -linkages depending on the configuration at the anomeric carbon and the stereogenic centre furthest from the anomeric carbon: α - (same configuration) and β - (different configuration) anomers^{55, 57}. The two linkages confer very different structural and biological properties to the overall sequence⁵⁵. The most frequent glycan-protein linkages are N- and O-linkages⁵⁵. In N-glycans, a GlcNAc unit forms a β -linkage with an Asn residue in a polypeptide chain (**Fig. 2A**)⁵⁹. N-glycans occur at the consensus peptide sequence Asn-X-Ser/Thr (where “X” is any amino acid except Pro) and they share a common core sequence consisting of two GlcNAc units followed by three mannose (Man) residues (**Fig. 2B**)⁵⁹. The core sequence can then be further elaborated to form three main classes of N-glycans: oligomannose (or high-mannose) type, complex type, or hybrid type (**Fig. 2B**)⁵⁵.

Several types of O-glycans exist, with the most common being O-GalNAc and O-GlcNAc glycans⁵⁵. In O-GalNAc glycans, a GalNAc unit forms an α -linkage to Ser or Thr residues in a polypeptide chain (**Fig. 2C**)⁵⁵. The initiating GalNAc unit can be extended to form one of four common core structures (**Fig. 2D**), and these can be further elongated to form linear or branched chains resembling those on N-glycans⁶⁰. Additional core structures (cores 5-8) exist, however these are rare and have restricted occurrence and low abundance⁶¹. In O-GlcNAc glycans, a GlcNAc unit forms a β -linkage to Ser or Thr residues in a polypeptide chain⁶². The GlcNAc unit in O-GlcNAc glycans is generally not elongated or modified⁶².

In eukaryotic cells, glycosylation of proteins usually takes place in the endoplasmic reticulum and Golgi apparatus⁶³. Glycosyltransferases catalyse glycosylation reactions using activated forms of monosaccharides (often nucleotide sugars) as substrates⁶⁴. Monosaccharide precursors can be salvaged from degraded glycans and imported into the cell via sugar transporters as nucleotide sugars⁶³. Most glycan degradation takes place within lysosomes, and the monosaccharide units are then exported into the cytoplasm via lysosomal carriers so they can be reused⁶³. Monosaccharide precursors can be alternatively created enzymatically from other sugars within the cell, where glucose can be directly or indirectly converted to all other sugars⁶³. Monosaccharide activation and interconversions normally take place within the cytoplasm of cells⁶³. The nucleotide sugars are subsequently transported into the lumen of the endoplasmic reticulum and Golgi via nucleotide sugar transporters⁶³.

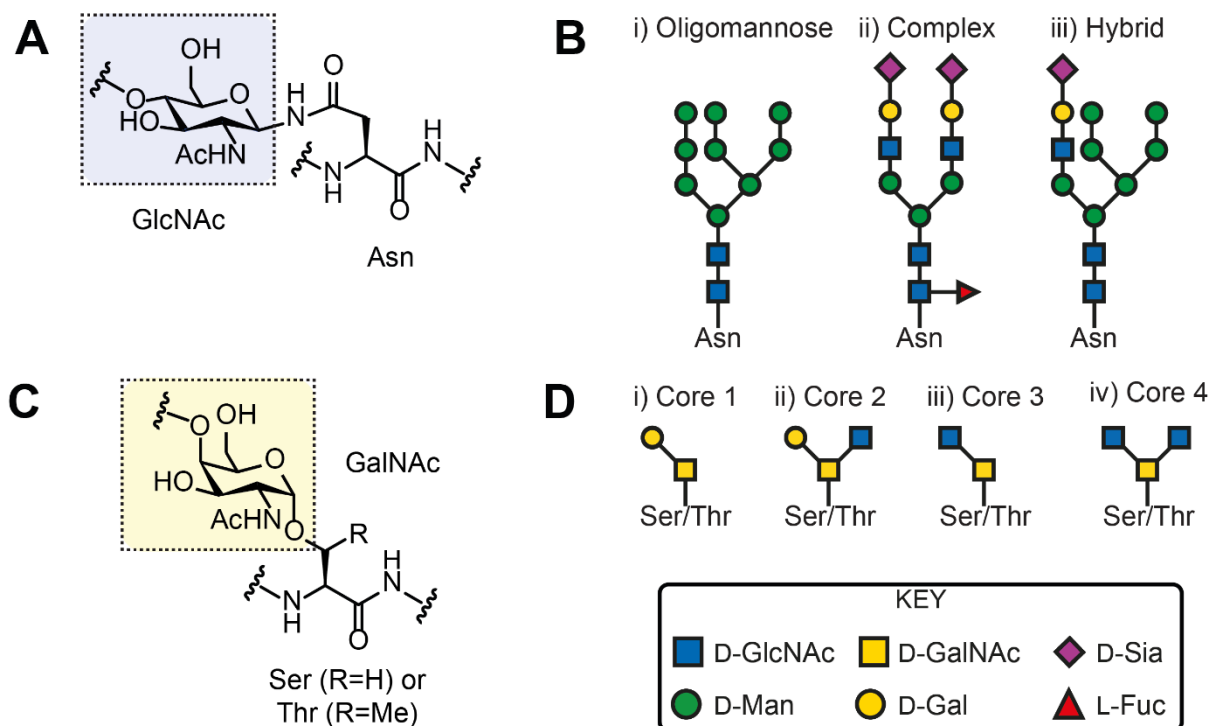


Fig. 2. N- and O-GalNAc glycans. **A)** In N-glycans, a GlcNAc unit forms a β -linkage with an Asn residue. **B)** Main classes of N-glycans. **C)** In O-GalNAc glycans, a GalNAc unit forms an α -linkage to Ser or Thr residues. **D)** Common core O-GalNAc structures.

Newly synthesised proteins contain signal sequences (often at the N-terminus of the polypeptide chain) which specify a particular destination within the cell⁶⁵. Proteins destined for the secretory pathway will contain a signal sequence which will target them to the endoplasmic reticulum, where they will be co-translationally translocated and folded before being trafficked to the Golgi and on to target membranes⁶⁴. Proteins can be co-translationally or post-translationally glycosylated at various stages in the secretory pathway⁵⁵. Cytosolic proteins lack a signal sequence and therefore remain in the cytosol after being synthesised⁶⁵. O-GlcNAc glycosylation is distinct from other types of glycosylation in that it predominantly occurs in the nucleocytoplasm and mitochondria of cells⁶². Furthermore, whilst glycans from the secretory pathway show a relatively slow turnover⁵⁵, the O-GlcNAc modification is more dynamic and O-GlcNAc glycans are added and removed multiple times throughout a protein's lifetime⁶².

O-GalNAc glycosylation and the GalNAc-T family

The first step of O-GalNAc glycosylation is catalysed by the family of polypeptide GalNAc transferases (GalNAc-Ts), where GalNAc is transferred from uridine diphosphate (UDP)-GalNAc to Ser or Thr residues in acceptor proteins^{60, 66}. The human genome encodes 20 GalNAc-Ts, making the GalNAc-T family one of the largest glycosyltransferase families⁶⁷. These enzymes are Golgi-resident type II transmembrane proteins⁶⁸. All GalNAc-Ts except GalNAc-T20 have a sugar-binding lectin domain at the C-terminus connected to the catalytic domain by an interdomain flexible linker (**Fig. 3A**)⁶⁸. To this date, the peptide sequence Ser/Thr-Pro-X-Pro (where X stands for a small hydrophobic residue such as Gly or Ala) is the only motif that most GalNAc-Ts seem to glycosylate^{68, 69}. This selectivity is conferred by three highly conserved hydrophobic Phe and Trp residues in the peptide-binding site within the catalytic domain⁶⁸. GalNAc-Ts lacking these conserved residues, such as GalNAc-T7 and -T10, do not display this substrate preference^{68, 69}. Similarly, when these residues were mutated to Ala in GalNAc-T2, the activity of the enzyme was practically eliminated⁷⁰.

Despite sharing a degree of substrate redundancy, there is increasing evidence that different members of the GalNAc-T family have distinct glycosylation preferences (**Fig. 3B**), where some GalNAc-Ts preferentially glycosylate “naked” unglycosylated peptides (e.g. GalNAc-T2) whilst others selectively (e.g. GalNAc-T4) or exclusively (e.g. GalNAc-T10) target partially glycosylated substrates⁶⁸. This distinction suggests that GalNAc-Ts may be organised as an “assembly line”^{55, 60, 66} and has led to the classification of GalNAc-Ts as early, intermediate or late glycosyltransferases (**Fig. 4**)⁶⁸. Recent findings have also demonstrated that GalNAc-Ts have different substrate charge specificities determined by each enzyme’s surface electrostatic charge distribution (**Fig. 3B**)⁷¹, adding an additional layer of specificity to the distinct substrate preferences of the different members of the GalNAc-T family.

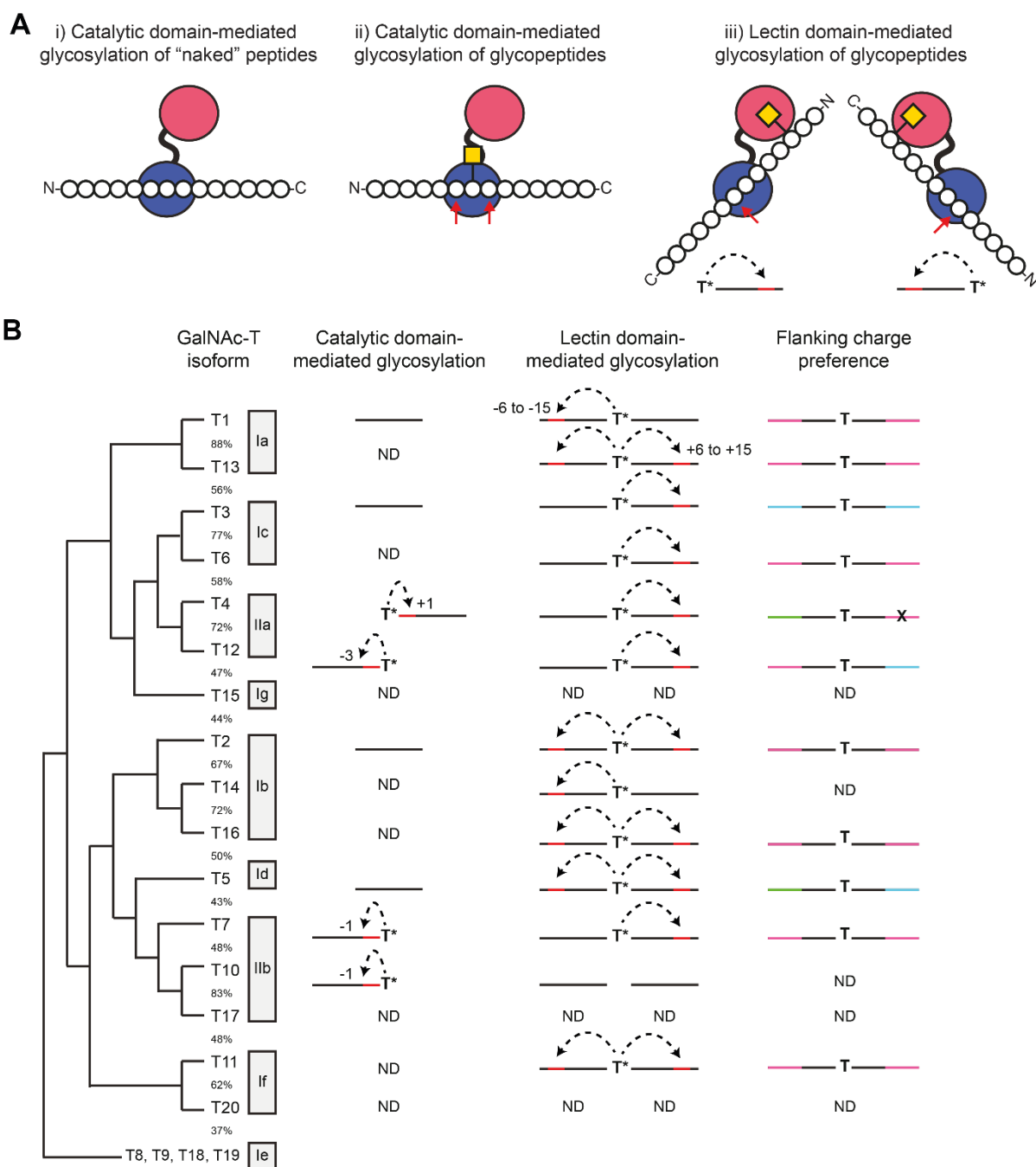


Fig. 3. Glycosylation modes and substrate preferences of the GalNAc-T family. **A)** Schematic representation of the three different glycosylation modes of GalNAc-Ts. The catalytic and lectin domains are represented as oval-shaped figures in blue and red respectively. The red arrows represent the acceptor site. Glycosylation of glycopeptides can occur in an N- or C-terminal direction. Figure adapted from de las Rivas et al. (2019)⁶⁸. **B)** Phylogenetic tree of GalNAc-Ts showing their prior glycosylation and flanking charge substrate preferences. The first column shows the GalNAc-T subfamilies and the percentage sequence identities⁷². The second and third columns show the prior glycosylation substrate preferences, where T* represents the position of the prior GalNAc unit, the red line represents the acceptor site, and the arrows represent the direction of glycosylation. The fourth column shows the flanking charge substrate preference, where pink, green, and blue represent negative, neutral, and positive flanking residues respectively. “X” illustrates the inverse preference of GalNAc-T4 with peptides with a C-terminal negative charge due to its weakly negative electrostatic surface⁷¹. “ND” stands for “not determined”. “-----” indicates no or weak activity. GalNAc-T8, -T9, -T18 and -T19 are not well characterised but exhibit negligible *in vitro* activities with most substrates, possibly due to mutations in the UDP-GalNAc binding pocket⁷³. Figure adapted from Ballard et al. (2023)⁷¹.

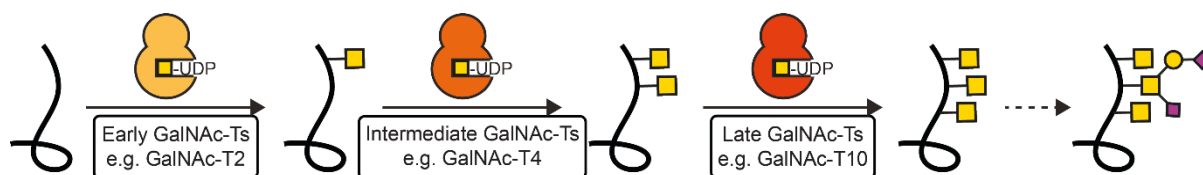


Fig. 4. The GalNAc-T family initiates O-GalNAc glycosylation. Depending on their substrate preference, GalNAc-Ts can be classified as early, intermediate, or late GalNAc-Ts. The initiating GalNAc unit can then be further glycosylated by downstream glycosyltransferases.

Glycosylation of glycopeptides by GalNAc-Ts can occur in one of two ways: 1) catalytic domain-dependent glycosylation, where acceptor sites are located 1-3 residues away from prior GalNAc units; and 2) lectin domain-dependent glycosylation, where binding of remote GalNAc units by the lectin domain directs distant acceptor sites (6-17 residues away) to the catalytic domain for glycosylation (**Fig. 3A**)⁶⁸. Both glycosylation modes can operate in an N- or C-terminal direction depending on the GalNAc-T (**Fig. 3A**)⁶⁸. The interdomain flexible linker (which varies in length and sequence between GalNAc-Ts) determines the relative orientation of the catalytic and lectin domains, hence contributing to the distinct lectin domain-mediated glycosylation preferences observed between the different isoenzymes (**Fig. 3B**)^{68, 74}.

The catalytic domain of all GalNAc-Ts adopts a GT-A fold⁶⁰ which is typical in most nucleotide binding proteins and is characterised by alternating α -helices and β -sheets⁷⁵. For optimal UDP-GalNAc binding, a metal ion (often Mn^{2+}) is required⁷⁶. All GalNAc-Ts contain conserved Asp and His residues (DXH motif) responsible for divalent cation binding⁷⁷. The coordination of the metal is octahedral, binding two phosphate oxygen atoms from UDP, the DXH motif, a His residue which is conserved in all GalNAc-Ts, and one water molecule⁷⁸. UDP-GalNAc is stabilised by multiple hydrogen bonds between amino acids located in the catalytic domain of GalNAc-Ts^{79, 80}. These amino acids are highly conserved in all GalNAc-Ts⁷⁹. The NH group of the acetamide moiety also forms hydrogen bonds with a β -phosphate oxygen atom from UDP, which may explain why GalNAc-Ts show a strong preference towards UDP-GalNAc instead of UDP-galactose (UDP-Gal)⁸⁰. GalNAc-T4, -T7, -T10 and -T12 were predicted to contain an additional GalNAc-binding site in the catalytic domain to account for their ability to glycosylate acceptor sites close to prior GalNAc units⁸¹. A crystal structure of GalNAc-T4 with a diglycopeptide revealed that there was no discernible pocket for binding of GalNAc - instead binding was mediated by hydrogen bonds and hydrophobic interactions between the GalNAc unit and residues that are not conserved in all GalNAc-Ts⁶⁸.

Binding of the peptide substrates to the catalytic domain of GalNAc-Ts is mediated by direct or water-mediated hydrogen bonds and hydrophobic interactions⁶⁸. Crystal structures of GalNAc-Ts bound to their peptide substrates revealed that the Thr acceptor sites were in a hydrophobic pocket formed by conserved Trp and Phe residues, which may explain why GalNAc-Ts often prefer to glycosylate Thr over Ser residues⁶⁸. These structures also revealed that the hydroxy group of the Thr residue was close to the anomeric carbon of the GalNAc moiety and to two of the phosphate oxygen atoms from UDP, and that there was no obvious nucleophilic residue in the active site of the enzyme on the opposite face of the ring⁸⁰. All these features were consistent with a front-faced substitution nucleophilic internal (S_Ni) mechanism, which was subsequently confirmed by computational simulations - a phosphate oxygen atom from UDP accepts a proton from the Thr hydroxy group and guides the nucleophilic residue to the same face of the anomeric carbon, resulting in retention of the anomeric configuration upon formation of the glycosidic bond (**Fig. 5A**)⁸⁰. The hydrogen bond between the NH group of the acetamide moiety and the β -phosphate oxygen atom from UDP helps stabilise the negative charge being developed at the phosphate group and helps facilitate the departure of the UDP leaving group⁸⁰.

At the surface of the UDP-GalNAc binding site within the catalytic domain there is a dynamic flexible loop which can adopt one of two conformations: an open conformation (inactive form) or a closed conformation (active form)⁸⁰. The Michaelis complexes of GalNAc-T2 revealed that GalNAc-Ts possess a Bi-Bi kinetic mechanism, where the flexible loop adopts open and closed conformations until UDP-GalNAc binds, which drives the loop to adopt a close conformation (**Fig. 5B**)⁸⁰. This conformation is stabilised by the presence of UDP-GalNAc, it excludes bulk water from the active site and facilitates binding of the peptide substrate⁸⁰. Once glycosylation is completed, the glycopeptide product departs, the loop adopts an open conformation and UDP is released, allowing the catalytic cycle to restart (**Fig. 5B**)⁸⁰.

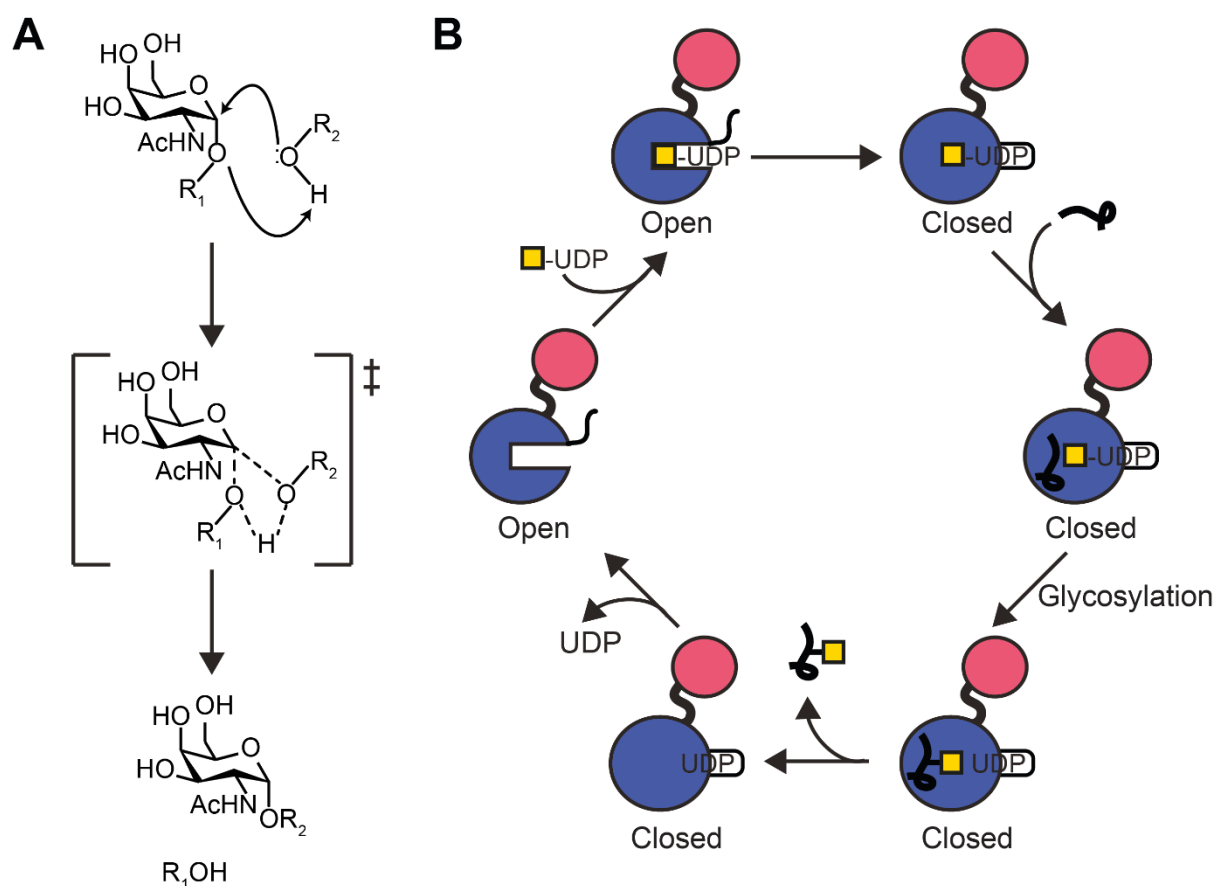


Fig. 5. Catalytic mechanisms of the GalNAc-T family **A)** Schematic representation of a S_Ni mechanism, where R₁ represents the UDP leaving group and R₂ represents the Ser or Thr acceptor residues. The leaving group accepts a proton from the hydroxy acceptor group and guides the nucleophilic residue to the same face of the anomeric carbon, resulting in retention of the anomeric configuration upon formation of the glycosidic bond. Figure adapted from Rini et al. (2022)⁷⁷. **B)** Schematic representation of the Bi-Bi kinetic mechanism. Figure adapted from Lira-Navarrete et al. (2014)⁸⁰.

The lectin domain of GalNAc-Ts contains three potential GalNAc-binding sites (the α -, β - and γ -subdomains)⁷⁶. Mutational and structural studies demonstrated that in most GalNAc-Ts only one of the binding sites is functional, and that an Asp residue within a Cys-Leu-Asp GalNAc binding motif is critical for lectin domain-mediated glycosylation⁸¹. Whilst the interaction between the lectin domain and the glycopeptide is predominantly mediated via hydrogen bonds and CH- π interactions between conserved residues within the enzyme and the remote GalNAc unit⁸², it has been demonstrated that lectin domain-mediated glycosylation can also be affected by the peptide sequence surrounding the GalNAc unit⁸³. Crystal structures of fly PGANT9-A and PGANT9-B splice variants may offer a possible explanation for this observation, where a loop protruding from the lectin domain towards the catalytic domain differs in charge state between splice variants, which in turn correlates with their preference towards differently charged substrates, suggesting that electrostatic interactions may be at play⁶⁸. A similar lectin flexible loop was observed for the first time in the crystal structure of GalNAc-T4⁸². Mutational studies demonstrated that this loop assists both GalNAc binding and release and is essential for the optimal activity of the catalytic domain⁸².

O-GalNAc glycosylation is arguably one of the most abundant post-translational modifications, predicted to occur on over 80% of proteins passing through the secretory pathway⁸². O-GalNAc glycosylation is also called mucin-type O-glycosylation since the mucin family of glycoproteins carry the greatest number of O-GalNAc glycans (>50% by weight)⁶⁸. O-GalNAc glycosylation is often found in densely clustered, heavily glycosylated domains within mucins and mucin domain-containing glycoproteins⁷⁶. Nevertheless, it is also present at single sites on most secreted and membrane-bound proteins^{60, 69}. Concomitant with their abundance, the biological functions of O-GalNAc glycans are many and can vary depending on their structure, their density, and the underlying protein⁶⁰. Most of the functions of O-GalNAc glycans have been originally attributed to their role in dense clusters within mucins⁶⁰. O-GalNAc glycans present in mucins, which line the epithelial surfaces of the body, provide physical and chemical protection to the underlying epithelial cells⁶⁰. Furthermore, since O-GalNAc glycans are hydrophilic and negatively charged, they also contribute to the viscosity of mucins by promoting binding of water and salts⁶⁰. O-GalNAc glycans can also provide structural support and stability by protecting glycoproteins from proteolytic cleavage⁸⁴.

The functions of O-GalNAc glycans at single sites are less well defined, nevertheless several studies have shown that O-GalNAc glycans can mediate cell-cell and cell-matrix interactions, affecting tissue formation and differentiation⁸⁵⁻⁸⁸. Studies have also shown that O-GalNAc glycosylation at one or a few sites of stem regions in cell membrane proteins can protect from ectodomain shedding⁸⁹. Furthermore, O-GalNAc glycans may act as carriers of ligands for lectins such as selectins, Siglecs or galectins⁸⁴.

Selectins are a family of cell-adhesion molecules which are expressed on platelets (P-selectin), endothelial cells (E-selectin), and leukocytes (L-selectin) and they recognise the sialyl Lewis X (SLe^x) structure often attached to core 2 O-GalNAc glycans (**Fig. 6A**)⁹⁰. Selectins mediate the migration of leukocytes into inflammatory sites⁹⁰. In normal vessels, leukocytes flow without adhesive interactions with endothelial cells⁹⁰. Upon inflammation, activated endothelial cells express E-selectin which engages with selectin ligands on circulating leukocytes and results in tethering, rolling and arrest of the leukocytes and their subsequent extravasation into underlying tissues⁹⁰. Binding of activated platelets expressing P-selectin to circulating leukocytes also contributes to the recruitment of leukocytes to sites of injury⁹⁰. PSGL-1 is the major ligand for P- and L-selectin and is also a ligand for E-selectin⁹⁰. PSGL-1 is expressed on most leukocytes (including neutrophils, monocytes, eosinophils, basophils, and subsets of T cells), haematopoietic stem cells and some activated endothelial cells⁹⁰.

Siglecs are a family of cell surface transmembrane sialic acid-binding lectins expressed by immune cells⁹¹. Siglecs recognise endogenous sialic acid-containing glycans as “self-ligands” and can help immune cells discriminate between “self-” and “non-self-ligands”⁹². Most Siglecs contain an immunoreceptor tyrosine-based inhibition motif within their cytoplasmic tails^{91, 92}. Engagement of Siglecs with their ligands results in downregulation of immune cell responses⁹².

Galectins recognise terminal β -galactose residues and are extracellularly and intracellularly expressed by many cell types^{93, 94}. Galectins have many biological functions including but not limited to cell adhesion, migration, growth, apoptosis, and immune cell regulation⁹⁵.

Breast cancer-associated changes in glycosylation

Changes in O-GalNAc glycosylation are seen in over 90% of BCs⁴⁴. Cancer-associated O-GalNAc glycans typically display truncated sugar branches and increased sialylation, as demonstrated by the expression of the Tn and T antigens and their sialylated forms (STn and ST antigens respectively) by breast tumours (**Fig. 6A**)⁷. The ER status has been shown to affect the type of O-GalNAc glycans displayed by BCCs, where ER+ tumours predominantly carry core 1 glycans, whilst ER- tumours carry primarily core 2 glycans instead, often modified with the SLe^x antigen (**Fig. 6A**), which is not normally found on core 2 glycans in healthy breast tissue⁷.

Changes in N-glycans are also detected in BC, where there is an increase in sialylated, fucosylated and branched (tri- and tetraantennary) N-glycans compared to healthy breast tissue (**Fig. 6B**)^{96, 97}. In contrast, there is a decrease in bisecting GlcNAc N-glycans (**Fig. 6B**)⁹⁶. Branching of N-glycans is mediated by the GlcNAc-transferases MGAT4 and MGAT5, while the generation of bisecting N-glycans is catalysed by MGAT3 instead (**Fig. 6B**)⁵⁹. Cancer-associated N-glycans are frequently decorated with the SLe^x antigen and poly-LacNAc units (repeats of the Gal-GlcNAc disaccharide) and capped with sialic acid and fucose (**Fig. 6B**)^{24, 97}. Core fucosylation, which involves the addition of fucose to the innermost GlcNAc residue in N-glycans by the fucosyltransferase FUT8, has been observed in multiple receptors implicated in BC⁹⁷. For example, core fucosylation of EGFR is linked to enhanced dimerization and phosphorylation, which promotes EGFR signalling and leads to tumour growth and malignancy^{7, 24, 97, 98}. Similarly, core fucosylation of the TGF- β receptor (TGFBR) promoted dimerisation and thereby TGF- β -induced EMT in BCCs^{97, 99, 100}.

Increased O-GlcNAcylation is also observed in BCs compared to healthy breast tissue^{96, 101}. Furthermore, higher levels of O-GlcNAc glycans have been detected in metastatic lymph nodes compared to primary tumour sites¹⁰¹. O-GlcNAcylation can modulate the functions of proteins by regulating protein phosphorylation, degradation, localisation and transcription^{24, 98}. Indeed, many BC-associated proteins, including Myc, cyclin D1, Rb and p53 are O-GlcNAcylated^{101, 102}. For example, Myc can be O-GlcNAcylated at Thr⁵⁸ which is also a phosphorylation site – increased O-GlcNAcylation of Myc increases its stabilisation and leads to BC progression²⁴. O-GlcNAcylation can also regulate glucose metabolism of BCCs – O-GlcNAcylation of the enzyme PFK1, which is a rate limiting enzyme of glycolysis, can redirect glucose flux into the pentose phosphate pathway for nucleic acid synthesis^{98, 102}.

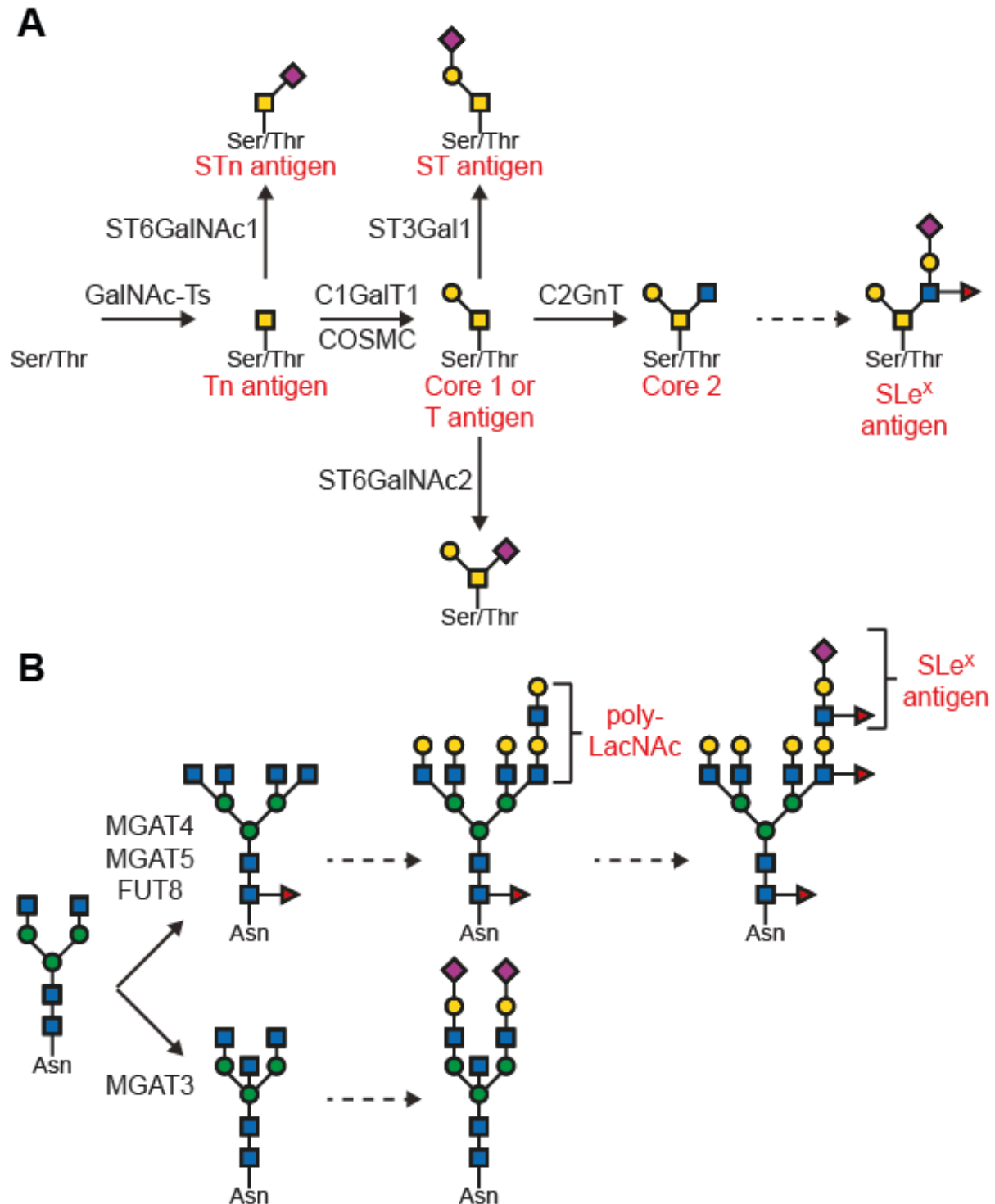


Fig. 6. BC-associated changes in glycosylation. **A)** Biosynthesis of BC-associated O-GalNAc glycans. Figure adapted from Magalhaes et al. (2021)¹⁰³. **B)** BC-associated N-glycans exhibit increased sialylation, fucosylation, and branching. They are also elongated with poly-LacNAc units and the SLe^x antigen, which can be recognized by galectins and selectins, respectively. A decrease in bisecting GlcNAc N-glycans is also observed in breast tumours. Figure adapted from Bellis et al. (2022)¹⁰².

Increased sialylation of glycoproteins and gangliosides is another frequent occurrence in BC^{7, 96, 99, 102, 104}. Expression of polysialic acid, a linear polymer of sialic acid residues, is associated with higher grade breast tumours and metastasis^{7, 24, 96, 105}. Since sialic acids are negatively charged, the increased sialylation of BCCs can disrupt cell-cell adhesions and facilitate the detachment of the primary tumour, hence promoting the metastatic potential of BCCs^{99, 102}. Overexpression and altered expression of gangliosides has also been associated with BC malignancy^{7, 24, 99, 102, 104}. Gangliosides are thought to regulate signal transduction by interacting with membrane-associated receptors, where changes in the gangliosides affect the organization of the receptors in lipid rafts within the plasma membrane, thereby influencing their downstream signalling²⁴. While 5-*N*-glycolylneuraminic acid (Neu5Gc) is typically absent in humans due to the lack of the enzyme responsible for converting cytidine monophosphate-5-*N*-acetylneuraminic acid (CMP-Neu5Ac) to Neu5Gc, Neu5Gc has been detected on glycoproteins and gangliosides in BCs^{102, 104, 105}. This is thought to result from the metabolic incorporation of diet-derived Neu5Gc into glycans, which triggers an immune response that promotes tumour growth by enhancing chronic inflammation and angiogenesis^{102, 105}.

Cancer-associated changes in glycosylation modulate new interactions with lectins expressed by immune cells, often triggering inhibitory pathways and immune evasion⁴⁴. The increased sialylation seen in BC results in immunosuppression through engagement with Siglecs^{7, 99, 102, 106}. Binding of MUC1 carrying ST antigens (MUC1-ST) to Siglec-9 (which is expressed on neutrophils, monocytes, macrophages, and dendritic cells (DCs)⁹¹) results in the secretion of factors involved in tumour progression and induces macrophages to display a tumour-associated macrophage-like phenotype, with increased expression of the PD-L1 ligand¹⁰⁷. Increased expression of PD-L1 promotes engagement to the inhibitory PD-1 receptor on activated T cells, which inhibits their function¹⁰⁷. Binding of MUC1-ST to Siglec-9 also results in recruitment of β -catenin to the cytoplasmic tail of MUC1, which induces its translocation to the nucleus and promotes tumour growth¹⁰⁸. Similarly, engagement of Siglec-7 (which is the major Siglec on natural killer (NK) cells and is also expressed at lower levels on monocytes, macrophages, DCs and a subset of CD8+ cytotoxic T cells⁹¹) inhibits NK-mediated cytotoxicity¹⁰⁹. The macrophage galactose-specific lectin (MGL) recognises terminal GalNAc residues and is expressed on macrophages and DCs¹¹⁰. MGL can bind the Tn and STn antigens¹¹¹ and can facilitate tumour progression by creating an immunosuppressive environment⁸⁴. Binding of MGL to CD45 carrying the Tn antigen on T cells decreases T cell proliferation and increases T cell death¹¹². Engagement of MGL on immature DCs impairs the migration of DCs and results in their retention in lymphatic tissues¹¹². Only upon DC maturation when MGL expression is down-regulated can the cells migrate out of the lymphatic tissues for efficient antigen presentation to T cells¹¹².

Cancer cells overexpress selectin ligands and hijack selectin-selectin ligand interactions to facilitate several aspects of cancer progression¹⁰². Expression of selectin ligands on circulating tumour cells can facilitate extravasation of BCCs from the circulation into the surrounding tissue and metastasis^{7, 44, 84, 103}. Furthermore, binding of selectin ligands on BCCs to activated platelets promotes platelet aggregation which shields tumour cells from immune elimination within the vasculature and contributes to the arrest of circulating BCCs to the endothelium and their subsequent extravasation¹¹³.

Overexpression of galectins is often seen in BC, and it correlates with tumour aggressiveness and metastatic potential of tumours⁹³. Different members of the galectin family have been shown to impact different cancer hallmarks, including immune evasion, angiogenesis, cancer cell growth and metastasis¹¹⁴. A few examples are listed below, such as overexpression of the STn antigen resulting in intracellular accumulation of galectin-3 which induces resistance of cancer cells to chemotherapeutic drugs¹¹⁵. Binding of galectin-3 to the extracellular domain of MUC1 carrying the T antigen (MUC1-T) induces MUC1 to lose its apical membrane polarisation, instead becoming expressed over the whole cell surface, thus promoting its interaction with EGFR¹¹⁶. The MUC1-EGFR interaction results in prolonged EGFR dimerisation and downstream signalling, which is associated with tumorigenesis and metastasis¹¹⁶. The cell surface polarisation induced by the interaction between MUC1-T and galectin-3 has also been shown to expose selectin ligands, which promotes adhesion of MUC1-expressing cells to the endothelium, thus facilitating extravasation and metastasis of BCCs¹¹⁷. Similarly, the interaction between galectin-3 and MUC1-T-expressing disseminating BCCs results in cancer cell aggregation, which increases the survival of BCCs by preventing suspension-induced apoptosis (anoikis)¹¹⁸. Branched N-glycans carrying poly-LacNAc units are also ligands for galectins, and together they can form a cross-linked lattice that inhibits receptor endocytosis, thereby leading to prolonged signalling^{24, 102}.

Cell-matrix interactions play a crucial role in the metastatic potential of BCCs²⁴. Integrins are the main family of cell surface transmembrane receptors mediating cell-matrix interactions¹¹⁹. Engagement of integrins results in downstream signalling which can influence many aspects of a cell's behaviour including cell proliferation and survival¹¹⁹. Heavily O-glycosylated glycoproteins such as mucins are often OE in BC⁴⁴. Mucin overexpression contributes to a bulky glycocalyx which can facilitate tumour progression by enhancing clustering of integrins and thus promoting integrin-mediated cell growth and survival^{24, 120}. Furthermore, aberrant N-glycosylation and α 2,6 sialylation of integrins enhances integrin-mediated BCC migration and invasion^{98, 99}.

The content and distribution of PGs also changes during carcinogenesis¹⁰². Heparan sulfate PGs (HSPGs) are present on the plasma membrane and in the extracellular matrix of breast tumours, and they have been implicated in regulating the growth, invasion, and metastasis of BCCs^{24, 99}. HS is a polymer of the repeating disaccharide unit GlcNAc-GlcA and is modified by uronic acid epimerisation and N- and O-sulfation⁵⁸. Unlike other GAGs, once displayed at the plasma membrane HS can undergo additional modifications, where a family of endosulfatases can remove sulfate groups and the enzyme heparanase can cleave the HS chains⁵⁸. HSPGs can promote cell-cell and cell-matrix interactions, as well as bind growth factors, thereby acting as co-receptors for various growth factor receptors and promoting downstream signalling^{24, 99, 102}. Hyaluronan (also called hyaluronic acid) is a secreted GAG consisting of the repeating disaccharide unit GlcNAc-GlcA, which primarily exists as a freestanding entity rather than being covalently linked to a protein core^{55, 58}. Nevertheless, hyaluronan can interact noncovalently with several proteins including multiple cell surface receptors, such as CD44, thereby regulating key signalling pathways associated with oncogenesis^{102, 121}. CD44 is a transmembrane receptor expressed by many cell types which is known to be implicated in BC development and progression^{24, 102}. Binding of hyaluronan to CD44 promotes HER2 signalling and multidrug resistance^{102, 121}.

Multiple mechanisms have been attributed to the BC-associated changes in glycosylation. One of the mechanisms that has been proposed is changes to the pH of the Golgi apparatus¹²². The Golgi apparatus requires an acidic pH for optimal activity and for correct localisation of glycosyltransferases - under normal conditions the pH of the Golgi ranges between 5.9 and 6.5¹²². The pH of the Golgi in many cancer cells is considerably more alkaline (pH $\geq$ 6.75), and it has been demonstrated that only a 0.2 pH unit increase in the pH of the Golgi is sufficient to induce increased expression of the T antigen¹²², via the pH-dependent mislocalisation of glycosyltransferases¹²³. Another possible mechanism is the availability of nucleotide sugars, where increased levels of UDP-GalNAc and UDP-GlcNAc have been reported in breast tumours¹²⁴ and BCCs¹²⁵. The increased availability of nucleotide sugars in BCCs is likely a result of the enhanced glucose uptake and its subsequent diversion into metabolic pathways, such as the hexosamine biosynthetic pathway which produces UDP-GlcNAc^{24, 98}.

A third mechanism that has been proposed is changes to the levels of the glycosyltransferases and glycosidases involved in the biosynthesis of cancer-associated glycans⁴⁴. The increased sialylation observed in BC can be attributed to the overexpression of sialyltransferases, such as ST3Gal1, which catalyses the conversion of the T antigen to the ST antigen and is OE in 90% of BCs, or ST6GalNAc1, which catalyses the conversion of the Tn antigen to the STn antigen and is OE in 25% of BCs (**Fig. 6A**)^{44,126}. The T antigen is also a substrate of C2GnT, which converts core 1 to core 2 O-glycans (**Fig. 6A**)⁴⁴. The relative expression of the different glycosyltransferases determines the glycan structures expressed by BCCs⁴⁴. For example, a higher expression of ST3Gal1 and ST6GalNAc2 have been detected in ER+ compared to ER- tumours, which can explain why ER+ tumours predominantly express sialylated core 1 O-glycans whilst ER- tumours express core 2 O-glycans instead⁴⁴. BCCs increase the expression of MGAT5 and decrease the expression of MGAT3 (**Fig. 6B**)^{24, 102}. Changes in the expression levels of core and terminal fucosyltransferases are also observed in BC, particularly the overexpression of the core fucosyltransferase FUT8 (**Fig. 6B**)^{7, 24, 98, 99, 102, 106}. Increased levels of the enzyme O-GlcNAc transferase (OGT), which catalyses the synthesis of O-GlcNAc glycans, have been found in breast tumours^{24, 96, 98}. The expression of glycosyltransferases involved in ganglioside synthesis are also altered in BC^{104, 106}. Overexpression of heparanase has been implicated in BC progression, where secreted heparanase can act in parallel with MMPs to remodel the extracellular matrix and facilitate BCC metastasis^{58, 102}.

GalNAc-Ts implicated in breast cancer

Changes in the expression levels and the localisation of many members of the GalNAc-T family have also been associated with BC⁴⁴. GalNAc-T1¹²⁷⁻¹²⁹, -T2¹²⁸⁻¹³⁰, -T3¹²⁹⁻¹³¹, -T4^{129, 132, 133}, -T5^{129, 130}, -T6^{129, 130, 134-143}, -T7^{128-130, 144-146}, -T10¹²⁸⁻¹³⁰, -T13¹⁴⁷, and -T14^{129, 143, 148, 149} are OE in BC. In contrast, GalNAc-T8^{129, 150}, -T11¹³⁰, -T12^{129, 130}, -T15¹²⁹, -T16¹²⁹ and -T20¹²⁹ are down-regulated in BC.

The highest GalNAc-T1 expression was reported in TNBC¹⁵¹ and metastatic BC¹²⁷. High GalNAc-T1 expression was associated with a shorter overall patient survival^{127, 152}. Furthermore, high GalNAc-T1 expression was associated with increased EMT, angiogenesis, protein secretion, low levels of CD8+ T-cell infiltration and worse survival in adolescent and young adult population, but not in older patients¹⁵¹. In one study, LAMTOR5 (which is upregulated in BC) was shown to drive the transcription of GalNAc-T1 in MCF7 cells as well as promoting the activation of c-Src, which results in translocation of GalNAc-T1 to the endoplasmic reticulum¹²⁷. LAMTOR5-mediated GalNAc-T1 overactivity resulted in increased expression of the Tn antigen on MUC1 and osteopontin (an extracellular protein associated with cell adhesion and cancer progression), which promoted BC metastasis¹²⁷. In a separate study, downregulation of GalNAc-T1 reduced sialylation on MMP-14, which increased the sensitivity of MMP-14 to autoproteolysis and reduced invasiveness of BCCs, suggesting that GalNAc-T1 can also mediate the metastatic potential of BCCs via MMP-14¹²⁸. In the same study, downregulation of GalNAc-T1 also enhanced intracellular drug uptake and anti-tumour drug effects, suggesting that GalNAc-T1 can mediate the invasiveness and chemotherapeutic sensitivity of BCCs¹²⁸.

A potential mechanism by which GalNAc-Ts are mislocalised to the endoplasmic reticulum is via COP1-mediated transport, and COP1 retrograde transport of GalNAc-Ts was reported to be upregulated in BCCs¹⁵³. GalNAc-T1 and GalNAc-T2 were detected in the endoplasmic reticulum of multiple BCC lines, with a concomitant increase in the expression of the Tn antigen¹⁵³. Loss of C1GalT1 activity is a rare event in BC^{44, 153}, suggesting that Tn expression in BC is driven by the mislocalisation of GalNAc-Ts to the endoplasmic reticulum and not C1GalT1 deficiency¹⁵³. Cells expressing endoplasmic reticulum-localised GalNAc-T2 showed increased adherence to fibronectin and collagen in a lectin domain-dependent manner, and resulted in increased cell adhesion, motility, and invasion of BCCs¹⁵³. In the same study, cells expressing endoplasmic reticulum-localised GalNAc-T2 also promoted efficient lung colonisation *in vivo*¹⁵³.

When assessing the expression of a range of GalNAc-Ts in a panel of benign and malignant mammary epithelial cell lines, those derived from benign breast epithelium expressed predominantly GalNAc-T1 and GalNAc-T2, whilst the malignant cell lines displayed a broader expression of GalNAc-Ts¹⁵⁴. For example, whilst GalNAc-T3 and GalNAc-T6 were weakly detectable or undetectable in benign breast epithelium, they were readily detected in all malignant BCCs¹⁵⁴. In a separate study, E-cadherin was proposed as a GalNAc-T3 substrate, where GalNAc-T3-mediated O-GalNAcylation of E-cadherin was required for E-cadherin's cell surface localisation¹⁵⁵. Loss of GalNAc-T3 resulted in retention of E-cadherin in the Golgi and induction of EMT in trophoblast stem cells, highlighting the role of GalNAc-T3 on protecting the epithelial state of cells¹⁵⁵.

It is well accepted that GalNAc-T1, -T2 and -T3 are the main GalNAc-Ts initiating O-GalNAcylation of MUC1¹⁵⁶. The extracellular domain of MUC1 consists of 20-120 repeats (called tandem repeats or TRs) of a conserved sequence composed of 20 amino acids with five potential glycosylation sites (in bold): GVT³**S**⁴APDT⁸RPAPGS¹⁴**T**¹⁵APPAH¹⁵⁶. The initial O-GalNAcylation of MUC1 at Thr³, Ser¹⁴ and Thr¹⁵ is driven by GalNAc-T1, -T2 and -T3¹⁵⁶. GalNAc-T4 can glycosylate Ser⁴ and Thr⁸ *in vitro*, nevertheless upon KO of GalNAc-T4 in HEK293 cells Ser⁴ was O-glycosylated, suggesting that other GalNAc-Ts can also glycosylate this site^{156, 157}. In contrast, GalNAc-T4 is the only GalNAc-T capable of glycosylating Thr⁸ and completing the glycosylation of MUC1¹⁵⁶⁻¹⁵⁸. MUC1 purified from milk was found to be partially glycosylated, with approximately 2.6 glycans per TR¹⁵⁹. In contrast, in MUC1 purified from T47D cells all five sites were occupied, with approximately 4.8 glycans per TR¹⁶⁰. Since only GalNAc-T4 can complete the glycosylation of MUC1⁸², this suggests an important role for GalNAc-T4 in the increased glycosylation density of MUC1 observed in BC. Furthermore, the PDTRP sequence is highly immunogenic, with most anti-MUC1 antibodies targeting this region¹⁵⁶, suggesting that GalNAc-T4-mediated glycosylation may be important in the immunogenicity of MUC1.

GalNAc-T4 was identified as an estrogen-regulated gene¹⁶¹, suggesting that GalNAc-T4 may play a role in the estrogen regulatory pathway, however the exact role remained unclear. In a subsequent study, the expression of GalNAc-T4 in a panel of mammary epithelial cells was investigated¹³². Whilst the expression of GalNAc-T4 was lower compared to other members of the GalNAc-T family, GalNAc-T4 was the only isoenzyme showing differential expression depending on the ER status of BCCs, with a high expression in ER+ cells and a low expression in ER- cells, indicating a connection between GalNAc-T4 and the estrogen receptivity of BCCs¹³².

Knock-down (KD) of GalNAc-T4 in T47D cells resulted in a growth defect and cell cycle arrest, suggesting that GalNAc-T4 had a pro-tumorigenic role in T47D cells by promoting proliferation and cell cycle progression¹³². KD of GalNAc-T4 decreased the expression of ER α (but not ER β) and FOXA1, a transcription factor which mediates the last step of the estrogen regulatory pathway by inducing the chromatin remodelling required for ER α binding to DNA¹³². FOXA1 was confirmed to be O-GalNAcylated by *Vicia Villosa* Lectin (VVL, which recognises terminal GalNAc residues) blot, and O-GalNAcylation of FOXA1 was decreased upon KD of GalNAc-T4, suggesting that FOXA1 may be a substrate of GalNAc-T4¹³². KD of GalNAc-T4 also decreased the interaction between FOXA1 and ER α , suggesting that GalNAc-T4-mediated O-GalNAcylation of FOXA1 may be important in the estrogen regulatory pathway¹³². In a follow-up study, GalNAc-T2 was also shown to glycosylate FOXA1 *in vitro* at Ser³⁵⁵, which may contribute to the role of O-GalNAcylation in the function of FOXA1 in the estrogen regulatory pathway¹⁶².

In a separate study, alternative mechanisms by which GalNAc-T4 mediated the proliferation of BCCs were investigated¹³³. The expression of GalNAc-T4 was shown to be the highest in MCF7 cells and the lowest in MDA-MB-231 cells¹³³. Knock-out (KO) of GalNAc-T4 in MCF7 cells promoted cell growth, increased colony formation, and increased the number of cells in S phase, whilst the opposite effect was observed when GalNAc-T4 was OE in MDA-MB-231 cells, suggesting that GalNAc-T4 had an anti-tumorigenic role in these cells by impairing proliferation and cell cycle progression¹³³. The group subsequently performed immunoprecipitation using VVL-agarose beads in wild-type (WT) and GalNAc-T4 KO MCF7 cells and subjected the immunoprecipitated products to Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) proteomics analysis¹³³. Analysis of VVL-binding proteins showed 117 proteins to be upregulated and 1172 proteins to be downregulated upon GalNAc-T4 KO¹³³. Pathway enrichment analysis demonstrated that most dysregulated pathways were involved in the proliferation of BCCs, including the Notch signalling pathway¹³³. Whilst Notch was not found as a VVL-binding protein, VVL-binding of Cullin-1 (which is a key regulatory protein of the Notch signalling pathway) was decreased upon GalNAc-T4 KO, suggesting that Cullin-1 may be a substrate of GalNAc-T4¹³³. Overall, these findings suggested that GalNAc-T4 may regulate the proliferation of BCCs by glycosylating members of the Notch signalling pathway such as Cullin-1¹³³.

In another study, a possible role of GalNAc-T4 in the metastatic potential of BCCs was reported¹²⁹. The expression of GalNAc-T4 was highest in low-metastatic luminal BC tumours, whilst the levels of GalNAc-T4 in highly invasive basal-like tumours was comparable to benign tissue, indicating a connection between GalNAc-T4 and the metastatic potential of BCCs¹²⁹. KO of GalNAc-T4 in MCF7 cells promoted migration, invasion, and EMT of BCCs, whilst the opposite effect was observed when GalNAc-T4 was OE in MDA-MB-231 cells¹²⁹. The TGF- β signalling pathway was impaired when GalNAc-T4 was OE in MDA-MB-231 cells, which is an upstream pathway of EMT¹²⁹. O-GalNAcylation of TGFBR1 and TGFBR2 was confirmed by VVL blot¹²⁹. O-GalNAcylation of both receptors was decreased upon GalNAc-T4 KO which promoted receptor dimerisation, suggesting that GalNAc-T4 may play a role in the TGF- β -induced EMT of BCCs by O-GalNAcylation of TGFBR1 and TGFBR2¹²⁹. *In vitro* glycosylation experiments were subsequently performed using peptides based on the extracellular domains on both receptors, which suggested that Ser³¹ in TGFBR2 was the glycosylation site of GalNAc-T4, and mutation of this glycosite to Ala enhanced TGF- β -induced EMT of BCCs¹²⁹. Overall, these findings suggested that GalNAc-T4-mediated O-GalNAcylation of TGFBR2 prevents receptor dimerisation which suppresses EMT and BCC invasion and migration, thus regulating the metastatic potential of BCCs¹²⁹.

In a separate study, GalNAc-T4 was shown to be upregulated in endothelial cells exposed to pro-inflammatory stimuli, inducing the expression of adhesion molecules such as ICAM-1 and VCAM-1 as well as chemokines such as MCP-1, which promoted adhesion of leukocytes to endothelial cells and trans-endothelial migration¹⁶³. In another study, GalNAc-T4 was also shown to be upregulated in stimulated monocytes¹⁶⁴. VVL pulldown on monocyte cell lysates demonstrated that PSGL-1 is O-GalNAcylated by GalNAc-T4¹⁶⁴. Previous work by other groups had already suggested that GalNAc-T4 can glycosylate Thr⁵⁷ in PSGL-1 *in vitro*⁷², and that an O-glycan at this site plus sulfation was required for PSGL-1 binding to P-selectin¹⁶⁵. Increased expression of GalNAc-T4 promoted O-GalNAcylation of PSGL-1, as well as monocyte adhesion and migration via PSGL-1 engagement¹⁶⁴. BCCs may hijack these selectin-selectin ligand interactions, suggesting an alternative mechanism by which GalNAc-T4 may regulate invasion and migration of BCCs.

GalNAc-T5 was identified as a commonly mutated gene in BC¹⁶⁶. Two missense BC-specific mutations (E507D and L692F) were identified¹⁶⁶, and later confirmed to inactivate GalNAc-T5's enzymatic activity¹⁶⁷, suggesting that inactivation of GalNAc-T5 may facilitate BC progression. GalNAc-T5 has also been positively correlated with chemotherapeutic resistance of tumours¹⁶⁸.

Freire et al. investigated the usefulness of GalNAc-Ts in detecting disseminated cells in bone marrow samples and they showed that the best specificity was obtained with GalNAc-T6, where GalNAc-T6 mRNA was detected in 36% of bone marrow samples obtained prior to surgery but absent in control samples, suggesting that GalNAc-T6 could serve as a biomarker of BCC dissemination¹³⁸. The role of GalNAc-T6 in BC has been associated with O-GalNAcylation of CCDC88C¹⁶⁹, ER α ¹⁷⁰, MUC1¹³⁴, fibronectin¹⁷¹, α 2M¹³⁶, GRP78¹⁷², and LGALS3BP¹⁴⁰.

CCDC88C is a component of the non-canonical Wnt signalling pathway, and its high expression was associated with a poor prognosis of lymph node metastatic BC patients¹⁶⁹. GalNAc-T6-mediated O-GalNAcylation of CCDC88C increased the stability of CCDC88C which drove BCC motility, suggesting a potential role of GalNAc-T6 in the metastatic potential of BCCs¹⁶⁹.

In a separate study, KD of GalNAc-T6 decreased the expression of ER α , its nuclear localisation, and the levels of downstream genes in BCCs¹⁷⁰. Ser⁵⁷³ was identified as the GalNAc-T6-mediated glycosylation site in ER α by LC-MS/MS analysis¹⁷⁰, suggesting that GalNAc-T6 as well as GalNAc-T4 may play a role in the estrogen regulatory pathway.

In another study, KD of GalNAc-T6 reduced the proliferation of BCCs as well as the levels of cytoplasmic MUC1¹³⁴ and nuclear β -catenin¹³⁷. A co-immunoprecipitation assay demonstrated that GalNAc-T6 interacts with the N-terminal portion of MUC1 (MUC1-N), whilst the C-terminal portion (MUC1-C) interacts with β -catenin¹³⁷. Overall, these findings suggested GalNAc-T6 may regulate the proliferation of BCCs via the MUC1-C/ β -catenin signalling pathway¹³⁷.

Fibronectin was shown to be O-GalNAcylated by VVL blot¹⁷¹. The levels of fibronectin were increased when WT but not catalytically inactive GalNAc-T6 was OE, whilst no difference in the transcript levels of fibronectin were observed between both cell lines, suggesting that O-GalNAcylation by GalNAc-T6 may contribute to the stability of fibronectin¹⁷¹. When WT but not catalytically inactive GalNAc-T6 was OE, the acinar structure of benign mammary epithelial MCF10A cells was disrupted and EMT induced, and these effects were abolished upon KD of fibronectin, suggesting that O-GalNAcylation of fibronectin by GalNAc-T6 may regulate the malignant transformation of BCCs¹⁷¹.

In a separate study, LC-MS/MS analysis of the secreted supernatant of WT and GalNAc-T6 KD MDA-MB-231 cells revealed α 2M as a potential substrate of GalNAc-T6¹³⁶. α 2M is a pan-proteinase inhibitor in humans¹³⁶. O-GalNAcylation of α 2M was confirmed by VVL blot¹³⁶. When GalNAc-T6 was OE in MDA-MB-468 the invasive and migratory abilities of BCCs increased, and these effects were suppressed upon KD of α 2M, suggesting that O-GalNAcylation of α 2M by GalNAc-T6 may regulate the migration and invasion of BCCs¹³⁶.

GRP78 is upregulated in cancer cells and is thought to play important roles in various cellular processes including endoplasmic reticulum stress and apoptosis¹⁷². VVL pulldown was performed on HeLa cells transfected with WT or catalytically inactive GalNAc-T6 followed by LC-MS/MS analysis, which resulted in the identification of GRP78 as a potential GalNAc-T6 substrate¹⁷². GRP78 was confirmed to be O-GalNAcylated, and the level of O-GalNAcylation decreased upon GalNAc-T6 KD¹⁷². The stability of GRP78 after treatment with cycloheximide (CHX, a protein synthesis inhibitor) was greater in cells which OE GalNAc-T6 compared to mock-transfected controls, and the levels of GRP78 decreased after CHX treatment when three potential glycosylation sites were mutated to Ala, suggesting that O-GalNAcylation of GRP78 by GalNAc-T6 is important for its stability¹⁷². Furthermore, O-GalNAcylation of GRP78 was also shown to be important for its localisation in the endoplasmic reticulum and anti-apoptotic function¹⁷², suggesting that GalNAc-T6 may mediate the survival of BCCs by O-GalNAcylation of GRP78.

LGALS3BP is a secreted growth-promoting glycoprotein whose levels are elevated in the serum of BC patients¹⁴⁰. LC-MS analysis for GalNAc-T6-binding proteins in GalNAc-T6-expressing vs mock-transfected HeLa cells identified LGALS3BP as a potential GalNAc-T6 substrate¹⁴⁰. O-GalNAcylation of LGALS3BP by WT but not catalytically inactive GalNAc-T6 was confirmed by VVL blot¹⁴⁰. KD of GalNAc-T6 reduced the protein level and secretion of LGALS3BP¹⁴⁰. Furthermore, when BCCs were treated with conditioned media from HeLa cells expressing O-GalNAcylation site mutants of LGALS3BP, the growth-promoting effect was impaired, suggesting that O-GalNAcylation by GalNAc-T6 is required for the growth-promoting function of LGALS3BP¹⁴⁰.

GalNAc-T7 expression was associated with hormone receptor status, with the strongest association observed with HER2 status^{145, 146}. Higher levels of GalNAc-T7 were detected in the serum of luminal BC patients compared to healthy individuals¹⁴⁶. Furthermore, combined detection of serum GalNAc-T7 with traditional BC markers enhanced the diagnostic sensitivity and specificity of luminal BC, suggesting that GalNAc-T7 may be a promising non-invasive BC biomarker¹⁴⁶. In a separate study, GalNAc-T7 was reported to mediate the proliferation, apoptosis, and autophagy of BCCs by inactivation of STAT3 signalling¹⁴⁴.

The role of GalNAc-T8 in BC has been associated with O-GalNAcylation of BMPR1A¹⁵⁰ and EGFR¹⁷³. GalNAc-T8 was OE in MDA-MB-231 cells which reduced BCC proliferation and activated the BMP signalling pathway. VVL pulldown confirmed that BMPR1A is O-GalNAcylated, and O-GalNAcylation was reduced upon KD of GalNAc-T8¹⁵⁰, suggesting that BMPR1A was a GalNAc-T8 substrate. WT BMPR1A was OE in MCF7 cells which reduced the proliferation of BCCs, and this effect was abrogated when O-GalNAcylation site mutants were OE instead¹⁵⁰, suggesting that O-GalNAcylation of BMPR1A by GalNAc-T8 regulates the proliferation of BCCs. In a separate study, GalNAc-T8 was OE in MDA-MB-231 cells which inhibited migration, EMT, and EGFR signalling of BCCs¹⁷³. VVL pulldown was performed on cell lysates from MDA-MB-231 cells and the results demonstrated that O-GalNAcylation of EGFR increased when GalNAc-T8 was OE¹⁷³. Overall, these findings suggested that GalNAc-T8 may regulate the metastatic ability of BCCs by O-GalNAcylation of EGFR¹⁷³.

GalNAc-T9 was shown to be frequently methylated and silenced in brain metastases but not in the associated primary breast tumours¹⁷⁴. KD of GalNAc-T9 was shown to increase the invasive potential of BCCs¹⁷⁴, suggesting that silencing of GalNAc-T9 may increase the metastatic potential of BCCs. GalNAc-T11 has been associated with tumour resistance to therapeutic agents¹⁷⁵. Guda et al. identified 8 mutations which inactivate the normal function of GalNAc-T12 and imparted susceptibility to cancer, as demonstrated by patients harbouring inactivating germ-line GalNAc-T12 mutations which developed BC in addition to a primary colon cancer¹⁶⁷. Tao et al. investigated DNA methylation data to screen for candidate BC biomarkers, and identified GalNAc-T13 as one of six genes whose methylation state could be used to predict BC outcomes¹⁷⁶. In a separate study, GalNAc-T13 and GalNAc-T17 were identified as the two most frequently hypermethylated and mutated genes in different cancer types¹⁷⁷, although the effect of these modifications requires further investigation.

Expression of GalNAc-T14 was positively correlated with stemness and drug resistance markers^{143, 148}. KD of GalNAc-T14 in MCF7 cells reduced the expression of β -catenin, the stemness markers OCT4 and SOX2, and the drug resistance marker ABCC5, suggesting that GalNAc-T14 may regulate stemness and drug resistance of BCCs through the β -catenin signalling pathway^{143, 148}. In a separate study, the expression of GalNAc-T14 was associated with multidrug resistance in BC¹⁷⁸. The expression of GalNAc-T14 positively correlated with the expression of P-gp, a cell membrane efflux pump which is often OE in cancer¹⁷⁸. O-GalNAcylation of P-gp by GalNAc-T14 was confirmed by lectin blot, and the stability of P-gp upon CHX treatment increased when GalNAc-T14 was OE¹⁷⁸. Overall, these findings suggested that GalNAc-T14 may regulate BCC drug resistance by promoting the expression and stability of P-gp¹⁷⁸.

A separate study demonstrated that the serum levels of GalNAc-T14 were upregulated in BC patients compared to healthy controls¹⁴⁸, suggesting that GalNAc-T14 may be a promising non-invasive BC marker in addition to GalNAc-T7¹⁴⁶. The role of GalNAc-T14 in BC has also been associated with O-GalNAcylation of EFEMP2¹⁷⁹. A yeast two-hybrid screening assay was performed and identified EFEMP2 as a novel binding partner of GalNAc-T14¹⁷⁹. KD of GalNAc-T14 decreased the expression and O-GalNAcylation of EFEMP2¹⁷⁹. Furthermore, the stability of EFEMP2 upon CHX treatment was greater when GalNAc-T14 was OE, suggesting that O-GalNAcylation of EFEMP2 by GalNAc-T14 is important for its stability¹⁷⁹. EFEMP2-expressing cells had a greater invasive ability than control cells, and this effect was abrogated upon KD of GalNAc-T14, suggesting that GalNAc-T14 regulates EFEMP2-mediated invasion of BCCs¹⁷⁹.

Song et al. demonstrated that GalNAc-T14 specifically promotes BC metastasis to the lung¹⁸⁰. KD of GalNAc-T14 reduced metastatic lung colonisation of BCCs *in vivo*, impaired sphere formation *in vitro*, and activated BMP signalling¹⁸⁰. Altogether these results suggested that GalNAc-T14 may promote the metastatic potential of BCCs by suppressing BMP signalling, which in turn enhances the self-renewal capability of BCCs¹⁸⁰. VVL pulldown confirmed that BMPR1A is O-GalNAcylated, and its O-GalNAcylation was reduced upon KD of GalNAc-T14¹⁸⁰. KD of GalNAc-T14 also reduced macrophage recruitment to lung nodules and reduced the expression of several cyto/chemokines, suggesting that GalNAc-T14 may promote macrophage recruitment by regulating the expression of cyto/chemokines¹⁸⁰.

In the same study they also demonstrated that macrophage-induced growth of BCCs was maintained even under conditions where the macrophages and BCCs were separated by a permeable membrane, suggesting that this effect was mediated by secreted factor(s) from macrophages¹⁸⁰. Genome wide RNA sequencing analysis of differentially expressed genes when GalNAc-T14 was KD or OE revealed that FGF-responsive genes were downstream targets of GalNAc-T14¹⁸⁰. Treatment of BCCs with an FGFR1/3-specific inhibitor suppressed the macrophage-induced growth of BCCs¹⁸⁰. VVL pulldown confirmed that FGFR1 (but not FGFR3) is O-GalNAcylated, and its O-GalNAcylation was reduced upon KD of GalNAc-T14¹⁸⁰. Overall, these findings suggested that GalNAc-T14 promotes lung-specific metastasis of BCCs by two mechanisms: (I) O-GalNAcylation of BMPR1A by GalNAc-T14 suppresses the anti-metastatic effect of BMPs in the lung stroma, which enables BCCs to establish metastatic colonies in the lung, and (II) GalNAc-T14 facilitates the development of macrometastases by promoting macrophage recruitment to the site of metastases and enhancing macrophage-induced growth of BCCs upon receiving macrophage derived FGFs by O-GalNAcylation of FGFR1¹⁸⁰.

Wu et al. demonstrated that the GalNAc-T16 single-nucleotide polymorphisms rs2105269 (A at position 69280517) and rs72625676 (T at position 69307268) were associated with an increased BC susceptibility in Chinese Han women aged 51 years or less, whilst in women aged over 51 years the rs1275678 polymorphism (A at position 69335147) was associated with a decreased BC risk, however larger population-based studies and functional investigation into the effect of these polymorphisms is required¹⁸¹.

Substrate identification of individual GalNAc-Ts

Whilst the role of certain GalNAc-Ts in BC has been linked to glycosylation of some proteins, limited information is currently available regarding the exact GalNAc-T-specific glycosylation sites that affect disease and can inform on therapeutic strategies. Further investigation is therefore required to confirm the proposed protein substrates, better define the exact mechanism by which GalNAc-T-mediated O-GalNAcylation of these proteins contributes to BC, and identify other substrates playing a role in BC. However, there are multiple challenges in discerning the substrates and glycosylation sites of a single GalNAc-T, including partial substrate redundancy and overlapping glycosylation sites⁶⁷. Classical methods of molecular biology are of limited use to study which sites are glycosylated on a glycoprotein as well as which glycan structures are present at each site (i.e. the glycoproteome). This is because glycans are secondary gene products – unlike protein sequences, they are not encoded directly in the genome and instead depend on many competing and sequential reactions between glycosyltransferases and glycosidases⁵⁵. Even with full knowledge of the expression levels of all relevant glycosyltransferases it is impossible to predict the presence or absence of glycans at a particular glycosite nor the final structure of glycans⁵⁵.

On top of this, O-GalNAc glycosylation is one the most challenging PTMs to study by MS¹⁸². The fact that O-GalNAc glycans lack a peptide consensus motif and that a single glycosite may carry multiple different glycoforms (microheterogeneity) or no glycans at all (macroheterogeneity) complicates glycosite prediction and often leads to inaccurate assignment of O-glycopeptides by search algorithms¹⁸³. For this reason, manual validation of the spectra is still required, which is challenging and time-consuming¹⁸³. Since current methodologies have limited ability to provide simultaneous structural and site information, the O-glycome and O-glycoproteome are often analysed separately¹⁸⁴. In glycomics, glycans are released enzymatically or chemically from the peptides they are attached to and analysed by LC-MS¹⁸⁵. However, while N-glycans can be removed by PNGase F, an amidase which cleaves the bond between Asn residues and the innermost GlcNAc unit of N-glycans¹⁸⁵, no endoglycosidase has been identified that can release all types of O-GalNAc glycans^{183, 186}. Furthermore, whilst N-glycopeptides can be sequenced by traditional collision-based methods (e.g. higher energy collisional dissociation or HCD), the more labile nature of O-glycopeptides means that glycans are often detached from the peptide upon MS fragmentation, and thus the O-glycosylation site cannot be identified¹⁸⁵.

Additional MS challenges include dissecting monosaccharides with the same molecular mass (e.g. GlcNAc or GalNAc), dissecting the nature of glycosidic linkages, or potentially missing important labile modifications such as sulfation and acetylation¹⁸⁵. All these issues are aggravated by the poor ionizability of O-glycopeptides and ion suppression by unmodified peptides¹⁸³. Moreover, even though O-GalNAc glycans can be present in multiple glycoproteins, they are abundantly present in mucin domains, whose complex nature intensifies the issues described above^{182, 183, 187-190}. The large number of Ser and Thr residues present in these domains makes site localisation even more challenging¹⁸³. Furthermore, the dense glycosylation of mucin domains limits access of proteases to the peptide backbone, making them generally resistant to digestion¹⁸³.

Most knowledge on the substrate specificity of individual GalNAc-Ts has been obtained through *in vitro* glycosylation experiments with synthetic peptides and recombinantly expressed GalNAc-Ts⁶⁶, however these experiments do not reflect the complex interplay between the different members of the GalNAc-T family in cells¹⁸⁴. Furthermore, it is unclear whether the recognition of small peptides reflects the native activity of GalNAc-Ts on folded proteins within cells^{66, 184}. Several major advances in the field of LC-MS/MS analysis have facilitated proteome-wide mapping of O-GalNAc glycosylation sites, particularly the development of the so-called SimpleCells¹⁸⁶. KO of the core 1 chaperone COSMC in these cells results in truncated O-GalNAc glycans bearing the Tn or STn antigen (**Fig. 6A**), which are easier to enrich (e.g. using VVL) and study by MS^{186, 191, 192}. However, glycan elaboration information beyond the initiating GalNAc is lost with these cells¹⁹³. Furthermore, since there is no competition between downstream glycan elongating glycosyltransferases and the follow-up activity of GalNAc-Ts, these cells are expected to yield denser glycosylation and hence may not fully reflect the native glycosylation state of non-engineered cells^{184, 191}. Another significant advance was the development of MS fragmentation methods other than collision-based methods, such as electron transfer dissociation (ETD) which triggers fragmentation of the peptide backbone but leaves the attached glycans intact, allowing O-GalNAc site localisation^{183, 185}. However, O-glycopeptides often have a low charge density which impairs ETD fragmentation¹⁸². Additional advances in proteolytic digestion techniques, enrichment strategies and bioinformatic platforms are reviewed by Rangel-Angarita and Malaker, especially with a focus on the LC-MS/MS analysis of mucin-like domains¹⁸³. Altogether, these advances have allowed the identification of more than 10000 O-GalNAc glycosylation sites on more than 2000 proteins¹⁸⁴.

To profile the GalNAc-T-specific glycosylation sites, KO of GalNAc-Ts with and without the SimpleCell background and LC-MS/MS analysis of differentially expressed O-glycoproteins between WT and KO samples has been performed with several GalNAc-Ts and has even been extended to *ex vivo* analysis of tissues from WT or GalNAc-T-null mice^{129, 133, 136, 191, 194-197}. Nevertheless, the partially redundant and competitive activity of GalNAc-Ts can lead to compensation and/or shift of glycosylation sites upon KO of a single GalNAc-T¹⁹³, as illustrated by findings that show an increase of VVL-binding proteins upon KO of GalNAc-T4¹³³. Considering all the above, it is challenging to assign protein substrates and glycosylation sites to specific GalNAc-Ts in a biologically relevant context.

GalNAc-T Bump-and-Hole Engineering

“Bump-and-hole” (BH) engineering is a chemical biology approach used to study a specific enzyme of a protein family without perturbing the remaining members¹⁹⁸. In this approach, the enzyme is engineered to contain an enlarged active site (“hole”) that can accept a substrate analogue bearing a complementary bulky chemical group (“bump”)(**Fig. 7A**)⁶⁷. Successful BH engineering requires the BH enzyme-substrate pair to remain biochemically competent whilst orthogonal to the native enzyme-substrate pair, where the WT enzyme should only accept the native substrate and the BH enzyme should only accept the bumped analogue¹⁹⁸. BH engineering was originally established in protein kinases¹⁹⁹ and has since been applied to a wide range of enzymes including methyltransferases²⁰⁰ and ADP-ribosyltransferases²⁰¹ amongst others¹⁹⁸. Nevertheless, substrate delivery of the bumped analogues to the cells has remained a challenge due to the inability of negatively charged substrates to cross the plasma membrane – for this reason few BH systems have been successfully established in living cells¹⁹³.

To study the substrate specificity and map the glycosylation sites of individual GalNAc-Ts, the Bertozzi Lab extended the BH approach to the GalNAc-T family, initially focusing on GalNAc-T2⁶⁷. To generate the hole-modified GalNAc-T2, they mutated three highly conserved bulky gatekeeper residues (I253, L310 and F361)(**Fig. 7B**) to Ala and then compared the enzymatic activities of WT and mutant GalNAc-T2 with a panel of UDP-GalNAc analogues by *in vitro* glycosylation experiments⁶⁷. The results demonstrated that the single mutants still used UDP-GalNAc as a substrate whilst the double mutant (I253A/L310A, BH GalNAc-T2) displayed a switch in UDP-sugar substrate specificity from the native substrate to the bumped analogues⁶⁷.

To see if the GalNAc-T approach could be applied to other members of the family, they subsequently and successfully extended this approach to GalNAc-T1 and GalNAc-T10⁶⁷. From the panel of UDP-GalNAc analogues investigated, UDP-GalN6yne (which contains a 6-carbon chain and a terminal alkyne at the acetamide moiety, **Fig. 7C**) displayed the most promising results, as orthogonality was observed with all BH GalNAc-T/UDP-GalN6yne pairs, with kinetic parameters of the same order of magnitude as the native enzyme-substrate pairs⁶⁷.

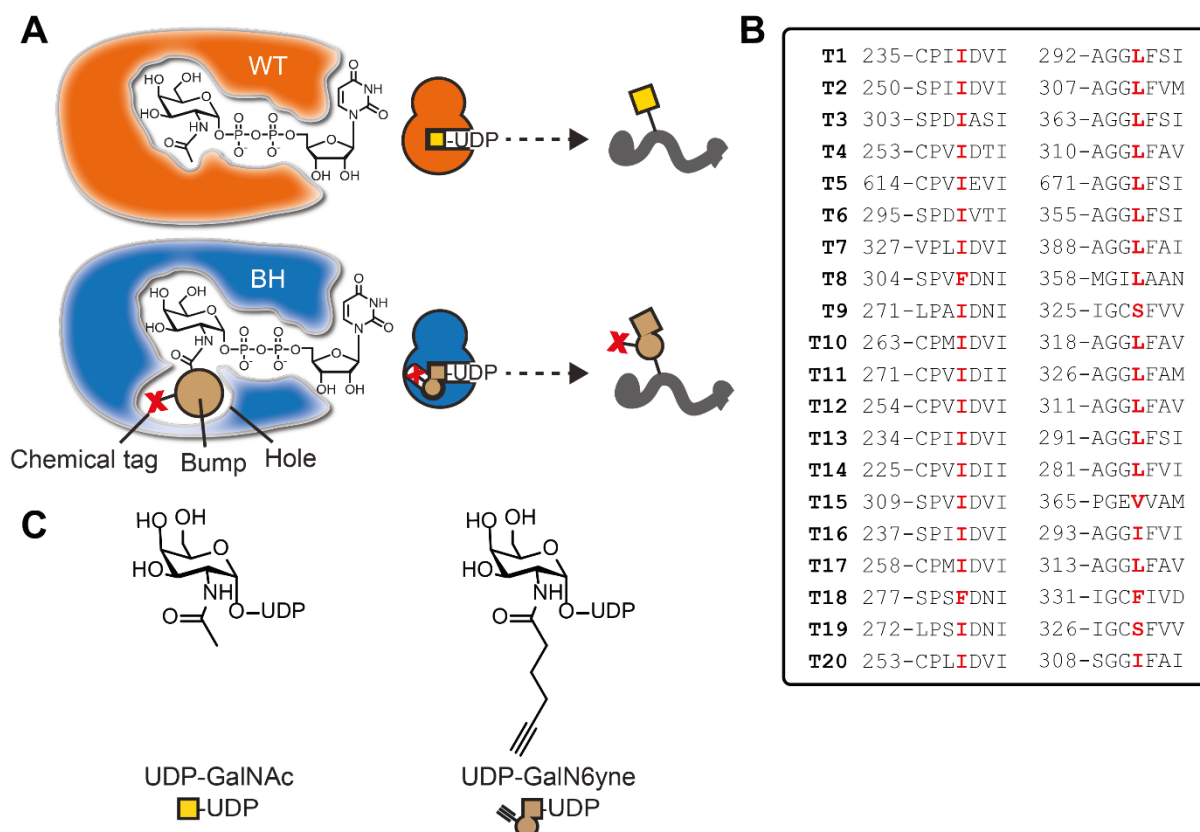


Fig. 7. GalNAc-T BH Engineering. **A)** Schematic representation of GalNAc-T BH engineering. Figure adapted from Cioce, Malaker and Schumann (2021)²⁰². **B)** Sequence alignment of gatekeeper residues in the GalNAc-T family. In red are the gatekeeper residues that can be mutated to Ala to generate the hole-modified GalNAc-T. Figure adapted from Calle et al. (2023).²⁰³ **C)** Chemical structures of the native substrate of GalNAc-Ts (UDP-GalNAc) and the bumped analogue UDP-GalN6yne.

In a follow up study, a co-crystal structure of BH GalNAc-T2 with UDP-GalN6yne revealed that UDP-sugar binding was completely conserved (**Fig. 8A**)¹⁹³. Similarly, a crystal structure of BH GalNAc-T2 with the peptide substrate EA2 confirmed that the three-dimensional enzyme architecture and bound ligand structure of WT and BH GalNAc-T2 were also conserved¹⁹³. The substrate specificity of BH GalNAc-T1 and GalNAc-T2 was subsequently investigated by *in vitro* glycosylation experiments¹⁹³. The results demonstrated that the BH GalNAc-Ts recapitulated the glycosylation profiles of the WT GalNAc-Ts, although the BH GalNAc-Ts glycosylated fewer peptides than the corresponding WT enzymes¹⁹³. Using fluorescence microscopy, they also demonstrated that BH GalNAc-Ts localised to the Golgi compartment, confirming their native localisation¹⁹³.

In the same study, the GalNAc-T BH approach was extended to living cells¹⁹³. Biosynthesis of UDP-GalNAc in cells occurs by two mechanisms: epimerisation of UDP-GlcNAc to UDP-GalNAc by the enzyme UDP-glucose 4-epimerase (GALE), or the GalNAc salvage pathway, where the kinase GALK2 converts the salvaged GalNAc into GalNAc-1-P, which is then transformed into UDP-GalNAc by the pyrophosphorylase AGX1 (**Fig. 9A**)^{63, 204}. In this study, a membrane-permeable peracetylated GalNAc-1-P analogue (Ac₃GalN6yne-1-P(SATE)₂) was used to overcome substrate delivery issues and to bypass the phosphorylation step catalysed by GALK2, such that only AGX1 was required to biosynthesise the bumped UDP-GalNAc analogue (**Fig. 8B**)¹⁹³.

Previous work by Yu et al. had demonstrated that a GlcNAc-1-P analogue was a poor substrate of WT AGX1, whilst engineered AGX1 allowed biosynthesis of the corresponding UDP-GlcNAc analogue²⁰⁵. Inspired by these results, Schumann et al. mutated the F381 and F383 gatekeeper residues of AGX1 to Ala or Gly¹⁹³. The biosynthesis of UDP-GalN6yne from the corresponding GalNAc-1-P analogue with both F383A and F383G single mutants but neither F381 single mutant nor the F381/F383 double mutant was confirmed in cells¹⁹³. In these experiments they also realised that UDP-GalN6yne was epimerised to the corresponding UDP-GlcNAc analogue by GALE, which could result in undesired labelling of GlcNAc-containing glycoconjugates such as extracellular N-glycans¹⁹³. AGX1^{F383A} was chosen for future experiments since it accepted a broader range of GalNAc-1-P analogues compared to the F383G mutant²⁰².

K562 cells were stably transfected with plasmids encoding FLAG-tagged WT AGX1 or AGX1^{F383A} and Vesicular Stomatitis Virus Glycoprotein (VSV-G)-tagged WT or BH GalNAc-T1 or GalNAc-T2¹⁹³. Stably transfected cells were treated with Ac₃GalN6yne-1-P(SATE)₂¹⁹³. Upon entering the cytosol, esterases cleaved the protecting groups which allowed UDP-sugar substrate biosynthesis to take place within the cells¹⁹³. The BH GalNAc-T then transferred the modified sugar to its substrates, followed by uptake of the chemically-tagged glycoproteins by the secretory pathway¹⁹³. The chemically-tagged cell surface glycoproteins were then derivatised by copper-catalysed alkyne-azide cycloaddition (CuAAC) with a fluorophore (**Fig. 8B**)¹⁹³. This reaction occurred in the presence of BTAA, a biocompatible Cu(I)-stabilising ligand which accelerates the CuAAC reaction and protects cells from oxidative damage²⁰⁶. The fluorophore used for these experiments was a picolyl azide derivative since these Cu(I)-chelating organic azides also markedly accelerate the CuAAC reaction²⁰⁷. PNGase F treatment was performed on the cell lysates to remove N-glycans and reduce background fluorescence¹⁹³. The chemically-tagged glycoproteins were visualised by in-gel fluorescence¹⁹³. In the absence of AGX1^{F383A} or BH GalNAc-T no fluorescent signal was observed, demonstrating that the experiments probed enzymatic glycosylation rather non-specific protein modification¹⁹³. Furthermore, BH GalNAc-T1 and GalNAc-T2 produced different band patterns, reflecting the different substrate specificities¹⁹³.

To characterise the substrates and glycosylation sites of the two GalNAc-Ts, GalNAc-T1 KO and GalNAc-T2 KO HepG2 cells were transfected with AGX1^{F383A} and the BH GalNAc-Ts¹⁹³. The stably transfected cells were then treated with dimethyl sulfoxide (DMSO) or Ac₃GalN6yne-1-P(SATE)₂¹⁹³. The secretome was collected, the chemically-tagged glycoproteins derivatised with biotin-dialkoxydiphenylsilane (DADPS)-picolyl azide and enriched¹⁹³. Since biotin-DADPS-picolyl azide contains an acid-labile diphenyldisiloxane moiety, elution of the tagged glycopeptides from the beads can be achieved by acid treatment¹⁹³. Following on-bead digest to release non-glycosylated peptides, glycopeptides were released by acid treatment and analysed by LC-MS/MS (**Fig. 8C**)¹⁹³. No glycopeptide hits were obtained in the cells treated with DMSO¹⁹³. In contrast, 37 peptides with 27 glycosylation sites and up to 4 different glycan compositions per glycopeptide were detected with BH GalNAc-T2, whilst 4 glycopeptides were detected with BH GalNAc-T1¹⁹³. Several known glycosylation sites from the SimpleCell MS-glycoproteomics dataset were confirmed, and new sites revealed¹⁹³. Overall, this approach allowed the determination of GalNAc-T-specific structural and site information in a single experiment for the first time¹⁹³.

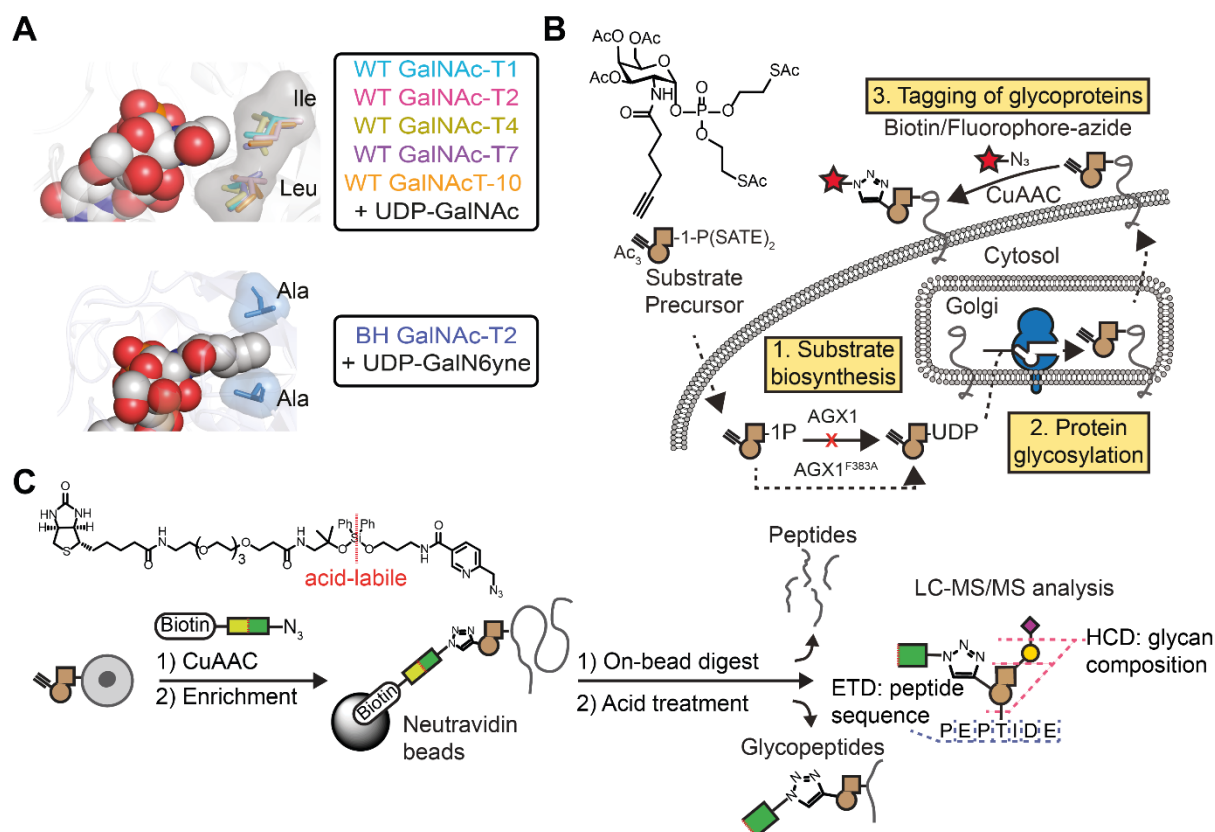


Fig. 8. Extending the GalNAc-T BH approach to living cells. **A)** UDP-sugar binding is conserved between WT and BH GalNAc-T/UDP-sugar pairs. The figure shows the gatekeeper residues identified in the crystal structures of WT GalNAc-T1 (PDB 1XHB), -T2 (PDB 4D0T), -T4 (PDB 5NQA), -T7 (PDB 6IWR), -T10 (PDB 2D7I) and BH GalNAc-T2 (PDB 6NQT). Figure adapted from Calle et al. (2023)²⁰³. **B)** Schematic representation of the steps required to achieve the tagging of glycoproteins by BH GalNAc-Ts in cells. Figure adapted from Schumann et al. (2020)¹⁹³. **C)** Characterisation of the GalNAc-T-specific substrates and glycosylation sites by MS-glycoproteomics. Figure adapted from Schumann et al. (2020)¹⁹³.

Since the GalNAc-T BH system was established in cells, efforts in the Schumann lab were focused on developing alternative substrate delivery strategies because Ac₃GalN6yne-1-P(SATE)₂ had a poor stability in cells and a time-consuming synthesis²⁰⁸. Other groups had successfully used peracetylated GalNAc analogues bearing small modifications (e.g. GalNAz) for the biosynthesis of the corresponding GalNAc-1-P analogues by the endogenous GALK2^{209, 210}. However, analogues with bigger side chains at the acetamide moiety showed reduced activity with GALK2²¹¹, so alternative kinases were investigated instead by the Schumann lab which would allow biosynthesis of GalN6yne-1-P from GalN6yne. The bacterial *N*-acetylhexosamine 1-kinase (NahK) from *Bifidobacterium longum*²¹² was identified as a promising candidate for this purpose, as this enzyme shows similar specificity and kinetic parameters for GalNAc as GALK2²¹¹, as well as a broad substrate promiscuity towards GalNAc analogues²¹². Indeed, NahK converted GalN6yne to GalN6yne-1-P almost quantitatively (**Fig. 9A**)²⁰⁴.

K562 cells were subsequently stably transfected with plasmids encoding hemagglutinin (HA)-tagged NahK, FLAG-tagged WT or AGX1^{F383A}, and VSV-tagged WT or BH GalNAc-T1 or GalNAc-T2²⁰⁴. The cells were treated with peracetylated GalN6yne (Ac₄GalN6yne)(**Fig. 9B**), which contains acetyl protecting groups on the hydroxyl groups, rendering the compound membrane-permeable²⁰⁴. The chemically-tagged cell surface glycoproteins were derivatised by CuAAC with a fluorophore, and then visualised by in-gel fluorescence²⁰⁴. No fluorescent signal was observed in the absence of NahK or AGX1^{F383A}, demonstrating that the signal was dependent on the biosynthesis of UDP-GalN6yne²⁰⁴. The fluorescent signal intensity in BH GalNAc-T samples was over seven-fold greater than that of the WT GalNAc-T samples²⁰⁴. Furthermore, the fluorescent signal in the latter was sensitive to PNGase F treatment, demonstrating that the signal in the WT GalNAc-T samples was primarily due to background fluorescence from N-glycans²⁰⁴. Incomplete signal abrogation by PNGase F could be due to incorporation of the GlcNAc analogue into the protein-proximal core, making N-glycans PNGase F resistant²⁰⁴. Successful identification of the substrates and glycosylation sites of BH GalNAc-T2 was subsequently performed by MS-glycoproteomics analysis in monolayer and co-culture cellular models²⁰⁴.

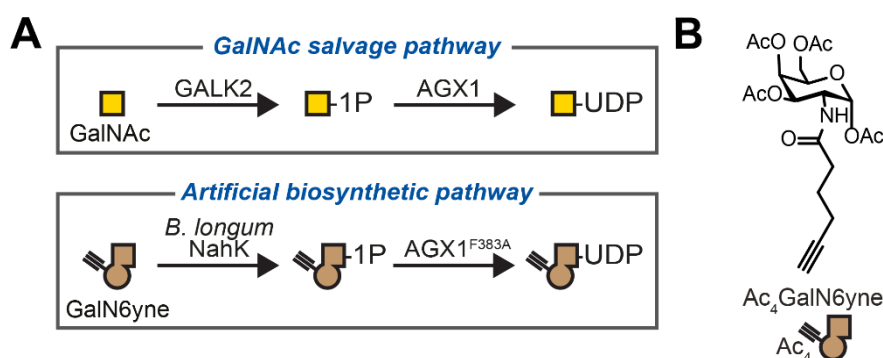


Fig. 9. UDP-GalN6yne biosynthesis using peracetylated GalNAc analogues. A) The GalNAc salvage pathway and the artificial biosynthetic pathway for UDP-GalN6yne biosynthesis. Figure adapted from Gonzalez-Rodriguez et al. (2023)²¹³. **B)** Chemical structure of Ac₄GalN6yne.

Cioce et al. also demonstrated that neither the presence of the artificial biosynthetic pathway nor compound feeding had substantial effects on the transcriptome of transfected cells, the ratio of UDP-GlcNAc/GalNAc, or native UDP-sugar levels²⁰⁴. Furthermore, no differences in binding patterns were observed with the lectins *Concanavalin A* (binds to mannose residues), *Sambucus Nigra* (binds to α 2-6-linked sialic acid) and *Aleuria Aurantia lectin* (binds to α 1-2-linked fucose) or the RL2 antibody (binds to nucleocytoplasmic O-linked O-GlcNAc glycans)²⁰⁴. In contrast, reduced binding was observed with *Maackia Amurensis II* (binds to α 2-3-linked sialic acid) upon BH GalNAc-T2 overexpression, indicating that overexpression of GalNAc-Ts results in slight alterations to the cellular glycome²⁰⁴.

During my MSci project, I extended the GalNAc-T BH system to GalNAc-T4, one of the glycopeptide-preferring GalNAc-Ts⁶⁸. Since the GalNAc-T family has a well conserved UDP-GalNAc binding site (**Fig. 8A**)⁶⁷, I mutated the same amino acids that were previously mutated for GalNAc-T1, -T2 and -T10 to generate BH GalNAc-T4 (I256A/L313A - the amino acid positions are based on the full-length sequence)(**Fig. 7B**)²⁰⁸. I performed the cloning on full-length (containing the transmembrane, catalytic and lectin domains) and truncated (lacking amino acids 1-46 including the transmembrane domain) constructs²⁰⁸. I used the full-length construct to establish the BH system in living cells as it was more representative of the native form of GalNAc-T4, whilst I used the truncated construct for recombinant protein expression and *in vitro* glycosylation experiments as secreted proteins offer easier expression in mammalian cells compared to transmembrane proteins²⁰⁸.

To recombinantly express truncated WT and BH GalNAc-T4, I transiently transfected Expi293F cells (derived from a line of human embryonic kidney cells) with plasmids encoding truncated WT or BH GalNAc-T4 tagged with a green fluorescent protein (GFP) (**Fig. 10A**)²⁰⁸. I collected the secretome of the cells and performed GFP-purification using sepharose beads coupled to a GFP-clamp consisting of two fused GFP-specific designed ankyrin repeat proteins²¹⁴. To cleave the protein from the beads, I performed an incubation with Tobacco etch virus (TEV) protease, followed by two washes to collect all the eluted protein²⁰⁸.

To compare the enzymatic activities of WT and BH GalNAc-T4, I performed an endpoint *in vitro* glycosylation experiment²⁰⁸. The reaction mixtures contained WT or BH GalNAc-T4, a synthetic glycopeptide substrate (glycopeptide **1**), and UDP-GalNAc or UDP-GalN6yne (**Fig. 10B**)²⁰⁸. Glycopeptide **1** was chosen as GalNAc-T4 showed the highest activity with this glycopeptide in a previous study by de las Rivas et al. which investigated the *in vitro* activity of GalNAc-T4⁸². The results from the endpoint *in vitro* glycosylation experiment demonstrated that comparable glycopeptide formation was observed between the BH GalNAc-T4/UDP-GalN6yne and WT GalNAc-T4/UDP-GalNAc pairs, whilst negligible activity was observed with the opposite GalNAc-T4/UDP-sugar pairs, demonstrating a switch in nucleotide-sugar specificity from the native substrate to the bumped analogue by BH GalNAc-T4²⁰⁸.

To confirm that BH engineering did not affect GalNAc-T4's peptide binding ability and specificity, I performed *in vitro* glycosylation experiments using the panel of synthetic glycopeptide substrates previously used by de las Rivas et al.⁸² (**Fig. 10B**). Since the results from this study demonstrated that GalNAc-T4 shows poor activity with naked peptides or glycopeptides containing C-terminal prior glycosylation⁸², I only used the glycopeptides containing N-terminal prior glycosylation for these experiments²⁰⁸. Glycopeptide **1** was a monoglycopeptide containing a remote prior GalNAc unit, glycopeptide **2** was a diglycopeptide containing remote and neighbouring prior GalNAc units, and glycopeptide **3** was a monoglycopeptide containing a neighbouring prior GalNAc unit (**Fig. 10B**)^{82, 208}. Glycosylation was observed with all glycopeptides, demonstrating that GalNAc-T4 displays catalytic domain-dependent and lectin domain-dependent glycosylation of glycopeptides (**Fig. 3A**)²⁰⁸. The lowest activity was observed with glycopeptide **3** which could only be glycosylated by catalytic domain-dependent glycosylation, confirming that GalNAc-T4's predominant mode of glycosylation is lectin domain-dependent glycosylation²⁰⁸. A lower activity was observed with glycopeptide **2** than glycopeptide **1**²⁰⁸. A possible reason for the reduced activity with glycopeptide **2** could be substrate inhibition due to nonproductive binding of the neighbouring GalNAc competing with productive binding of the remote GalNAc to the lectin domain, which would not occur with glycopeptide **1** which only contains the remote GalNAc unit²⁰⁸. A slightly lower activity was seen with the native enzyme-substrate pair compared to the BH pair with all glycopeptides tested²⁰⁸. A possible explanation could be product inhibition taking place with the WT pair due to nonproductive binding of the lectin domain to the GalNAc unit transferred by WT GalNAc-T4 to the acceptor site²⁰⁸. Since GalN6yne should only be accepted by the catalytic domain but not the lectin domain, product inhibition due to nonproductive binding of the lectin domain to GalN6yne should not occur with the BH pair²⁰⁸. Altogether, the results supported the findings reported by de las Rivas et al.⁸² and demonstrated that BH GalNAc-T4 emulated the known glycopeptide substrate preference of WT GalNAc-T4²⁰⁸.

With an orthogonal BH GalNAc-T4/UDP-sugar pair in hand, the next aim was to probe GalNAc-T4-mediated O-GalNAcylation in living cells²⁰⁸. I therefore transfected K562 cells with plasmids encoding HA-tagged NahK, FLAG-tagged WT or AGX1^{F383A}, and VSV-G-tagged full-length WT or BH GalNAc-T4 (**Fig. 10C**)²⁰⁸. These plasmids also contained a hygromycin resistance gene, so generation of stable colonies was achieved by hygromycin treatment²⁰⁸. I fed the stably transfected cells with Ac₄GalN6yne, derivatised the chemically-tagged cell surface glycoproteins by CuAAC with a fluorophore, and then visualised the chemically-tagged glycoproteins by in-gel fluorescence after treatment with PNGase F (**Fig. 10D**)²⁰⁸. No fluorescent signal was observed in DMSO-treated cells or in Ac₄GalN6yne-treated cells where NahK or AGX1^{F383A} were absent (**Fig. 10E**)²⁰⁸. In contrast, multiple intense bands were observed with BH GalNAc-T4 and not the WT enzyme (**Fig. 10E**), demonstrating successful and selective tagging of glycoproteins in cells by GalNAc-T4²⁰⁸.

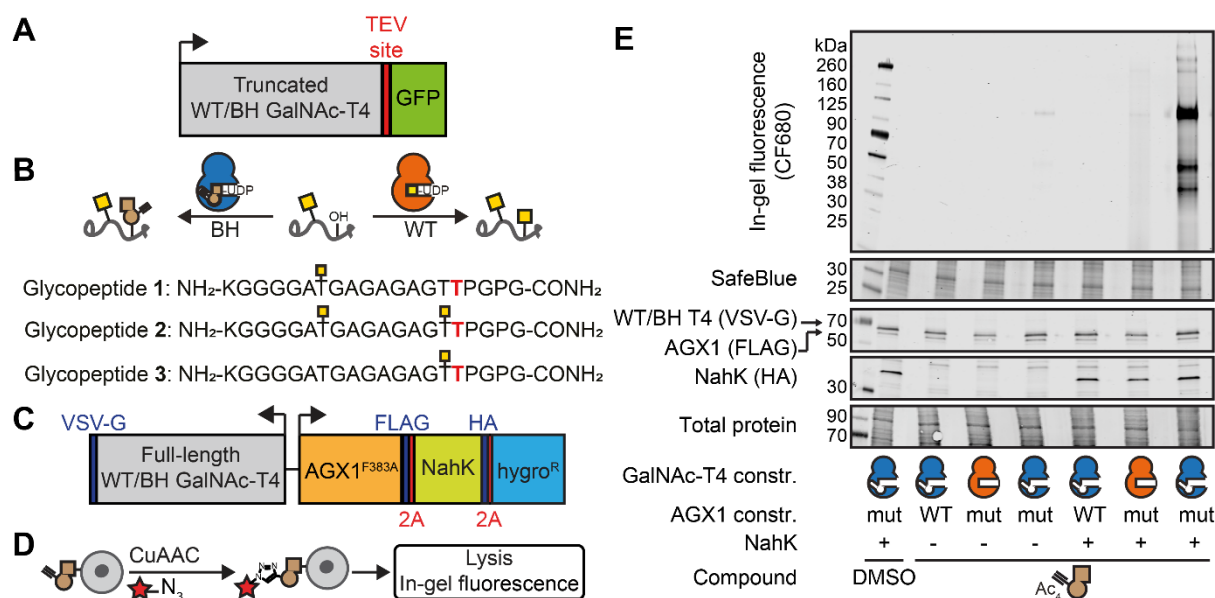


Fig. 10. Extending the GalNAc-T BH system to GalNAc-T4. **A)** Expression construct used to recombinantly express WT and BH GalNAc-T in Expi293F cells. **B)** Schematic representation of *in vitro* glycosylation experiments performed to evaluate the enzymatic activities of WT and BH GalNAc-T4. Red T represents the peptide acceptor site. The yellow square depicts prior GalNAc units. **C)** Expression construct used to establish the BH GalNAc-T4 system in K562 cells. The 2A self-cleaving peptides (2A) allow efficient expression of multiple proteins from a single plasmid²¹⁵. Figure adapted from Schumann et al. (2020)¹⁹³. **D)** Schematic representation of the workflow for the cell surface tagging experiments. The red star represents a fluorophore. Figure adapted from Schumann et al. (2020)¹⁹³. **E)** Cell surface tagging of glycoproteins in transfected PNGase F-treated K562 cell lysates as assessed by in-gel fluorescence. SafeBlue staining and the REVERT protein staining kit were used to assess total protein content. DMSO-treated cells were included as a negative control. The predicted molecular weight of GalNAc-T4, AGX1 and NahK is 67, 58 and 41 kDa respectively. Figure adapted from Calle (2020)²⁰⁸. AGX1 mut = AGX1^{F383A}.

PROJECT AIMS AND OBJECTIVES

With a strategy in hand to probe GalNAc-T4-specific glycosylation in living cells, the overall aim of this project was to use the BH GalNAc-T4 system to characterise the substrates and glycosylation sites of GalNAc-T4 in BCCs to better understand the role of GalNAc-T4 in BC. We aimed to use the BH GalNAc-T approach to validate the proteins that have been previously proposed as substrates of GalNAc-T4, as well as identifying other substrates playing a role in BC. Furthermore, we also aimed to determine the GalNAc-T4-mediated glycosylation sites in the BC-relevant substrates, as this could help inform on therapeutic strategies and could help draw structural conclusions on the role of GalNAc-T4-mediated glycosylation in BC.

To achieve these aims, the objectives of this project were:

1. To validate that GalNAc-T4 was implicated in BC by generating GalNAc-T4 KO T47D and MCF7 cells and comparing the growth of WT and GalNAc-T4 KO cells *in vitro* and *in vivo*.
2. To evaluate the enzymatic activities of WT and BH GalNAc-T4 *in vitro* before extending the BH GalNAc-T4 system to BCCs.
3. To establish the BH GalNAc-T4 system in BCCs.
4. To characterise the substrates and glycosylation sites of GalNAc-T4 in BCCs by MS-glycoproteomics.
5. To evaluate the role of GalNAc-T4-mediated glycosylation in BC using *in silico*, *in cellulo* and *in vivo* methods.

METHODS

A list with all the reagents and resources used in this project as well as the supplier information can be found in **SI Table 1**.

1. Generation of GalNAc-T4 KO BCCs

1.1. Cloning of plasmids for CRISPR-Cas9 KO of GalNAc-T4

Two single guide RNAs (sgRNAs) oligonucleotides targeting the gene encoding human GalNAc-T4 were cloned into a pSpCas9(BB)-2A-GFP plasmid. The pSpCas9(BB)-2A-GFP plasmid (1 µg) was digested using BbsI in 10X CutSmart Buffer (50 µL reaction volume) for 30 min at 37 °C. The digested plasmid was run on an agarose gel and then purified using the QIAquick Gel Extraction Kit.

The two sgRNA oligonucleotides were phosphorylated and annealed by mixing together the forward (10 µM) and reverse (10 µM) oligonucleotides with T4 Polynucleotide Kinase in 10X T4 Ligation Buffer (20 µL reaction volume) and incubating the samples in a thermocycler using the following conditions: incubation at 37 °C for 30 min, followed by incubation at 95 °C for 5 min, ramp down to 25 °C at 0.1 °C/s and then hold at 4 °C. The annealed oligonucleotides (1:250 dilution) were then ligated to the digested plasmid (50 ng) by incubation with Quick Ligase at room temperature (RT) for 10 min. Sanger sequencing confirmed the successful insertion of the sgRNA into the plasmid.

1.2. Generation of GalNAc-T4 KO BCCs by CRISPR-Cas9 genome editing

T47D and MCF7 cells were seeded in 2.5 mL growth medium per well of a 6-well plate (cell density of T47D cells: 5×10^5 viable cells/mL; MCF7 cells: 4×10^5 viable cells/mL). For T47D cells the growth medium used was RPMI with 10% (v/v) foetal bovine serum (FBS), 100 U/mL penicillin and 100 mg/mL streptomycin. For MCF7 cells the growth medium used was DMEM with 10% (v/v) FBS, 100 U/mL penicillin and 100 mg/mL streptomycin. After 24 h, the cells were transfected with sgRNA-coding plasmids (2.5 µg) using the TransIT-BrCa Transfection Reagent (5 µL) according to the manufacturer's instructions. After 72 h, GFP+ cells were isolated as single cells into 96-well plates by the Flow Cytometry STP at the Francis Crick Institute. Clones were expanded and GalNAc-T4 KO was confirmed by Western blot and amplicon sequencing. The sequencing data was analysed using the online tool CRISPResso2²¹⁶.

2. *In vitro* growth assays with WT and GalNAc-T4 KO BCCs

WT and GalNAc-T4 KO T47D and MCF7 cells were seeded at a density of 4×10^4 viable cells/mL in 100 μ L growth medium per well of a 96-well plate. After 24 h, the cells were treated with 5 μ M of DEVD-NucView488 peptide and cell growth was monitored using an Incucyte® Live-Cell Analysis System. 5 phase contrast and fluorescent images per well were captured every 12 h for 4 days using a 10X objective. Confluence measurements and fluorescent object quantification was performed using the integrated analysis software. Statistical analysis was performed using GraphPad Prism software. A one-way ANOVA test was performed for the T47D cells. An unpaired t-test was performed for the MCF7 cells.

3. *In vivo* growth assays with WT and GalNAc-T4 KO BCCs

An 850 μ g/mL solution of β -oestradiol was prepared in 20% ethanol/sterile water by Dr Victoria Bridgeman. 1 mL of this solution was added to 99 mL of mice drinking water two days before tumour implantation by Dr Victoria Bridgeman. WT and GalNAc-T4 KO T47D and MCF7 cells were resuspended in 50 μ L of growth-factor-reduced (GFR) Matrigel and injected into the hind fat pad of female NOD-SCID IL2Rgnull (NSG) mice (cells per injection for T47D cells: 3×10^6 cells; for MCF7 cells: 5×10^6 cells; 5 mice per condition) by Dr Victoria Bridgeman. Body weight and tumour dimensions were monitored twice weekly by Dr Victoria Bridgeman. Tumour volume was calculated using **Equation 1** by Dr Victoria Bridgeman. Once the tumours measured $>200 \text{ mm}^3$ the mice were harvested, and the tumour weight recorded by Dr Victoria Bridgeman. Statistical analysis was performed using GraphPad Prism software. An unpaired t-test was performed on the tumour weights at endpoint.

Equation 1. Formula for tumour volume. The value “Length” corresponds to the longer of the two perpendicular axes²¹⁷.

$$\text{Tumour volume (mm}^3\text{)} = \frac{\text{Length} \times \text{Width}^2}{2}$$

4. *In vitro* glycosylation experiments with WT/BH GalNAc-T4 and UDP-sugars 1-16

4.1. *In vitro* glycosylation reaction with WT/BH GalNAc-T4 and UDP-sugars 1-16

Reaction mixtures (25 μ L) were prepared containing 20 mM Tris-HCl pH 7.5, 50 mM sodium chloride (NaCl), 10 mM manganese chloride (MnCl_2), 50 μ M glycopeptide **1**, 250 μ M UDP-sugar **1-16** and WT or BH GalNAc-T4 (100 nM). The reaction was incubated at 37 °C and quenched by heating at 95 °C for 10 s. Glycopeptide formation was determined by LC-MS and MS+ peak integration using **Equation 2**.

Equation 2. Determining glycopeptide formation from the *in vitro* glycosylation experiments²⁰³.

$$\text{Glycopeptide formation (\%)} = \frac{\text{Peak area of product}}{\text{Peak area of starting material} + \text{Peak area of product}} \times 100$$

4.2. LC-MS analysis of synthetic glycopeptides

The samples were analysed using an ACQUITY™ QDa™ Detector connected to an ACQUITY™ H-Class PLUS (CH-A) Core System. The samples were loaded onto a ACQUITY™ UPLC® BEH C18 column (1.7 µm particle size, 2.1 mm x 50 mm). The solvents used were **A** (0.1% (v/v) formic acid in LC-MS grade water) and **B** (0.1% (v/v) formic acid in LC-MS grade acetonitrile). The elution profile used was a linear gradient from 3% to 95% **B** over 3.6 min, followed by wash and re-equilibration at 3% **B** for 0.4 min. The flow rate was 0.5 mL/min. Starting material and product species were detected as doubly charged protonated ions.

5. Michaelis-Menten Kinetics

Reaction mixtures (25 µL) were prepared containing 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MnCl₂, 100 µM glycopeptide **1**, UDP-GalNAc or UDP-GalN6yne (15.6, 32.2, 62.5, 125, 250, 500 µM), and WT or BH GalNAc-T4 (4 nM). The reaction was incubated at 37 °C for 1.5 h and quenched by heating at 95 °C for 10 s. Glycopeptide formation was determined by LC-MS and MS+ peak integration using **Equation 2**. LC-MS analysis was performed as described in **Section 4.2**. The reaction rate was calculated by dividing the measured glycopeptide formation over the time of the reaction (5400 s). The non-linear Michaelis–Menten and k_{cat} fitting programs found in GraphPad Prism were used to calculate the kinetic constants of K_M and k_{cat} .

6. Hygromycin kill curve of BCCs

T47D^{WT}, T47D^{T4-KO 1}, T47D^{T4-KO 6}, and MDA-MB-231^{WT} cells were seeded in 1 mL growth medium per well of a 12-well plate (cell density of T47D cells: 2.5x10⁵ viable cells/mL; MDA-MB-231 cells: 4x10⁵ viable cells/mL). For MDA-MB-231 cells the growth medium used was DMEM with 10% (v/v) FBS, 100 U/mL penicillin and 100 mg/mL streptomycin. After 24 h, the medium was replaced with fresh growth medium containing hygromycin B (50-500 µg/mL). The medium was replaced every 2-3 days with fresh selection medium. After 10-15 days of antibiotic treatment, visual toxicity was assessed and the lowest antibiotic concentration at which all cells were dead determined.

7. Stable transfection of BH GalNAc-T4 into BCCs

BCCs were seeded in 2.5 mL growth medium per well of a 6-well plate (cell density of T47D cells: 5×10^5 viable cells/mL; MCF7 cells: 4×10^5 viable cells/mL; MDA-MB-231 cells: 3.5×10^5 viable cells/mL). After 24 h, the cells were transfected with pSBbi-based plasmids (2.4 μ g) and pCMV(CAT)T7-SB100 plasmid (125 ng) using the TransIT-BrCa Transfection Reagent (5 μ L) according to the manufacturer's instructions. The pSBbi plasmids used herein were: pSBbi-AGX1^{F383A}-NahK, pSBbi-AGX1^{F383A}-NahK-GalNAc-T4^{WT} and pSBbi-AGX1^{F383A}-NahK-GalNAc-T4^{BH}. After 48-72 h, the medium was replaced with fresh growth medium containing hygromycin B (for T47D cells: 50 μ g/mL hygromycin B; for MCF7 cells: 100 μ g/mL hygromycin B; for MDA-MB-231 cells: 150 μ g/mL hygromycin B). The antibiotic selection was maintained until stable colonies were generated.

8. Cell surface tagging of glycoproteins in transfected BCCs and in-gel fluorescence analysis

BCCs (transfected T47D, MCF7 and MDA-MB-231 cells from **Section 7**) were seeded at a density of 5×10^5 viable cells/mL in 1.6 mL growth medium per well of a 6-well plate. After 24 h, the cells were treated with Ac₄GalN6yne (1, 5, 10 or 25 μ M), Ac₄ManNAI (10 μ M) or DMSO and incubated overnight at 37 °C. The cells were washed once with PBS without calcium chloride (CaCl₂) and magnesium chloride (MgCl₂), incubated for 5 min with 4 mM ethylenediaminetetraacetic acid (EDTA) in 2% FBS in PBS (Cell Buffer) and detached with a 1 mL pipette. The cell suspension was collected, centrifuged (500 $\times$ g, 5 min, 4 °C) and the pellet resuspended in 200 μ L of Cell Buffer. The samples were washed twice with 200 μ L of Cell Buffer and resuspended in 35 μ L of Cell Buffer. The samples were treated with 35 μ L of a 2X CuAAC solution (1 mM BTAA, 200 μ M copper sulfate (CuSO₄), 10 mM sodium ascorbate, 10 mM aminoguanidine hydrochloride and 200 μ M CF680-picolyl azide) and incubated for 7 min at RT on an orbital shaker. The reaction was quenched by addition of 35 μ L of 3 mM bathocuproinedisulfonic acid in PBS. The samples were centrifuged (500 $\times$ g, 3 min, 4 °C), washed twice with 200 μ L of Cell Buffer and resuspended in 100 μ L of ice-cold Lysis Buffer **A** (RIPA buffer, 100 mU/ μ L benzonase and 1X HaltTM protease inhibitors). The cells were lysed for 20 min at 4 °C on an orbital shaker and centrifuged (1500 $\times$ g, 20 min, 4 °C). The supernatant was collected, and the protein concentration determined using the PierceTM BCA Protein Assay Kit.

Loading buffer (a 2:2:1:1 (v/v/v/v) mixture of 1 M Tris-HCl pH 6.5, 80% (v/v) glycerol, 20% (w/v) sodium dodecyl sulfate (SDS) and 1 M dithiothreitol (DTT)) was added to the cell lysates and the samples were run for SDS-polyacrylamide gel electrophoresis (PAGE) using a 4-20% Criterion™ TGX™ Precast gel. Total protein content was assessed by SafeBlue Protein Stain. A technical replicate of the SDS-PAGE gel was run and used for protein expression control by Western blot. All imaging was performed on an Odyssey CLx. The images were visualized and processed on ImageStudioLite software and cropped using Adobe Illustrator.

9. Tagging of glycoproteins in lysates of transfected T47D cells and streptavidin blot analysis

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells were seeded at a density of 5×10^5 viable cells/mL in 1.6 mL growth medium per well of a 6-well plate. After 24 h, the cells were treated with either 1 μ M Ac₄GalN6yne, 10 μ M Ac₄ManNAI or DMSO and incubated overnight at 37 °C. The cells were washed once with PBS without CaCl₂ and MgCl₂, incubated for 5 min with 4 mM EDTA in PBS without CaCl₂ and MgCl₂, and detached with a 1 mL pipette. The cell suspension was collected, centrifuged ($500 \times g$, 5 min, 4 °C) and the pellet resuspended in 200 μ L of PBS without CaCl₂ and MgCl₂. The samples were washed twice with 200 μ L of PBS without CaCl₂ and MgCl₂ and the pellet resuspended in 100 μ L of ice-cold Lysis Buffer **A** (Section 8). The cells were lysed for 20 min at 4 °C on an orbital shaker and centrifuged ($1500 \times g$, 20 min, 4 °C). The supernatant was collected, and the protein concentration was determined using the Pierce™ BCA Protein Assay Kit.

The samples were incubated with 5 U of PNGase F overnight at 37 °C, treated with a 10X CuAAC solution (6 mM BTAA, 3 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-picolyl azide) and incubated overnight at RT on an orbital shaker at 400 rpm. Loading buffer (Section 8) was added and the samples were run for SDS- PAGE using a 4-20% Criterion™ TGX™ Precast gel. The gel was transferred to a nitrocellulose membrane using the Trans-Blot Turbo RTA Midi 0.2 μ m Nitrocellulose Transfer Kit. Total protein content was assessed using the Revert™ 700 Total Protein Stain Kit. Glycoprotein tagging was assessed using IRDye 800CW Streptavidin according to the manufacturer's instructions. All imaging was performed on an Odyssey CLx. The images were visualized and processed on ImageStudioLite software and cropped using Adobe Illustrator.

10. Tagging of glycoproteins in the secretome of transfected T47D cells and streptavidin blot analysis

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells were seeded at a density of 5×10^5 viable cells/mL in 1.6 mL growth medium per well of a 6-well plate. After 24 h, the growth medium was replaced with FBS-free medium and the cells were treated with either 1 μ M Ac₄GalN6yne, 10 μ M Ac₄ManNAI or DMSO. After overnight incubation at 37 °C, the secretome was collected and centrifuged ($500 \times g$, 5 min, 4 °C). The samples were concentrated and the buffer exchanged with PBS using Amicon Ultra Centrifugal Filters (3 kDa MWCO). Protein concentration was determined using the PierceTM BCA Protein Assay Kit. PNGase F treatment, CuAAC reaction and glycoprotein tagging analysis by streptavidin blot was performed as described in **Section 9**.

11. Label-free MS-proteomics experiment on lysates and secretome of transfected T47D cells

11.1. Sample preparation for label-free MS-proteomics

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells were seeded at a density of 4×10^5 viable cells/mL in 20 mL growth medium into 15 cm dishes (one dish for each lysate sample, two dishes for each secretome sample). After 24 h, the cells were treated with 1 μ M Ac₄GalN6yne or DMSO and incubated overnight at 37 °C. Lysates and secretome samples were prepared as described in **Section 9** and **10** respectively. The samples (500 μ g) were diluted to 300 μ L in PBS and incubated with 300 μ L of neutravidin-coated magnetic beads for 2 h at RT under rotation to remove endogenous biotinylated proteins. The supernatant was collected, the samples incubated with 5 U of PNGase F overnight at 37 °C, and the reaction quenched by heating at 95 °C for 10 s. The samples were then treated with a 10X CuAAC solution (12 mM BTAA, 6 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-DADPS-picolyl azide) and incubated at RT on an orbital shaker at 400 rpm. After overnight incubation, the samples were treated with a fresh 10X CuAAC solution and incubated overnight again. The samples were subsequently treated with 10-fold excess ice-cold methanol and incubated overnight at -80 °C for protein precipitation. The samples were centrifuged ($3700 \times g$, 20 min, 4 °C), washed twice with 2.5 mL of ice-cold methanol, the supernatant discarded, and the pellet air-dried. The pellets were resuspended in 250 μ L of 0.1% (w/v) Rapigest in PBS, sonicated for 25 min in a water bath, centrifuged ($3700 \times g$, 20 min, 4 °C) and the supernatant collected. The pellets were resuspended in 250 μ L of 6M urea in PBS, sonicated for 25 min in a water bath, centrifuged ($3700 \times g$, 20 min, 4 °C) and the supernatant collected.

The pellets were resuspended in 250 μ L of PBS, sonicated for 25 min in a water bath, centrifuged ($3700 \times g$, 20 min, 4 °C) and the supernatant collected. The three supernatants were combined and incubated with 300 μ L of Lys-dimethylated neutravidin-coated magnetic beads²¹⁸ for 2 h at RT under rotation. The supernatant was collected, and the beads washed 3 times with 300 μ L of 1% (w/v) Rapigest in PBS, 6 times with 300 μ L of 6M urea in PBS, 6 times with 300 μ L of 50 mM ammonium bicarbonate (AmBic) in LC-MS grade water, and 6 times with 100 μ L of 40% (v/v) LC-MS grade acetonitrile in LC-MS grade water. The beads were resuspended in 100 μ L of 10 mM DTT in 50 mM AmBic and incubated for 15 min at 50 °C on an orbital shaker at 400 rpm to reduce disulfide bonds. The beads were washed twice with 100 μ L of 50 mM AmBic and incubated with 100 μ L of 20 mM iodoacetamide (IAA) in 50 mM AmBic for 30 min at RT in the dark to alkylate bonds. The beads were treated with 100 μ L of 10 mM DTT in 50 mM AmBic to quench the reaction, washed twice with 100 μ L of 50 mM AmBic, and resuspended in 100 μ L of 50 mM AmBic. The samples were incubated with 300 ng of Lys-C overnight at 37 °C on an orbital shaker at 400 rpm. The supernatant was collected, and the beads washed with 100 μ L of 50 mM AmBic. The supernatant and the wash were combined and incubated with 200 ng of trypsin for 8h at 37 °C on an orbital shaker at 400 rpm. The samples were vacuum-dried by Speedvac.

11.2. Peptide desalting using UltraMicroSpin™ columns

UltraMicroSpin™ columns were conditioned once with 100 μ L of LC-MS grade acetonitrile. The samples from **Section 11.1** were reconstituted in 100 μ L of 0.1% (v/v) formic acid/ 2% (v/v) LC-MS grade acetonitrile in 50 mM AmBic and loaded onto the column. The column was washed with 50 μ L of 2% (v/v) LC-MS grade acetonitrile in LC-MS grade water. To elute the peptides 50 μ L of 0.1% (v/v) formic acid / 80% (v/v) LC-MS grade acetonitrile in LC-MS grade water was added to the column. The samples were vacuum-dried by Speedvac.

11.3. LC-MS/MS analysis for label-free MS-proteomics

The samples from **Section 11.2** were reconstituted in 16 μ L of 0.1% (v/v) formic acid in LC-MS grade water, sonicated for 15 min in a water bath, centrifuged ($18000 \times g$, 5 min, 4 °C) and the supernatant injected in triplicate into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The samples were loaded onto an Acclaim™ PepMap™ 100 C18 HPLC Column (2 μ m particle size, 75 μ m x 150 cm). The solvents used were **A** (0.1% (v/v) formic acid, 5% (v/v) DMSO in LC-MS grade water) and **B** (0.1% (v/v) formic acid, 5% (v/v) DMSO, 20% LC-MS grade water in LC-MS grade acetonitrile). The elution profile used was as follows - 0 min: 2% **B**, 5 min: 2% **B**, 5.5 min: 8% **B**, 93 min: 40% **B**, 94 min: 95% **B**, 104 min: 95% **B**, 105 min: 2% **B**. The flow rate was 275 nL/min.

MS1 scans were collected with a mass range from 350 to 1500 m/z, 120000 resolution, 4×10^5 ion inject target, and 50 ms maximum inject time. Dynamic exclusion was set to exclude for 20 s with a repeat count of 1. Charge states 2-6 with an intensity greater than 1×10^4 were selected for fragmentation. Selected precursors were fragmented using HCD at 30% nCE with 1.2 m/z isolation window, 1×10^4 ion inject target, and 100 ms maximum inject time.

11.4. Analysis of label-free MS-proteomics peptide data using MaxQuant and Perseus

Raw MS files were processed using MaxQuant software (version 2.0.2.0). In the “Group-specific parameters” section the multiplicity was set to 1. Methionine oxidation and N-terminal acetylation were selected as variable modifications and cysteine carbamidomethylation was selected as a fixed modification, with a total common maximum of 5. In the “Digestion” tab the digestion mode was set to Specific and in the Enzyme section Trypsin/P was selected, with a maximum of 2 missed cleavages allowed. In the “Global parameters” section, the *Homo sapiens* and *Avidin* FASTA files (downloaded from UniProt²¹⁹) were uploaded.

The peptides.txt file was loaded into Perseus (version 1.6.14.0) for data visualization. In the “Main” section the “Intensity” value for each sample was selected. In the “Categorical” section “Potential contaminant” and “Reverse” were selected. The matrix was filtered to remove matching rows based on the “Reverse” and “Potential contaminant” categorical columns. The replicates of each sample were grouped using the “Categorical annotation rows” tool. The “Intensity” values were then transformed using the “log2(x)” function. The matrix was filtered to remove matching rows with a minimum of 70% of valid values using the “Filter rows based on valid values” tool. Missing values were imputed using the “Replace missing values from normal distribution” tool. The data was normalised by subtracting the median of each column using the “Subtract” tool. A two-sided two-sample Welch’s t-test was performed, choosing a Benjamini-Hochberg False Discovery Rate (FDR)-adjusted p-value of 0.05 as a cut-off.

12. SILAC-based MS-(glyco)proteomics experiments on transfected BCCs

12.1. SILAC metabolic labelling and compound administration

T47D/MCF7/MDA-MB-231^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D/MCF7/MDA-MB-231^{WT}-AGX1^{F383A}-NahK cells were cultured in heavy or light media for 5-6 doublings. The heavy medium contained RPMI/DMEM for SILAC with 10% (v/v) dialysed FBS, 50 mg K8 L-Lysine, 50 mg R10 L-Arginine, and 50 mg P0 L-Proline. The light medium contained RPMI/DMEM for SILAC with 10% (v/v) dialysed FBS, 50 mg K0 L-Lysine, 50 mg R0 L-Arginine and 50 mg P0 L-Proline.

The cells were seeded in 20 mL growth medium into 15 cm dishes (one dish for each lysate sample, two dishes for each secretome sample)(cell density of T47D cells: 4×10^5 viable cells/mL; MCF7 cells: 3.5×10^5 viable cells/mL; MDA-MB-231 cells: 3.5×10^5 viable cells/mL). After 24 h, the cells were treated with 1 μ M Ac₄GalN6yne or DMSO and incubated overnight at 37 °C. Lysates and secretome samples were prepared as described in **Section 9** (lysing the cells in 150 μ L of Lysis Buffer **A** for 45 min) and **10**, respectively.

12.2. Sample preparation to check for SILAC heavy label incorporation

An aliquot from each heavy sample (25 μ g) was used to check the SILAC heavy label incorporation using the SP3 technology²²⁰. The samples were diluted to 25 μ L solution in reconstitution solution (50 mM HEPES pH 7.5, 1% (v/v) SDS, 1% (v/v) Triton X-100, 1% (v/v) Tween-20, 1% (v/v) sodium deoxycholate, 5 mM EDTA, 50 mM NaCl, 1% (v/v) glycerol, 5 mM DTT and 1X Halt™ protease inhibitors) and incubated for 30 min at 60 °C on an orbital shaker at 400 rpm to reduce disulfide bonds. The samples were then treated with 2.5 μ L of 200 mM IAA in water and incubated for 30 min at RT in the dark to alkylate bonds, followed by incubation with 1.25 μ L of 100 mM DTT in water at RT for 15 min to quench the reaction.

Two vendor stocks of Sera-Mag™ SpeedBead Carboxylate-Modified Magnetic Particles were combined in a 1:1 ratio, washed with 200 μ L of water and then reconstituted in water to a concentration of 50 mg/mL. 5 μ L of the combined Sera-Mag SpeedBeads (250 μ g) were added to each sample and diluted to 50 μ L in reconstitution solution. To induce binding of the proteins to the beads, the samples were treated with 50 μ L of ethanol and incubated for 5 min at RT on an orbital shaker at 1000 rpm. The samples were washed 3 times with 180 μ L of 80% ethanol in water and the beads resuspended in 100 μ L of 50 mM AmBic in water containing 300 ng Lys-C and 200 ng of trypsin. The samples were sonicated for 30 s in a water bath to fully disaggregate the beads and incubated at 37 °C on an orbital shaker at 1000 rpm overnight. The samples were centrifuged ($18213 \times g$, 1 min, RT), the supernatant collected and filtered using Ultrafree-MC centrifugal filters ($12000 \times g$, 4 min, RT). The flow-through was collected and analysed by LC-MS/MS.

12.3. LC-MS/MS analysis to check for SILAC heavy label incorporation

The T47D samples from **Section 12.2** were injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno using the same LC-MS/MS method as described in **Section 11.3**.

The MCF7 and MDA-MB-231 samples from **Section 12.2** were injected into an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The elution profile used was as follows - 0 min: 5% **B**, 6 min: 5% **B**, 7 min: 8% **B**, 32 min: 25% **B**, 45 min: 45% **B**, 47 min: 60% **B**, 49 min: 95% **B**, 54 min: 95% **B**, 55 min: 5% **B**. The flow rate was 250 nL/min.

MS1 scans were collected with a mass range from 375 to 1500 m/z, 120000 resolution, 4×10^5 ion inject target, and 50 ms maximum inject time. Dynamic exclusion was set to exclude for 20 s with a repeat count of 1. Charge states 2-5 with an intensity greater than 1×10^4 were selected for fragmentation. Selected precursors were fragmented using HCD at 32% nCE with 1.2 m/z isolation window, 1×10^4 ion inject target, and 300 ms maximum inject time.

12.4. Checking SILAC heavy label incorporation using MaxQuant and Perseus

Raw MS files were processed using MaxQuant software (version 2.0.2.0). In the “Group-specific parameters” section the multiplicity was set to 2, the maximum labelled to 3 and in the heavy labels table Arg10 and Lys8 were selected. Methionine oxidation and N-terminal acetylation were selected as variable modifications and cysteine carbamidomethylation was selected as a fixed modification, with a total common maximum of 5. In the “Digestion” tab the digestion mode was set to Specific and in the Enzyme section Trypsin/P was selected, with a maximum of 2 missed cleavages allowed. In the “Global parameters” section, the *Homo sapiens* and *Avidin* FASTA files (downloaded from UniProt²¹⁹) were uploaded. In the “Protein quantification” tab the label minimum ratio count was set to 1.

The peptides.txt file was loaded into Perseus (version 1.6.14.0) for data visualization. In the “Main” section the “Ratio H/L” value for each sample was selected. In the “Categorical” section “Potential contaminant” and “Reverse” were selected. The matrix was filtered to remove matching rows based on the “Reverse” and “Potential contaminant” categorical columns. The “Ratio H/L” values were then transformed using the “log2(x)” function. Using the Summary statistics (columns) tool the median of the “Ratio H/L” values was calculated. This value was then sequentially transformed with the “2^x” function, followed by the “x/(1+x)” function and the “x*100” function to convert the median Ratio H/L value to a percentage.

12.5. Sample preparation for SILAC-based MS-(glyco)proteomics

Heavy and light samples from **Section 12.1** were mixed in a 1:1 ratio (final protein content: 900 µg for lysates, 500 µg for secretome samples) according to **Fig. 24A (Results Section 3.3)** and diluted to 300 µL in PBS. The samples were incubated with 350 µL of neutravidin-coated magnetic beads for 2 h at RT under rotation to remove endogenous biotinylated proteins. The supernatant was collected, the samples incubated with 5 U of PNGase F overnight at 37 °C, and the reaction quenched by heating at 95 °C for 10 s. The samples were then treated with a 10X CuAAC solution (12 mM BTAA, 6 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-DADPS-picolyl azide) and incubated for 72 h at RT on an orbital shaker at 400 rpm. The samples were concentrated, and the buffer exchanged with PBS using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO). The samples were incubated with 350 µL of Lys-dimethylated neutravidin-coated magnetic beads²¹⁸ for 2 h at RT under rotation. The supernatant was collected, and the beads washed 3 times with 350 µL of 1% (w/v) Rapigest in PBS, 6 times with 350 µL of 6M urea in PBS, 6 times with 350 µL of 50 mM AmBic in LC-MS grade water, and 6 times with 100 µL of 40% (v/v) LC-MS grade acetonitrile in LC-MS grade water.

The beads were resuspended in 100 µL of 10 mM DTT in 50 mM AmBic and incubated for 15 min at 50 °C on an orbital shaker at 400 rpm to reduce disulfide bonds. The beads were washed twice with 100 µL of 50 mM AmBic and incubated with 100 µL of 20 mM IAA in 50 mM AmBic for 30 min at RT in the dark to alkylate bonds. The beads were treated with 100 µL of 10 mM DTT in 50 mM AmBic to quench the reaction, washed twice with 100 µL of 50 mM AmBic, and resuspended in 100 µL of 50 mM AmBic. The samples were incubated with 300 ng of Lys-C overnight at 37 °C on an orbital shaker at 400 rpm. The supernatant was collected, and the beads washed with 100 µL of 50 mM AmBic. The supernatant and the wash were combined and incubated with 200 ng of trypsin for 8h at 37 °C on an orbital shaker at 400 rpm. The peptide samples were vacuum-dried by Speedvac. Peptide desalting of the T47D secretome samples was performed as described in **Section 11.2**. For the lysate samples, peptide fractionation was performed as described in **Section 12.6**.

To prepare the T47D glycopeptide samples, after overnight incubation with Lys-C the beads were incubated with 150 μ L of 1% (v/v) formic acid in LC-MS grade water for 30 min at RT on an orbital shaker at 400 rpm. The supernatant was collected and the incubation with 1% (v/v) formic acid in LC-MS grade water repeated. The supernatant was collected, and the beads washed with 150 μ L of LC-MS grade acetonitrile. The wash and the collected supernatants were combined and incubated with 200 ng of trypsin for 8h at 37 °C on an orbital shaker at 400 rpm. The glycopeptide samples were vacuum-dried by Speedvac. Glycopeptide desalting was performed as described in **Section 11.2**.

12.6. Peptide fractionation using UltraMicroSpin™ columns

UltraMicroSpin™ columns were conditioned once with 100 μ L of 100% LC-MS grade acetonitrile, once with 100 μ L of 80% (v/v) LC-MS grade acetonitrile in 10 mM AmBic, once with 100 μ L of 40% (v/v) LC-MS grade acetonitrile in 10 mM AmBic, once with 100 μ L of 20% (v/v) LC-MS grade acetonitrile in 10 mM AmBic, and twice with 100 μ L of 10 mM AmBic. The lysate peptide samples from **Section 12.5** were reconstituted in 100 μ L of 10 mM AmBic and loaded onto the column. The column was washed with 50 μ L of 10 mM AmBic. To elute the peptides the column was treated with 50 μ L of 10% (v/v) LC-MS grade acetonitrile in 10 mM AmBic and the flow-through collected (fraction 1). This step was repeated with 50 μ L of 15% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 2), 50 μ L of 20% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 3), 50 μ L of 25% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 4), 50 μ L of 30% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 5), 50 μ L of 35% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 6), 50 μ L of 40% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 7), 50 μ L of 80% (v/v) LC-MS grade acetonitrile in 10 mM AmBic (fraction 8). All fractions were vacuum-dried by Speedvac.

12.7. LC-MS/MS analysis for SILAC-based MS-proteomics

The T47D secretome samples from **Section 12.5** were reconstituted in 16 μ L of 0.1% (v/v) formic acid in LC-MS grade water, sonicated for 15 min in a water bath, centrifuged (18000 $\times$ g, 5 min, 4 °C) and the supernatant injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The elution profile used was as follows - 0 min: 2% **B**, 5 min: 2% **B**, 5.5 min: 4% **B**, 93 min: 40% **B**, 94 min: 95% **B**, 104 min: 95% **B**, 105 min: 2% **B**. The flow rate was 275 nL/min. The MS/MS method used was the same one as described in **Section 11.3**.

The lysate peptide samples from **Section 12.6** were reconstituted in 50 μ L of 1% (v/v) formic acid/ 1% (v/v) LC-MS grade acetonitrile in LC-MS grade water, sonicated for 15 min in a water bath, centrifuged ($18000 \times g$, 5 min, 4 °C) and the supernatant injected in triplicate into and Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno using the same LC-MS/MS method as described in **Section 12.3**.

12.8. Analysis of SILAC-based MS-proteomics peptide data using MaxQuant and Perseus

Raw MS files from the peptide samples were processed using MaxQuant software (version 2.0.2.0) as described in **Section 12.4**. The fractions from the same sample were marked with the same experiment name to group and process the data as a single sample²²¹.

The peptides.txt file was loaded into Perseus (version 1.6.14.0) for data visualization. In the “Main” section the “Ratio H/L normalized” value for each sample was selected. In the “Categorical” section “Potential contaminant” and “Reverse” were selected. The matrix was filtered to remove matching rows based on the “Reverse” and “Potential contaminant” categorical columns. The replicates of each sample were grouped using the “Categorical annotation rows” tool. The Ratio H/L normalized” values of samples **1** and **2** (**Fig. 24-25** and **SI Fig. 9**) were grouped using the “Categorical annotation rows” tool (“treatment comparison”). The Ratio H/L normalized” values of samples **3** and **4** (**Fig. 24-25** and **SI Fig. 9**) were grouped using the “Categorical annotation rows” tool (“cell line comparison”). The “Ratio H/L normalized” values of samples **2** and **4** (**Fig. 24-25** and **SI Fig. 9**) were inverted using the “1/x” function. The “Ratio H/L normalized” values were then transformed using the “log2(x)” function. A two-sided one-sample t-test was performed, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off.

12.9. LC-MS/MS analysis for SILAC-based MS-glycoproteomics

The T47D glycopeptide samples from **Section 12.5** were reconstituted in 16 μ L of 0.1% (v/v) formic acid in LC-MS grade water, sonicated for 15 min in a water bath, centrifuged ($18000 \times g$, 5 min, 4 °C) and the supernatant injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The elution profile used was as follows - 0 min: 2% **B**, 6 min: 2% **B**, 114 min: 40% **B**, 115 min: 95% **B**, 119 min: 95% **B**, 120 min: 2% **B**. The flow rate was 275 nL/min.

MS1 scans were collected with a mass range from 300 to 1500 m/z, 120000 resolution, 4×10^5 ion inject target, and 50 ms maximum inject time. Dynamic exclusion was set to exclude for 10 s with a repeat count of 3. Charge states 2-6 were selected for fragmentation. Selected precursors were fragmented using HCD at 28% nCE with 2 m/z isolation window, 5×10^4 ion inject target, and 54 ms maximum inject time before collection at 30000 resolution in the Orbitrap. For precursors from 300 to 1000 m/z, presence of selected oxonium ions over 10% relative abundance triggered a charge calibrated ETD scan to be collected in the ion trap with a 3 m/z isolation window, 1×10^4 ion inject target, and 100 ms maximum injection time. Raw MS files from the glycopeptide samples were processed using Byonic software as described by Calle et al²⁰³.

13. Validation of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown

T47D/MCF7/MDA-MB-231^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D/MCF7/MDA-MB-231^{WT}-AGX1^{F383A}-NahK cells were seeded at a density of 4×10^5 viable cells/mL in 20 mL growth medium into 15 cm dishes (one dish for each sample). After 24 h, the cells were treated with 1 μ M Ac₄GalN6yne and incubated overnight at 37 °C. Lysates were prepared as described in **Section 9** (lysing the cells in 150 μ L of Lysis Buffer **A** for 45 min). The samples (900 μ g) were diluted to 300 μ L in PBS and incubated with 350 μ L of neutravidin-coated magnetic beads for 2 h at RT under rotation to remove endogenous biotinylated proteins. The supernatant was collected, the samples incubated with 5 U of PNGase F overnight at 37 °C, and the reaction quenched by heating at 95 °C for 10 s. The samples were then treated with a 10X CuAAC solution (12 mM BTAA, 6 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-picolyl azide) and incubated for 72 h at RT on an orbital shaker at 400 rpm. The samples were concentrated and the buffer exchanged with PBS using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO). An aliquot of each sample was taken at this point (**Input**).

The samples were incubated with 350 µL of Lys-dimethylated neutravidin-coated magnetic beads²¹⁸ for 2 h at RT under rotation. The supernatant was collected, and the beads washed 3 times with 350 µL of 1% (w/v) Rapigest in PBS, 6 times with 350 µL of 6M urea in PBS, 6 times with 350 µL of 50 mM AmBic in LC-MS grade water, and 6 times with 100 µL of 40% (v/v) LC-MS grade acetonitrile in LC-MS grade water. The proteins were eluted by incubating the beads with 40 µL of a 2:2:1 (v/v/v/v) mixture of 1 M Tris-HCl pH 6.5, 80% (v/v) glycerol, 20% (w/v) SDS) for 10 min at 95 °C. The supernatant was collected (**Eluate**). The **Input** and **Eluate** samples were run for SDS-PAGE using a 4-20% Criterion™ TGX™ Precast gel. The gel was transferred to a nitrocellulose membrane using the Trans-Blot Turbo RTA Midi 0.2 µm Nitrocellulose Transfer Kit. Protein expression control was performed by Western blot.

14. Testing different biorthogonal reporter reagents to facilitate MS-glycoproteomic analyses

The methods below have been published in Calle et al. (2021)¹⁸² so they have been adapted for this section.

14.1. *In vitro* glycosylation reaction with WT GalNAc-T2 and UDP-sugars **1**, **3** and **11**

Reaction mixtures (25 µL) were prepared containing 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MnCl₂, 5 U Calf Intestinal Alkaline Phosphatase, 50 µM peptide **1**, 51 µM UDP-sugar **1**, **3** or **11** and WT GalNAc-T2 (100 nM). The reaction was incubated at 37 °C for 42 h and quenched by heating at 95 °C for 10 s. Reaction progress was monitored by taking aliquots, adding the same volume of ethanol to quench the reaction, and glycopeptide formation determined by LC-MS and MS⁺ peak integration using **Equation 2**. LC-MS analysis was performed as described in **Section 4.2**.

14.2. CuAAC reaction with reporter reagents **A**, **D**, **E** (by Ganka Bineva-Todd)

The samples from **Section 14.1** were treated with a 1X CuAAC solution (1200 µM BTAA, 600 µM CuSO₄, 10 mM sodium ascorbate and 100 µM reagents **A**, **E** or **D**) and incubated at RT overnight. Reaction progress was monitored by taking aliquots, adding the same volume of ethanol to quench the reaction, and the supernatant analysed by LC-MS as described in **Section 4.2**. The reaction mixtures were then treated with 2% (v/v) formic acid, incubated for 1 h at RT and vacuum-dried by Speedvac.

14.3. CuAAC reaction with reporter reagents **B, C, F**

The samples from **Section 14.1** were treated with a 1X CuAAC solution (600 μ M BTAA, 300 μ M CuSO₄, 5 mM sodium ascorbate, 5 mM aminoguanidine hydrochloride and 75 μ M reagent **B, C** or **F**) and incubated at RT. After 4 h, the samples were treated with a fresh 1X CuAAC solution and incubated overnight. Reaction progress was monitored by taking aliquots, adding the same volume of ethanol to quench the reaction, and the supernatant analysed by LC-MS as described in **Section 4.2**. The reaction mixtures were then treated with 2% (v/v) formic acid, incubated for 1 h at RT and vacuum-dried by Speedvac.

14.4. Solid Phase Extraction using Strata-X columns

Strata-X columns were conditioned once with 1 mL of LC-MS grade acetonitrile and once with 1 mL of 0.1% (v/v) formic acid in water. The samples from **Section 14.2 and 14.3** were reconstituted in 1 mL of 0.1% (v/v) formic acid in LC-MS grade water and loaded onto the column. The column was washed with 1 mL of 0.1% (v/v) formic acid in LC-MS grade water. To elute the synthetic glycopeptides 300 μ L of 80% (v/v) LC-MS grade acetonitrile/ 0.1% (v/v) formic acid in LC-MS grade water was added to the column. The samples were aliquoted into 100 pmol aliquots and vacuum-dried by Speedvac.

14.5. LC-MS/MS analysis of synthetic glycopeptides tagged with the reporter reagents **A-F**

The samples from **Section 14.4** were reconstituted in 200 μ L of 0.1% (v/v) formic acid in LC-MS grade water, sonicated for 15 min in a water bath and injected into an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Helen Flynn. The samples were loaded onto an Acclaim™ PepMap™ 100 C18 HPLC Column (2 μ m particle size, 75 μ m x 150 cm). The solvents used were **A** (0.1% (v/v) formic acid, 5% (v/v) DMSO in LC-MS grade water) and **B** (0.1% (v/v) formic acid, 5% (v/v) DMSO, 20% LC-MS grade water in LC-MS grade acetonitrile). The elution profile used was as follows - 0 min: 2% **B**, 6 min: 2% **B**, 33 min: 40% **B**, 34 min: 95% **B**, 44 min: 95% **B**, 45 min: 2% **B**. The flow rate was 275 nL/min. The MS/MS method used was the same one as described in **Section 12.9**. All glycopeptides were manually analysed using Thermo Xcalibur software by Dr Stacy Malaker.

15. Puromycin kill curve of transfected T47D cells

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells were seeded at a density of 5×10^5 viable cells/mL in 1 mL growth medium per well of a 12-well plate. After 24 h, the medium was replaced with fresh growth medium containing puromycin (0.5-10 μ g/mL). The medium was replaced every 2-3 days with fresh selection medium. After one week of antibiotic treatment, visual toxicity was assessed and the lowest antibiotic concentration at which all cells were dead determined.

16. Lentiviral transduction of transfected T47D cells with tagged CD46 and GOLM1

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells were seeded at a density of 5×10^5 viable cells/mL in 2 mL growth medium per well of a 6-well plate. After 24 h, the cells were transduced with 2×10^6 lentiviral particles expressing tagged CD46 or GOLM1 and 10 μ g/mL TransducelT Transduction Reagent. After 48 h, the medium was replaced with fresh growth medium containing 1 μ g/mL puromycin. The antibiotic selection was maintained until stable colonies were generated.

17. Protein expression and Ni-NTA/GFP-affinity purification of recombinant CD46

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells transduced with tagged CD46 BC or tagged CD46 C were seeded at a density of 5×10^5 viable cells/mL in 20 mL growth medium into 15 cm dishes (five dishes per sample). After 24 h, the cells were treated with 1 μ M Ac₄GalN₆yne and incubated overnight at 37 °C. The secretome was collected, centrifuged (350 × g, 5 min, 4 °C) and filtered using a Stericup® Quick Release Vacuum Filtration System. 1X Halt™ protease inhibitors and 1 mM nickel sulfate (NiSO₄) were added to the filtered supernatant and incubated for 15 min at 4 °C on a tube roller shaker to prevent non-specific binding to the beads. Ni-NTA® Agarose beads (1 mL) were conditioned 4 times with 5 mL of water and 2 times with 5 mL of Wash Buffer **A** (50 mM Tris-HCl pH 7.5, 125 mM NaCl, 20 mM imidazole). The beads were added to the filtered supernatant and incubated overnight at 4 °C on a tube roller shaker. The samples were centrifuged (2000 × g, 3 min, 4 °C) and washed 3 times with 10 mL of Wash Buffer **A**. To elute the protein, 2 mL of Elution Buffer (50 mM Tris-HCl pH 7.5, 125 mM NaCl, 500 mM imidazole) diluted with Wash Buffer **A** to obtain different imidazole concentrations (100 mM, 2x 200 mM, 300 mM) was added to the samples. The eluted fractions were pooled together and further purified by GFP-affinity purification.

GFP-clamps (100 μ L) were conditioned 4 times with 500 μ L of water and 2 times with 500 μ L of Wash Buffer **B** (50 mM Tris-HCl pH 7.5, 125 mM NaCl). The beads were added to the filtered supernatant and incubated for 2 h at 4 °C on a tube roller shaker. The beads were centrifuged (2000 $\times g$, 3 min, 4 °C) and washed 2 times with 1 mL of Wash Buffer **B**. To elute the protein, the beads were resuspended in 200 μ L of Wash Buffer **B** and incubated with 75 μ g of TEV protease overnight at 4 °C under rotation. The supernatant was collected, and the beads washed 2 times with 200 μ L of Wash Buffer **B**, collecting the supernatant after each wash. The eluted fractions were pooled together, concentrated using Amicon Ultra-0.5 Centrifugal Filters (30 kDa MWCO) and the buffer exchanged twice with PBS. Protein concentration was determined by Nanodrop. The samples were flash-frozen in liquid nitrogen and stored at -80 °C.

18. MS-glycoproteomics experiment with recombinant CD46 BC and CD46 C

18.1. Sample preparation for MS-glycoproteomics with recombinant CD46 using iTag-azide for the CuAAC reaction

The samples from **Section 17** (3 μ g) were incubated with 5 U of PNGase F for 4 h at 37 °C, and the reaction quenched by heating at 95 °C for 10 s. The samples were then treated with a 1X CuAAC solution (2.4 mM BTAA, 1.2 mM CuSO₄, 10 mM sodium ascorbate, 10 mM aminoguanidine hydrochloride and 1 mM iTag-azide) and incubated overnight at RT on an orbital shaker at 400 rpm. The mucinase/glycosidase/protease treatment below was performed according to **Table 6** and **7 (Results Section 5.1.2)**. The samples were incubated with 300 ng of SmE and incubated for 4 h at 37 °C. 3 U of SialEXO and 3 U of OglyZOR were added to the samples and incubated for 4 h at 37 °C. The samples were vacuum-dried by Speedvac, resuspended in 23 μ L of Lysis Buffer **B** (50 mM Tris-HCl pH 8, 5% (v/v) SDS in water) and sonicated for 5 min in a water bath. The samples were treated with 1 μ L of 120 mM DTT in 50 mM AmBic and incubated for 15 min at 55 °C on an orbital shaker at 400 rpm to reduce disulfide bonds. The samples were treated with 1 μ L of 500 mM IAA in 50 mM AmBic and incubated for 10 min at RT in the dark to alkylate bonds. The samples were treated with 2.5 μ L of 27.5% (v/v) phosphoric acid in water to completely denature the proteins and trap them efficiently.

The samples were treated with 165 μ L of Wash Buffer **C** (100 mM Tris-HCl pH 7.5, 90% (v/v) methanol in water) and added to S-Trap™ micro spin columns. The columns were centrifuged (4000 $\times g$, 30 sec, RT) and washed 4 times with 150 μ L of Wash Buffer **C**. The columns were centrifuged (4000 $\times g$, 1 min, RT) and incubated with 50 mM Tris-HCl pH 8 containing 300 ng of trypsin overnight at 37 °C. The samples were incubated with 50 mM Tris-HCl pH 8 containing 600 ng of chymotrypsin for 4 h at 37 °C. To elute the peptides, the columns were treated with 40 μ L of 50 mM Tris-HCl pH 8, centrifuged (4000 $\times g$, 1 min, RT) and the flow-through collected. This step was repeated with 40 μ L of 0.2% (v/v) formic acid in LC-MS grade water and 40 μ L of 50% (v/v) LC-MS grade acetonitrile in LC-MS grade water. The eluted fractions were combined and vacuum-dried by Speedvac.

18.2. LC-MS/MS analysis for MS-glycoproteomics

The samples from **Section 18.1** were reconstituted in 16 μ L of 0.1% (v/v) formic acid in LC-MS grade water, sonicated for 15 min in a water bath, centrifuged (18000 $\times g$, 5 min, 4 °C) and the supernatant injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno using the LC-MS/MS method described in **Section 12.9**. Raw MS files from the glycopeptide samples were processed using Byonic software as described by Calle et al²⁰³.

19. Protein expression and Ni-NTA purification of recombinant CD46 BC and GOLM1

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells and T47D^{WT}-AGX1^{F383A}-NahK cells transduced with tagged CD46 BC or tagged GOLM1 were seeded at a density of 4x10⁵ viable cells/mL in 20 mL growth medium into 15 cm dishes (8 dishes for the CD46 sample, 6 dishes for the GOLM1 sample). After 24 h, the cells were treated with 1 μ M Ac₄GalN6yne and incubated overnight at 37 °C. The secretome was collected, centrifuged (350 $\times g$, 5 min, 4 °C) and filtered using a Stericup® Quick Release Vacuum Filtration System. 1X Halt™ protease inhibitors were added to the filtered supernatant. Lysates were prepared as described in **Section 9** (lysing the cells in 1 mL of Lysis Buffer **A** for 1 h). 1 mM NiSO₄ was added to the samples and incubated for 15 min at 4 °C on a tube roller shaker to prevent non-specific binding to the beads.

Ni-NTA® Agarose beads (1 column volume (CV) = 1 mL for lysate samples; 1 CV = 3 mL for secretome samples) were conditioned 4 times with 5 CV of water and 2 times with 5 CV of Wash Buffer **D** (50 mM Tris-HCl pH 7.5, 300 mM NaCl). The beads were added to the samples and incubated overnight at 4 °C on a tube roller shaker. The samples were centrifuged ($2000 \times g$, 3 min, 4 °C) and washed 3 times with 10 CV of Wash Buffer **D**. To elute the protein 2 CV of Elution Buffer (50 mM Tris-HCl pH 7.5, 125 mM NaCl, 500 mM imidazole) diluted with Wash Buffer **D** to obtain different imidazole concentrations (2x 200 mM, 3x 300 mM) was added to the samples. The eluted fractions were pooled together, concentrated using Amicon Ultra Centrifugal Filters (10 kDa MWCO) and the buffer exchanged twice with PBS. Protein concentration was determined by Nanodrop. The samples were flash-frozen in liquid nitrogen and stored at -80 °C.

20. MS-glycoproteomics experiment with recombinant CD46 BC and GOLM1

20.1. Sample preparation for MS-glycoproteomics with recombinant CD46 and GOLM1 using biotin-DADPS-picolyl azide for the CuAAC reaction

The samples from **Section 19** (100 µg) were incubated with 5 U of PNGase F overnight at 37 °C, and the reaction quenched by heating at 95 °C for 10 s. The samples were then treated with a 10X CuAAC solution (12 mM BTAA, 6 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-DADPS-picolyl azide) and incubated for 72 h at RT on an orbital shaker at 400 rpm. The buffer was exchanged with PBS using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO). 100 U of SialEXO and 100 U of OglyZOR were added to the samples and incubated for 4 h at 37 °C. The reaction was quenched by heating at 95 °C for 10 s and the samples diluted to 200 µL in PBS.

The samples were incubated with 200 µL of Lys-dimethylated neutravidin-coated magnetic beads²¹⁸ for 1 h at RT under rotation. The supernatant was collected, and the beads washed 3 times with 250 µL of 1% (w/v) SDS in LC-MS grade water, 6 times with 250 µL of 6M urea in PBS and 6 times with 250 µL of 50 mM AmBic in LC-MS grade water. The beads were resuspended in 100 µL of 10 mM DTT in 50 mM AmBic and incubated for 15 min at 50 °C on an orbital shaker at 400 rpm to reduce disulfide bonds. The beads were washed twice with 100 µL of 50 mM AmBic and incubated with 100 µL of 20 mM IAA in 50 mM AmBic for 30 min at RT in the dark to alkylate bonds. The beads were treated with 100 µL of 10 mM DTT in 50 mM AmBic to quench the reaction, washed 3 times with 100 µL of 50 mM AmBic, and resuspended in 100 µL of 50 mM AmBic.

The samples were incubated with 100 ng of Lys-C overnight at 37 °C on an orbital shaker at 400 rpm. The supernatant was collected, the beads washed with 100 µL of 50 mM AmBic and incubated with 150 µL of 1% (v/v) formic acid in LC-MS grade water for 30 min at RT on an orbital shaker at 400 rpm. The supernatant was collected and the incubation with 1% (v/v) formic acid in LC-MS grade water repeated. The supernatants were combined and incubated with 100 ng of trypsin overnight at 37 °C on an orbital shaker at 400 rpm. The CD46 sample was then incubated with 500 ng of chymotrypsin for 4 h at 37 °C. The digestion in both samples was quenched by addition of 1% (v/v) formic acid. The samples were vacuum-dried by Speedvac.

20.2. Glycopeptide desalting using UltraMicroSpin™ columns

UltraMicroSpin™ columns were conditioned twice with 100 µL of 100% LC-MS grade acetonitrile and twice with 100 µL of 0.1% (v/v) formic acid in 50 mM AmBic. The samples from **Section 20.1** were reconstituted in 100 µL of 0.1% (v/v) formic acid in 50 mM AmBic and loaded onto the column. The column was washed with 50 µL of 1% (v/v) formic acid in 50 mM AmBic. To elute the glycopeptides, the column was treated with 50 µL of 0.1% (v/v) formic acid/ 80% (v/v) LC-MS grade acetonitrile in LC-MS grade water. The samples were vacuum-dried by Speedvac. LC-MS/MS analysis for MS-glycoproteomics was performed as described in **Section 18.2**.

21. *In vivo* glycoprotein tagging with BH GalNAc-T2

The methods below have been published in Cioce et al. (2022)²⁰⁴ so they have been adapted for this section.

21.1. Mammary fat pad tumour induction and compound administration

4T1-AGX1^{F383A}-NahK-GalNAc-T2^{BH} cells were resuspended in 50 µL of GFR Matrigel and injected into the left hind fat pad of female NSG mice (1x10⁶ cells per injection) by Dr Victoria Bridgeman. 4T1 cells transfected with empty vector were injected into the right hind fat pad of the mice as an internal control by Dr Victoria Bridgeman. Eleven days post tumour cell injection, mice harbouring tumours of roughly 200-300 mm³ were randomly assigned to receive intratumoural injection with 50 µL of vehicle (5% (v/v) DMSO/PEG-400; n = 1) or 30 mg/kg Ac₄GalN6yne (6 mg/mL in vehicle; n = 2) once daily for three consecutive days by Dr Victoria Bridgeman.

Thirteen days post tumour cell injection, mice were randomly assigned to receive intraperitoneal injection with 100 μ L of vehicle ($n = 1$), 200 mg/kg Ac₄GalN6yne (40 mg/mL in vehicle; $n = 3$) or 200 mg/kg Ac₄ManNAI (40 mg/mL in vehicle; $n = 2$) once daily for five consecutive days by Dr Victoria Bridgeman. 24 h after the last injection, the tumours were harvested and the BH GalNAc-T2 tumours halved by Dr Victoria Bridgeman.

21.2. Tumour dissociation for cell isolation and protein expression control by Western blot

Half of the BH GalNAc-T2 tumours were manually minced and enzymatically digested with a freshly prepared digestion solution (a 1:3:3:193 mixture of DNase I, Liberase TM, Liberase TH and HBSS without CaCl₂ or MgCl₂) for 2 h at 37 °C. The samples were filtered using a 100 μ m strainer, the digestion quenched with growth medium (DMEM with 10% (v/v) FBS, 100 U/mL penicillin and 100 mg/mL streptomycin) and centrifuged (300 $\times g$, 10 min, 4 °C). The pellets were resuspended in 10 mL growth medium, plated on a 10 cm dish, and incubated at 37 °C. After overnight incubation, the cells were washed once with PBS and the medium replaced with fresh growth medium. The cells were subsequently propagated for 10 days in growth medium containing 50 μ g/mL hygromycin B by Dr Anna Cioce. Lysates were prepared as described in **Section 9** (using 8 mM EDTA to detach the cells) followed by protein expression control by Western blot by Dr Anna Cioce.

21.3. Tumour homogenization and glycoprotein tagging analysis by streptavidin blot

Tumour pieces were homogenized in ice-cold Lysis Buffer **C** (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% (v/v) Triton X-100, 0.5% (v/v) sodium deoxycholate, 0.1% (v/v) SDS, 1 mM MgCl₂, 50 μ M PUGNAc, 100 mU/ μ L benzonase and 1X Halt™ protease inhibitors) using a Precellys© homogeniser. The samples were centrifuged (10000 $\times g$, 10 min, 4 °C), the supernatant collected, and the protein concentration determined using the Pierce™ BCA Protein Assay Kit. The samples (1 mg) were diluted to 300 μ L in PBS and incubated with 375 μ L of neutravidin-coated magnetic beads for 2 h at 4 °C under rotation to remove endogenous biotinylated proteins. The supernatant was collected and incubated with a 10X CuAAC solution (6 mM BTAA, 3 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-picolyl azide) overnight at RT on an orbital shaker at 400 rpm. Glycoprotein tagging analysis by streptavidin blot was performed as described in **Section 9**.

22. *In vivo* glycoprotein tagging with BH GalNAc-T4

An 850 µg/mL solution of β -oestradiol was prepared in 20% ethanol/sterile water by Dr Victoria Bridgeman. 1 mL of this solution was added to 99 mL of mice drinking water two days before tumour implantation by Dr Victoria Bridgeman. T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells were resuspended in 50 µL of GFR Matrigel and injected into the left hind fat pad of female NSG mice (1×10^6 cells per injection) by Dr Victoria Bridgeman. T47D^{WT}-AGX1^{F383A}-NahK cells were injected into the right hind fat pad of the mice as an internal control by Dr Victoria Bridgeman. When the tumours measured roughly 150 mm³, the mice were randomly assigned to receive intraperitoneal injection with 100 µL of vehicle (n = 3), 200 mg/kg Ac₄GalN6yne (50 mg/mL in vehicle; n = 4), 100 mg/kg Ac₄GalN6yne (25 mg/mL in vehicle; n = 4), 50 mg/kg Ac₄GalN6yne (12.5 mg/mL in vehicle; n = 4), or 100 mg/kg Ac₄ManNAI (25 mg/mL in vehicle; n = 2) once daily for three consecutive days by Dr Victoria Bridgeman. 24 h after the last injection, the tumours were harvested and enzymatically digested as described in **Section 21.2** by Dr Victoria Bridgeman. Magnetic activated cell sorting (MACS) was used to enrich for CD298+ cells from the enzymatically digested tumours by Dr Victoria Bridgeman. The CD298+ cells were lysed in ice-cold Lysis Buffer **A** (**Section 8**) for 20 min at 4 °C on an orbital shaker and centrifuged (1500 × g, 20 min, 4 °C) by Dr Victoria Bridgeman. The supernatant was collected, and the protein concentration was determined using the PierceTM BCA Protein Assay Kit. The samples (10 µg) were treated with a 10X CuAAC solution (6 mM BTAA, 3 mM CuSO₄, 50 mM sodium ascorbate, 50 mM aminoguanidine hydrochloride and 1 mM biotin-picolyl azide) and incubated overnight at RT on an orbital shaker at 400 rpm. Glycoprotein tagging analysis by streptavidin blot was performed as described in **Section 9**. Protein expression control was performed by Western blot.

RESULTS

CHAPTER 1: Investigating the role of GalNAc-T4 in breast cancer cells

There is still much ambiguity on the actual function of GalNAc-T4 in BC, since previous work by other labs had published contradicting results regarding the role of GalNAc-T4 in BCCs. Niang et al. generated GalNAc-T4 KD T47D cells and saw a growth defect upon GalNAc-T4 KD when assessing cell proliferation using the Cell Counting Kit-8 (Keygen Biotechnology)¹³². Wu et al. generated GalNAc-T4 KO MCF7 cells and saw a boost in growth upon GalNAc-T4 KO when assessing cell proliferation using the MTT Cell Proliferation Assay Kit (Beyotime)¹³³. Even though T47D and MCF7 cells are both ER+ luminal BCCs³⁶ expressing comparable levels of GalNAc-T4¹³³, these results suggested that GalNAc-T4 may have opposite roles in different BCCs – a pro-tumorigenic role in T47D cells and an anti-tumorigenic role in MCF7 cells. However, since different methodologies were used to assess the potential role of GalNAc-T4 in BCCs, this could explain the contradicting results published by the two groups. The first objective of this project was therefore to validate that GalNAc-T4 was implicated in BC by generating GalNAc-T4 KO T47D and MCF7 cells and comparing the growth of WT and GalNAc-T4 KO cells *in vitro* and *in vivo* using the same methodology for both cell lines.

1.1. Establishing GalNAc-T4 KO BCCs

The first step was to generate GalNAc-T4 KO T47D and MCF7 cells using CRISPR-Cas9. CRISPR is a gene editing strategy consisting of a guide RNA (gRNA), which is complementary to the target DNA sequence, and a CRISPR-associated endonuclease (Cas), which makes a double strand break (DSB) in the gene of interest^{222, 223}. The error-prone non-homologous end joining mechanism will subsequently repair the DSB but may insert or delete nucleotides (indels) in the process^{222, 223}. If the indels occur within a coding exon, and the number of indels isn't divisible by 3, this will induce frameshift mutations, resulting in premature stop codons and disruption of gene function^{222, 223}. The gRNA consists of two segments of RNA: CRISPR RNA (which binds to the target DNA) and transactivating CRISPR RNA (which interacts with the Cas protein)^{222, 223}. Both segments can be connected to form a single gRNA (sgRNA) molecule^{222, 223}. Binding of the Cas protein to the genomic DNA requires a protospacer adjacent motif (PAM) to be present downstream of the target site^{222, 223}. Different Cas proteins recognise different PAM motifs^{222, 223}. The most common Cas protein is Cas9, which recognises 5'-NGG-3' PAM motifs (where N is any nucleotide) and then causes a DSB 3-4 nucleotides upstream of the PAM site^{222, 223}.

The coding region of *GALNT4* consists of a single exon located within an intron of the *POC1B* gene²²⁴. Appropriate sgRNA sequences have been validated for the whole GalNAc-T family²²⁵. The validated sgRNA for *GALNT4* targets the first third of the *GALNT4* coding region and it was selected due to its high cutting efficiency, no predicted off targets, and because it induces +1/-1 indels, hence ensuring in-exon frameshifts²²⁵.

To generate the GalNAc-T4 KO cells, during my MSci project I designed sgRNA oligonucleotides following the Target Sequence Cloning Protocol from Addgene²⁰⁸. The designed sgRNA oligonucleotides consisted of the validated sgRNA targeting *GALNT4*²²⁵, with an additional guanine at the 5' end to increase targeting efficiency²²², and overhang sequences to allow annealing of the sgRNA to plasmids co-expressing *Streptococcus pyogenes* Cas9 (SpCas9) nuclease and GFP²⁰⁸. The sgRNA oligonucleotides were cloned into the plasmid encoding spCas9 and GFP followed by transient transfection of this plasmid into T47D and MCF7 cells (**Fig. 11A**). GFP+ cells were sorted as single cells into 96-well plates by the Flow Cytometry STP at the Francis Crick Institute. Clones were expanded, and GalNAc-T4 KO was initially assessed by Western blot. No bands at the expected molecular weight (MW) of GalNAc-T4 (67 kDa) were observed for the T47D GalNAc-T4 KO (T47D^{T4-KO}) clones **1/4/6/7** (**Fig. 11B**) and for the MCF7^{T4-KO} clone **4** (**Fig. 11C**) by Western blot. GalNAc-T4 KO in these clones was then validated by PCR-based amplicon sequencing (**Fig. 11D**). The MCF7^{T4-KO} clone **4** (MCF7^{T4-KO} **4**) and the T47D^{T4-KO} clones **1** (T47D^{T4-KO} **1**) and **6** (T47D^{T4-KO} **6**) were selected for downstream experiments as heterozygous KOs with +1/-1 indels resulting in premature stop codons (**SI Fig. 2-4**).

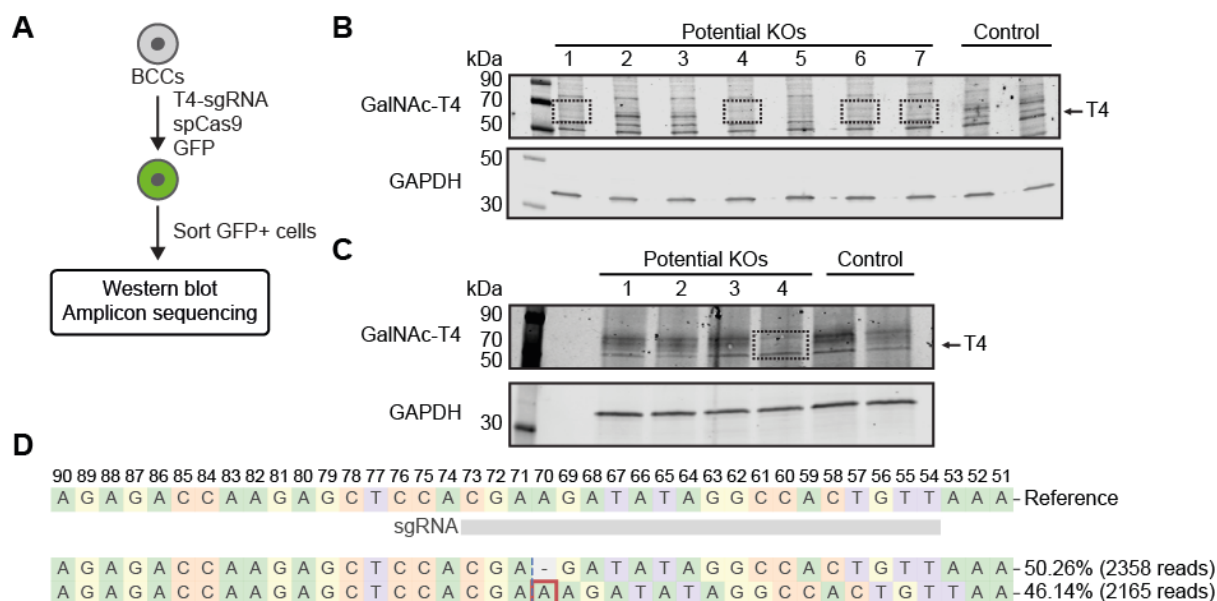


Fig. 11. Establishing GalNAc-T4 KO BCCs. **A)** Schematic representation of the workflow to generate GalNAc-T4 KO BCCs. Western blot against GalNAc-T4 to assess KO of endogenous GalNAc-T4 from **B)** T47D and **C)** MCF7 cells. The predicted MW of GalNAc-T4 and GAPDH is 67 (arrow pointing) and 36 kDa, respectively. Dashed boxes indicate the clones in which a band at the MW of GalNAc-T4 was missing. GAPDH was used as a loading control. 25 μ g of protein were used per lane. Antibody staining was performed using antibodies against GalNAc-T4 (1:200, 12897-1-AP) and GAPDH (1:10000, ab181602). T4 = GalNAc-T4. **D)** Representative allele frequency table of clone T47D^{T4-KO}1 generated by CRISPResso2²¹⁶ after PCR-based amplicon sequencing. Nucleotides are indicated by specific colours (A = green; C = red; G = yellow; T = purple). The vertical dashed line indicates the predicted cleavage site. Red rectangles highlight inserted sequences. Horizontal dashed lines indicate deleted sequences. The allele frequency tables of the other KO clones used in these experiments are shown in **SI Fig. 1**.

1.2. *In vitro* growth assays of WT and GalNAc-T4 KO BCCs

To assess if GalNAc-T4 KO exerts a growth phenotype, *in vitro* growth assays were subsequently performed using the WT and GalNAc-T4 KO T47D and MCF7 cells. Cell growth was monitored for 4 days using an Incucyte® Live-Cell Analysis System (**Fig. 12A**). Apoptosis was monitored by assessing caspase-3 activation using the DEVD-NucView488 peptide (**SI Fig. 5**)²²⁶. This peptide is a non-toxic, highly sensitive, membrane permeable caspase-3 substrate which consists of 2 components: Ac-DEVD (a caspase-3 recognition sequence) and NucView488 (a DNA intercalating dye)²²⁶. NucView488 is only fluorescent when bound to DNA²²⁶. When both components are linked together, the negatively charged nature of the Ac-DEVD group renders the dye incapable of binding to DNA²²⁶. Upon cleavage by caspase-3, the NucView488 is released and becomes fluorescent on binding to DNA²²⁶. Confluence measurements and fluorescent object quantification was performed using the integrated analysis software.

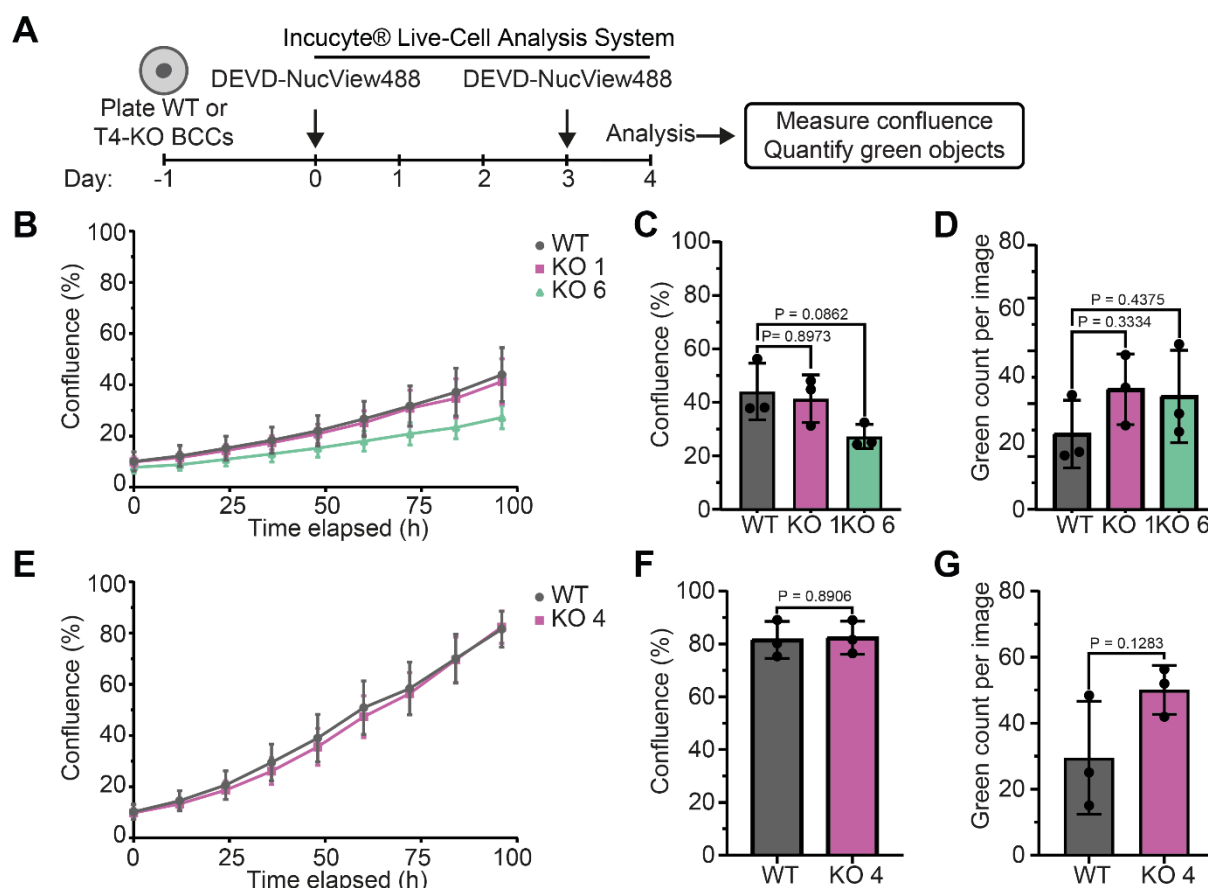


Fig. 12. *In vitro* growth assays of WT and GalNAc-T4 KO BCCs. **A)** Schematic representation of the workflow to compare the growth of WT and GalNAc-T4 KO BCCs *in vitro*. The confluence of WT and GalNAc-T4 KO **B)** T47D and **E)** MCF7 cells was measured for 4 days to assess differences in growth upon GalNAc-T4 KO. A one-way ANOVA test was performed on the **C)** confluence and **D)** green count per image at endpoint of WT and GalNAc-T4 KO T47D cells. An unpaired t-test was performed on the **F)** confluence and **G)** green count per image at endpoint of WT and GalNAc-T4 KO MCF7 cells. Data are individual points and means from three independent replicates, with each independent replicate consisting of three technical replicates $\pm$ standard deviation (SD). P values are indicated in the figure.

Looking at the results of the confluence measurements, there seemed to be a trend towards a lower confluency in T47D^{T4-KO} cells compared to the controls (**Fig. 12B-C**), whilst comparable confluency was observed between MCF7^{WT} and MCF7^{T4-KO} cells (**Fig. 12E-F**). Looking at the results of the green, fluorescent object quantification, there seemed to be a trend towards a higher apoptotic signal in the GalNAc-T4 KO cells compared to the controls in both cell lines (**Fig. 12D** and **Fig. 12G**). However, none of the differences were significant, most likely due to the technical variability in the experiment. For example, since T47D and MCF7 cells are adherent, the initial cell seeding density could influence the overall growth of the cells, where cells that were initially seeded in closer proximity would display more favourable growth compared to those initially seeded further apart. The effect of the technical variability on the results could be minimised by performing additional replicates. Nevertheless, we decided to test alternative methods of assessing the growth of the WT and GalNAc-T4 KO BCCs.

1.3. *In vivo* growth assay of WT and GalNAc-T4 KO BCCs

Since the results of the *in vitro* growth assays were inconclusive, we decided to compare the growth of WT and GalNAc-T4 KO cells *in vivo*, which is more representative of the growth and progression of BC tumours. We decided to test the T47D^{WT} and T47D^{T4-KO} cells first. In these experiments, T47D^{WT} or T47D^{T4-KO} cells were injected into the fat pad of NSG mice and tumour growth was monitored over time by Dr Victoria Bridgeman (**Fig. 13A**). T47D cells are ER+ - these receptors are activated by estrogen, triggering downstream signalling which is key in the proliferation and survival of the cells³⁵. β -oestradiol administration was therefore performed two days prior to tumour implantation and throughout the experiment by Dr Victoria Bridgeman to maintain high estrogen levels for successful tumour growth (**Fig. 13A**)²²⁷. A statistically significant growth defect was seen in T47D^{T4-KO} tumours compared to T47D^{WT} tumours (**Fig. 13B-C**) whilst the body weight of the mice in both groups remained comparable (**Fig. 13D**). This result supports the trend towards a lower confluency observed with the T47D^{T4-KO} cells in the *in vitro* growth assay and the growth defect seen upon GalNAc-T4 KD by Niang et al.¹³² Overall, these results suggested that GalNAc-T4 has a pro-tumorigenic role in T47D cells.

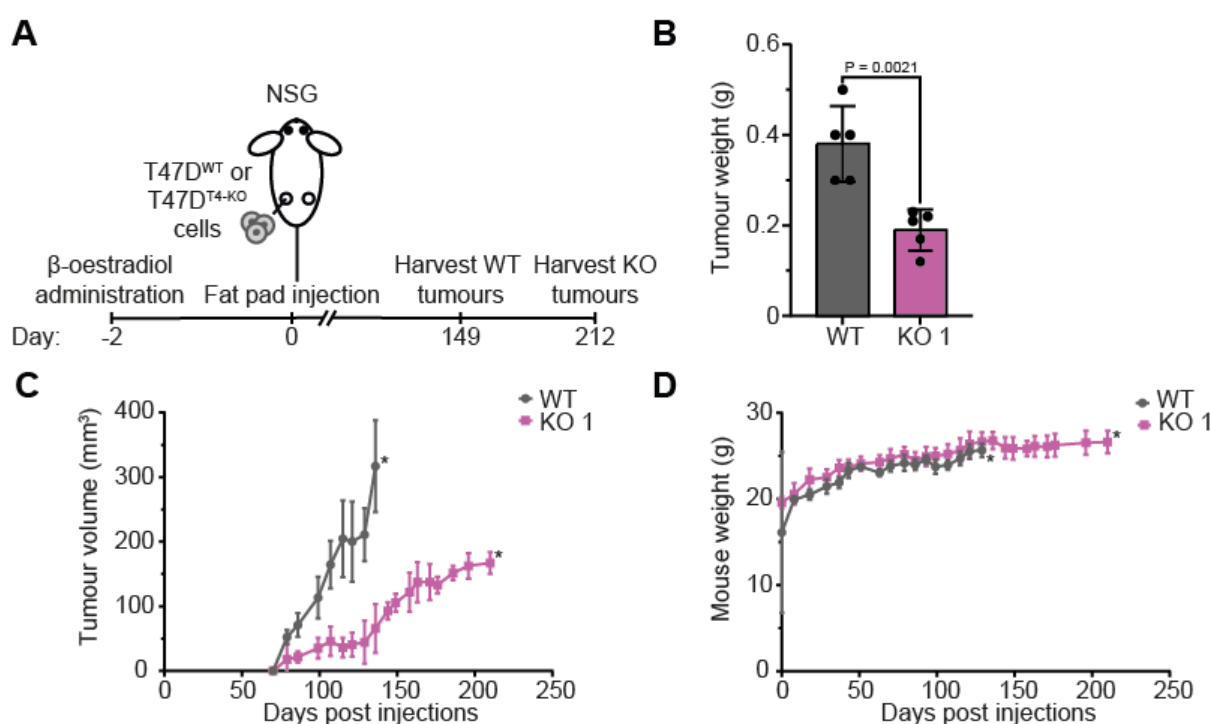


Fig. 13. *In vivo* growth assay of WT and GalNAc-T4 KO T47D cells. **A)** Schematic representation of the workflow for the *in vivo* growth assay with WT and GalNAc-T4 KO T47D cells. T47D^{WT} or T47D^{T4-KO} tumours were grown in the fat pads of NSG mice. Once the tumours measured >200 mm³, the mice were harvested, and the tumour weight was recorded by Dr Victoria Bridgeman. **B)** An unpaired t-test was performed on the weight of T47D^{WT} or T47D^{T4-KO} tumours after harvesting. **C)** The tumour dimensions of T47D^{WT} or T47D^{T4-KO} tumours were measured with a calliper and the tumour volume determined using **Equation 1** by Dr Victoria Bridgeman. **D)** The weight of the mice bearing T47D^{WT} or T47D^{T4-KO} tumours was recorded by Dr Victoria Bridgeman. Data are individual points and means from five mice per group $\pm$ SD. P values are indicated in the figure. The asterix (*) indicates when mice were culled.

We next decided to test the MCF7^{WT} and MCF7^{T4-KO} cells with the same methodology used for the T47D cells (**Fig. 14A**). A statistically significant growth defect was also seen in MCF7^{T4-KO} tumours compared to MCF7^{WT} tumours (**Fig. 14B-C**) whilst the body weight of the mice in both groups remained comparable (**Fig. 14D**). This result contradicts the boost in growth seen upon GalNAc-T4 KO by Wu et al.¹³³ Nevertheless, *in vivo* models are more representative of the growth and progression of BC tumours than one-dimensional cellular models. Overall, these results suggested that GalNAc-T4 also has a pro-tumorigenic role in MCF7 cells, indicating that the contradicting results published previously regarding the role of GalNAc-T4 in T47D and MCF7 cells was likely due to the different methodologies used by the two groups. The exact mechanism behind the pro-tumorigenic role of GalNAc-T4 in BCCs was unclear since the substrates of GalNAc-T4 are not well characterised, but we hypothesised that it could be due to the putative role of GalNAc-T4 in the estrogen regulatory network¹³² and the Notch signalling pathway¹³³. Identifying the substrates of GalNAc-T4 would help elucidate the mechanism by which GalNAc-T4 regulates the proliferation of BCCs. It therefore became important to profile the substrates of GalNAc-T4 in BCCs.

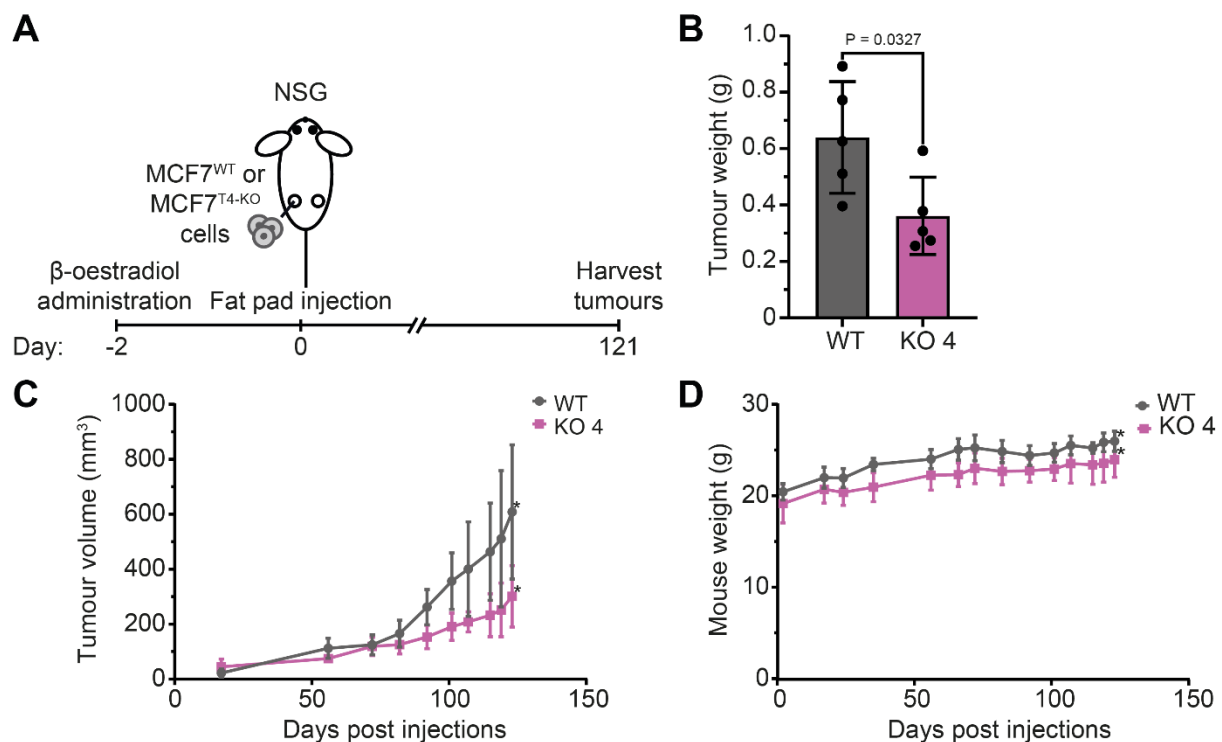


Fig. 14. *In vivo* growth assay of WT and GalNAc-T4 KO MCF7 cells. **A)** Schematic representation of the workflow for the *in vivo* growth assay with WT and GalNAc-T4 KO MCF7 cells. MCF7^{WT} or MCF7^{T4-KO} tumours were grown in the fat pads of NSG mice. Once the tumours measured >200 mm³, the mice were harvested, and the tumour weight was recorded by Dr Victoria Bridgeman. **B)** An unpaired t-test was performed on the weight of MCF7^{WT} or MCF7^{T4-KO} tumours after harvesting. **C)** The tumour dimensions of MCF7^{WT} or MCF7^{T4-KO} tumours were measured with a calliper and the tumour volume determined using **Equation 1** by Dr Victoria Bridgeman. **D)** The weight of mice bearing MCF7^{WT} or MCF7^{T4-KO} tumours was recorded by Dr Victoria Bridgeman. Data are individual points and means from five mice per group ± SD. P values are indicated in the figure. The asterisk (*) indicates when mice were culled.

CHAPTER 2: Developing BH GalNAc-T4 systems in breast cancer cells

To profile the substrates of GalNAc-T4 in BCCs, we decided to use the chemical biology bump-and-hole (BH) engineering approach. During my MSci project, I developed a BH GalNAc-T4 system and confirmed that BH GalNAc-T4 emulates the known glycopeptide substrate preference of WT GalNAc-T4 whilst displaying a UDP-sugar substrate switch from the native substrate to the bumped analogue UDP-GalN6yne²⁰⁸. Before extending the BH GalNAc-T4 system to BCCs, the first objective was to test a greater panel of UDP-GalNAc analogues to see if there were any alternative UDP-sugars with more favourable glycopeptide formation and orthogonality than UDP-GalN6yne. Orthogonality is important for *in cellulo* work – if only BH GalNAc-T4 can chemically-tag glycoproteins with the modified sugar (but not the WT enzymes), we can selectively characterise the substrates of GalNAc-T4 in the presence of all other members of the family. With an orthogonal GalNAc-T4/UDP-sugar pair in hand, the next objective was to extend the BH GalNAc-T4 system to BCCs and characterise the substrates and glycosylation sites of GalNAc-T4 by MS-glycoproteomics.

2.1. *In vitro* validation of BH GalNAc-T4

2.1.1. In vitro glycosylation experiments with a panel of UDP-GalNAc analogues

The soluble constructs of truncated WT and BH GalNAc-T4 tagged with GFP previously used during my MSci project (**Fig. 10A**)²⁰⁸ were herein expressed in Expi293F cells and purified by GFP-affinity purification by the Structural Biology STP at the Francis Crick Institute. Protein expression control by Western blot confirmed the identity of the recombinant proteins (**SI Fig. 6**).

I subsequently tested the UDP-sugar substrate preference of WT and BH GalNAc-T4 using an endpoint *in vitro* glycosylation experiment with a panel of UDP-GalNAc analogues to see if there were any alternative UDP-sugars with more favourable glycopeptide formation and orthogonality than UDP-GalN6yne. The reaction mixtures contained WT or BH GalNAc-T4, a synthetic glycopeptide substrate (glycopeptide **1**), and a panel of UDP-GalNAc analogues (UDP-sugar **1-16**)(**Fig. 15A**). Glycopeptide **1** was chosen as it was the one that showed the highest glycopeptide formation when I tested the glycopeptide substrate preference of WT and BH GalNAc-T4 during my MSci project²⁰⁸.

UDP-sugar **1** (UDP-GalNAc) is the native substrate of GalNAc-T4. UDP-sugars **2-14** are UDP-GalNAc analogues containing an azide or alkyne group at the acetamide moiety to allow derivatisation of the tagged glycoproteins by CuAAC. UDP-sugars **2-14** have been previously synthesised and tested with WT and BH GalNAc-T1, -T2 and -T10 by Choi et al.⁶⁷ I also tested these compounds using WT and BH GalNAc-T4 during my MSci project²⁰⁸. UDP-sugar **2** is UDP-GalN6yne – the UDP-GalNAc analogue showing the most promising results with all BH GalNAc-Ts tested to date^{67, 208}. UDP-sugar **16** was synthesised by Ganka Bineva-Todd and it contains a cyclopropene moiety to allow derivatisation of the tagged glycoproteins by an inverse electron demand Diels–Alder (IEDDA) reaction instead of by CuAAC, and is therefore an alternative for bioorthogonal derivatisation without employing cytotoxic Cu(I)²²⁸. UDP-sugar **15** was a gift from Dr Jennifer Kohler and it does not contain a moiety to allow bioorthogonal derivatisation of the tagged glycoproteins. Instead, it contains a diazirine group which allows photocrosslinking between the tagged glycoproteins and neighbouring molecules upon ultraviolet irradiation^{205, 229}. This compound could therefore be used to find interaction partners of the substrates of GalNAc-T4^{205, 229}.

Many glycosyltransferase assays rely on the detection of UDP as a side product of the glycosylation reaction, however background hydrolysis of UDP-sugars may skew the results of these assays⁶⁷. Quantification of glycopeptide formation by UV peak integration was unfeasible due to overlapping starting material and product peaks. Instead, the reaction mixtures were incubated overnight at 37 °C and glycopeptide formation was determined by MS+ peak integration using **Equation 2**. Accurate quantification by MS+ peak integration was possible due to comparable ionisation efficiencies of the starting material and product²⁰³.

The results supported the findings reported by Choi et al.⁶⁷ and the preliminary data obtained during my MSci project²⁰⁸, where the analogues with longer or branched side chains (UDP-sugar **2** and UDP-sugars **6-16**) were generally better substrates for BH GalNAc-T4 than the WT enzyme (**Fig. 15B**). The most promising results were still obtained with UDP-GalN6yne (UDP-sugar **2**) - comparable glycopeptide formation (>90%) was observed between the BH GalNAc-T4/UDP-GalN6yne and WT GalNAc-T4/UDP-GalNAc pairs whilst reduced activity (<30%) was observed with the opposite GalNAc-T4/UDP-sugar pairs. Encouraging results were also obtained with the UDP-sugars **14**, **15** and **16**, where substantial turnover (>50%) was seen with BH GalNAc-T4 and minimal activity was seen with the WT enzyme.

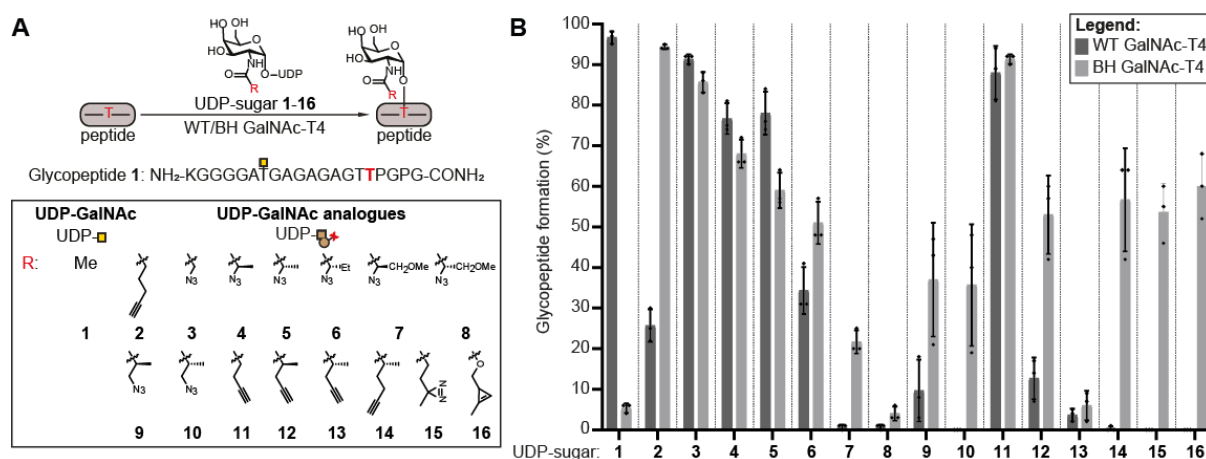


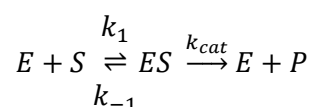
Fig. 15. Endpoint *in vitro* glycosylation experiments with WT/BH GalNAc-T4 and a panel of UDP-GalNAc analogues. **A)** Schematic representation of *in vitro* glycosylation experiments performed to evaluate the UDP-sugar substrate preference of WT and BH GalNAc-T4. Red T represents the peptide acceptor site used by GalNAc-T4. The yellow square depicts the prior GalNAc unit in the peptide. Figure adapted from Choi et al. (2019)⁶⁷. **B)** Glycopeptide formation by WT/BH GalNAc-T4 (100 nM) after overnight incubation at 37 °C with glycopeptide 1 (50 µM) and UDP-sugars 1-16 (250 µM). Data are individual points and means from three independent replicates, with each independent replicate consisting of two technical replicates ± SD.

To enable *in cellulo* work, WT GalNAc-Ts should not accept the UDP-GalNAc analogue. Choi et al. previously demonstrated that UDP-GalN6yne is not accepted by WT GalNAc-T1, -T2 and -T10⁶⁷, therefore we decided to pursue UDP-GalN6yne as a promising orthogonal GalNAc-T/UDP-sugar pair with the most favourable turnover compared to the other UDP-sugars tested.

2.1.2. Michaelis-Menten Kinetics

To determine whether BH GalNAc-T4 retains the kinetic parameters of the native system, the Michaelis-Menten kinetics of WT and BH GalNAc-T4 were investigated. Enzymatic reactions can be generally described by **Equation 3**, where a substrate (S) binds to an enzyme (E) to form an intermediate enzyme-substrate (ES) complex, which either dissociates to the unbound substrate and enzyme, or proceeds to yield the product (P) and regenerate the enzyme²³⁰.

Equation 3. General equation to describe a simple enzymatic reaction $S \rightarrow P$ ²³⁰.



The overall rate of product formation (v) is therefore directly proportional to the concentration of the ES complex as described by **Equation 4**²³⁰. This equation only holds true at the beginning of the reaction before the product accumulates and whilst the reverse reaction ($E + P \rightarrow ES$) can be assumed to be negligible so there is no product inhibition²³⁰.

Equation 4. Overall rate of product formation²³⁰.

$$v = k_{cat} [ES]$$

Under the steady-state approximation, the rates at which the ES complex forms and disappears are equal, so the concentration of the ES complex remains small and constant²³⁰. The concentration of the ES complex can therefore be described in terms of the concentration of the free enzyme and free substrate using **Equation 5**²³⁰. This approximation only holds true at the beginning of the reaction, before the concentration of the substrate, and hence of the ES complex, changes appreciably²³⁰.

Equation 5. Describing the concentration of the ES complex in terms of the concentrations of free enzyme and free substrate when the steady-state approximation is satisfied²³⁰.

$$k_1[E][S] = k_{-1}[ES] + k_{cat}[ES]$$

$$[ES] = \frac{k_1[E][S]}{k_{-1} + k_{cat}}$$

The concentration of free enzyme [E] and free substrate [S] will be equal to the total enzyme [E_T] or total substrate [S_T] concentration minus the concentration of the ES complex [ES]²³⁰. However, since the enzyme is a catalyst, the concentration of the substrate will be much greater than the concentration of the enzyme²³⁰. It can therefore be assumed that the free substrate concentration [S] will be almost the same as the total substrate concentration ([S] ≈ [S_T])²³⁰. Substituting for [E] and rearranging **Equation 5** results in **Equation 6**²³⁰.

Equation 6. Describing the concentration of the ES complex in terms of the concentrations of total enzyme and free substrate when the steady-state approximation is satisfied²³⁰.

$$[ES] = \frac{k_1 ([E_T] - [ES])[S]}{k_{-1} + k_{cat}}$$

$$[ES](k_{-1} + k_{cat} + k_1[S]) = k_1[E_T][S]$$

$$[ES] \left(\frac{k_{-1} + k_{cat}}{k_1} + [S] \right) = [E_T][S]$$

$$[ES] = \frac{[E_T][S]}{\left(\frac{k_{-1} + k_{cat}}{k_1} \right) + [S]}$$

Replacing **Equation 6** into **Equation 4** allows us to define the rate of product formation in terms of the total enzyme and free substrate concentrations using **Equation 7**²³⁰.

Equation 7. Rate of product formation in terms of total enzyme and free substrate concentration²³⁰.

$$v = \frac{k_{cat} [E_T][S]}{\left(\frac{k_{-1} + k_{cat}}{k_1}\right) + [S]}$$

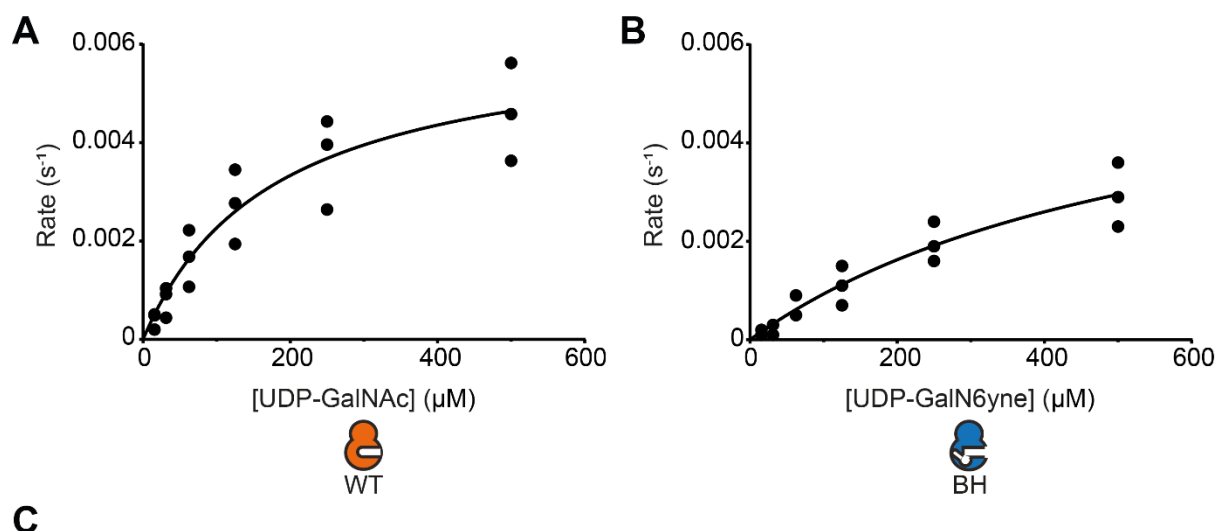
When the enzyme is saturated with the substrate, the total enzyme concentration $[E_T]$ will be equal to the ES complex concentration ($[E_T] = [ES]$), and the rate of product formation will reach a limiting maximum value (v_{max})²³⁰. Under these criteria, we can combine **Equation 4** and **Equation 7** and arrive at the Michaelis-Menten equation (**Equation 8**), where the term K_M is the Michaelis constant²³⁰. This constant represents the substrate concentration at half-maximal reaction rate, and it provides an indication of the enzyme's affinity for its substrate²³⁰.

Equation 8. Michaelis-Menten equation²³⁰.

$$v = \frac{v_{max} [S]}{K_M + [S]}$$

During my MSci project I determined the enzyme concentration at which 10-20% glycopeptide formation was achieved with the WT GalNAc-T4/UDP-GalNAc and BH GalNAc-T4/UDP-GalN6yne pairs with glycopeptide **1**²⁰⁸. This turnover was chosen as a reflection of the initial stage of the reaction (when minimal substrate consumption has occurred, but sufficient product is present to be detected by MS), thus fulfilling the prerequisites of Michaelis-Menten kinetics²³⁰. Incubation of the reaction mixtures for 1.5 h was chosen as a convenient endpoint to measure glycopeptide formation at the initial stage of the reaction²⁰⁸.

To determine the kinetic parameters of the GalNAc-T4/UDP-sugar pairs, the reaction mixtures were prepared containing fixed enzyme and glycopeptide **1** concentrations and different concentrations of UDP-GalNAc or UDP-GalN6yne. The concentration of glycopeptide **1** was in excess to ensure that the reaction was pseudo-first order with respect to the UDP-sugar concentration (the K_M of WT GalNAc-T4 with glycopeptide **1** has been reported as $41 \pm 19 \mu\text{M}$ ⁷⁴ and $50.7 \pm 6.3 \mu\text{M}$ ⁸²)²⁰⁸.



C

GalNAc-T4 / UDP-sugar	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($mM^{-1}s^{-1}$)	R-square fit
WT / UDP-GalNAc	1.6	0.2	8.9	0.87
BH / UDP-GalN6yne	1.6	0.6	2.7	0.91

Fig. 16. Michaelis-Menten kinetics of WT GalNAc-T4/UDP-GalNAc and BH GalNAc-T4/UDP-GalN6yne. Michaelis-Menten plots of **A)** WT GalNAc-T4/UDP-GalNAc and **B)** BH GalNAc-T4/UDP-GalN6yne. Initial rates were measured by incubating WT/BH GalNAc-T4 (4 nM) at 37 °C for 1.5 h with glycopeptide **1** (100 μ M) and varying concentrations of UDP-GalNAc or UDP-GalN6yne (15.6, 31.3, 62.5, 125, 250, 500 μ M). Data are individual points and means from three independent replicates, with each independent replicate consisting of two technical replicates. **C)** Michaelis-Menten kinetic parameters of WT and BH GalNAc-T4/UDP-sugar pairs. The non-linear Michaelis-Menten and k_{cat} fitting programs in GraphPad Prism were used to calculate the k_{cat} and the K_M of both enzyme/substrate pairs.

The catalytic constant (k_{cat}) of both enzyme/substrate pairs was the same ($1.6 s^{-1}$)(**Fig. 16C**). In contrast, the Michaelis constant (K_M) of the WT GalNAc-T4/UDP-GalNAc pair (0.2 mM) was 3-fold lower than that of the BH GalNAc-T4/UDP-GalN6yne pair (0.6 mM)(**Fig. 16C**), reflecting the higher affinity occurring within the active site of the WT enzyme. Consequently, the catalytic efficiency (k_{cat}/K_M) of the WT pair ($8.9 mM^{-1}s^{-1}$) was 3-fold higher than that of the BH GalNAc-T4/UDP-GalN6yne pair ($2.7 mM^{-1}s^{-1}$)(**Fig. 16C**). Nevertheless, since all the values are within the same order of magnitude, the kinetic parameters of both enzyme-substrate pairs were considered comparable.

2.1.3. Competition experiment with the native substrate (UDP-GalNAc) and the bumped analogue (UDP-GalN6yne)

In the endpoint *in vitro* glycosylation experiments, comparable glycopeptide formation was observed between the WT GalNAc-T4/UDP-GalNAc (97%) and BH GalNAc-T4/UDP-GalN6yne (94%) pairs (**Fig. 15B**). However, since some activity was also observed between the WT GalNAc-T4/UDP-GalN6yne (5%) and BH GalNAc-T4/UDP-GalNAc (26%) pairs (**Fig. 15B**), we decided to investigate the selectivity of the enzymes in the presence of a 1:1 mixture of the two UDP-sugars (**Fig. 17A**). This is important for *in cellulo* work as it better represents the substrate exposure in the secretory pathway²³¹.

To be able to assess the selectivity of the enzymes towards UDP-GalNAc or UDP-GalN6yne, we had to ensure that the reaction mixtures were under steady-state kinetics. I therefore used the enzyme concentration at which ~20% glycopeptide formation was achieved²⁰⁸ and incubated the reaction mixtures for 1.5 h as a convenient endpoint to measure glycopeptide formation at the initial stage of the reaction²⁰⁸.

Upon incubation of WT GalNAc-T4 with a 1:1 molar ratio of both UDP-GalNAc and UDP-GalN6yne, a 16:1 ratio of GalNAc- vs GalN6yne-modified products were found (**Fig. 17B**). Thus, WT GalNAc-T4 preferred the native substrate UDP-GalNAc. Upon incubation of BH GalNAc-T4 with a 1:1 molar ratio of both UDP-GalNAc and UDP-GalN6yne, only GalN6yne-modified products were found (**Fig. 17B**). Thus, BH GalNAc-T4 exclusively accepted the bumped analogue UDP-GalN6yne.

Taken together, these results demonstrated that in an environment that more closely mimics the secretory pathway, WT GalNAc-T4 preferentially reacts with UDP-GalNAc whilst BH GalNAc-T4 exclusively reacts with UDP-GalN6yne, confirming that the BH GalNAc-T4/UDP-GalN6yne pair is an orthogonal enzyme-substrate pair to the native pair (WT GalNAc-T4/UDP-GalNAc).

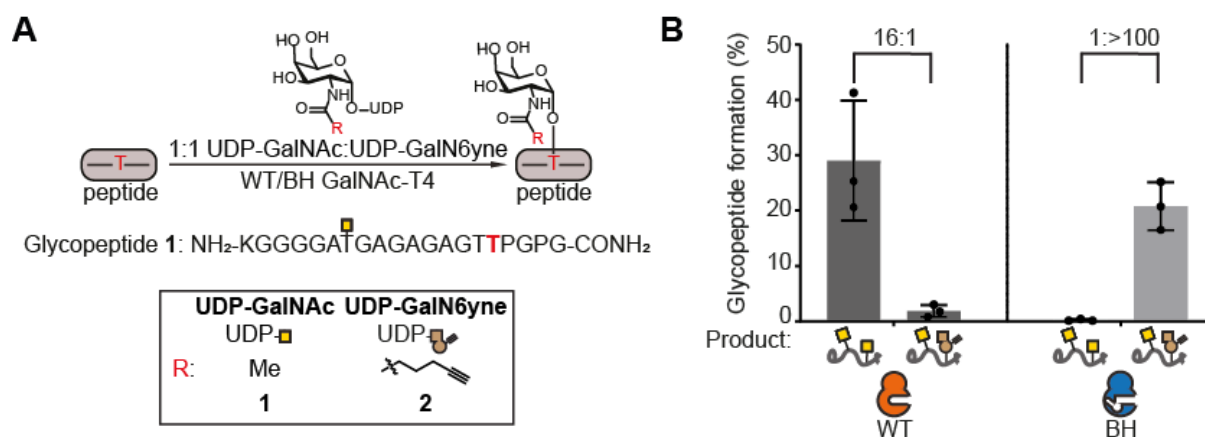


Fig. 17. Competition experiment with WT/BH GalNAc-T4 and UDP-GalNAc/UDP-GalN6yne. **A)** Schematic representation of competition experiments performed with WT/BH GalNAc-T4 and a 1:1 ratio of UDP-GalNAc and UDP-GalN6yne. Red T represents the peptide acceptor site used by GalNAc-T4. The yellow square depicts the prior GalNAc unit in the peptide. Figure adapted from Choi et al. (2019)⁶⁷. **B)** Glycopeptide formation by WT/BH GalNAc-T4 (8 nM) after incubation at 37 °C for 1.5 h with glycopeptide 1 (50 µM) and a 1:1 mixture of UDP-GalNAc and UDP-GalN6yne (250 µM). Data are individual points and means from three independent replicates, with each independent replicate consisting of two technical replicates ± SD.

Overall, the *in vitro* glycosylation experiments validated the BH GalNAc-T4/UDP-GalN6yne pair as an orthogonal enzyme/substrate pair showing comparable glycopeptide formation and kinetic parameters as the native pair (WT GalNAc-T4/UDP-GalNAc). During my MSci project, I also confirmed that BH GalNAc-T4 emulated the known glycopeptide substrate preference of WT GalNAc-T4²⁰⁸. Altogether, these results demonstrated that BH GalNAc-T4 recapitulates the activity of the WT enzyme *in vitro* whilst displaying a UDP-sugar substrate switch from the native substrate to the bumped analogue.

2.2. Establishing a BH GalNAc-T4 system in BCCs

2.2.1. Cell surface tagging of glycoproteins in T47D^{WT} and T47D^{T4-KO} cells

With an orthogonal BH GalNAc-T4/UDP-sugar pair in hand, the next aim was to establish a BH GalNAc-T4 system in BCCs. The T47D cell line, an ER+ luminal BCC line displaying high GalNAc-T4 expression¹³², was identified as a promising model system for these experiments. Previous work on T47D cells has demonstrated the use of this cell line as a representative model of BC-associated alterations of O-GalNAc glycosylation due to its resemblance to the structural and antigenic profile of primary breast tumours²³².

Two T47D^{T4-KO} clones (KO 1 and KO 6) and T47D^{WT} cells were stably transfected with plasmids encoding the two enzymes of the artificial biosynthetic pathway for UDP-sugar biosynthesis (AGX1^{F383A} and NahK) and WT or BH GalNAc-T4 (T47D-AGX1^{F383A}-NahK-GalNAc-T4^{WT} and T47D-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells respectively)(**Fig. 18A** and **SI Fig. 7**). To have a control without OE WT or BH GalNAc-T4, the cells were also transfected with plasmids only expressing the artificial biosynthetic pathway (T47D-AGX1^{F383A}-NahK cells). Since UDP-GalN6yne was shown to be epimerised by the enzyme GALE to the corresponding UDP-GlcNAc analogue¹⁹³, resulting in undesired tagging of GlcNAc-containing glycoconjugates such as nucleocytoplasmic O-GlcNAc or extracellular N-glycans, these cells would also help assess any potential background tagging of glycoproteins.

A dose-response experiment was initially performed to determine the concentration of Ac₄GalN6yne to use in subsequent experiments. The transfected T47D^{T4-KO} and T47D^{WT} cells were fed with different concentrations of Ac₄GalN6yne and reacted with the fluorophore CF680-picolyl azide, which undergoes a CuAAC reaction with the tagged glycoproteins (**Fig. 18B**). DMSO-treated cells were included as a negative control. Cells treated with Ac₄ManNAI, a well-characterised and promiscuous sialic acid precursor (**Fig. 18E**)²³³, were included as a positive control. Cell surface tagging of glycoproteins was assessed by in-gel fluorescence. Protein expression control by Western blot demonstrated the success of the transfection (**Fig. 18C**). As expected, no fluorescent signal was seen in the DMSO-treated cells whilst a strong signal was seen in all cells treated with Ac₄ManNAI regardless of whether BH GalNAc-T4 was present or not (except for sample 24 in **Fig. 18C** since there wasn't enough CuAAC solution left for this sample).

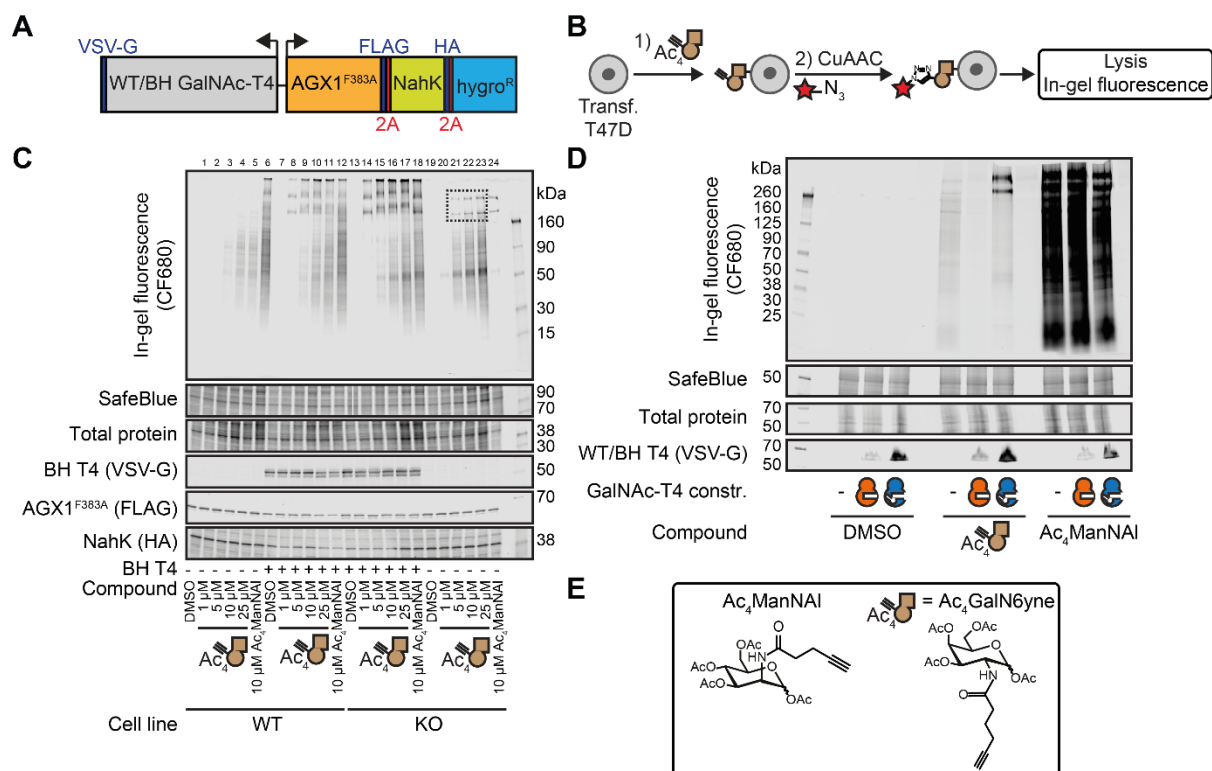


Fig. 18. Cell surface tagging of glycoproteins in transfected T47D^{WT} and T47D^{T4-KO} cells using BH GalNAc-T4. **A)** Expression construct encoding genes for VSV-G-tagged WT/BH GalNAc-T4, FLAG-tagged AGX1^{F383A} and HA-tagged NahK used to establish the BH GalNAc-T4 system in T47D^{WT} and T47D^{T4-KO} cells. **B)** Schematic representation of the workflow for the cell surface tagging experiments. The red star represents a fluorophore. Figure adapted from Schumann et al. (2020)¹⁹³. **C)** Dose-response experiment of cell surface tagging of glycoproteins in transfected T47D^{WT} and T47D^{T4-KO} cell lysates as assessed by in-gel fluorescence. The dashed box indicates bands seen in the T47D^{T4-KO} cells which were absent in the corresponding T47D^{WT} cells. 15 µg of protein were used per lane. Antibody staining was performed using antibodies against VSV-G (1:500, ab3861), FLAG (1:2000, PA1-984B) and HA (1:2000, ab130275). **D)** Cell surface tagging of glycoproteins in PNGase F-treated transfected T47D^{WT} cell lysates as assessed by in-gel fluorescence. 15 µg of protein were used per lane. Antibody staining was performed using an antibody against VSV-G (1:500, ab3861). Data are from one experiment. WT/BH T4 = WT/BH GalNAc-T4.

T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} and T47D^{T4-KO}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells showed a concentration-dependent increase in fluorescence with Ac₄GalN6yne, demonstrating that we can successfully tag glycoproteins using BH GalNAc-T4 in these cells. The strongest signal in all the BH GalNAc-T4 samples was observed in bands at a MW >160 kDa. We speculated that these bands could be due to mucins, since MUC1 is one of the few validated substrates of GalNAc-T4^{72, 156, 157, 234}. To test our hypothesis, we decided to treat the lysates with StcE prior to in-gel fluorescence. StcE is a protease that selectively cleaves mucins containing the peptide sequence Ser/Thr*-X-Ser/Thr, where “X” is any amino acid and the asterisk (*) represents a glycosylated Ser/Thr residue¹⁸⁷. Cleavage occurs before the second Ser/Thr only if the first Ser/Thr is glycosylated¹⁸⁷. The bands at a MW >160 kDa were sensitive to treatment with StcE (**SI Fig. 8B**), confirming their identity as mucins.

The signal in the T47D^{T4-KO}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells was higher than in the corresponding T47D^{WT} cells, likely because of the competition for glycosylation sites between BH and endogenous GalNAc-T4 reducing the tagging in the T47D^{WT} cells. A signal was observed in T47D^{WT}-AGX1^{F383A}-NahK and T47D^{T4-KO}-AGX1^{F383A}-NahK cells due to background tagging of glycoproteins. Based on these results, 1 μ M Ac₄GalN6yne was chosen for future experiments since at this sugar concentration a strong signal was observed in the cells expressing BH GalNAc-T4 whilst minimal signal was seen in the cells where BH GalNAc-T4 was absent.

Bands at a MW >160 kDa were seen in the T47D^{T4-KO}-AGX1^{F383A}-NahK cells that were not observed in the corresponding T47D^{WT} cells (**Fig. 18C**, dashed box), indicating that compensatory mechanisms may be taking place upon KO of GalNAc-T4, as previously suggested by Wu et al.¹³³ Since using these cells could result in misleading conclusions, only transfected T47D^{WT} cells would be used in future experiments, without KO of endogenous GalNAc-T4.

Having determined the optimal concentration of Ac₄GalN6yne to use (1 μ M Ac₄GalN6yne), the cell surface tagging experiments were repeated with T47D^{WT}-AGX1^{F383A}-NahK, T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{WT}, and T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. Compared to the previous experiment (**Fig. 18C**), more background signal was seen in T47D^{WT}-AGX1^{F383A}-NahK cells (**Fig. 18D**). To discern between N- and O-glycans and reduce background fluorescence, the lysates were treated with PNGase F prior to in-gel fluorescence analysis. The signal of all samples was reduced upon PNGase F treatment (**SI Fig. 8C**), demonstrating that tagged N-glycans were successfully removed. The remaining signal could be due to incorporation of the GlcNAc analogue product of the epimerisation by GALE into the protein-proximal core, making N-glycans PNGase F resistant²⁰⁴. Alternatively, the bulky fluorophore used to derivatise the tagged glycoproteins may reduce recognition by PNGase F. In the Ac₄GalN6yne-treated samples, the strongest signal was observed in the T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells (**Fig. 18D**), with a 1.5-fold and 4.7-fold increase in mean fluorescent intensity over the T47D^{WT}-AGX1^{F383A}-NahK and T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{WT} cells, respectively. The VSV-G blot demonstrated that BH GalNAc-T4 was more highly OE than WT GalNAc-T4 (**Fig. 18D**). Since this difference in expression levels would be a source of variation that could lead to artifacts, we decided not to use the T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{WT} cells in future experiments. Instead, we would use the T47D^{WT}-AGX1^{F383A}-NahK as a control to dissect BH GalNAc-T4-specific tagging from any potential background tagging of glycoproteins.

2.2.2. Tagging of glycoproteins in lysates and secretome of transfected T47D cells

We next tested if we could tag intracellular and secreted proteins as well as cell surface proteins in T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} and T47D^{WT}-AGX1^{F383A}-NahK cells. The cell surface tagging experiment was performed as described above (**Fig. 19A**). For the tagging of glycoproteins in lysates and the secretome of transfected T47D^{WT} cells, 24 h after sugar feeding, the secretome was collected and the cells were lysed. The samples were treated with PNGase F to remove N-glycans and the glycoproteins were derivatised with biotin-picolyl azide by CuAAC. Tagging of glycoproteins was assessed by streptavidin blot (**Fig. 19C**).

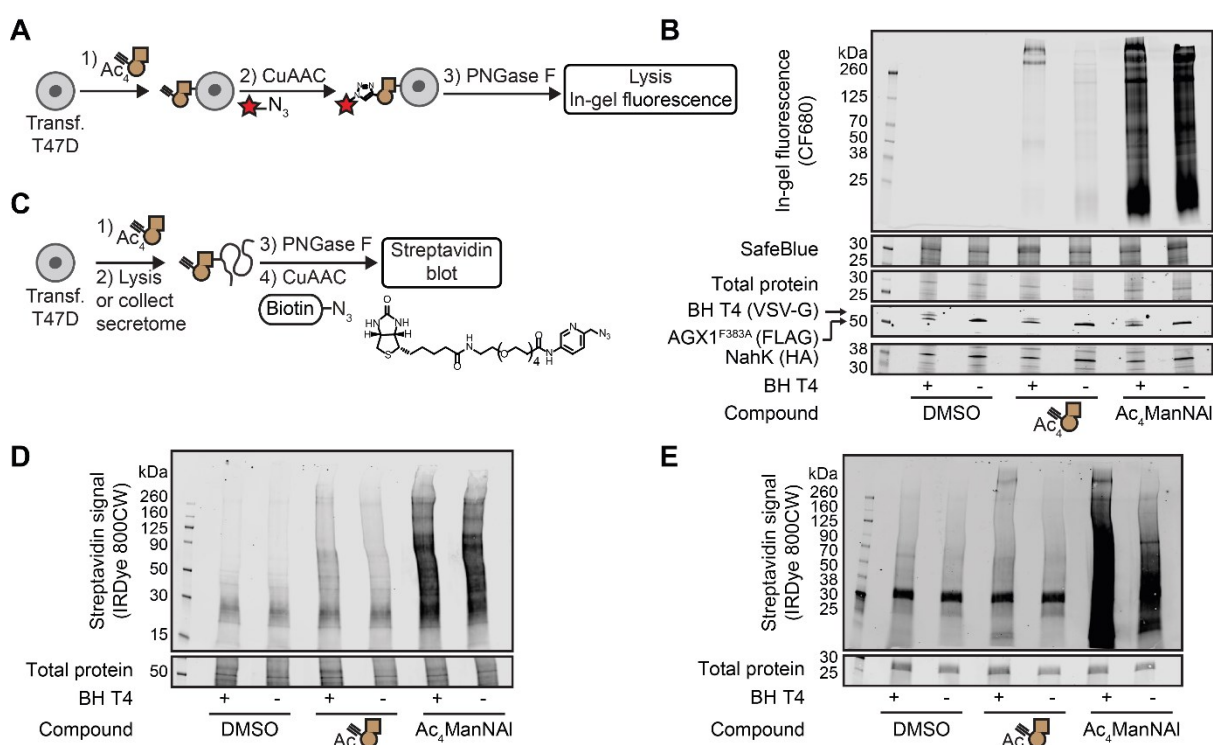


Fig. 19. Cell surface, lysate and secretome tagging of glycoproteins in transfected T47D^{WT} cells using BH GalNAc-T4. Schematic representation of the workflow for the **A**) cell surface, **C**) lysate and secretome tagging experiments. **B**) Cell surface tagging of glycoproteins in PNGase F-treated transfected T47D^{WT} cell lysates as assessed by in-gel fluorescence. Tagging of glycoproteins in PNGase F-treated **D**) lysates and **E**) the secretome of transfected T47D^{WT} cells as assessed by streptavidin blot. 15 µg of protein were used per lane for the cell surface and lysate samples and 5 µg for the secretome samples. Antibody staining was performed using antibodies against VSV-G (1:500, ab3861), FLAG (1:2000, PA1-984B) and HA (1:2000, ab130275). Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). Data are from one experiment. BH T4 (+) = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells, control (-) = T47D^{WT}-AGX1^{F383A}-NahK cells.

The results from the tagged cell surface glycoproteins supported the results from the cell surface tagging experiments described in **Fig. 18** - minimal signal was seen in the DMSO-treated samples, a strong signal was seen in both samples treated with Ac₄ManNAI, some background signal was seen in the T47D^{WT}-AGX1^{F383A}-NahK cells treated with Ac₄GalN6yne, but the strongest bands were still observed in the T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} samples, especially at a MW >160 kDa (**Fig. 19B**). A strong band around 25 kDa was seen in all the lysates (**Fig. 19D**) and secretome (**Fig. 19E**) samples, perhaps due to endogenous biotinylated proteins. A higher background signal was seen in the T47D^{WT}-AGX1^{F383A}-NahK lysates treated with Ac₄GalN6yne (**Fig. 19D**) compared to the corresponding cell surface (**Fig. 19B**) and secretome (**Fig. 19E**) samples, probably due to the high abundance of tagged nucleocytoplasmic O-GlcNAc glycans⁶². Nevertheless, the highest signal in the Ac₄GalN6yne-treated samples was consistently observed in the T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} samples (**Fig. 19B** and **Fig. 19D-E**) - demonstrating that we can successfully tag glycoproteins in the cell surface, lysates and secretome of T47D cells using BH GalNAc-T4.

2.2.3. Cell surface tagging of glycoproteins in transfected MCF7 and MDA-MB-231 cells

Once I had demonstrated that we could tag glycoproteins in T47D^{WT} cells with BH GalNAc-T4, we wanted to see if the BH GalNAc-T4 system would be applicable to other BCCs. For these experiments, we chose MCF7^{WT} cells as a second ER+ luminal BCC line expressing comparable levels of endogenous GalNAc-T4 to T47D^{WT} cells¹³³, and the MDA-MB-231^{WT} cell line as a TNBC cell line expressing GalNAc-T4 at much lower levels than luminal cells^{129, 133}. Both cell lines were transfected with the plasmids encoding NahK, AGX1^{F383A}, and WT or BH GalNAc-T4. Cell surface tagging experiments were performed as described for T47D cells (**Fig. 20A**).

Protein expression control by Western blot confirmed the expression of the transgenes (**Fig. 20B-C**). Some background signal was observed in the DMSO-treated samples (lanes 1 and 2, **Fig. 20B-C**), most likely due to leftover CuAAC solution which should be removed with additional washes. A strong signal was seen in all samples treated with Ac₄ManNAI (lanes 9 and 10, **Fig. 20B-C**). Some background signal was seen in the AGX1^{F383A}-NahK (lanes 3 and 4, **Fig. 20B-C**) and AGX1^{F383A}-NahK-GalNAc-T4^{WT} (lanes 5 and 6, **Fig. 20B-C**) cells fed with Ac₄GalN6yne despite PNGase F treatment. Interestingly, the signal from these samples was slightly different. A possible explanation for this observation could be that AGX1^{F383A} and NahK are OE in the AGX1^{F383A}-NahK cells compared to the AGX1^{F383A}-NahK-GalNAc-T4^{WT} cells, potentially due to the smaller former plasmid resulting in a higher transfection efficiency than the latter plasmid, which in turn leads to increased background tagging of glycoproteins.

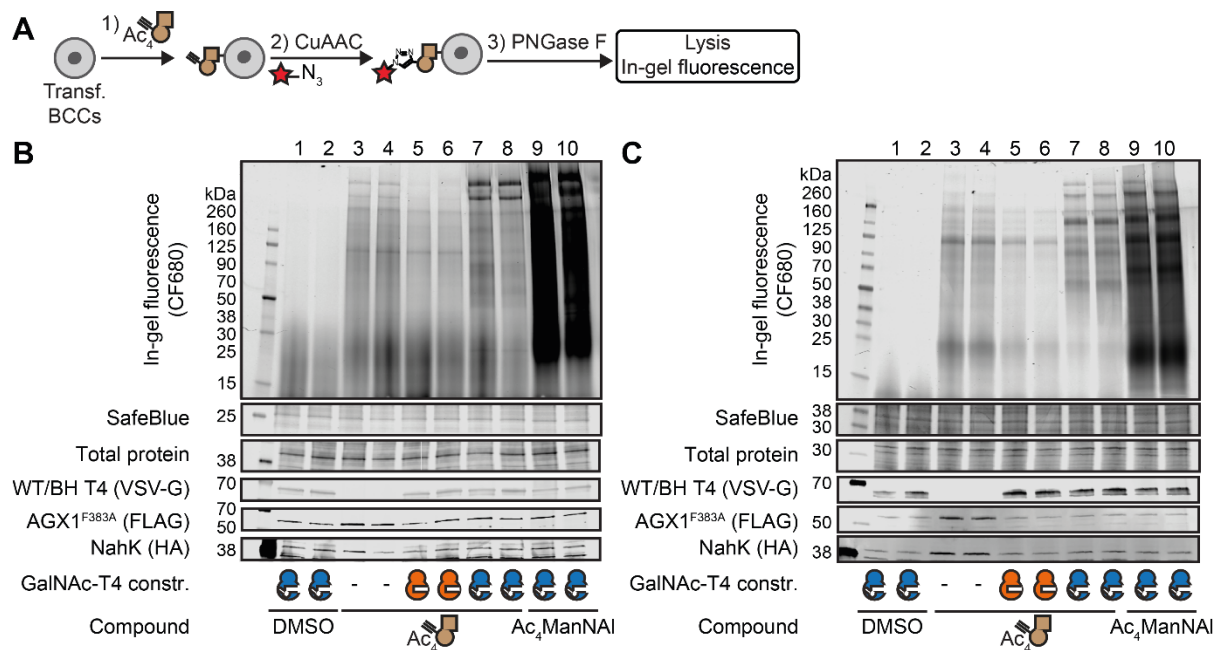


Fig. 20. Cell surface tagging of glycoproteins in transfected MCF7^{WT} and MDA-MB-231^{WT} cells using BH GalNAc-T4. **A)** Schematic representation of the workflow for the cell surface tagging experiments. Cell surface tagging of glycoproteins in PNGase F-treated transfected **B)** MCF7^{WT} and **C)** MDA-MB-231^{WT} cell lysates as assessed by in-gel fluorescence. 15 µg of protein were used per lane. Antibody staining was performed using antibodies against VSV-G (1:750, PA129903), FLAG (1:500, ab1257) and HA (1:500, ab18181). Data are representative of two independent replicates, with each independent replicate consisting of two technical replicates. WT/BH T4 = WT/BH GalNAc-T4.

Multiple intense bands due to the tagged glycoproteins were observed in the AGX1^{F383A}-NahK-GalNAc-T4^{BH} samples for both cell lines (lanes 7 and 8, **Fig. 20B-C**), with the most intense bands consistently observed at a MW >160 kDa. Overall, these results demonstrated that we could tag the substrates of GalNAc-T4 in T47D, MCF7 and MDA-MB-231 cells. The different band patterns observed with the different cell lines could reflect the differing substrate specificity of GalNAc-T4.

CHAPTER 3: Characterising the substrates of GalNAc-T4 in breast cancer cells by MS-proteomics

Having demonstrated that we could tag the substrates of GalNAc-T4 in T47D, MCF7 and MDA-MB-231 BCCs, the next step was to characterise the tagged glycoproteins by MS-proteomics.

3.1. Label-free MS-proteomics on lysates & secretome of transfected T47D^{WT} cells

I first performed a label-free MS-proteomics experiment to characterise the substrates of GalNAc-T4 in the lysates and secretome of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. For this experiment, I used the T47D^{WT}-AGX1^{F383A}-NahK cells as a control for background tagging of glycoproteins. I fed the cells with Ac₄GalN6yne or DMSO. After overnight incubation I collected the secretome and lysed the cells. I then incubated the samples with neutravidin beads to remove endogenous biotinylated proteins (pre-clear), followed by PNGase F treatment to remove N-glycans. Tagged glycoproteins were derivatised with biotin-DADPS-picolyl azide by CuAAC. To remove any excess CuAAC reagents, the proteins were precipitated with methanol overnight. The pellets were re-solubilised sequentially with Rapigest and urea, and the tagged glycoproteins enriched on neutravidin beads (**Fig. 21A**). Aliquots of the same volume were taken at different steps of the workflow to track the tagged glycoproteins by streptavidin blot. A strong band around 50-70 kDa was observed in the total protein of all secretome samples (lanes 4-6 and 10-12, **Fig. 21B**), perhaps due to leftover albumin (66.5 kDa) despite using FBS-free medium 24 h before collecting the secretome. This observation suggests that a longer incubation with FBS-free medium may be required to fully deplete albumin from the samples, however the viability of the cells decreased when the cells were grown in FBS-free medium for longer than 24 h. Some bands around 70-90 kDa were seen in the streptavidin blot of the lysate samples after the pre-clear step (lanes 1-3, **Fig. 21B**), suggesting that the removal of endogenous biotinylated proteins was incomplete for these samples. Mammalian cells have four known endogenously biotinylated proteins: the three mitochondrial proteins pyruvate carboxylase (130 kDa), methylcrotonyl-CoA carboxylase (75 kDa), and propionyl-CoA carboxylase (72 kDa); and the cytoplasmic protein acetyl-CoA carboxylase (220 kDa)²³⁵. The bands around 70-90 kDa could therefore be due to leftover methylcrotonyl-CoA carboxylase and propionyl-CoA carboxylase. Nevertheless, Chandler and Ballard have provided evidence of other unknown biotinylated mitochondrial proteins which could be an alternative source of the bands²³⁶. No bands were seen in the streptavidin blot for the secretome samples after pre-clear (lanes 4-6, **Fig. 21B**), as anticipated as none of the endogenous biotinylated proteins are expected to be secreted.

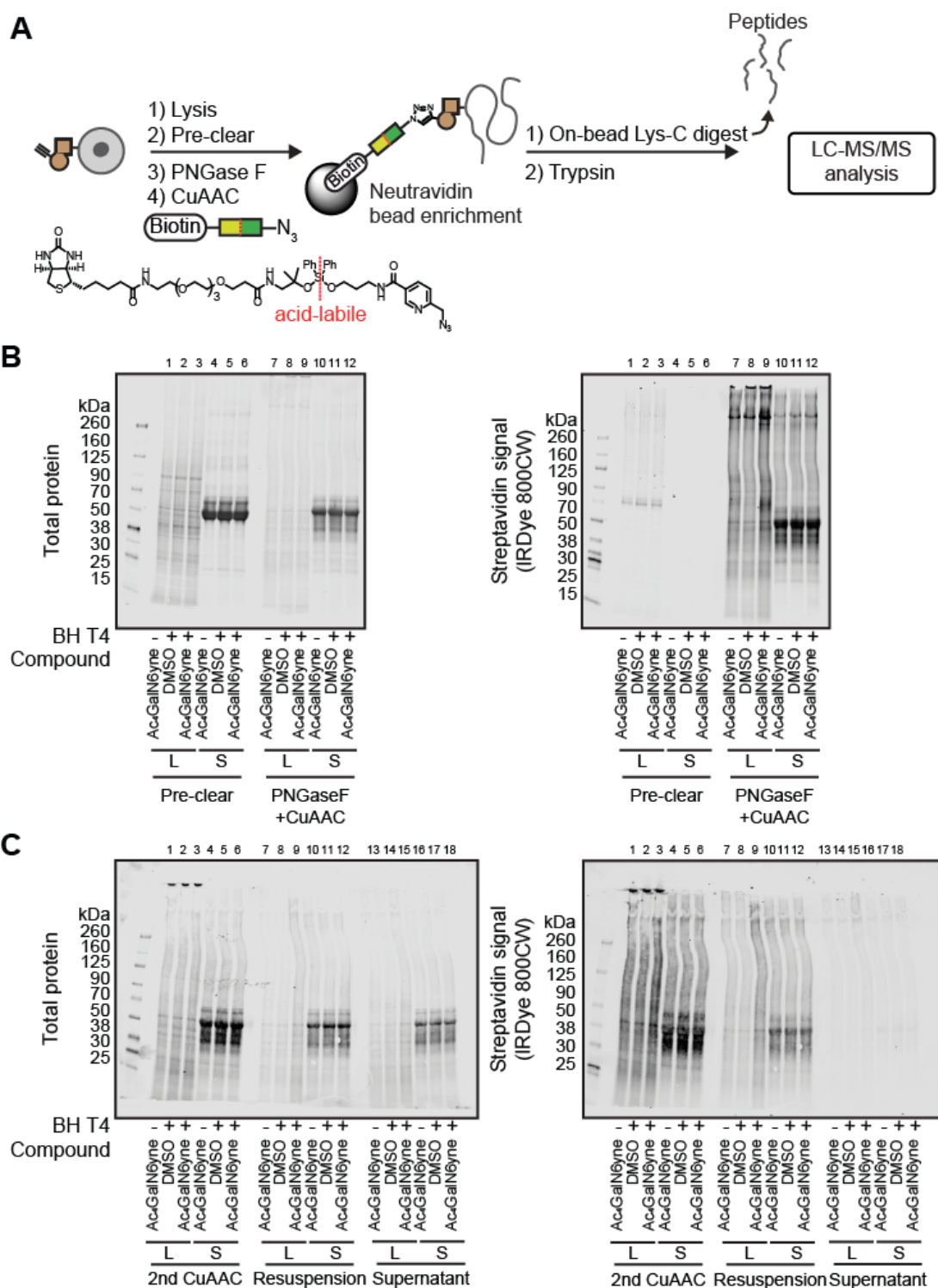


Fig. 21. Tracking tagged glycoproteins in lysates and secretome of transfected T47D^{WT} cells during neutravidin bead enrichment by streptavidin blot. **A)** Schematic representation of the workflow for the label-free MS-proteomics experiment. Figure adapted from Schumann et al. (2020)¹⁹³. **B)** Tracking tagged glycoproteins in lysates and secretome of transfected T47D^{WT} cells during neutravidin bead enrichment by streptavidin blot after removal of endogenous biotinylated proteins (pre-clear), PNGase F treatment and an initial CuAAC reaction overnight. **C)** Tracking tagged glycoproteins in lysates and secretome of transfected T47D^{WT} cells during neutravidin bead enrichment by streptavidin blot after a second CuAAC reaction overnight and sample clean-up by protein precipitation and resuspension to remove excess CuAAC reagents before neutravidin bead enrichment. The supernatant after enrichment was also included. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). Data are from one experiment. BH T4 (+) = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells, control (-) = T47D^{WT}-AGX1^{F383A}-NahK cells, L = lysate, S = secretome.

The streptavidin signal in all the samples after CuAAC was comparable (lanes 7-9 and 10-12, **Fig. 21B**). Based on the previous results from tagging glycoproteins in lysates and the secretome of T47D cells (**Fig. 19**) we would have expected to see minimal signal in the DMSO-treated samples and the strongest signal in the T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} samples treated with Ac₄GalN6yne. We therefore added more CuAAC solution to the samples and continued the incubation overnight to maximise the derivatisation of the tagged glycoproteins. Despite this, background signal was still observed in the DMSO-treated samples and the T47D^{WT}-AGX1^{F383A}-NahK control samples (lanes 1-2 and 4-5, **Fig. 21C**). We hypothesised that this could be due to leftover biotin-DADPS-picolyl azide from the CuAAC solution in the samples. Indeed, when we precipitated the proteins to remove excess CuAAC reagents and analysed the streptavidin signal after resuspension, a strong signal was observed in the lysate of Ac₄GalN6yne-treated T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells (lane 9, **Fig. 21C**) whilst minimal signal was seen in the control lysates (lanes 7-8, **Fig. 21C**). However, the total protein of the former was also higher than the total protein of the control lysates. Indeed, the degree of protein precipitation and resuspension varied significantly between samples (even visible to the naked eye) which could affect the results obtained. The streptavidin signal of all the secretome samples was comparable even after protein precipitation and resuspension (lanes 10-12, **Fig. 21C**), suggesting that the tagging of glycoproteins with BH GalNAc-T4 may not have been as successful in these samples as in **Fig. 19E**. No streptavidin signal was seen in any of the supernatants after neutravidin bead enrichment (lanes 13-18, **Fig. 21C**), suggesting that the tagged glycoproteins remained bound to the beads.

After neutravidin bead enrichment of the tagged glycoproteins, the unmodified peptides can be released from the beads by on-bead digestion with Lys-C, whilst tagged glycopeptides can be released from the beads by acid treatment due to the acid-labile diphenyldisiloxane moiety present in biotin-DADPS-picolyl azide. In this experiment we only focused on the peptide fraction as a preliminary test. After the on-bead digestion with Lys-C, the peptides were digested again with trypsin, desalted using UltraMicroSpinTM columns, and vacuum-dried by Speedvac. The samples were reconstituted in 0.1% (v/v) formic acid in LC-MS grade water and injected into an Orbitrap EclipseTM TribridTM Mass Spectrometer coupled to an UltiMateTM 3000 RSLCnano System by Lucia Di Vagno. The raw MS data was processed on MaxQuant and the output visualised in Perseus.

A two-sided two-sample Welch's t-test was performed, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off. Statistically-enriched peptides were detected in all the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls (**Fig. 22**) - however we couldn't confirm that these proteins were enriched due to them being substrates of GalNAc-T4 or due to differences in the levels of the proteins between samples caused by the variability in the protein precipitation and resuspension steps. Furthermore, we would have expected to detect statistically-enriched peptides due to MUC1 as the *bona fide* substrate of GalNAc-T4^{72, 156, 157, 234}, however this was not the case in any of the samples. We therefore decided to optimise the workflow to obtain more confident and reproducible results.

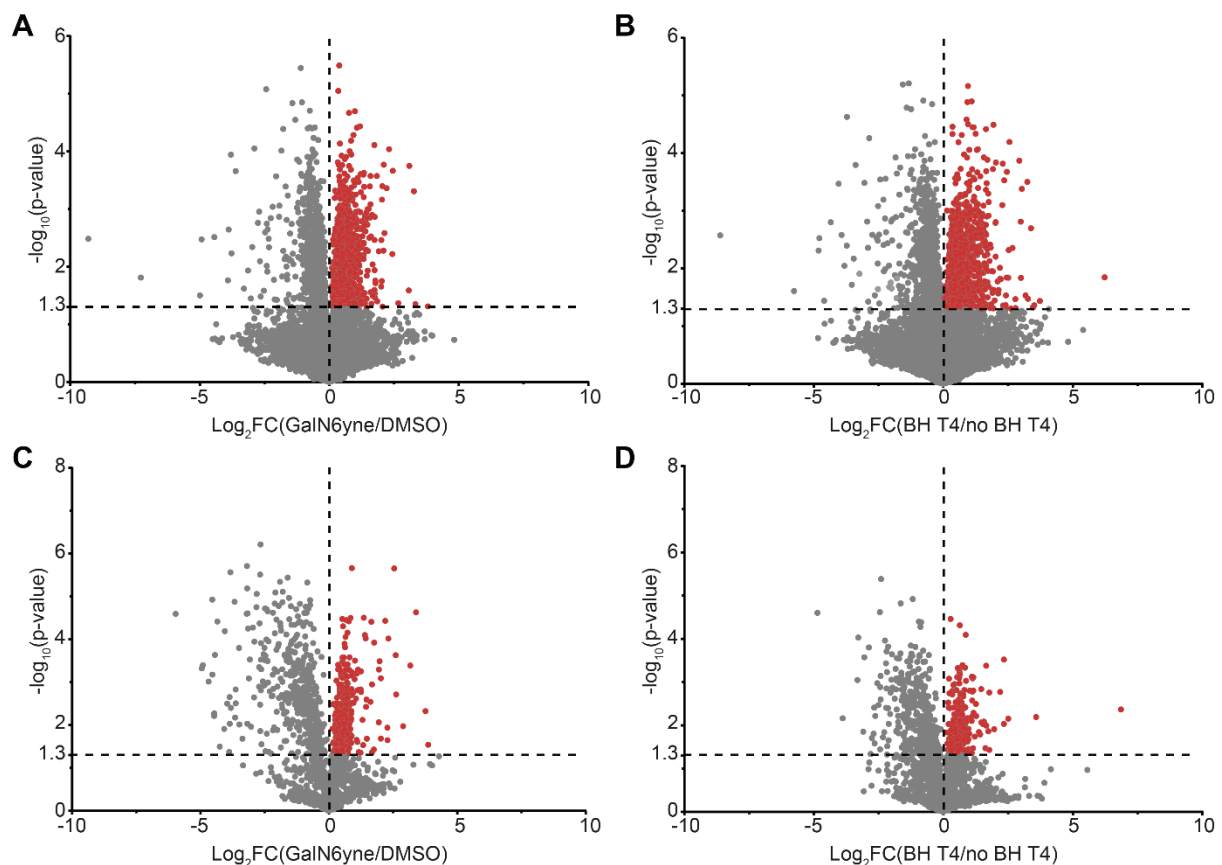


Fig. 22. Statistical analysis of the label-free MS-proteomics experiments on lysates and secretome of transfected T47D^{WT} cells. MS analysis of the peptide fractions from **A**) lysates or **C**) the secretome of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells treated with Ac₄GalN6yne or DMSO as a control. MS analysis of the peptide fractions from **B**) lysates or **D**) the secretome of GalN6yne-treated T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells or T47D^{WT}-AGX1^{F383A}-NahK cells as a control. Differentially enriched peptides were determined by performing a two-sided two-sample Welch's T-test using Perseus software. Data are visualised as a scatter plot, with each dot representing a different peptide detected, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off. Data are representative of three technical replicates. BH T4 = BH GalNAc-T4.

3.2. Optimisation of MS-proteomics workflow

Since the protein precipitation and resuspension steps were the main source of inter-sample variability in the MS-proteomics workflow, the Malaker group (Yale University) recommended testing Amicon filters for sample clean-up and buffer exchange after the CuAAC reaction instead. T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} or T47D^{WT}-AGX1^{F383A}-NahK cells were fed with Ac₄GalN6yne and lysed. Tagged glycoproteins were derivatised with biotin-picolyl azide by CuAAC, followed by sample clean-up with Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO). Protein recovery and tagging of glycoproteins was assessed by streptavidin blot (**Fig. 23A**).

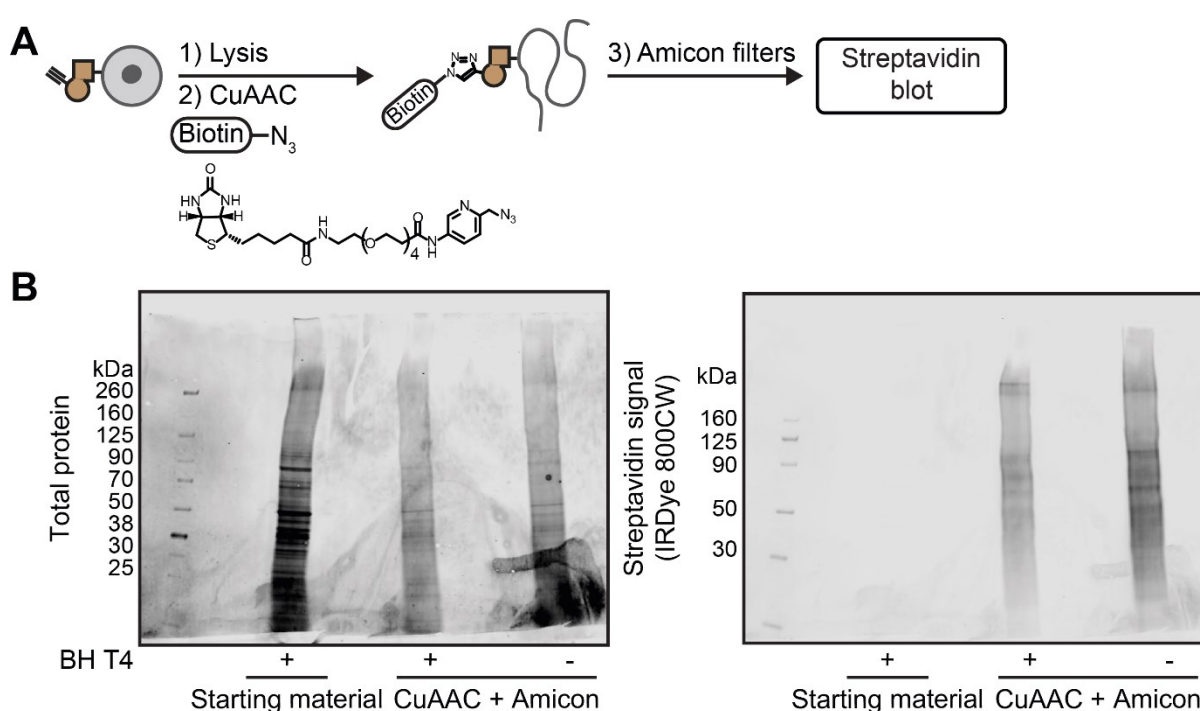


Fig. 23. Testing Amicon filters as an alternative approach to protein precipitation and resuspension for sample clean-up and buffer exchange after the CuAAC reaction. **A)** Schematic representation of the experiment. **B)** Protein recovery and tagging of glycoproteins was assessed by streptavidin blot. 50 µg of protein were used per lane. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). Data are representative of two independent replicates. BH T4 (+) = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells, control (-) = T47D^{WT}-AGX1^{F383A}-NahK cells.

The total protein signal was reduced after the CuAAC reaction and sample clean-up with the Amicon filters (**Fig. 23B**), perhaps due to the proteins sticking to the filters. Nevertheless, the protein recovery was comparable between the two samples and the signal observed in the streptavidin blot looked analogous to the results previously obtained in the cell surface tagging experiments (**Fig. 18-20**). Based on these results, we decided to use Amicon filters instead of protein precipitation and resuspension in future MS-proteomics experiments.

We also decided to shift from label-free MS-proteomics to Stable Isotope Labelling by Amino acids in Cell culture (SILAC)-based MS-proteomics. In this approach, cells are grown in medium containing isotopically heavy or light Lys and Arg amino acids²³⁷. Since trypsin cleaves at the C-terminus of these amino acids, all tryptic peptides will contain at least one labelled amino acid²³⁸. Heavy and light samples are combined at the beginning of the workflow, reducing variability due to sample preparation²³⁷. The source of the sample can be identified from the mass difference caused by the presence of the heavy or light isotopes. The relative abundance of the proteins can be measured by calculating the difference in the intensities between the heavy and light peptides, allowing relative protein quantitation²³⁷.

3.3. SILAC-based MS-(glyco)proteomics on lysates of transfected T47D^{WT} cells

I first attempted the SILAC-based MS-proteomics strategy using lysates of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. For this experiment, I used the T47D^{WT}-AGX1^{F383A}-NahK cells as a control for background tagging of glycoproteins. The cells were first cultured in growth medium containing isotopically heavy or light amino acids for 5-6 doublings to maximise the heavy SILAC label incorporation²³⁷. The growth medium also contained light Pro to minimise the metabolic conversion of Arg to Pro²³⁷. The cells were then treated with Ac₄GalN6yne or DMSO and lysed after overnight incubation. An aliquot from the heavy lysates was processed using the SP3 technology²²⁰ and injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The raw MS data was processed on MaxQuant and the incorporation of the heavy SILAC label was determined using Perseus. The incorporation efficiency of all heavy samples was >95% (**Table 2**), as required for SILAC experiments²³⁷.

Table 2. Determining the incorporation efficiency of the heavy SILAC label on lysates of transfected T47D^{WT} cells.

Cell line	Treatment	Incorporation efficiency (%)			
		R1	R2	R3	Average
T47D ^{WT} -AGX1 ^{F383A} -NahK	Ac ₄ GalN6yne	97	96	96	96
T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	DMSO	96	97	98	97
T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne	97	96	96	96

1:1 mixtures of the heavy and light lysates were prepared according to **Fig. 24A**. The samples were first incubated with neutravidin beads to remove endogenous biotinylated proteins (pre-clear), followed by PNGase F treatment to remove N-glycans. The tagged glycoproteins were then derivatised with biotin-DADPS-picolyl azide by CuAAC for 72 h to maximise the derivatisation of the tagged glycoproteins. Sample clean-up and buffer exchange to PBS was subsequently performed using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO), followed by enrichment of the tagged glycoproteins with neutravidin beads (**Fig. 24B**). Aliquots of the same volume were taken at different steps of the workflow to track the tagged glycoproteins by streptavidin blot (**Fig. 24C**). As seen in the label-free MS-proteomics experiment, some bands around 70-90 kDa were seen in the streptavidin blot of the lysate samples after the pre-clear step (**Fig. 24C**), suggesting that the removal of endogenous biotinylated proteins was incomplete. A boost in streptavidin signal was seen in all samples after the CuAAC reaction (**Fig. 24C**), as expected since all samples contained a BH GalNAc-T4/Ac₄GalN₆yne component, demonstrating that the reaction was successful. Minimal streptavidin signal was seen in the supernatants after neutravidin bead enrichment (**Fig. 24C**), suggesting that the tagged glycoproteins remained bound to the beads.

After neutravidin bead enrichment of the tagged glycoproteins, the unmodified peptides were released from the beads by on-bead digestion with Lys-C, followed by a second digestion with trypsin. The samples were fractionated using UltraMicroSpin™ columns, vacuum-dried by Speedvac, reconstituted in 1% (v/v) formic acid/ 1% (v/v) LC-MS grade acetonitrile in LC-MS grade water, and injected into an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno (**Fig. 24B**). The tagged glycopeptides were eluted from the beads by acid treatment, digested with trypsin, desalted using UltraMicroSpin™ columns, vacuum-dried by Speedvac, reconstituted in 0.1% (v/v) formic acid in LC-MS grade water, and injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno (**Fig. 24B**).

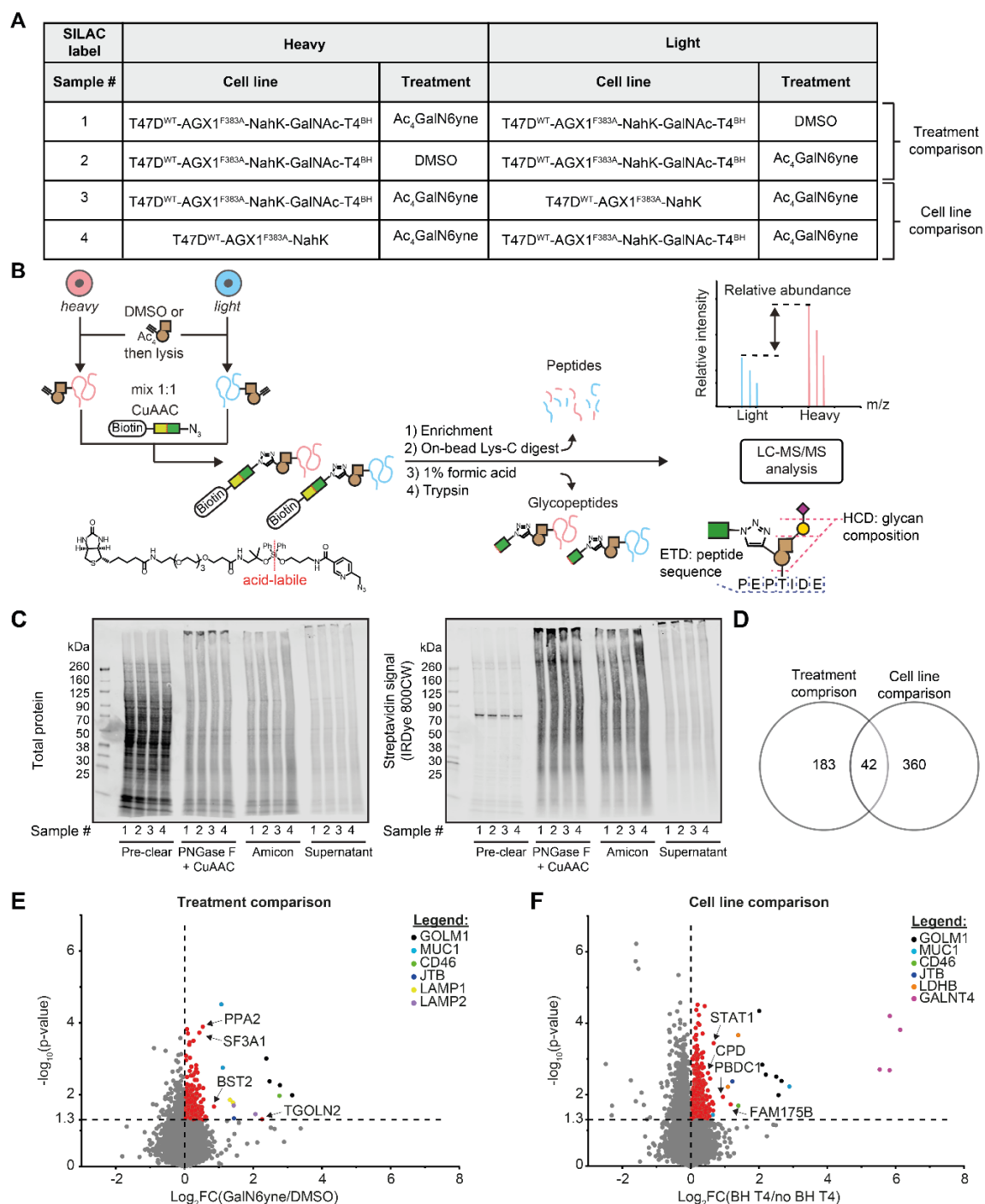


Fig. 24. Characterising the substrates of GalNAc-T4 in lysates of transfected T47D^{WT} cells by SILAC-based MS-proteomics. **A)** 1:1 mixtures of heavy and light T47D lysates. **B)** Schematic representation of the workflow for the SILAC-based MS-proteomics experiments using lysates of transfected T47D^{WT} cells. Figure adapted from Cioce et al. (2022)²⁰⁴. **C)** Tracking tagged glycoproteins during neutravidin bead enrichment by streptavidin blot. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). **D)** Venn Diagram showing the number of proteins enriched in each comparison from transfected T47D^{WT} lysates. Peptides that were assigned more than one protein group were excluded from this calculation. **E)** MS analysis of the peptide fractions from lysates of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells treated with Ac₄GalN6yne or DMSO as a control. **F)** MS analysis of the peptide fractions from lysates of Ac₄GalN6yne-treated T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} or T47D^{WT}-AGX1^{F383A}-NahK cells as a control. Differentially enriched peptides were determined by performing a two-sided one-sample T-test using Perseus software. Data are visualised as a scatter plot, with each dot representing a different peptide detected, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off. Only peptides with non-zero values were plotted. Data are from three independent experiments, each independent replicate consisting of three technical replicates. BH T4 = BH GalNAc-T4.

The raw data from the peptide samples was processed on MaxQuant and the output visualised in Perseus. The data were grouped in MaxQuant to obtain two distinct comparisons. In the first comparison, the data from samples **1** and **2** were grouped to assess the effect of using different treatments on the same cell line (“treatment comparison”: GalN6yne vs DMSO)(**Fig. 24A**). In the second comparison, the data from samples **3** and **4** were grouped to assess the effect of using the same treatment on two different cell lines (“cell line comparison”: T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} vs T47D^{WT}-AGX1^{F383A}-NahK cells)(**Fig. 24A**). A two-sided one sample t-test was performed on the two comparisons separately, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off. Statistically-enriched peptides were detected in all the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls (**Fig. 24E and 24F**). Promisingly, one of the proteins with statistically-enriched peptides was MUC1 (**Fig. 24E and 24F**), which is the *bona fide* substrate of GalNAc-T4^{72, 156, 157, 234}, hence validating the results obtained.

The Venn diagram in **Fig. 24D** shows the number of proteins statistically-enriched in each comparison. In the treatment comparison, the statistically-enriched proteins will include proteins tagged by BH GalNAc-T4 but also enriched due to background tagging of glycoproteins (e.g. from GlcNAc-containing glycoproteins after epimerisation of UDP-GalN6yne by GALE). In the cell line comparison, the statistically-enriched proteins will include proteins tagged by BH GalNAc-T4 but also enriched due to differences in the protein expression levels between the two cell lines. For example, since BH GalNAc-T4 is OE compared to the endogenous enzyme (**SI Fig. 8A**), peptides due to GalNAc-T4 were enriched in the T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} samples compared to the T47D^{WT}-AGX1^{F383A}-NahK samples. The commonly enriched hits from both comparisons should therefore represent the substrates of GalNAc-T4.

Peptides from 42 proteins were commonly enriched in both comparisons, with 98% (41/42) of these proteins previously annotated as O-glycosylated in the Glygen database²³⁹. From these proteins, MUC1, GOLM1, CD46 and JTB were enriched over 2-fold in all the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls. These results demonstrated that we can successfully characterise the substrates of GalNAc-T4 in lysates of T47D cells by SILAC-based MS-proteomics. The identified substrates are listed in **SI Table 2**.

The raw data from the glycopeptide samples was processed on Byonic. The data was first searched to generate a FASTA file containing only the proteins found in the sample. This reduced the search time of the subsequent glycan search, where the data was searched again with the following glycan modifications in the “Glycans” tab: HexNAc(1), HexNAc(1)-NeuAc(1), HexNAc(1)-Hex(1), HexNAc(1)-Hex(1)-NeuAc(1), HexNAc(1)-Hex(1)-NeuAc(2), HexNAc(2)-Hex(1)-NeuAc(1), HexNAc(2)-Hex(2) and HexNAc(2)-Hex(2)-NeuAc(1) and an additional 287.1389 m/z to account for the chemical modification^{182, 193, 203, 204}. The output excel file was filtered for glycans and specific cleavage sites. If any hits satisfied these criteria, the MS spectra was manually examined by Lucia Di Vagno using Xcalibur software to validate the glycosite hits. However, none of the expected HCD signature ions due to the presence of the chemically modified sugar were detected, suggesting that the peptides were incorrectly annotated by the software²⁰³. There are multiple challenges in determining glycosylation site information as described in the “**Substrate identification of individual GalNAc-Ts**” section. We therefore decided to focus only on the peptide fraction for future experiments – this would allow us to identify the substrates of GalNAc-T4 in our samples, whilst further optimisation on the glycopeptide sample processing would be required for glycosylation site localisation.

3.4. SILAC-based MS-proteomics on the secretome of transfected T47D^{WT} cells

I next repeated the SILAC-based MS-proteomics experiment using the secretome of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} or T47D^{WT}-AGX1^{F383A}-NahK cells and only focusing on the peptide fraction. The workflow was the same as the one used for the T47D lysates, the only differences were that on the day of sugar treatment the medium was replaced to FBS-free medium to minimise interference due to serum proteins and that after overnight incubation with the sugar the secretome was collected, concentrated and the medium replaced with PBS (**Fig. 25B**).

An aliquot from the heavy secretome samples was processed using the SP3 technology²²⁰ and injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The raw MS data was processed on MaxQuant and the incorporation of the heavy SILAC label determined using Perseus. The incorporation efficiency of the heavy SILAC label in the heavy secretome samples (**Table 3**) was lower than in the corresponding lysate samples (**Table 2**), which could affect the results obtained. To overcome this, I should have allowed the cells to grow for longer.

Table 3. Determining the incorporation efficiency of the heavy SILAC label on the secretome of transfected T47D^{WT} cells.

Cell line	Treatment	Incorporation efficiency (%)		
		R1	R2	Average
T47D ^{WT} -AGX1 ^{F383A} -NahK	Ac ₄ GalN6yne	91	81	86
T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	DMSO	92	80	86
T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne	91	82	87

When tracking the tagged glycoproteins by streptavidin blot, no bands were observed in the streptavidin blot after the pre-clear step (**Fig. 25F**), demonstrating the absence of endogenous biotinylated proteins in the samples. A strong band around 50-70 kDa was observed in the total protein stain of all samples (**Fig. 25F**), likely due to leftover albumin (66.5 kDa) despite using FBS-free medium 24 h before collecting the secretome. A boost in streptavidin signal was seen in all samples after the CuAAC reaction (**Fig. 25F**), as expected since all samples contained a BH GalNAc-T4/Ac₄GalN6yne component, demonstrating that the reaction was successful. Minimal streptavidin signal was seen in the supernatants after neutravidin bead enrichment (**Fig. 25F**), demonstrating that the tagged glycoproteins remained bound to the beads.

Statistically-enriched peptides were detected in all the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls (**Fig. 25D and 25E**), with 3 proteins (AGRN, C3 and EFEMP1) commonly enriched in both comparisons (**Fig. 25C**). These 3 proteins have been previously annotated as O-glycosylated in the GlyGen database²³⁹ (**SI Table 3**). These results demonstrated that we can also characterise the substrates of GalNAc-T4 in the secretome of T47D cells.

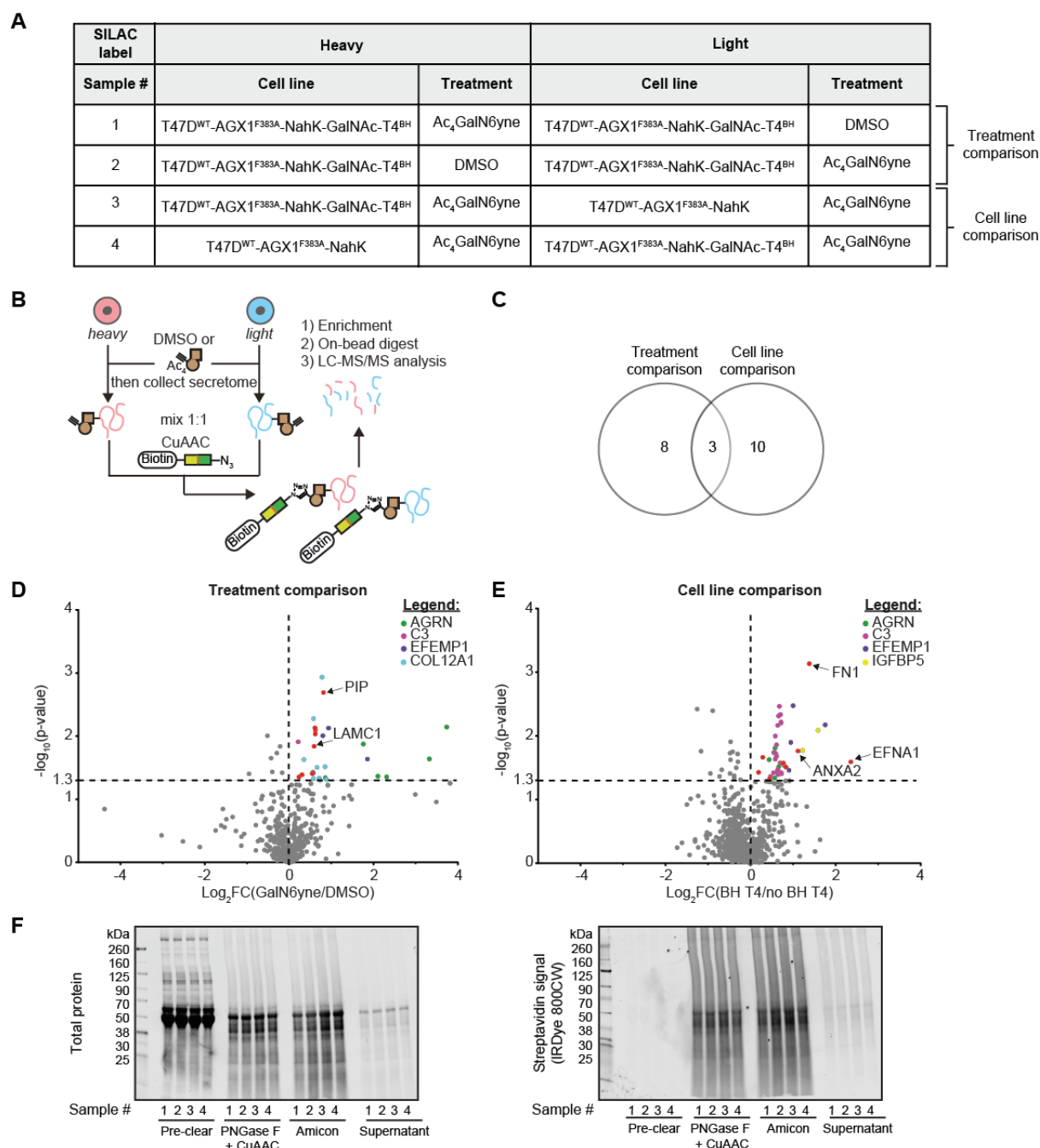


Fig. 25. Characterising the substrates of GalNAc-T4 in the secretome of transfected T47D^{WT} cells by SILAC-based MS-proteomics. **A)** 1:1 mixtures of heavy and light T47D secretome samples. **B)** Schematic representation of the workflow for the SILAC-based MS-proteomics experiments using the secretome of transfected T47D^{WT} cells. Figure adapted from Cioce et al. (2022)²⁰⁴. **C)** Venn Diagram showing the number of proteins enriched in each comparison from the secretome of transfected T47D^{WT} cells. Peptides that were assigned more than one protein group were excluded from this calculation **D)** MS analysis of the peptide fractions from the secretome of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells treated with Ac₄GalN6yne or DMSO as a control. **E)** MS analysis of the peptide fractions from the secretome of Ac₄GalN6yne-treated T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} or T47D^{WT}-AGX1^{F383A}-NahK cells as a control. Differentially enriched peptides were determined by performing a two-sided one-sample T-test using Perseus software. Data are visualised as a scatter plot, with each dot representing a different peptide detected, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off. Only peptides with non-zero values were plotted. Data are from two independent experiments. **F)** Tracking tagged glycoproteins during the SILAC-based MS-proteomics experiment on the secretome of transfected T47D^{WT} cells by streptavidin blot. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). BH T4 = BH GalNAc-T4.

3.5. SILAC-based MS-proteomics on lysates of transfected MCF7^{WT} and MDA-MB-231^{WT} cells

Having characterised the substrates of GalNAc-T4 in T47D cells, we next repeated the experiments on the MCF7^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} and MDA-MB-231^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. For this experiment, I used the MCF7^{WT}-AGX1^{F383A}-NahK and MDA-MB-231^{WT}-AGX1^{F383A}-NahK cells as a control for background tagging of glycoproteins. Since the secretome analysis proved to be more technically challenging as a greater number of cells was required to maximise the starting protein content, and the cells had to be cultured for longer to maximise the incorporation of the heavy SILAC label, we decided to focus on the lysate analysis instead.

The workflow of the experiment was the same as the one used for the T47D lysates but only analysing the peptide samples. The incorporation efficiency of the heavy SILAC label was >95% in all heavy samples for both cell lines (**SI Table 4** and **SI Table 5**), satisfying the SILAC requirements²³⁷. The results when tracking the tagged glycoproteins by streptavidin blot were comparable to those obtained with the T47D lysates (**Fig. 24C**) – some bands were observed after the pre-clear step (**SI Fig. 9B-C**), suggesting incomplete removal of endogenous biotinylated proteins; a strong signal was seen in all samples after CuAAC reaction (**SI Fig. 9B-C**), confirming the success of the reaction; and minimal signal was observed in the supernatant after enrichment (**SI Fig. 9B-C**), suggesting that the tagged glycoproteins remained bound to the beads.

After neutravidin bead enrichment of the tagged glycoproteins, the unmodified peptides were released from the beads by on-bead digestion with Lys-C, followed by a second digestion with trypsin. The samples were fractionated using UltraMicroSpinTM columns, vacuum-dried by Speedvac, reconstituted in 1% (v/v) formic acid/ 1% (v/v) LC-MS grade acetonitrile in LC-MS grade water, and injected into an Orbitrap FusionTM LumosTM TribridTM Mass Spectrometer coupled to an UltiMateTM 3000 RSLCnano System by Lucia Di Vagno. The raw data from the peptide fraction was processed on MaxQuant, and statistical analysis was performed in Perseus after grouping the data to obtain the same two distinct comparisons as with T47D cells - the “treatment comparison” and the “cell line comparison” (**Fig. 26A**). Statistically-enriched peptides were detected in all the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls (**Fig. 26B-E**), including peptides due to MUC1 in the Ac₄GalN6yne-treated BH GalNAc-T4 MCF7 samples (**Fig. 26B-C**). Commonly enriched hits from both comparisons were obtained for both cell lines (**Fig. 26F-G**).

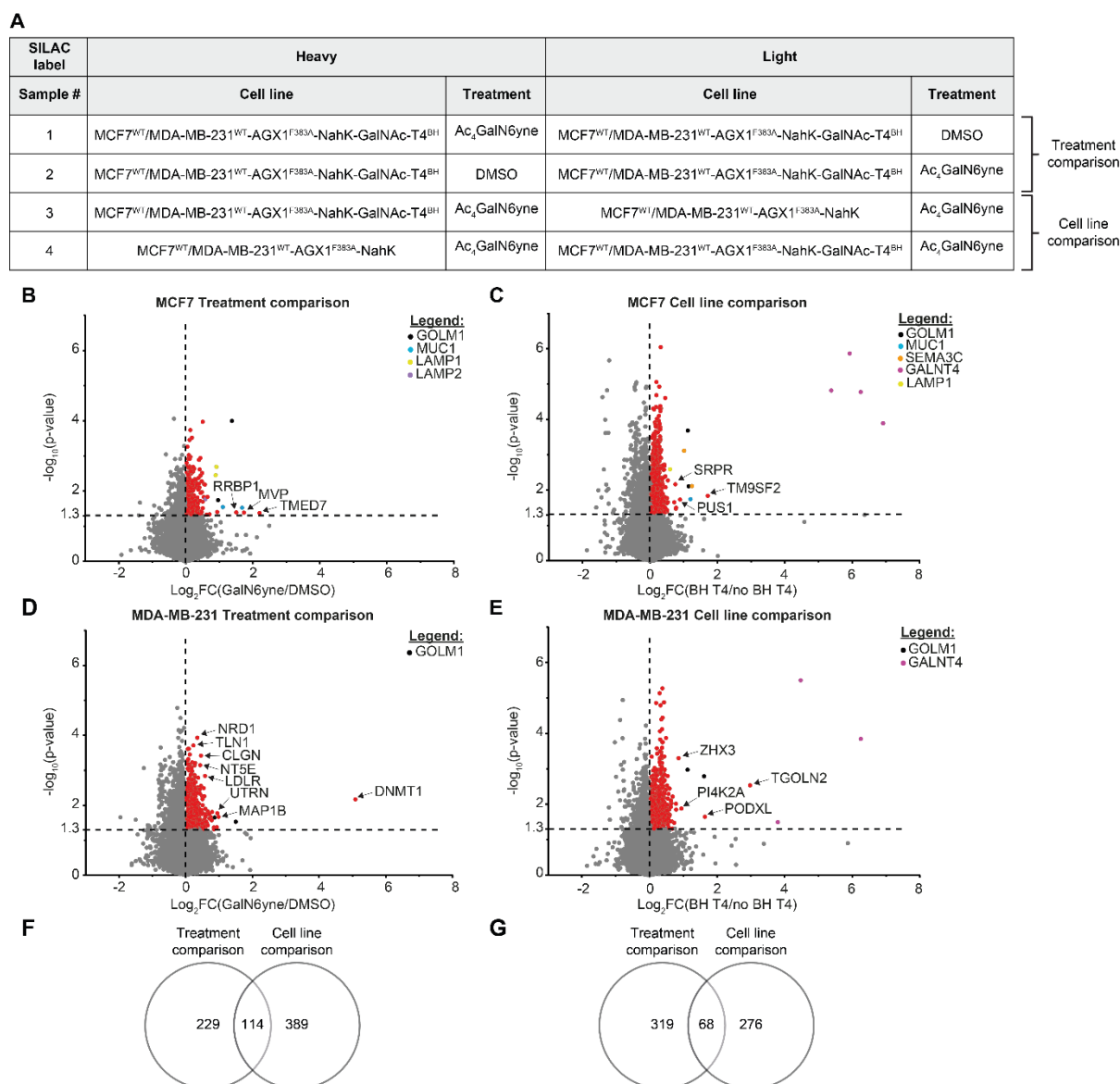


Fig. 26. Characterising the substrates of GalNAc-T4 in lysates of transfected MCF7^{WT} and MDA-MB-231^{WT} cells by SILAC-based MS-proteomics. **A)** 1:1 mixtures of heavy and light MCF7 and MDA-MB-231 lysates. MS analysis of the peptide fractions from lysates of **B)** MCF7^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} and **D)** MDA-MB-231^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells treated with Ac₄GalN6yne or DMSO as a control. MS analysis of the peptide fractions from lysates of Ac₄GalN6yne-treated AGX1^{F383A}-NahK-GalNAc-T4^{BH} or control AGX1^{F383A}-NahK **C)** MCF7 and **E)** MDA-MB-231 cells. Differentially enriched peptides were determined by performing a two-sided one-sample T-test using Perseus software. Data are visualised as a scatter plot, with each dot representing a different peptide detected, choosing a Benjamini-Hochberg FDR-adjusted p-value of 0.05 as a cut-off. Only peptides with non-zero values were plotted. Data are from three independent experiments, each independent replicate consisting of three technical replicates. Venn Diagram showing the number of proteins enriched in each comparison from **F)** MCF7 and **G)** MDA-MB-231 lysates. Peptides that were assigned more than one protein group were excluded from this calculation. BH T4 = BH GalNAc-T4.

Peptides from 114 proteins were commonly enriched in the Ac₄GalN6yne-treated BH GalNAc-T4 MCF7 samples, with 98% (112/114) of these proteins previously annotated as O-glycosylated in the Glygen database²³⁹. From these proteins, MUC1 and GOLM1 were enriched over 2-fold in the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls. Peptides from 68 proteins were commonly enriched in the Ac₄GalN6yne-treated BH GalNAc-T4 MDA-MB-231 samples, with 96% (65/68) of these proteins previously annotated as O-glycosylated glycosylated in the Glygen database²³⁹. From these proteins, only GOLM1 was enriched over 2-fold in the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls. These results demonstrated that we can also successfully characterise the substrates of GalNAc-T4 in lysates of MCF7 and MDA-MB-231 cells by SILAC-based MS-proteomics. The identified substrates in MCF7 and MDA-MB-231 cells are listed in **SI Table 6** and **SI Table 7** respectively.

Comparing the data from the SILAC-based MS-proteomics experiments with the 3 BCCs, most hits were cell-line specific, nevertheless some substrates were conserved between the different BCCs as shown in **Table 4**. MUC1, the *bona fide* substrate of GalNAc-T4^{72, 156, 157, 234}, was conserved in both luminal T47D and MCF7 cells but not the basal-like MDA-MB-231 cells. This was anticipated since MUC1 is physiologically expressed in the apical surface of luminal epithelial cells of the mammary gland²⁴⁰. Upon malignant transformation, MUC1 loses its apical polarity and instead becomes aberrantly expressed over the whole cell surface and within the cytoplasm of BCCs²⁴¹. Indeed, MUC1 overexpression has been associated with lower grade, ER+ tumours and the aberrant localisation of MUC1 (rather than its level of expression) has been associated with worse prognosis^{241, 242}. Only GOLM1 was conserved in all 3 BCCs. GOLM1 is also upregulated in BC tissue compared to healthy mammary tissue²⁴³.

Table 4. GalNAc-T4 substrates conserved between BCCs based on the SILAC-based MS-proteomics experiments. GOLM1 is highlighted in red, MUC1 is highlighted in blue.

T47D & MCF7	T47D & MDA-MB-231	MCF7 & MDA-MB-231
GOLM1	CPD	FLNB
LAMP2	GOLM1	GOLM1
LRPPRC	TLN1	HDLBP
MUC1		NUCKS1
PLEC		NUMA1
POLR2A		UGDH
RANBP2		
SLC5A6		
SNRNP200		
VPS35		

3.6. Validation of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown

When performing statistical analysis on MS-proteomics data, a separate statistical test is performed on each protein²⁴⁴. For each independent test, there is a probability of a false-positive result (5% when the level of significance (α) is set to 0.05)²⁴⁴. The probability of false-positives increases with the number of independent tests performed²⁴⁴. To control the number of false-positives, multiple testing correction methods must therefore be performed²⁴⁴. In the SILAC-based MS-proteomics experiments described above, the Benjamini-Hochberg FDR correction was performed, where the FDR controls the proportion of significant tests that are false-positives (i.e. an FDR adjusted p-value of 0.05 implies that 5% of significant tests will result in false-positives)²⁴⁴. In the Benjamini-Hochberg correction, the p-values of all tests are first ranked in increasing order²⁴⁴. An adjusted p-value (q-value) is then calculated for each item on the list using **Equation 9** depending on the number of tests (n) and the position in the rank (k)²⁴⁴. If the q-value is smaller than a pre-defined α (0.05 in this case), the protein will be considered statistically significant²⁴⁴.

Equation 9. Multiple testing correction using the Benjamini-Hochberg FDR method. Adjusted q-values (q) are calculated from the corresponding p-values (p), the number of tests (n) and the position of each item in the rank (k)²⁴⁴.

$$q = p \left(\frac{n}{k} \right)$$

Despite multiple testing corrections, a proportion of the tests will still be false-positives. We therefore decided to validate some of the MS-proteomics results by performing neutravidin bead pulldown experiments. For this experiment I tested GOLM1 as the only protein that was conserved in all three cell lines (**Fig. 27A, Table 4**); MUC1 as the *bona fide* substrate of GalNAc-T4^{72, 156, 157, 234} which was a common hit in both T47D and MCF7 cells but not in the MDA-MB-231 cells (**Fig. 27A, Table 4**); and CD46 as a putative T47D-specific hit, as this protein was enriched over 2-fold in the BH GalNAc-T4 T47D samples compared to the controls but it wasn't a hit in the other cell lines (**Fig. 27A**).

T47D^{WT}/MCF7^{WT}/MDA-MB-231^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} and T47D^{WT}/MCF7^{WT}/MDA-MB-231^{WT}-AGX1^{F383A}-NahK cells were fed with Ac₄GalN6yne (**Fig. 27C**). After overnight incubation, the cells were lysed (**Fig. 27C**). The samples were first incubated with neutravidin beads to remove endogenous biotinylated proteins (pre-clear), followed by PNGase F treatment to remove N-glycans (**Fig. 27C**). The tagged glycoproteins were then derivatised with biotin-picolyl azide by CuAAC for 72 h (**Fig. 27C**). Sample clean-up and buffer exchange to PBS was subsequently performed using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO) (**Fig. 27C**). An aliquot of each sample was taken at this point (**Input**). Tagged glycoproteins were enriched with neutravidin beads (**Fig. 27C**). The glycoproteins were eluted by boiling the beads (**Eluate**).

Aliquots of the same volume were taken at different steps of the workflow to track the tagged glycoproteins by streptavidin blot. No bands were observed in the streptavidin blot after the pre-clear step, indicating the successful removal of endogenous biotinylated proteins (lanes 1-6, **Fig. 27D**). A boost in streptavidin signal was seen in all samples after the CuAAC reaction, demonstrating that the reaction was successful (lanes 7-12, **Fig. 27D**). Interestingly, the intensity of the signal after the CuAAC reaction was different between the different cell lines - the highest in the T47D cells (lanes 7-8, **Fig. 27D**) and the lowest in the MCF7 cells (lanes 9-10, **Fig. 27D**). This difference could reflect different protein expression levels between cell lines, or different efficiencies of the CuAAC reaction between samples. Minimal streptavidin signal was seen in the supernatants after neutravidin bead enrichment (lanes 13-18, **Fig. 27D**), demonstrating that the tagged glycoproteins remained bound to the beads.

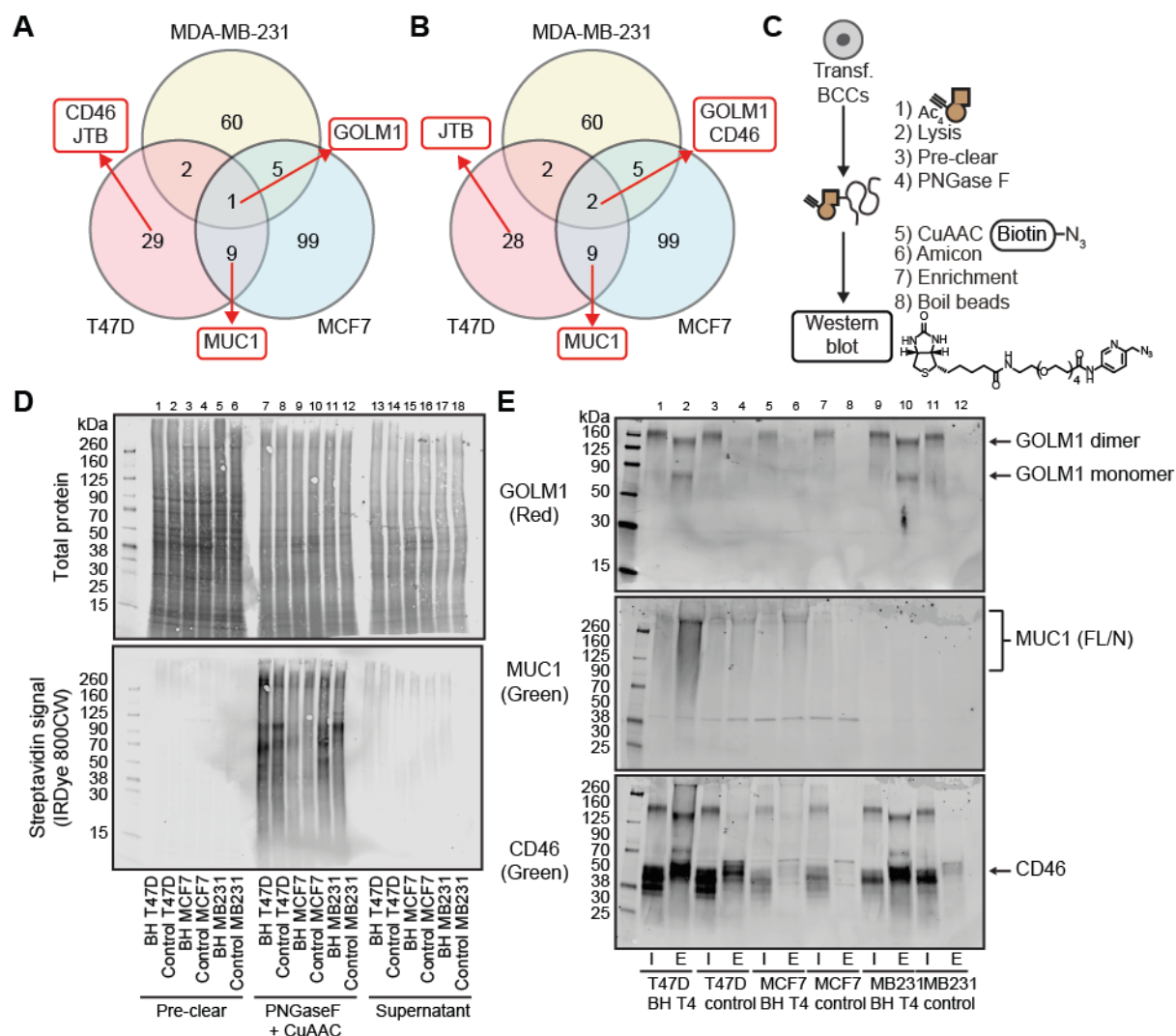


Fig. 27. Validating the results of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown. **A)** Venn diagram summarising the results of the SILAC-based MS-proteomics experiments. The numbers represent the number of proteins identified in each section. Peptides that were assigned more than one protein group were excluded from this calculation. **B)** Updated Venn diagram following the validation of the SILAC-based MS-proteomics experiments by neutravidin bead pulldown. The numbers represent the number of proteins identified in each section. Peptides that were assigned more than one protein group were excluded from this calculation. **C)** Schematic representation of the workflow for the neutravidin bead pulldown experiments to validate the SILAC-based MS-proteomics results. **D)** Tracking tagged glycoproteins during the neutravidin bead pulldown experiments by streptavidin blot. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). **E)** Protein expression control by Western blot from the neutravidin bead pulldown experiments. The antibody stainings were performed on the same membrane in the order displayed. Red/Green refers to the imaging channel. Antibody staining was performed using antibodies against GOLM1 (1:500, HPA010638), MUC1 (1:500, MA106503) and CD46 (1:500, 12494-1-AP). Data are representative of two independent replicates. BH T4 = AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells; control = AGX1^{F383A}-NahK cells; MB231 = MDA-MB-231; I = input; E = eluate.

The **Input** and **Eluate** samples were analysed by Western blot (**Fig. 27E**). The intensity of the bands in the **Input** samples was always comparable between BH GalNAc-T4 (AGX1^{F383A}-NahK-GalNAc-T4^{BH}) and control (AGX1^{F383A}-NahK) samples (lanes 1 & 3/5 & 7/9 & 11, **Fig. 27E**). Some signal was always observed in the control **Eluate** samples (lanes 4/8/12, **Fig. 27E**), likely due to background tagging of glycoproteins.

When staining against GOLM1, intense bands were observed between 50-90 kDa and between 125-160 kDa due to the presence of the GOLM1 monomer and dimer in the samples (**Fig. 27E**)²⁴⁵. A 1.1-fold, 1.2-fold, and 1.3-fold increase in mean fluorescent intensity was observed in the **Eluate** BH GalNAc-T4 T47D, MCF7 and MDA-MB-231 samples over the controls, respectively (lanes 2 vs 4/6 vs 8/10 vs 12, **Fig. 27E**), validating GOLM1 as a GalNAc-T4 substrate in the three cell lines. The intensity of the bands due to GOLM1 was stronger in the T47D and MDA-MB-231 samples compared to the MCF7 samples, most likely due to differences in expression levels between the cells.

When staining against MUC1, bands were observed in the T47D and MCF7 samples at a MW >260 kDa due to full-length MUC1 (lanes 1-8, **Fig. 27E**)²⁴⁶. No MUC1 bands were observed in the MDA-MB-231 samples (lanes 9-12, **Fig. 27E**), validating the MS-proteomics results with this cell line. A 2.5-fold and 1.4-fold increase in mean fluorescent intensity was observed in the **Eluate** BH GalNAc-T4 T47D and MCF7 samples over the controls, respectively (lanes 2 vs 4/6 vs 8, **Fig. 27E**), validating MUC1 as a GalNAc-T4 substrate in T47D and MCF7 cells.

When staining against CD46, multiple bands were observed between 38-50 kDa due to the existence of multiple CD46 splice variants (**Fig. 27E**)²⁴⁷⁻²⁴⁹. As a putative T47D-specific hit, we would have expected to see CD46 bands only in the **Eluate** T47D samples (lane 2, **Fig. 27E**). Instead CD46 bands were also observed in the **Eluate** MCF7 (lane 6, **Fig. 27E**) and MDA-MB-231 (lane 10, **Fig. 27E**) samples. A 1.5-fold, 1.4-fold, and 3.3-fold increase in mean fluorescent intensity was observed in the **Eluate** BH GalNAc-T4 T47D, MCF7 and MDA-MB-231 samples over the controls, respectively (lanes 2 vs 4/6 vs 8/10 vs 12, **Fig. 27E**), suggesting that CD46 was a GalNAc-T4 substrate in the three cell lines (**Fig. 27B**).

Looking back at the data from the MS-proteomics experiments, no peptides were detected for CD46 in the MCF7 or MDA-MB-231 samples, suggesting an issue with the detection of the protein in these cells due to low expression levels. We therefore compared the expression levels of CD46, MUC1 and GOLM1 in the panel of WT and transfected T47D, MCF7 and MDA-MB-231 cells. Staining against VSV-G and HA confirmed the expression of BH GalNAc-T4 and NahK respectively, and hence the identity of the cells (**Fig. 28**). The levels of CD46, MUC1 and GOLM1 were comparable between WT and transfected cells, demonstrating that transfection did not affect the protein expression levels (**Fig. 28**). The most intense bands for all three proteins were seen in the T47D cells (**Fig. 28**), suggesting that these cells express the highest levels of the proteins. The weakest bands for CD46 (**Fig. 28A**) and GOLM1 (**Fig. 28C**) were seen in the MCF7 cells, as previously observed in the neutravidin bead pulldown experiments. Very weak bands were seen in the MDA-MB-231 samples when staining against MUC1 (**Fig. 28B**), suggesting that these cells express very low levels of MUC1. This observation could explain why no MUC1 peptides were detected in the MDA-MB-231 samples in the MS-proteomics experiments. The results from this experiment agreed with the data published on the Human Protein Atlas²⁵⁰. The results obtained with MUC1 also agreed with previous findings by other groups^{246, 251}. Overall, this experiment highlighted how low expression levels may affect the detection of proteins by MS, as evidenced for CD46 in MCF7 and MDA-MB-231 cells and for MUC1 in MDA-MB-231 cells. Nevertheless, the neutravidin bead pulldown experiments validated that MUC1 is a GalNAc-T4 substrate in both luminal cells whilst CD46 and GOLM1 are substrates of GalNAc-T4 in the three BCCs tested (**Fig. 27B**).

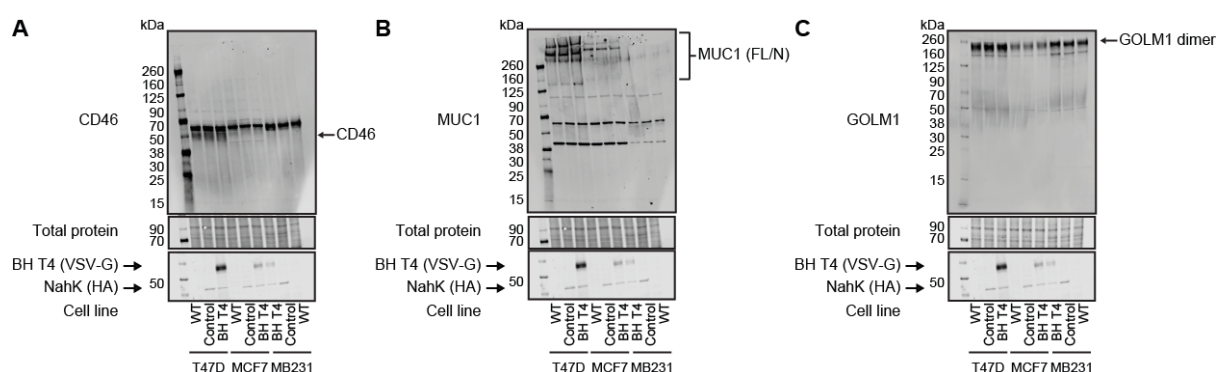


Fig. 28. Comparing the expression levels of CD46, MUC1 and GOLM1 in WT and transfected BCCs. Western blots against **A**) CD46, **B**) MUC1 and **C**) GOLM1. The same samples were used for all blots but antibody stainings were performed on three separate membranes: CD46 and MUC1 on one membrane (in the order stated), VSV-G and HA on a second membrane (in the order stated), and GOLM1 on a third membrane. The VSV-G and HA blot shown in the three panels is the same blot. 10 µg of protein were used per lane. Antibody staining was performed using antibodies against CD46 (1:500, 12494-1-AP), MUC1 (1:500, MA106503), GOLM1 (1:500, HPA010638), VSV-G (1:500, ab50549) and HA (1:500, ab18181). Data are representative of two independent replicates. BH T4 = AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells; control = AGX1^{F383A}-NahK cells; MB231 = MDA-MB-231.

Statistical significance implies that the mean abundance of a protein is different between two experimental groups²⁴⁴. However, statistical significance does not necessarily equate to biological relevance²⁴⁴. Biologically relevant results lie above the technical variation of an experiment (e.g. due to sample handling or LC-MS/MS instruments), which is why fold change cut-offs are typically applied in MS-proteomics experiments²⁴⁴. As mentioned above, CD46, MUC1 and GOLM1 were enriched over 2-fold in the Ac₄GalN6yne-treated BH GalNAc-T4 samples compared to the controls, suggesting that these hits are likely biologically relevant. We therefore decided to focus on these 3 proteins for downstream experiments. Nevertheless, future work should also validate some of the other hits obtained in the MS-proteomics experiments, such as JTB as another protein enriched over 2-fold in the Ac₄GalN6yne-treated BH GalNAc-T4 T47D samples compared to the controls (**Fig. 27B**).

3.7. Analysing the expression of CD46, MUC1 and GOLM1 in WT and GalNAc-T4 KO BCCs

Since there was a statistically significant growth defect upon GalNAc-T4 KO *in vivo* (**Fig. 13** and **Fig. 14**) we wondered if this could be due to differences in the expression levels of its substrates upon KO of GalNAc-T4. I therefore compared the expression of CD46, MUC1 and GOLM1 in WT and GalNAc-T4 KO cells by Western blot. However, comparable expression levels of the three proteins were observed between WT and GalNAc-T4 KO cells (**Fig. 29**). These results suggested that GalNAc-T4-mediated glycosylation must serve an alternative purpose other than affecting the expression levels of these proteins. To help draw structural conclusions on the role of GalNAc-T4-mediated glycosylation, the next objective was therefore to determine the GalNAc-T4-mediated glycosylation site(s) by MS-glycoproteomics.

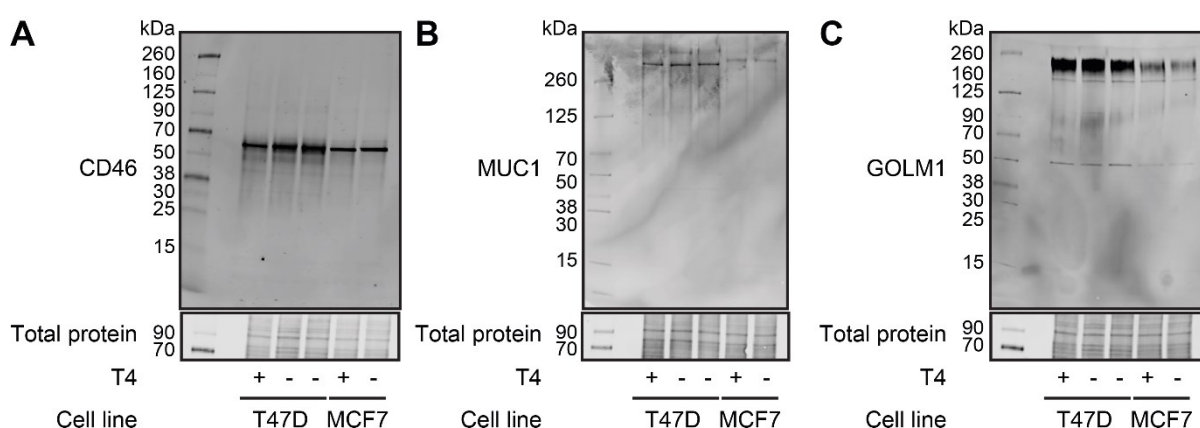


Fig. 29. Comparing the expression levels of CD46, MUC1 and GOLM1 in WT and GalNAc-T4 KO BCCs. Western blots against **A)** CD46, **B)** MUC1 and **C)** GOLM1. The same samples were used for all blots but antibody stainings were performed on three separate membranes. 10 µg of protein were used per lane. Antibody staining was performed using antibodies against CD46 (1:500, 12494-1-AP), MUC1 (1:500, MA106503), and GOLM1 (1:500, HPA010638). Data are representative of three independent replicates. T4 = endogenous GalNAc-T4, (+) = WT BCCs, (-) = GalNAc-T4 KO BCCs.

CHAPTER 4: Facilitating MS-glycoproteomic analyses on O-GalNAc glycoproteins using bioorthogonal reporter reagents

The work described in this chapter has been published in Calle et al. (2021)¹⁸².

Having characterised the substrates of GalNAc-T4 in a panel of BCCs, the next step was to identify the GalNAc-T4-mediated glycosylation site(s). Determining the exact glycosylation sites that affect disease can help inform on therapeutic strategies and can help draw structural conclusions on the role of GalNAc-T4-mediated glycosylation in BC. However, as described in the “**Substrate identification of individual GalNAc-Ts**” section, there are multiple challenges when performing glycosylation site-localisation of O-GalNAc glycans by MS-glycoproteomics, including but not limited to the lack of a consensus sequence, the poor ionisation efficiency of O-glycopeptides, and ion suppression due to unmodified peptides^{182, 183, 187-190}. These issues are worsened with glycoproteins containing mucin-like domains due to their complex and highly O-glycosylated nature¹⁸³. All these challenges make O-GalNAc glycosylation one of the most difficult PTMs to study by MS¹⁸². We believe that these challenges are likely why no glycopeptide hits were observed in the SILAC-based MS-(glyco)proteomics experiment on the lysates of transfected T47D^{WT} cells (**Section 3.3**).

Despite these challenges, successful glycosylation site-identification by MS-glycoproteomics was previously achieved by Schumann et al. with BH GalNAc-T1 and -T2¹⁹³, as described in the “**GalNAc-T Bump-and-Hole Engineering**” section. In these experiments, cells expressing the artificial biosynthetic pathway for UDP-sugar biosynthesis and BH GalNAc-T1 or -T2 were fed with a membrane-permeable substrate precursor, the tagged glycoproteins were derivatised with biotin-DADPS-picolyl azide, enriched with neutravidin beads, and analysed by LC-MS/MS¹⁹³. The analyte solution was converted into gaseous ions by electron spray ionisation (ESI) on positive ion mode¹⁹³. Upon entering the mass analyser, the gaseous ions were separated by their mass/charge ratio (m/z)²⁵². The most abundant peak in the MS1 spectrum is termed the base peak²⁵². The abundance of every peak in the spectrum is then determined relative to the base peak to obtain an MS1 ion relative abundance²⁵². Since ESI is a soft ionisation technique, little dissociation of the molecules takes place, and the peptides are predominantly detected as a protonated molecular ion $[M+H]^+$ ²⁵². The protons will be localised on the most basic amino acids present in the peptide (e.g. Arg, Lys or His)²⁵². Depending on the number of basic groups present, the peptide may be detected as a multiply protonated species instead.

Fragmentation of the MS1 precursor ions was then performed by HCD and ETD¹⁹³. In HCD, glycopeptides are bombarded with nitrogen atoms¹⁸². Kinetic energy is converted to internal energy which results in the most labile bonds being cleaved¹⁸³. HCD often leads to cleavage of the amide bonds (which results in b- and y-type fragment ions, **Fig. 30**) and loss of the sugar from the glycopeptide due to the labile nature of glycosidic bonds^{183, 253}. HCD fragmentation of GalNAc-modified glycopeptides generates a canonical “HCD HexNAc fingerprint”, consisting of the HexNAc oxonium ion (204.086 m/z) and five fragments from this ion at 126.055 m/z ([HexNAc – C₂H₆O₃]⁺), 138.055 m/z ([HexNAc – CH₆O₃]⁺), 144.065 m/z ([HexNAc – C₂H₄O₂]⁺), 168.066 m/z ([HexNAc – 2H₂O]⁺), and 186.076 m/z ([HexNAc – H₂O]⁺)²⁵⁴.

In ETD, an electron is transferred from an anionic reagent to the positively-charged peptides, resulting in charge reduction and fragmentation²⁵³. ETD often leads to cleavage of the N-C α bonds, which results in c- and z-type fragment ions (**Fig. 30**)^{183, 253}. The higher the charge density of the peptide, the more efficient the electron transfer and hence the fragmentation induced by ETD¹⁸³. Since ETD occurs with little vibrational energy redistribution, glycans often remain intact during fragmentation, enabling glycosylation site-localisation^{183, 253}.

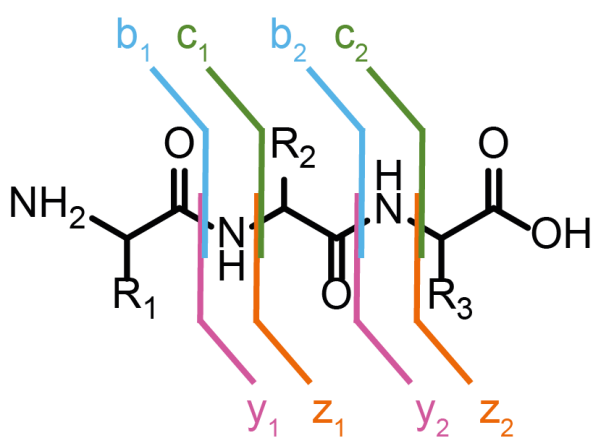


Fig. 30. Nomenclature for peptide fragmentation. HCD results in the formation of b- (in blue) and y- (in pink) type fragment ions. ETD results in the formation of c- (in green) and z- (in orange) type fragment ions. The subscript indicates the number of amino acids contained in the fragment ion relative to the N- or C-terminus. Figure adapted from Macias et al. (2020)²⁵³.

When analysing the data from the MS-glycoproteomics experiments with BH GalNAc-T1 and -T2, the group noticed that the tagged glycoproteins displayed unique MS signatures. Instead of the canonical “HCD HexNAc fingerprint”, new HCD oxonium ions were detected at 491.225 m/z, 330.156 m/z, 413.133 m/z and 302.150 m/z (indicated in green, **Fig. 31B**)^{182, 193}. These new HCD oxonium ions were used to trigger product-dependent ETD (HCD-pd-ETD)^{182, 193}. A unique fingerprint ion was consistently observed in all the ETD spectra at 194.10 m/z when the MS1 precursor charge was >3, and weakly detected in most ETD spectra when the MS1 precursor charge was 2 (**Fig. 31D**)¹⁸². The presence of these new HCD oxonium ions and ETD fingerprint ions facilitated glycopeptide assignment^{182, 193}.

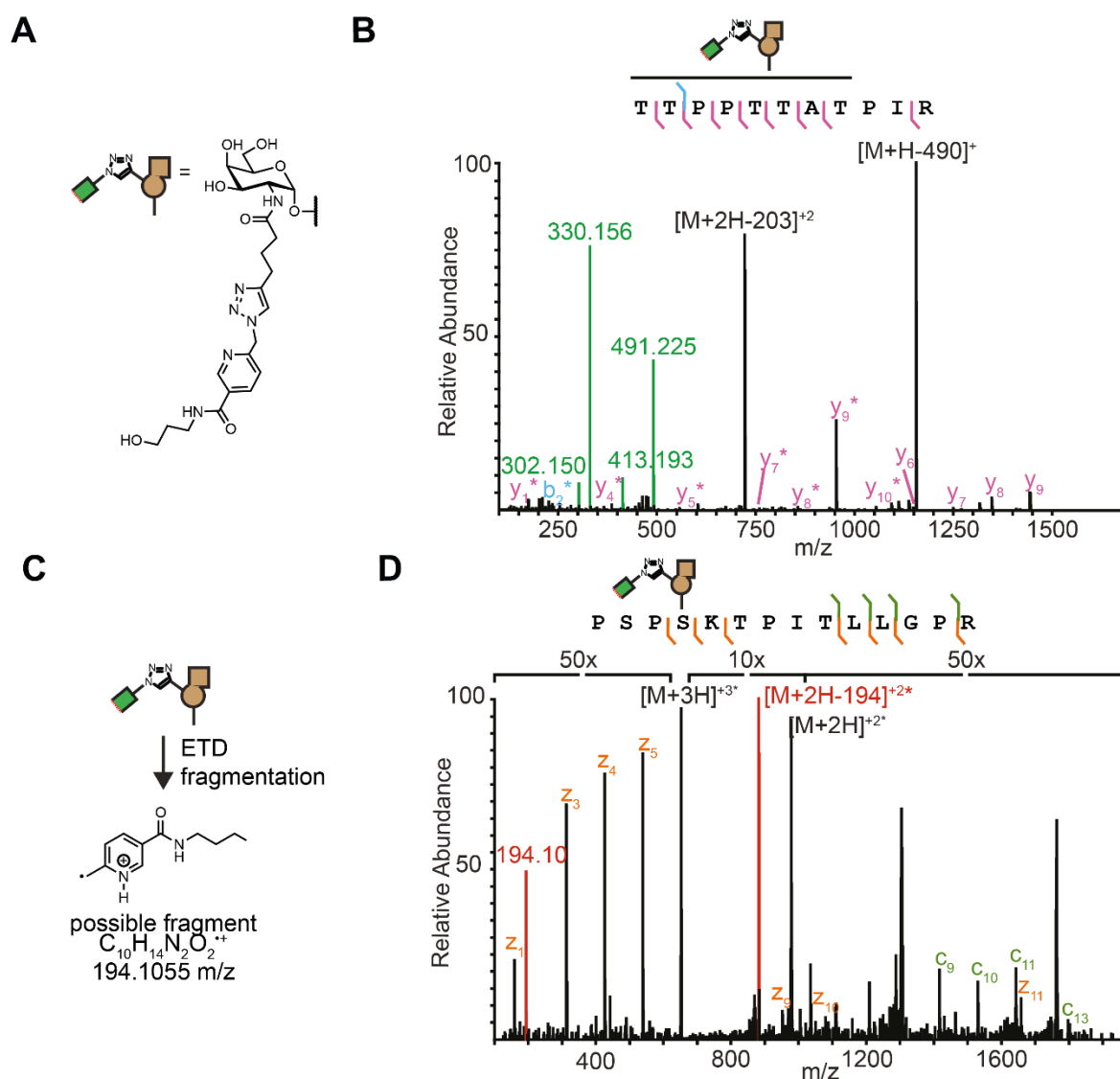


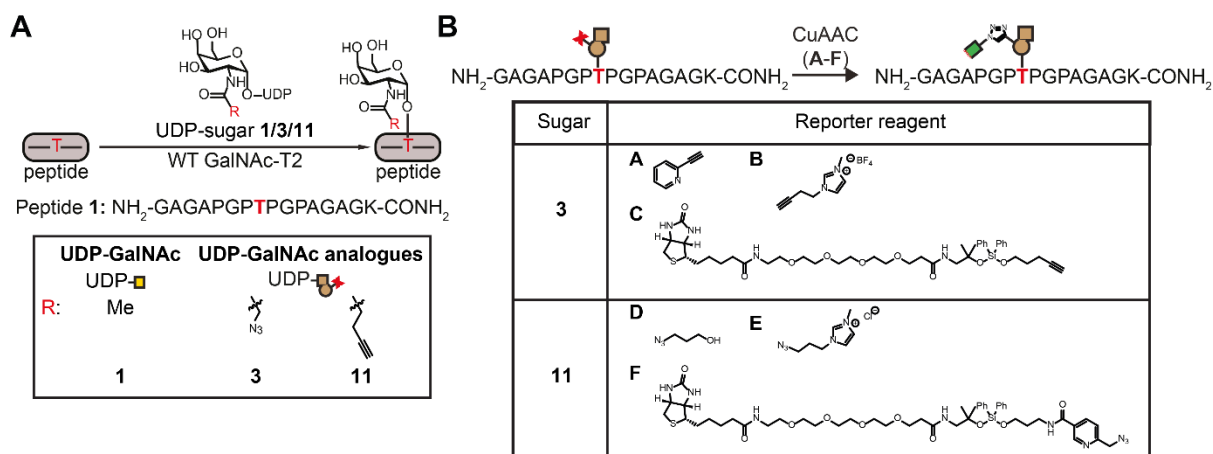
Fig. 31. Unique MS signatures of GalN6yne-modified glycopeptides derivatised with biotin-DADPS-picolyl azide. **A)** GalN6yne-modified glycopeptides derivatised with biotin-DADPS-picolyl azide after acid treatment and cleavage of the acid-labile diphenyldisiloxane moiety. **B)** Representative HCD spectrum of a chemically-tagged glycopeptide showing the unique HCD oxonium ions (in green). B ions are represented in blue, and y ions in pink. **C)** Proposed structure of the ETD fingerprint ion at 194.10 m/z. **D)** Representative ETD spectrum of a chemically-tagged glycopeptide showing the ETD fingerprint ion (in red). C ions are represented in green, and z ions in orange. Figure adapted from Calle et al. (2021)¹⁸².

Encouraged by these results, we decided to investigate whether alternative reporter reagents could impart glycopeptides with unique HCD and ETD fingerprint ions as mentioned above, as well as an increased charge density to facilitate ETD fragmentation - altogether improving site-localisation of glycopeptides¹⁸². We built up on previous work performed by the Schumann lab with GalNAc-T2 and set up *in vitro* glycosylation experiments with WT GalNAc-T2, a synthetic peptide substrate (peptide **1**) and either UDP-GalNAc (UDP-sugar **1**), UDP-GalNAz (UDP-sugar **3**) or UDP-GalNAIk (UDP-sugar **11**). Soluble His-tagged WT GalNAc-T2 (lacking the transmembrane domain but with the catalytic and lectin domains intact) was expressed in Expi293F cells and purified by Ni-NTA purification. Peptide **1** was identified by Gerken et al. as an optimal and selective GalNAc-T2 substrate²⁵⁵. We chose UDP-GalNAz and UDP-GalNAIk as azide- and alkyne-containing UDP-GalNAc analogues, respectively, showing comparable glycopeptide formation with WT GalNAc-T2 as the native substrate UDP-GalNAc (based on unpublished data from the Schumann lab). Ganka Bineva-Todd and I then derivatised the tagged glycopeptides by CuAAC with a panel of reporter reagents with increasingly basic or charged groups to improve the ionisation efficiency and increase the charge density of the tagged glycopeptides¹⁸².

The commercial reagent **F** (Biotin-DADPS-picolyl azide) is the biotin-based reporter reagent originally used by Schumann et al. to characterise the substrates of BH GalNAc-T1 and BH GalNAc-T2¹⁹³, and it is also the probe used herein in the SILAC-based MS-proteomics experiments (**Results Chapter 3**). The commercial reagent **C** (Biotin-DADPS-alkyne) is another biotin-based reporter reagent used by the Schumann lab in related studies with azide-containing sugars²⁵⁶. Both reagents contain an acid-labile diphenyldisiloxane moiety. Upon acid treatment, the corresponding alcohol is produced. The commercial reagent **A** (2-Ethynylpyridine) was previously used by Amant et al. to derivatise azide-containing peptide nucleic acid sequences by CuAAC to generate pyridyl-triazole ligands which can reversibly bind metal cations *in vitro*²⁵⁷. The commercial reagent **D** (3-Azido-1-propanol) was previously used by Bumpus and Baskin to profile Phospholipase D enzyme activity by CuAAC in cells²⁵⁸. Imidazolium tag (iTag)-containing reagents have been successfully used by other groups to analyse and purify glycans *in vitro* and *in cellulo*²⁵⁹⁻²⁶⁶. The iTag-containing reagents used herein were a gift from Prof. Carmen Galan (reagent **B**) and synthesised by Dr Andrea Marchesi (Schumann lab)(reagent **D**). Since these compounds contain a permanent positive charge, we envisioned that the iTag would impart a higher charge density to the tagged glycopeptides and hence result in greater ETD fragmentation efficiency¹⁸².

The *in vitro* glycosylation reaction mixtures containing WT GalNAc-T2, peptide **1**, and UDP-sugar **1**, **3** or **11** were incubated at 37 °C for 25 h (**Fig. 32A**). An aliquot was taken for LC-MS analysis and glycopeptide formation was determined by MS⁺ peak integration using **Equation 2**. Glycopeptide formation (>75%) was confirmed in all samples (**SI Fig. 10B**). To maximise glycosylation, the reaction mixtures were incubated for an additional 17 h. The samples were then treated with CuAAC solutions containing the different reporter reagents **A-F** (**Fig. 32B**). The success of the CuAAC reaction was confirmed in all samples, although the reaction efficiency varied between reagents (**SI Fig. 10D**). Nevertheless, all samples were injected into an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Helen Flynn from the Proteomics STP at the Francis Crick Institute. Manual analysis of the MS spectra was performed with Xcalibur software by Dr Stacy Malaker.

Looking at the MS1 intensities, the glycopeptides containing the underivatized modified sugars (samples **1**, **2** and **6**) were exclusively detected as a +2 ion (**Fig. 32C**)¹⁸², where protonation occurs in the Lys amino acid and the terminal NH₂ group from the peptide. In contrast, for the glycopeptides derivatised with the reporter reagents, the abundance of the +3 ion correlated with the basicity of the chemical tag¹⁸². The lowest 3+ ion abundance was observed in the glycopeptides derivatised with the reporter reagents containing the aliphatic alcohols (samples **5** and **7**, **Fig. 32C**), where the weakly basic triazole (pKaH_(H2O) = 1.17)²⁶⁷ product of the CuAAC reaction accepts an additional proton¹⁸². The pyridine and picolyl groups in reagents **A** and **F**, respectively, contain another basic nitrogen which increased the abundance of the 3+ ion (samples **3** and **9**, **Fig. 32C**)¹⁸². The basic nitrogen in the picolyl group is located within a nicotinamide moiety (pKaH_(H2O) = 3.35)²⁶⁸, where an electron-withdrawing amide reduces the basicity of the nitrogen compared to the pyridine group (pKaH_(H2O) = 5.23)²⁶⁷, which is reflected on the abundance of the 3+ ion¹⁸². The samples containing the iTag-containing reagents (samples **4** and **8**, **Fig. 32C**) showed 100% abundance of the +3 ion, likely due to the permanent positive charge of the iTag¹⁸².



C

Sample #	Sugar	Reporter reagent	MS1 ion relative abundance (%)	
			+2 ion	+3 ion
1	1	-	100	n.d.
2	3	-	100	n.d.
3		A	100	45
4		B	50	100
5		C	100	0.8
6	11	-	100	n.d.
7		D	100	0.8
8		E	40	100
9		F	100	10

Fig. 32. Testing different bioorthogonal reporter reagents to facilitate MS-glycoproteomic analyses. **A)** Schematic representation of *in vitro* glycosylation experiments performed with WT GalNAc-T2, peptide 1, and UDP-sugars 1, 3 and 11. Red T represents the peptide acceptor site used by GalNAc-T2. Figure adapted from Choi et al. (2019)⁶⁷. **B)** Schematic representation of the CuAAC reaction between the tagged glycopeptides and a panel of bioorthogonal reporter reagents A-F. **C)** Relative abundance of the MS1 ions from the glycopeptides tagged with sugars 1, 3 and 11 and derivatised with reporter reagents A-F. Table adapted from Calle et al. (2021)¹⁸². n.d. = not detected.

We next investigated the fragmentation behaviour of the synthetic glycopeptides by HCD. When analysing the HCD spectra, one of the most abundant ions present in all samples was the peptide ion without the glycan (i.e. $[M+H\text{-glycan}]^+$)(**Fig. 33**)¹⁸². As seen in the MS-glycoproteomics experiments with BH GalNAc-T1 and -T2¹⁹³, unique HCD oxonium ions that were distinct from the canonical HexNAc fingerprint ions were seen in all the samples containing the modified sugars derivatised with the reporter reagents (indicated in green, **Fig. 33**)¹⁸². For reagent **A** the most abundant oxonium ion was at 348.131 m/z ($M + H$), with a second less abundant ion at 330.120 m/z due to the loss of water (**Fig. 33A**)¹⁸². For reagent **B** a single oxonium ion was detected at 379.173 m/z ($M + H$)(**Fig. 33B**)¹⁸². For reagent **C** the most abundant ion was at 329.146 m/z ($M + H$) with a second less abundant ion at 311.136 m/z due to the loss of water (**Fig. 33C**)¹⁸². For reagent **D** two abundant ions were detected at 343.161 m/z ($M + H$) and at 182.092 m/z due to the loss of hexosamine, as well as two less abundant ions detected at 325.151 m/z due to the loss of water and at 265.130 m/z (**Fig. 33D**)¹⁸². For reagent **E** the most abundant ion was at 407.204 m/z ($M + H$) with a second less abundant ion at 325.151 m/z due to the loss of water (**Fig. 33E**)¹⁸². For reagent **F** an abundant ion was detected at 477.209 m/z ($M + H$), although the most abundant ion was detected at 316.140 m/z due to the loss of hexosamine, as well as less abundant ions detected at 399.177, 288.134 and 178.203 m/z (**Fig. 33F**)¹⁸². The presence of at least one of these unique HCD ions was used to trigger HCD-pd-ETD in all samples¹⁸².

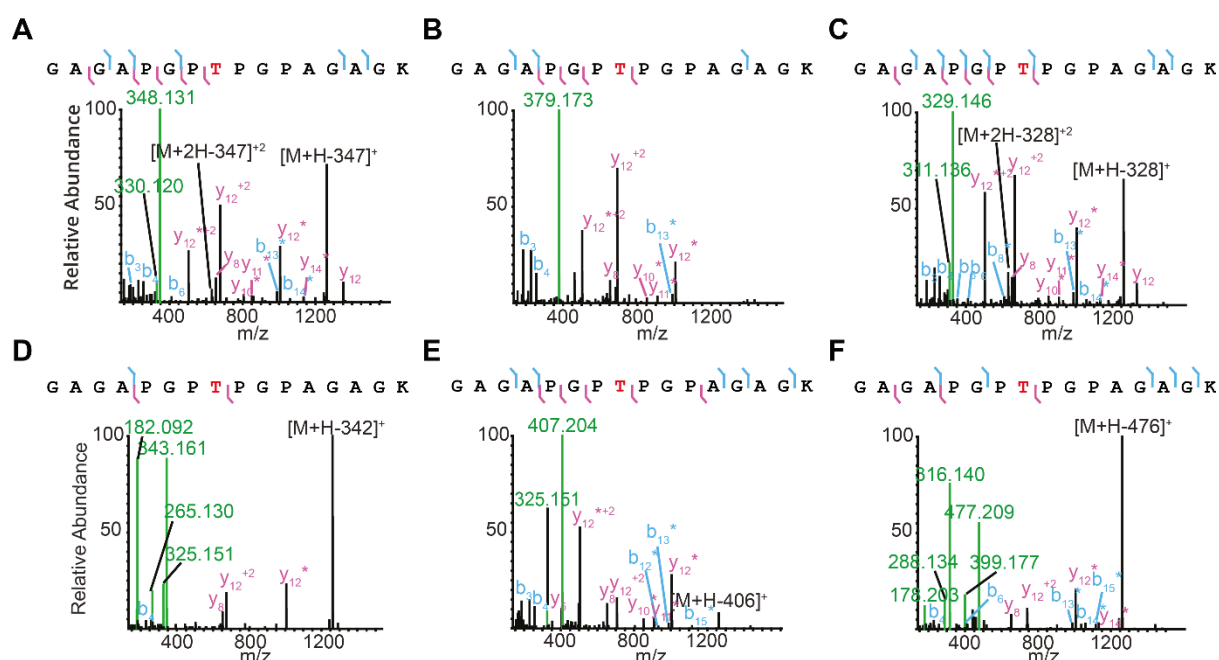


Fig. 33. HCD spectra of the synthetic glycopeptides derivatised with reporter reagents A-F. B ions are represented in blue, y ions in pink, and unique HCD oxonium ions in green. Red T represents the peptide acceptor site used by GalNAc-T2. Figure adapted from Calle et al. (2021)¹⁸².

We next investigated the fragmentation behaviour of the synthetic glycopeptides by ETD. We were able to trigger ETD on all glycopeptides¹⁸². The greatest number of c- and z-type fragment ions (and hence the highest sequence coverage) was detected with the iTag-containing reagents **B** and **E**, most likely due to the increased charge density imparted by the iTag inducing high ETD fragmentation efficiency¹⁸². An ETD fingerprint ion was observed at 178.21 m/z with reagent **B** (**Fig. 34B**)¹⁸², comparable to the one previously observed at 194.10 m/z with biotin-DADPS-picolyl azide in the MS-glycoproteomics experiments with BH GalNAc-T1 and -T2 (**Fig. 31D**)¹⁹³, allowing confident glycopeptide assignment. A possible mechanism for the formation of this ion could be electron transfer to the iTag-containing reagent causing dissociation of the triazole from the hexosamine (**Fig. 34A**)¹⁸².

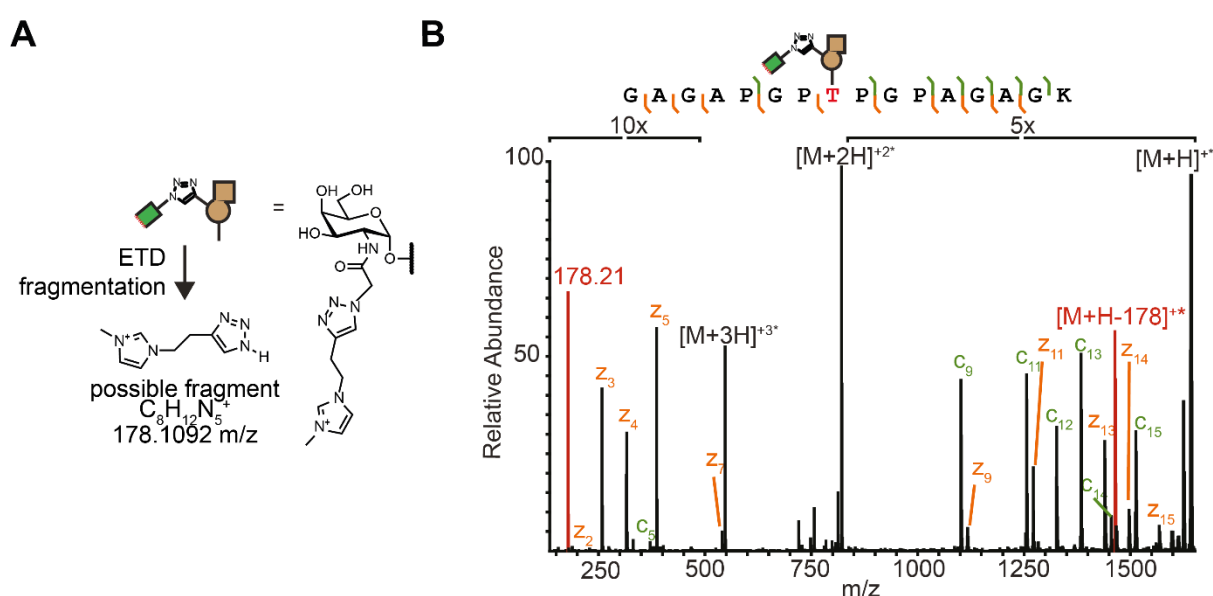


Fig. 34. ETD spectrum of the synthetic glycopeptide modified with sugar 3 and derivatised with the iTag-containing reagent B. **A)** Proposed structure of the ETD fingerprint ion at 178.21 m/z. **B)** ETD spectrum of the glycopeptide modified with sugar 3 and derivatised with reagent B showing the ETD fingerprint ion (in red). C ions are represented in green, and z ions in orange. Red T represents the peptide acceptor site used by GalNAc-T2. Figure adapted from Calle et al. (2021)¹⁸².

Overall, these experiments highlighted the unique MS properties that modified sugars and bioorthogonal reporter reagents can impart on glycopeptides to facilitate MS-glycoproteomics analyses, with the most promising results obtained with the iTag-containing reagents¹⁸².

CHAPTER 5: Determining the GalNAc-T4-mediated glycosylation site(s) in CD46 and GOLM1 by MS-glycoproteomics

GOLM1 and CD46 were the two hits conserved amongst the three BCCs investigated in the MS-proteomics experiments (**Fig. 27B**). As mentioned above, GOLM1 is upregulated in BC tissue compared to healthy mammary tissue and it has been shown to promote proliferation and invasion of BCCs *in vitro* and *in vivo*, putatively through regulation of the EMT process and expression of MMP-13²⁴³. GOLM1 is glycosylated²⁴⁵, and Jiang et al. demonstrated that removal of N-glycans at Asn¹⁴⁴ enhances metastasis of hepatocellular cancer cells²⁶⁹ - revealing a role of GOLM1 glycosylation in cancer. GOLM1 is also predicted to carry O-GalNAc glycans as reported in GlycoDomainViewer¹⁹², however the role of O-GalNAc glycans on the function of GOLM1 is unknown.

CD46 is also upregulated in BC tissue compared to healthy mammary tissue^{270, 271}. CD46 acts as a cofactor for complement factor I, which cleaves and inactivates the complement proteins C3b and C4b, hence protecting cells from complement-mediated damage (**Fig. 35A**)^{272, 273}. CD46 is also thought to regulate the adaptive immune response by controlling T cell activation, contraction, and homeostasis (**Fig. 35B**)²⁷⁴. Upon T cell stimulation, autocrine engagement of CD46 leads to the MMP-mediated shedding of the CD46 ectodomain and cleavage of its cytoplasmic tail by γ -secretase (**Fig. 35B**)²⁷⁴. The cytoplasmic tail is then translocated into the nucleus where it triggers downstream signalling (**Fig. 35B**)²⁷⁴. Alternative splicing of CD46 results in 2 possible cytoplasmic tails termed CYT-1 or CYT-2²⁷⁴. Resting CD4+ T cells predominantly express CYT-2-containing CD46 isoforms²⁷⁴. During T cell activation, CYT-1-containing CD46 isoforms are upregulated²⁷⁴. Nuclear translocation of CYT-1 results in IFN- γ production and a T helper 1 effector cell (T_H1) phenotype (**Fig. 35B**)²⁷⁴. Upon T cell contraction, T cells revert to predominantly expressing CYT-2-containing CD46 isoforms²⁷⁴. Nuclear translocation of CYT-2 results in IL-10 production and a type 1 regulatory T cell (Tr1) phenotype (**Fig. 35B**)²⁷⁴. CD46 contains a domain which is enriched in Ser, Thr and Pro amino acids termed the STP domain²⁷². Liszewski et al. demonstrated that deletion of the STP domain reduced the cofactor activity of CD46²⁷⁵. Similarly, Ni Choileain et al. demonstrated that deletion of the STP domain reduced CD46 shedding, impaired the recruitment of CD46 to the immune synapse, and the ensuing switch from T_H1 into Tr1 cells following T cell stimulation²⁷⁶. Altogether, these results demonstrated that the STP domain is crucial for fine-tuning the function of CD46. Since the STP domain is heavily O-glycosylated²⁷³, it also indirectly inferred a role of O-glycosylation in the function of CD46.

We therefore decided to focus on determining the GalNAc-T4-mediated glycosylation site(s) of GOLM1 and CD46 due to their clinical relevance and the putative role of glycosylation on the function of these proteins. The T47D cell line was chosen for these experiments since Müller et al. demonstrated that the repertoire of glycan structures found in T47D cells is limited¹⁶⁰, which should facilitate MS-glycoproteomics analyses on these cells. We transduced the transfected T47D^{WT} cells with viruses expressing full-length GOLM1 or secreted CD46 tagged with GFP and a His-tag (**Fig. 36A**). After feeding the cells with the modified sugar the tagged proteins can be purified by Ni-NTA and/or GFP-affinity purification (**Fig. 36B**). Simplifying the complexity of the sample by performing the experiment on purified recombinant proteins should facilitate the glycosylation site identification by MS-glycoproteomics.



Fig. 36. Determining the GalNAc-T4-mediated glycosylation site(s) in CD46 and GOLM1 by MS-glycoproteomics. **A**) Expression construct encoding the protein of interest (POI) tagged with GFP and a His-tag used in the MS-glycoproteomics experiments. **B**) Schematic representation of the workflow for the MS-glycoproteomics experiments with purified recombinant CD46 and GOLM1.

5.1. Identifying the GalNAc-T4-mediated glycosylation site(s) in CD46

I first attempted to obtain the GalNAc-T4-mediated glycosylation site(s) in CD46. CD46 is a type I membrane protein that is expressed by all cells except erythrocytes²⁷². The N-terminal portion of CD46 consists of four short consensus repeats (SCR1-4) which constitute the bulk of the extracellular region of CD46 and mediate C3b and C4b binding (**Fig. 37**)²⁴⁷. CD46 is N-glycosylated at Asn⁴⁹ (SCR1), Asn⁸⁰ (SCR2) and Asn²³⁹ (SCR4)²⁴⁷. The four SCRs are followed by a short region which is rich in Ser, Thr, and Pro amino acids termed the STP domain which is heavily O-glycosylated (**Fig. 37**)²⁷³. The STP domain is followed by a transmembrane domain and a cytoplasmic tail (**Fig. 37**)²⁷². Alternative splicing of the CD46 gene results in multiple splice variants, however most cells primarily express four of these isoforms in approximately the same ratio²⁷⁷. The four isoforms (BC1/BC2/C1/C2) have identical SCRs but differ in the STP domain and the cytoplasmic tail (**Fig. 37**)^{248, 272}. The BC isoforms differ from the C isoforms in the presence of an additional 15 amino acids within the STP domain (**Fig. 37**)^{248, 272}. BC1 and C1 isoforms express the cytoplasmic tail CYT-1, whilst the BC2 and C2 isoforms express the cytoplasmic tail CYT-2^{248, 272}.

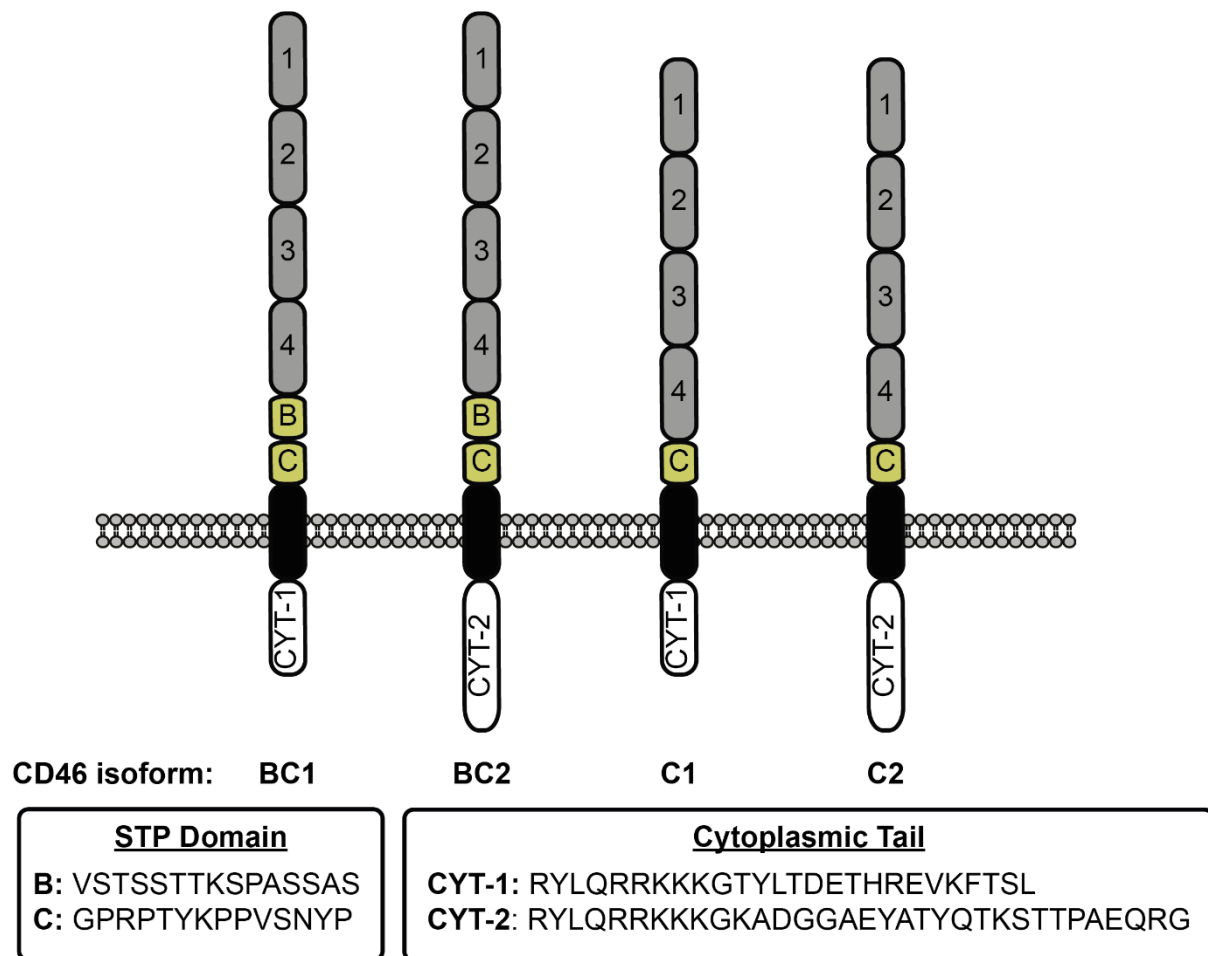


Fig. 37. Structure and isoforms of CD46. Schematic representation of the four most common isoforms of CD46 – BC1, BC2, C1 and C2. The four SCRs are represented in grey, the STP domain in green, the transmembrane domain in black and the cytoplasmic tail in white. The amino acid sequence of the STP domain and the cytoplasmic tail splice variants are boxed. Figure adapted from Manchester et al. (1994)²⁴⁸.

For these experiments we used T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells (“BH T4” cells), and the T47D^{WT}-AGX1^{F383A}-NahK cells as a control (“control” cells). I transduced the cells with viruses encoding soluble versions of the CD46 BC and C isoforms tagged with GFP and a His-tag, where the extracellular region of CD46 was intact but the transmembrane and cytoplasmic domains were absent. The sequences of the CD46 BC and C constructs used in these experiments are available in **SI Fig. 11** and **SI Fig. 12** respectively. We chose to express soluble versions of CD46 since secreted proteins offer easier expression in mammalian cells compared to transmembrane proteins and to simplify the issue of the two different cytoplasmic tails. Furthermore, previous work has successfully expressed soluble versions of CD46 and then confirmed its ability to inhibit complement activation *in vitro* and *in vivo*^{249, 278}. Transduction success was confirmed by checking for GFP expression using a fluorescence microscope and protein expression control by Western blot (**SI Fig. 13**).

5.1.1. Protein expression and Ni-NTA/GFP-affinity purification of recombinant CD46

The CD46 constructs were predominantly present in the secretome of the transduced cells (**SI Fig. 13B**). I therefore collected the secretome of the transduced cells fed with Ac₄GalN6yne and performed an initial Ni-NTA purification (**Fig. 38A**). In this experiment, I used normal growth medium instead of FBS-free medium to recapitulate the native glycosylation state of the cells. Aliquots of the same volume were taken at different steps of the Ni-NTA purification to track the presence of CD46 by SDS-PAGE. A strong band around 50-70 kDa was observed in all samples (**Fig. 38B**), likely due to leftover albumin (expected MW = 66.5 kDa) - the intensity of which decreased throughout the procedure. To elute the protein from the beads, I treated the samples with an increasing concentration of imidazole. The expected MW of the CD46 construct was 65 kDa. Weak bands at this MW were seen in all the eluted fractions (**Fig. 38B**). However, additional bands were also observed due to impurities co-eluting with CD46 (**Fig. 38B**). The eluted fractions were hence combined, and an additional GFP-affinity purification was performed using the same methodology I used during my MSci project with recombinant WT/BH GalNAc-T4 (**Fig. 38A**)²⁰⁸. Aliquots of the same volume were taken at different steps of the GFP-affinity purification to track the presence of CD46 by SDS-PAGE. The expected MW of the CD46 constructs after TEV cleavage was 37 kDa. Weak bands at this MW were seen in the eluted fractions for all samples (**Fig. 38C**). Bands between 50-70 kDa were also observed in the eluted fractions (**Fig. 38C**), most likely due to the uncleaved proteins (expected MW = 65 kDa). The weak band seen between 25-30 kDa in the eluted fractions was likely due to TEV (expected MW = 28 kDa). The E1-E3 samples were combined, and the protein concentration was determined by Nanodrop. The concentration of all the samples was quite low (**Table 5**), especially the BH T4 CD46 BC samples. Looking at the gels, there still seemed to be some uncleaved protein bound to the beads (**Fig. 38C**). After the beads had been left over the weekend at 4 °C, we performed a second TEV cleavage to collect as much recombinant CD46 as possible. The amount of protein obtained after this second TEV cleavage was much greater (**Table 5**).

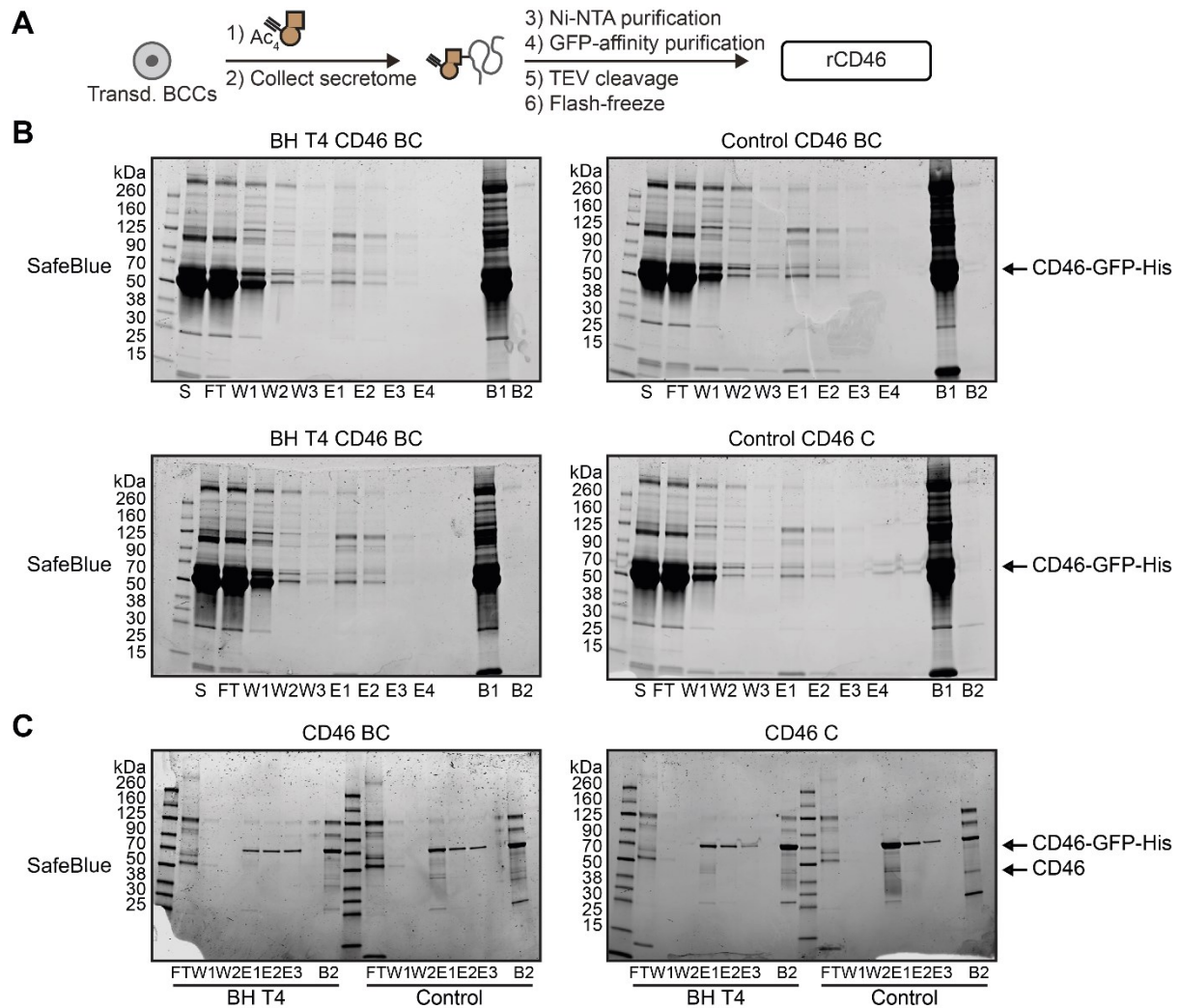


Fig. 38. Protein expression and Ni-NTA/GFP-affinity purification of recombinant CD46. **A)** Schematic representation of the protein expression and purification of recombinant CD46. **B)** SDS-PAGE gels of the fractions after Ni-NTA purification. The expected MW of the CD46 BC and C constructs before TEV cleavage is 66 and 64 kDa respectively (arrow pointing). The expected MW of albumin is 66.5 kDa. **C)** SDS-PAGE gels of the fractions after GFP-affinity purification. The expected MW of the CD46 BC and C constructs after TEV cleavage is 38 and 36 kDa respectively (arrow pointing). The expected MW of both TEV and the GFP-His construct is 28 kDa. Data are from one experiment. rCD46 = recombinant CD46; S = supernatant; FT = flow-through; W1-3 = wash 1-3; E1-4 = elution 1-4; B1 = beads before elution; B2 = beads after elution; BH T4 = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells; control = T47D^{WT}-AGX1^{F383A}-NahK cells.

Table 5. Protein concentration and amount in each sample (50 μ L) of the recombinant CD46 samples after Ni-NTA and GFP-affinity purification. BH T4 = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells, control = T47D^{WT}-AGX1^{F383A}-NahK cells

Sample	1st TEV cleavage		2nd TEV cleavage	
	Concentration (mg/mL)	Amount of protein (μ g)	Concentration (mg/mL)	Amount of protein (μ g)
BH T4 CD46 BC	0.07	3.5	2.85	142.5
Control CD46 BC	0.45	22.5	3.01	150.5
BH T4 CD46 C	1.86	93.0	3.61	180.5
Control CD46 C	0.56	28.0	3.29	164.5

To confirm whether recombinant CD46 was tagged by GalNAc-T4, I treated the recombinant CD46 samples with PNGase F to remove N-glycans, performed a CuAAC reaction with biotin-picolyl azide, and assessed the tagging of recombinant CD46 by streptavidin blot (**Fig. 39A**). Protein expression control was simultaneously performed by Western blot. Since the concentration of the samples from the 1st TEV cleavage was very low, I used the samples from the 2nd TEV cleavage instead (**Table 5**). Multiple bands were observed in the total protein stain for all samples due to impurities co-eluting with recombinant CD46 (**Fig. 39B**). This suggested that the second GFP-affinity purification step did not substantially improve the purity of the sample. When staining against CD46, bands were observed between 30-50 kDa in all samples (**Fig. 39B**), confirming the presence of CD46. The intensity of the bands varied between samples (**Fig. 39B**), reflecting the different levels of recombinant CD46 expressed. The expected MW of the CD46 BC and C constructs after TEV cleavage was 38 and 36 kDa, respectively, however the bands observed in the BC samples were closer to 50 kDa instead (**Fig. 39B**). A possible reason for this difference in MW could be due to O-glycosylation of the proteins, since the B domain has ten potential O-GalNAc glycosylation sites (in red: **VSTSSTTKSPASSAS**) whilst the C domain only has two (in red: **GPRPTYKPPVSNYP**). Two bands were observed for CD46 in all samples (**Fig. 39B**), perhaps due to non-uniform derivatisation of tagged recombinant CD46 during the CuAAC reaction. A signal was observed in the streptavidin blot for all samples; however, the signal was very low and not specific to the CD46 bands (**Fig. 39B**). The low intensity could be due to the low amount of protein used in the experiment (2 μ g) and the non-specific bands could be due to the high background signal caused by leftover total protein stain despite destaining the membrane before performing the antibody stainings.

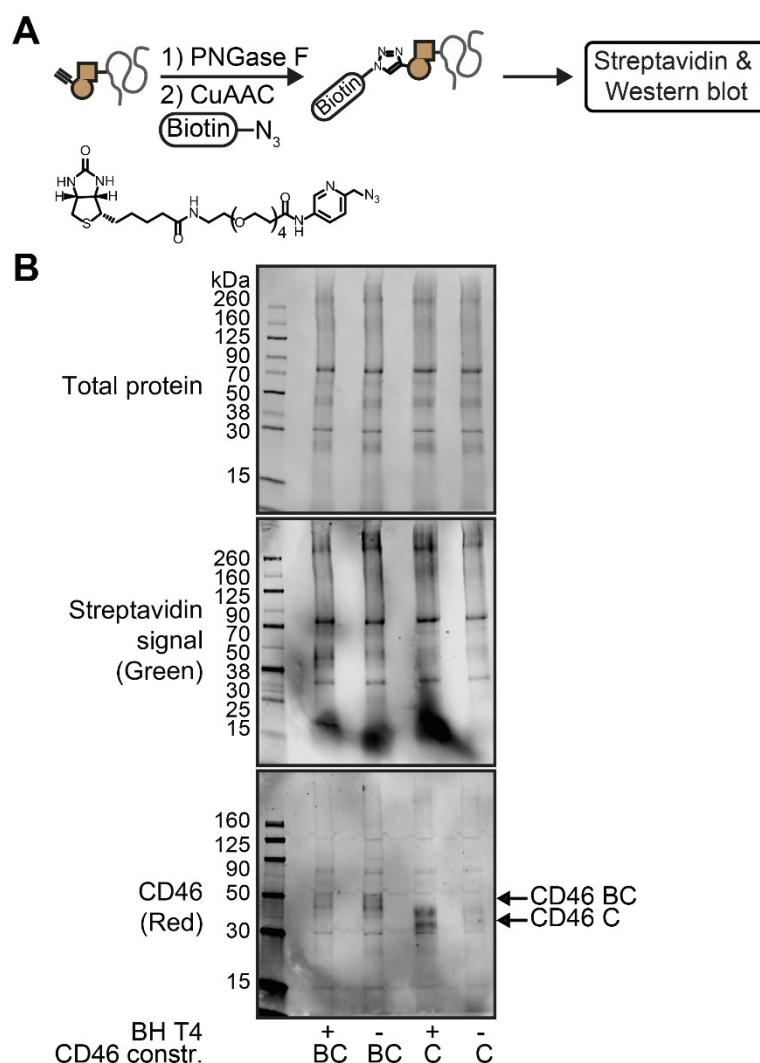


Fig. 39. Tagging and protein expression control of recombinant CD46. **A)** Schematic representation of the experiment. **B)** Tagging and protein expression control by Western blot after the eluted fractions were subjected to PNGase F treatment and CuAAC reaction with biotin-picolyl azide. The stainings were performed on the same membrane in the order displayed. Red/Green refers to the imaging channel. The expected MW of the CD46 BC and C constructs after TEV cleavage is 38 and 36 kDa respectively (arrows pointing). 2 μ g of protein were used per lane. Stainings were performed using IRDye® 800CW Streptavidin (1:5000, 926-32230) and an antibody against CD46 (1:500, 12494-1-AP). Data are from one experiment. BH T4 (+) = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells; control (-) = T47D^{WT}-AGX1^{F383A}-NahK cells.

To confirm successful tagging of recombinant CD46 by GalNAc-T4 I repeated the experiment using a higher amount of protein (25 µg). I also wanted to confirm whether the CD46 bands were observed at a higher MW than expected due to glycosylation of the protein. As mentioned above, CD46 is known to be N-glycosylated in the SCR domains²⁴⁷, and O-glycosylated in the STP domain^{272, 273}. I therefore treated the samples with the glycosidases SialEXO (digests sialic acids) and OglyZOR (digests core 1 and core 3 O-glycans in the presence of SialEXO) in addition to PNGase F (digests N-glycans) and compared the MW of the CD46 bands against undigested samples by Western blot (**Fig. 40A**). To dissect the contribution of N- and O-glycans to the shift in the MW of the bands, I also repeated the experiment and compared the MW of the CD46 bands in samples treated with PNGaseF/SialEXO/OglyZOR against samples treated with PNGase F but not SialEXO/OglyZOR (**Fig. 40B**) – the latter should contain O-glycans only.

When loading a higher amount of protein and staining against CD46, bands were observed in all samples (**Fig. 40B-C**), confirming the presence of the CD46 constructs. In the undigested samples, the CD46 bands were observed between 40-70 kDa (lanes 1/3/5/7, **Fig. 40B**). In the samples treated with PNGase F but not SialEXO/OglyZOR, the CD46 bands were observed between 30-50 kDa instead (lanes 1/3/5/7, **Fig. 40C**). Treating the samples with PNGase F/SialEXO/OglyZOR shifted the bands to the expected MW of CD46 (expected MW of the CD46 BC and C constructs after TEV cleavage is 40 and 36 kDa respectively)(lanes 2/4/6/8, **Fig. 40B-C**). This confirmed that the bands were previously observed at a higher MW than expected due to glycosylation of CD46, and that CD46 is N- and O-glycosylated. The shift in MW was greater in the BC samples (lanes 1-4, **Fig. 40C**) compared to the C samples (lanes 5-6, **Fig. 40C**), confirming that the former contain more O-glycans than the latter, as hypothesised above since the B domain has ten potential O-GalNAc glycosylation sites whilst the C domain only has two. As seen in the composite images, the streptavidin signal overlapped with the CD46 bands (**Fig. 40B-C**), suggesting that recombinant CD46 was tagged by BH GalNAc-T4. The streptavidin signal was maintained after SialEXO/OglyZOR treatment (lanes 2/4/6/8, **Fig. 40B-C**), suggesting that unmodified O-glycans can be removed whilst the modified sugar remains intact. Some signal was observed in the control samples (lanes 3/4/7/8, **Fig. 40B-C**), probably due to leftover tagged N-glycans despite PNGase F treatment. The signal in the BH T4 CD46 C samples (lanes 5/6, **Fig. 40B-C**) was comparable to the signal in the control samples (lanes 3/4/7/8, **Fig. 40B-C**). In contrast, there was a boost in the streptavidin signal in the BH T4 CD46 BC samples (lanes 1/2, **Fig. 40B-C**), suggesting that BH GalNAc-T4 may preferentially glycosylate the CD46 BC isoform. Since the only difference between the BC and C isoforms is the presence of the STP B domain in the former, this would suggest that the glycosylation site of GalNAc-T4 in CD46 is in the STP B domain.

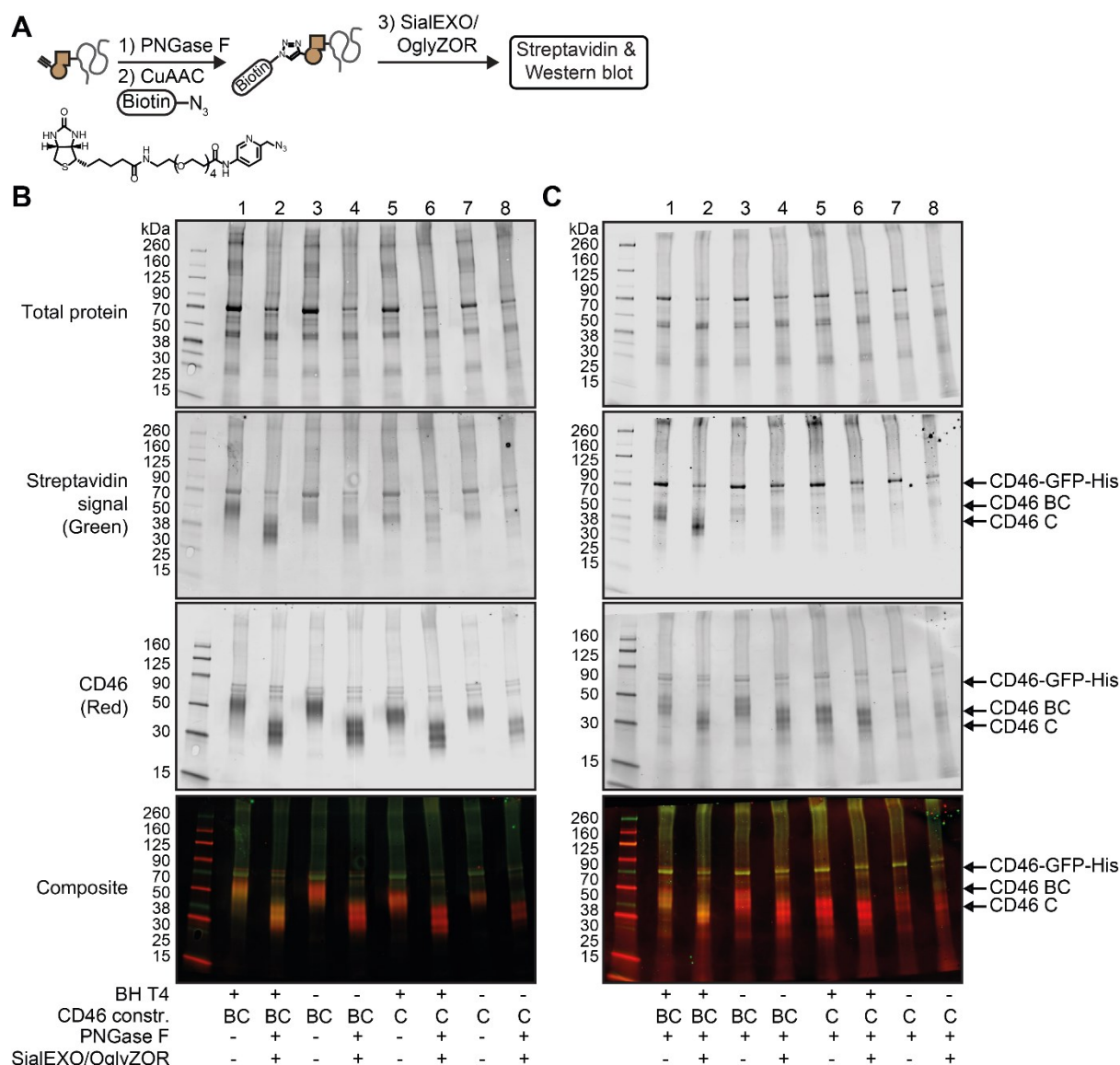


Fig. 40. Glycosidase treatment of recombinant CD46. **A)** Schematic representation of the experiment. **B and C)** Tagging and protein expression control by Western blot after the eluted fractions were subjected to CuAAC reaction with biotin-picolyl azide and glycosidase treatment. In **B)** the controls weren't treated with any glycosidase whilst in **C)** the controls were treated with PNGase F but not SialEXO or OglyZOR. The stainings were performed on the same membrane in the order displayed. Red/Green refers to the imaging channel. The expected MW of the CD46 BC and C constructs after TEV cleavage is 38 and 36 kDa, respectively (arrows pointing). 25 µg of protein were used per lane. Stainings were performed using IRDye® 800CW Streptavidin (1:5000, 926-32230) and an antibody against CD46 (1:500, 12494-1-AP). Data are from one experiment. BH T4 (+) = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells, control (-) = T47D^{WT}-AGX1^{F383A}-NahK cells.

5.1.2. Optimisation of the MS-glycoproteomics workflow for CD46

Having confirmed the identity and the tagging of the recombinant CD46 constructs by BH GalNAc-T4, we next sought to optimise the MS-glycoproteomics workflow to determine the GalNAc-T4-mediated glycosylation site(s) in CD46. For these experiments we used the recombinant CD46 samples from the 1st TEV cleavage since these samples had a greater amount of CD46 even if the total protein content was lower (**SI Fig. 14**). Since the BH T4 CD46 C sample had the highest protein concentration (**Table 5**), we used this sample for the optimisation experiment. To maximise the coverage of CD46, we chose to include chymotrypsin (68 cleavage sites) in the workflow in addition to trypsin (27 cleavage sites). Since our results suggested that the glycosylation site of GalNAc-T4 in CD46 was in the STP B domain, which is a mucin-like domain, we hypothesised that including a mucinase in the workflow might help maximise the coverage of this domain. We chose to use the mucinase SmE for this workflow, which cleaves N-terminally to a glycosylated Ser or Thr residue¹⁹⁰. SmE can accommodate a variety of O-glycans at the glycosylated Ser/Thr, as well as O-glycosylation at the adjacent C-terminal residue¹⁹⁰. We also included the glycosidases SialEXO and OglyZOR in the workflow, since the results of the glycosidase experiment demonstrated that we could remove unmodified sugars whilst leaving the modified sugar intact (**Fig. 40B-C**). Reducing the glycosylation state of the proteins should help with protease digestion by facilitating access to the peptide backbone, as well as helping with the ionisation and detection of the glycopeptides by MS. I therefore incubated the samples with PNGase F to remove N-glycans, followed by CuAAC reaction with the iTag-containing reagent **E** (iTag-azide) and treatment with SmE, SialEXO and OglyZOR. Sample clean-up and protease digestion was performed using S-TrapTM micro spin columns. The samples were vacuum-dried by Speedvac, reconstituted in 0.1% (v/v) formic acid in LC-MS grade water, and injected into an Orbitrap EclipseTM TribridTM Mass Spectrometer coupled to an UltiMateTM 3000 RSLCnano System by Lucia Di Vagno. All conditions tested in this experiment are shown in **Table 6**.

The raw MS data was processed using Byonic as described in **Section 3.3** but adding 217.1329 m/z instead of 287.1389 m/z to account for the chemical modification from the iTag-azide instead of the biotin-DADPS-picolyl azide¹⁸². The output excel files were filtered for CD46 (since many additional proteins were also detected due to the impure nature of the samples) and for specific cleavage sites. No glycopeptide hits satisfying these criteria were detected in any of the samples, however this wasn't totally unexpected since minimal streptavidin signal had been observed with the BH T4 CD46 C sample (lanes 5/6, **Fig. 40B-C**).

Table 6. Conditions tested to maximise the coverage of CD46 for MS-glycoproteomics. The sequence coverage achieved with each condition calculated by Byonic is also included in the table.

#	PNGaseF	CuAAC	SmE	OglyZOR	Trypsin	Chymotrypsin	Coverage (%)
1	✓	✓				✓	55
2	✓	✓	✓			✓	55
3	✓	✓		✓		✓	60
4	✓	✓			✓	✓	51
5	✓	✓	✓	✓		✓	49
6	✓	✓	✓		✓	✓	49
7	✓	✓		✓	✓	✓	50
8	✓	✓	✓	✓	✓	✓	51

Since the highest coverage was obtained with condition **3 (Table 6)**, I repeated the experiment with all the recombinant CD46 samples using this condition. We processed the raw MS data using Byonic with the human FASTA file downloaded from Uniprot²¹⁹ for the search. Between 70-99 proteins were detected in all the samples, especially different forms of keratin, highlighting the impurity of the samples. The output excel files were filtered for CD46 and for specific cleavage sites. No glycopeptide hits satisfying these criteria were detected in any of the samples.

The canonical CD46 sequence in Uniprot²¹⁹ is the sequence of the ABC1 isoform, which is predominantly expressed in EBV-B cells and leukemic cells²¹⁹ and contains an additional 15 amino acids (VLPPSSTKPPALSHS) in the STP domain²⁷². We therefore decided to repeat the raw data processing using Byonic with the FASTA files from the CD46 BC and C isoforms for the search instead. When analysing the CD46 C samples, we used the FASTA file from the CD46 C1 and C2 isoforms for the Byonic search. The output excel files were filtered for CD46 and specific cleavage sites. The Thr of the STP C domain appeared as a hit in all the CD46 C samples. However, when manual analysis of the MS spectra was performed with Xcalibur software by Lucia Di Vagno, none of the expected HCD signature ions due to the presence of the GalN6yne modified with the iTag-azide were detected, demonstrating that this hit was a false positive.

When analysing the CD46 BC samples, we used the FASTA file from the CD46 BC1 and BC2 isoforms for the Byonic search. The output excel files were filtered for CD46 and specific cleavage sites. No glycopeptide hits satisfying these criteria were detected in any of the samples. However, none of the samples detected the VSTSST peptide from the STP B domain, suggesting that the GalNAc-T4 glycosylation site could be in this region.

Since we had performed the workflow optimisation experiment with the BH T4 CD46 C sample, we wondered if the reason we missed the VSTSST peptide was due to unsuitable digestion conditions of the STP B domain. I therefore repeated the optimisation experiment using the control CD46 BC sample instead. The raw data was processed using Byonic with the FASTA files from the CD46 BC1 or BC2 isoforms for the search. Depending on the FASTA file used for the search, the results obtained were slightly different (**Table 7**). The coverage obtained in all samples was much lower compared to the previous experiment (**Table 6**). Nevertheless, the VSTSST sequence was detected with condition **4** and **7** (when using the BC2 FASTA file for the search). The output excel files were filtered for CD46 and for specific cleavage sites. No glycopeptide hits satisfying these criteria were detected in any of the samples.

Table 7. Conditions tested to maximise the coverage of CD46 for MS-glycoproteomics using the control CD46 BC construct and the CD46 BC1/BC2 FASTA files for the Byonic search. The sequence coverage achieved with each condition calculated by Byonic is also included in the table. The table also includes information on whether the VSTSST region of the STP B domain was detected or not.

#	PNGaseF	CuAAC	SmE	OglyZOR	Trypsin	Chymotrypsin	Coverage (%)		VSTSST?	
							BC1	BC2	BC1	BC2
1	✓	✓				✓	14	43	X	X
2	✓	✓	✓			✓	15	11	X	X
3	✓	✓		✓		✓	21	39	X	X
4	✓	✓			✓	✓	54	51	✓	✓
5	✓	✓	✓	✓		✓	4	4	X	X
6	✓	✓	✓		✓	✓	13	13	X	X
7	✓	✓		✓	✓	✓	36	49	X	✓
8	✓	✓	✓	✓	✓	✓	13	13	X	X

We wondered if the low abundance of the glycopeptide of interest and the impure nature of the samples was resulting in the glycopeptide of interest being suppressed by more abundant species. To overcome this issue, neutravidin bead enrichment could be performed to enrich the glycopeptide of interest. Since the abundant contaminant proteins were predominantly different forms of keratin, reduced sample handling should improve sample purity. Furthermore, since the second GFP-affinity purification did not substantially improve the purity of the sample, we decided to perform a single Ni-NTA purification. We also hypothesised that the glycopeptide of interest could be too hydrophilic due to its highly O-glycosylated nature to be released from the S-Trap™ micro spin columns when performing the sample clean-up. We therefore decided to use condition **7** for future experiments - not only did we obtain the second highest coverage and were able to detect the VSTSST region with this condition (**Table 7**), but also incorporating OglyZOR should help reduce the hydrophilicity of the glycopeptides by removing unmodified sugars.

5.1.3. Protein expression and Ni-NTA purification of recombinant CD46 BC

Since our results suggested that the CD46 BC isoform was the main substrate of GalNAc-T4, we decided to focus only on the BH T4 CD46 BC sample. I therefore expressed more protein and performed a single Ni-NTA purification (**Fig. 41A**). To minimise protein loss in the flow-through and washes, I excluded imidazole from the wash buffer. To minimise non-specific binding to the beads, I increased the salt concentration of the wash buffer from 125 mM to 300 mM. I also increased the number of elution steps to collect as much protein as possible. For this experiment, I performed the Ni-NTA purification from both the secretome and the lysate of transduced cells fed with Ac₄GalN6yne to increase the yield of tagged recombinant CD46. The expected MW of the CD46 construct was 66 kDa. A band around this MW was observed in the secretome samples but no clear band was observed in any of the lysate samples (**Fig. 41B**). Nevertheless, the eluted fractions (E1-E5) were combined, and the protein concentration of the samples purified from the lysate and the secretome was determined by Nanodrop (**SI Table 8**). Due to technical issues, I couldn't determine the protein concentration in the sample purified from the secretome. Since the samples purified from the secretome contained contaminations of what could be albumin which could disturb downstream analyses (**Fig. 41B**), we focused on the sample purified from the lysate instead.

To confirm whether recombinant CD46 BC was tagged by GalNAc-T4, I treated the recombinant CD46 sample purified from the lysate with PNGase F to remove N-glycans, performed a CuAAC reaction with biotin-picolyl azide, and assessed the tagging of recombinant CD46 by streptavidin blot (Fig. 41C). Protein expression control was simultaneously performed by Western blot. When staining against CD46, a strong band was observed at the expected MW of the construct (66 kDa)(Fig. 41D), confirming the presence of recombinant CD46 in the sample. A streptavidin signal was seen at the same MW as the CD46 bands (Fig. 41D), suggesting that CD46 was tagged by BH GalNAc-T4. Since the streptavidin signal overlapped with the CD46 bands (Fig. 41D), incubation with neutravidin beads should selectively enrich for CD46, helping overcome the impurity of the sample.

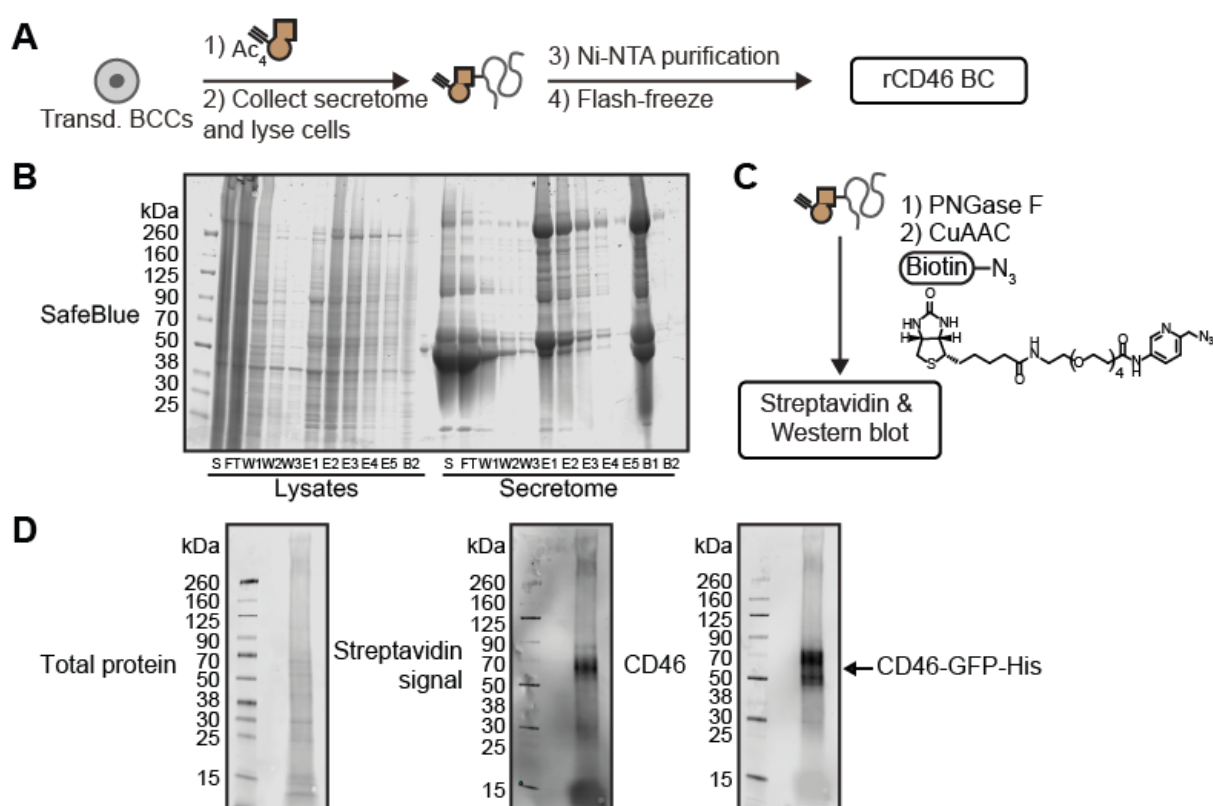


Fig. 41. Protein expression and Ni-NTA purification of recombinant CD46 BC. **A)** Schematic representation of the protein expression and Ni-NTA purification of recombinant CD46 BC. rCD46 = recombinant CD46. **B)** SDS-PAGE gel of the fractions after Ni-NTA purification of BH T4 CD46 BC. S = supernatant; FT = flow-through; W1-3 = wash 1-3; E1-5 = elution 1-5; B1 = beads before elution; B2 = beads after elution. **C)** Schematic representation of the tagging and protein expression control experiment. **D)** Tagging and protein expression control by Western blot after the eluted fractions were subjected to PNGase F treatment and CuAAC reaction with biotin-picolyl azide. The stainings were performed on the same membrane in the order displayed. The expected MW of the CD46 BC construct is 66 kDa (arrow pointing). 5 µg of protein were used per lane. Staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230) and an antibody against CD46 (1:500, 12494-1-AP). Data are from one experiment.

5.1.4. MS-glycoproteomics experiment on recombinant CD46 BC

To determine the GalNAc-T4-mediated glycosylation site(s) of CD46 BC, I incubated the recombinant CD46 BC sample with PNGase F to remove N-glycans, followed by CuAAC reaction with biotin-DADPS-picolyl azide for 72 h (**Fig. 42A**). Sample clean-up and buffer exchange to PBS was subsequently performed using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO), followed by SialEXO and OglyZOR glycosidase treatment to remove unmodified sugars whilst leaving the modified sugar intact (as demonstrated in **Fig. 40B-C**). Enrichment of the tagged recombinant CD46 was performed with neutravidin beads (**Fig. 42A**).

Aliquots of the same volume were taken at different steps of the workflow to track the tagged recombinant CD46 by Western blot. The intensity of the total protein stain from the supernatant after enrichment was lower compared to the other samples (**Fig. 42B**), suggesting unequal loading of the samples on the gel. When staining against CD46, a strong band was observed at the expected MW of the construct (66 kDa)(**Fig. 42B**), confirming the presence of recombinant CD46 in the samples. The CD46 bands were shifted to a lower MW after SialEXO and OglyZOR glycosidase treatment (**Fig. 42B**), confirming the successful removal of O-glycans. A band due to CD46 was detected in the supernatant after enrichment (**Fig. 42B**), suggesting the presence of untagged CD46 in the sample. This implied that the tagging of recombinant CD46 by CuAAC was not quantitative, as suggested by previous work in the field^{279, 280}.

A streptavidin signal was observed after the CuAAC reaction (**Fig. 42B**), demonstrating that the reaction was successful. The streptavidin signal was retained after SialEXO and OglyZOR glycosidase treatment whilst the bands were shifted to a lower MW (**Fig. 42B**), confirming the removal of unmodified sugars whilst the modified sugar was left intact. No streptavidin signal was seen in the supernatant after enrichment (**Fig. 42B**), confirming that the tagged proteins remained bound to the beads. The streptavidin signal overlapped with the CD46 bands (**Fig. 42B**), suggesting that recombinant CD46 was tagged by BH GalNAc-T4.

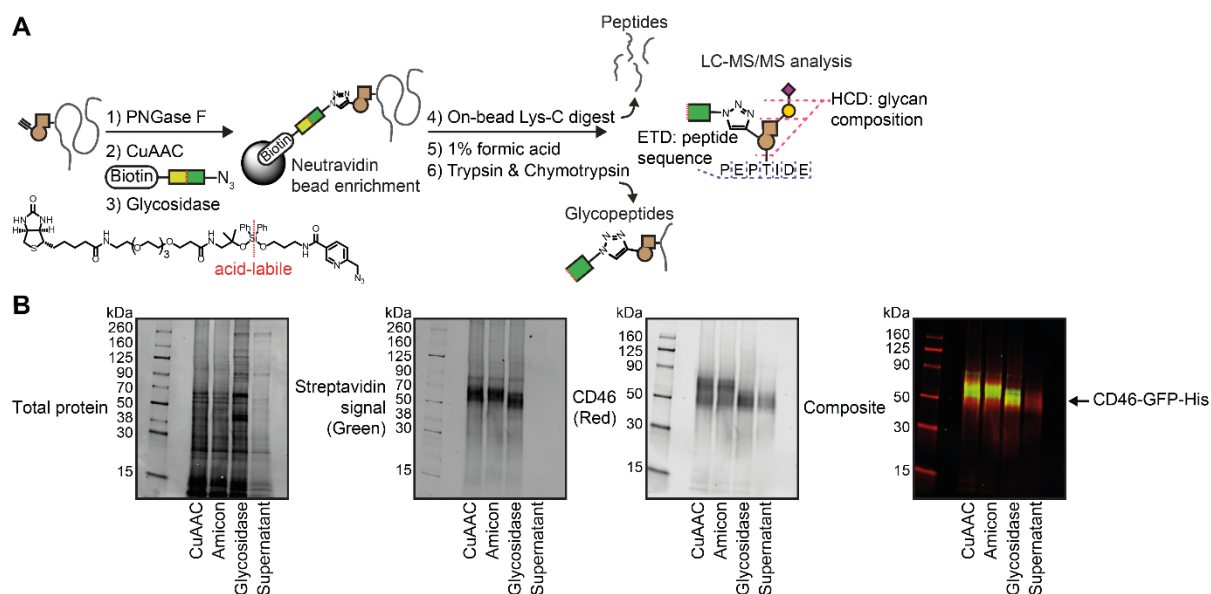


Fig. 42. Determining the GalNAc-T4-mediated glycosylation site(s) on recombinant CD46 BC by MS-glycoproteomics. A) Schematic representation of the workflow for the MS-glycoproteomics experiment on recombinant CD46 BC. **B)** Tracking tagged recombinant CD46 during the MS-glycoproteomics experiment by Western blot. The stainings were performed on the same membrane in the order displayed. Red/Green refers to the imaging channel. The expected MW of the CD46 BC construct is 66 kDa (arrow pointing). Staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230) and an antibody against CD46 (1:500, 12494-1-AP). Data are from one experiment. Glycosidase = SialEXO and OglyZOR.

After neutravidin bead enrichment, the unmodified peptides were released from the beads by on-bead digestion with Lys-C. The tagged glycopeptides were eluted from the beads by acid treatment, digested with trypsin and chymotrypsin, desalted using UltraMicroSpin™ columns and vacuum-dried by Speedvac. The samples were reconstituted in 0.1% (v/v) formic acid in LC-MS grade water, and injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The raw MS data was first processed using Byonic with the human FASTA file for the search to gain insight into the purity of the sample. 232 proteins were detected in this sample compared to the 88 detected in the previous experiment. In the previous experiment the samples had been purified by Ni-NTA and GFP-affinity purification, whilst in this experiment they were purified by a single Ni-NTA purification followed by neutravidin bead enrichment of the tagged glycoproteins. We would have expected selective enrichment of recombinant CD46 since the strongest streptavidin signal overlapped with the CD46 bands (**Fig. 42B**), suggesting that other proteins were binding non-specifically to the neutravidin beads. The coverage of the CD46 sequence was also much lower in this experiment regardless of the FASTA file used for the search. When using the human FASTA, the coverage was 11%, when using the CD46 BC1 FASTA, the coverage was 9%, and when using the CD46 BC2 FASTA the coverage was 31%, whilst in the previous experiment the lowest coverage was 59% when using the human FASTA.

The output excel files were filtered for CD46 and for specific cleavage sites. No glycopeptide hits satisfying these criteria were detected, nor was the VSTSST region detected in any of the searches. In the previous experiment, we used the iTag-azide reagent to derivatise the tagged glycoproteins by CuAAC, whilst in this experiment we used the biotin-DADPS-picolyl azide reagent instead to allow enrichment of the tagged glycoproteins. As seen in **Chapter 4**, the positive charge imparted by the iTag facilitated ETD fragmentation, which enhanced the peptide sequence coverage. This could explain why a lower sequence coverage was seen in this experiment compared to the previous experiment. Furthermore, the greater number of contaminant proteins detected in this experiment could lead to the glycopeptide of interest being suppressed by more abundant species. Alternatively, the glycopeptide of interest may be retained in the columns used for peptide desalting. Furthermore, no glycopeptide hits were detected in any of the experiments, even when we detected the VSTSST peptide in the optimisation experiment using the control CD46 BC sample, suggesting that alternative digestion conditions may be required to detect the glycopeptide of interest. Further optimisation would therefore be required to obtain the GalNAc-T4-mediated glycosylation site in CD46.

5.2. Identifying the GalNAc-T4-mediated glycosylation site(s) in GOLM1

Due to the difficulties in determining the GalNAc-T4-mediated glycosylation site in CD46, we decided to shift our focus to GOLM1 instead. The GOLM1 sequence contains 34 Thr and Ser residues as potential O-GalNAc glycosylation sites but no mucin-like domains, which should facilitate MS-glycoproteomics analyses on GOLM1 compared to CD46. Furthermore, GOLM1 has 52 tryptic sites compared to the 27 tryptic sites in CD46, which should also allow for greater sequence coverage of GOLM1 compared to CD46 without requiring an additional protease like chymotrypsin. GOLM1 is a type II transmembrane protein predominantly expressed by epithelial cells^{245, 281}. It consists of a short N-terminal cytoplasmic domain, followed by a transmembrane domain, a coiled-coil domain, a flexible loop, and a C-terminal luminal domain (**Fig. 43**)^{245, 282}. GOLM1 is N-glycosylated at Asn¹⁰⁹, Asn¹⁴⁴ and Asn³⁹⁸ – the first two N-glycosites are in the coiled-coil domain whilst the third is in the C-terminal luminal domain^{245, 282}. This protein resides in the cis- and medial-Golgi but also undergoes endosomal trafficking, which results in GOLM1 being proteolytically cleaved by furin and the luminal portion secreted to the extracellular space²⁸³. GOLM1 exists as a disulfide-bonded dimer²⁴⁵. The exact function of GOLM1 has not been clearly defined, but it is thought to play a role in the Golgi structure²⁸⁴ and in the secretory pathway - regulating expression, trafficking, secretion, and recycling of proteins^{282, 285}.

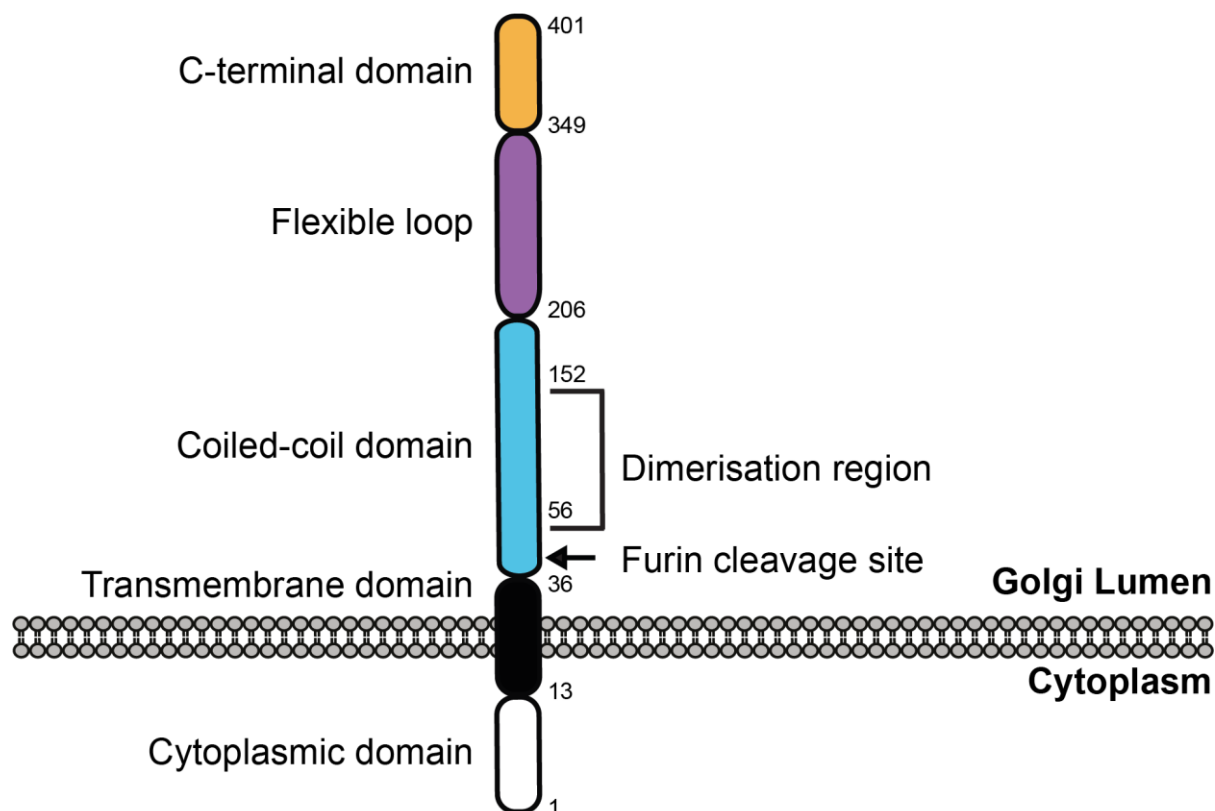


Fig. 43. Structure of GOLF1. GOLF1 consists of a short N-terminal cytoplasmic domain (in white), followed by a transmembrane domain (in black), a coiled-coil domain (in blue), a flexible loop (in purple), and a C-terminal luminal domain (in orange). GOLF1 has a furin cleavage site at Arg⁵⁵. The region from Ala⁵⁶ to Phe¹⁵² within the coiled-coil domain mediates the dimerisation of endogenous and secreted GOLF1. The dimer is stabilised by disulfide bonds at Cys¹⁵⁹ and Cys¹⁷⁰. Three N-glycosylation sites have been reported at Asn¹⁰⁹, Asn¹⁴⁴ and Asn³⁹⁸ and two potential O-glycosylation sites have been predicted at Thr²¹⁸ and Thr²³⁵. The numbers represent the amino acid positions. Figure adapted from Hu et al. (2011)²⁴⁵.

Previous work on GOLF1 has demonstrated that, since GOLF1 lacks a signal sequence²⁸⁶, the transmembrane domain and a positively charged residue at the cytoplasmic end are necessary for Golgi localisation, and that Golgi localisation is crucial for its proper glycosylation status²⁴⁵. I therefore transduced T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells ("BH T4" cells) with viruses encoding full-length GOLF1 tagged with GFP and a His-tag (**Fig. 36A**). The sequence of the GOLF1 construct used in this experiment is available in **SI Fig. 15**. Transduction success was confirmed by checking for GFP expression using a fluorescence microscope (**SI Fig. 16**).

5.2.1. Protein expression and Ni-NTA purification of recombinant GOLM1

Ni-NTA purification was performed on lysates of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells transduced with tagged GOLM1 and fed with Ac₄GalN6yne (**Fig. 44A**). Aliquots of the same volume were taken at different steps of the purification to track the presence of GOLM1 by SDS-PAGE. The expected MW of the GOLM1 construct was 74 kDa. No clear band was observed at this MW in any of the samples (**Fig. 44B**). Nevertheless, the eluted fractions (E1-E5) were combined, and the protein concentration was determined by Nanodrop (**SI Table 8**).

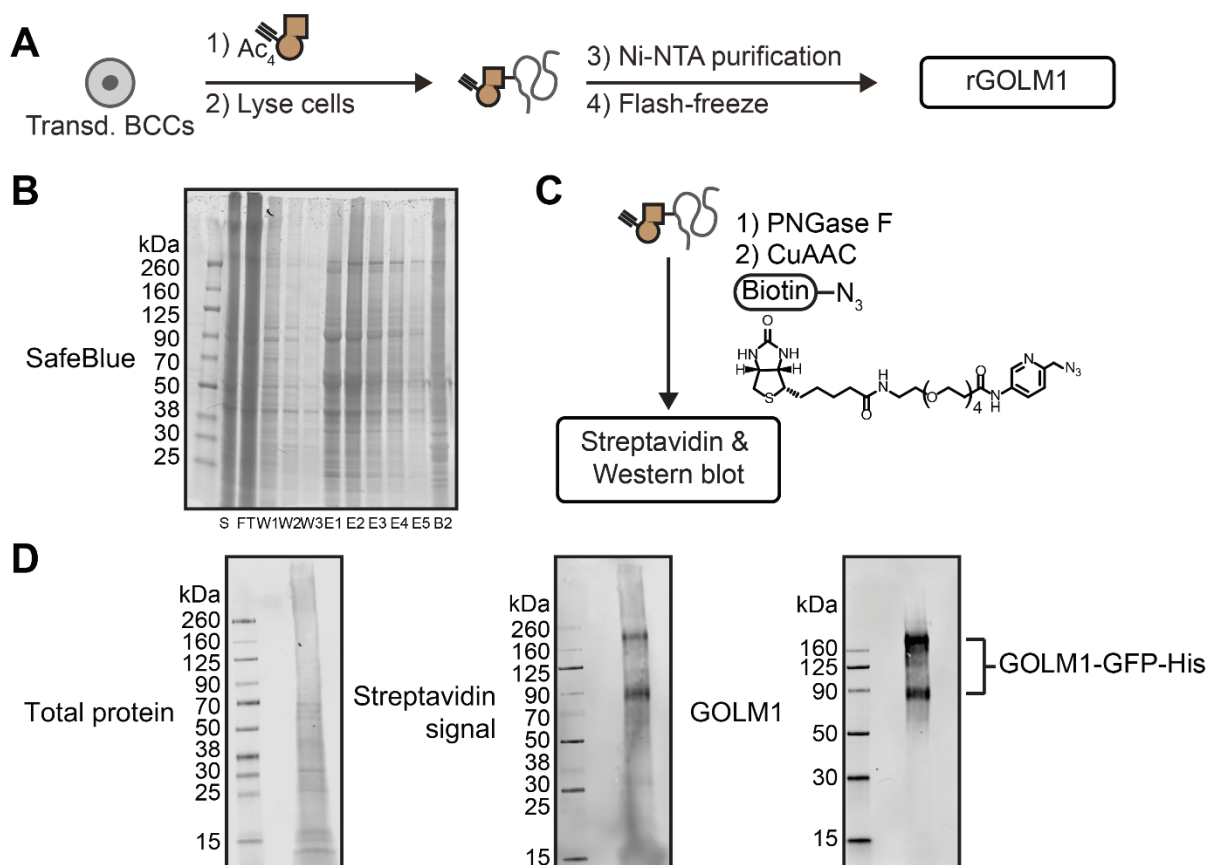


Fig. 44. Protein expression and Ni-NTA purification of recombinant GOLM1. **A)** Schematic representation of the protein expression and Ni-NTA purification of recombinant GOLM1. rGOLM1 = recombinant GOLM1. **B)** SDS-PAGE gel of the fractions after Ni-NTA purification of recombinant GOLM1. S = supernatant; FT = flow-through; W1-3 = wash 1- 3; E1-5 = elution 1-5; B2 = beads after elution. **C)** Schematic representation of the tagging and protein expression control experiment. **D)** Tagging and protein expression control by Western blot after the eluted fractions were subjected to PNGase F treatment and CuAAC reaction with biotin-picolyl azide. The stainings were performed on the same membrane in the order displayed. The expected MW of the recombinant GOLM1 monomer and dimer is 74 and 148 kDa respectively (arrows pointing). 5 µg of protein were used per lane. Staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230) and an antibody against GOLM1 (1:500, HPA010638). Data are from one experiment.

To confirm whether recombinant GOLM1 was tagged by GalNAc-T4, I treated the recombinant GOLM1 sample with PNGase F to remove N-glycans, performed a CuAAC reaction with biotin-picolyl azide, and assessed the tagging of recombinant GOLM1 by streptavidin blot (**Fig. 44C**). Protein expression control was simultaneously performed by Western blot. When staining against GOLM1 intense bands were observed between 50-90 kDa and between 125-160 kDa due to the presence of the GOLM1 monomer (expected MW = 74 kDa) and dimer (expected MW = 148 kDa) in the sample (**Fig. 44D**)²⁴⁵. The streptavidin signal overlapped with the GOLM1 bands (**Fig. 44D**), suggesting that recombinant GOLM1 was tagged by BH GalNAc-T4

5.2.2. MS-glycoproteomics experiment on recombinant GOLM1

To determine the GalNAc-T4-mediated glycosylation site of GOLM1, I incubated the recombinant GOLM1 sample with PNGase F to remove N-glycans, followed by CuAAC reaction with biotin-DADPS-picolyl azide for 72 h (**Fig. 45A**). Sample clean-up and buffer exchange to PBS was subsequently performed using Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO), followed by SialEXO and OglyZOR glycosidase treatment and enrichment of the tagged recombinant GOLM1 with neutravidin beads (**Fig. 45A**). Aliquots of the same volume were taken at different steps of the workflow to track tagged recombinant GOLM1 by Western blot. The results obtained were comparable to those obtained previously for CD46 (**Fig. 42B**). The intensity of the total protein stain from the supernatant after enrichment was lower compared to the other samples (**Fig. 45B**), suggesting unequal loading of the samples on the gel. When staining against GOLM1, strong bands at the expected MW of the monomer (74 kDa) and dimer (148 kDa) were observed (**Fig. 45B**), confirming the presence of recombinant GOLM1 in the samples. The GOLM1 bands were shifted to a lower MW after SialEXO and OglyZOR glycosidase treatment (**Fig. 45B**), confirming the successful removal of O-glycans. Bands due to GOLM1 were detected in the supernatant after enrichment (**Fig. 45B**), suggesting the presence of untagged GOLM1 in the sample, and that the tagging of recombinant GOLM1 by CuAAC was not quantitative. A streptavidin signal was observed after the CuAAC reaction (**Fig. 45B**), demonstrating that the reaction was successful. The streptavidin signal was retained after SialEXO and OglyZOR glycosidase treatment whilst the bands were shifted to a lower MW (**Fig. 45B**), confirming the removal of unmodified sugars whilst the modified sugar was left intact. No streptavidin signal was seen in the supernatant after enrichment (**Fig. 45B**), confirming that the tagged proteins remained bound to the beads. The streptavidin signal overlapped with the GOLM1 bands (**Fig. 45B**), suggesting that recombinant GOLM1 was tagged by BH GalNAc-T4. The strongest streptavidin signal was observed in the monomer bands (**Fig. 45B**), suggesting that dimerisation of recombinant GOLM1 could be disrupted when modified with the bulky biotin-DADPS-picolyl azide.

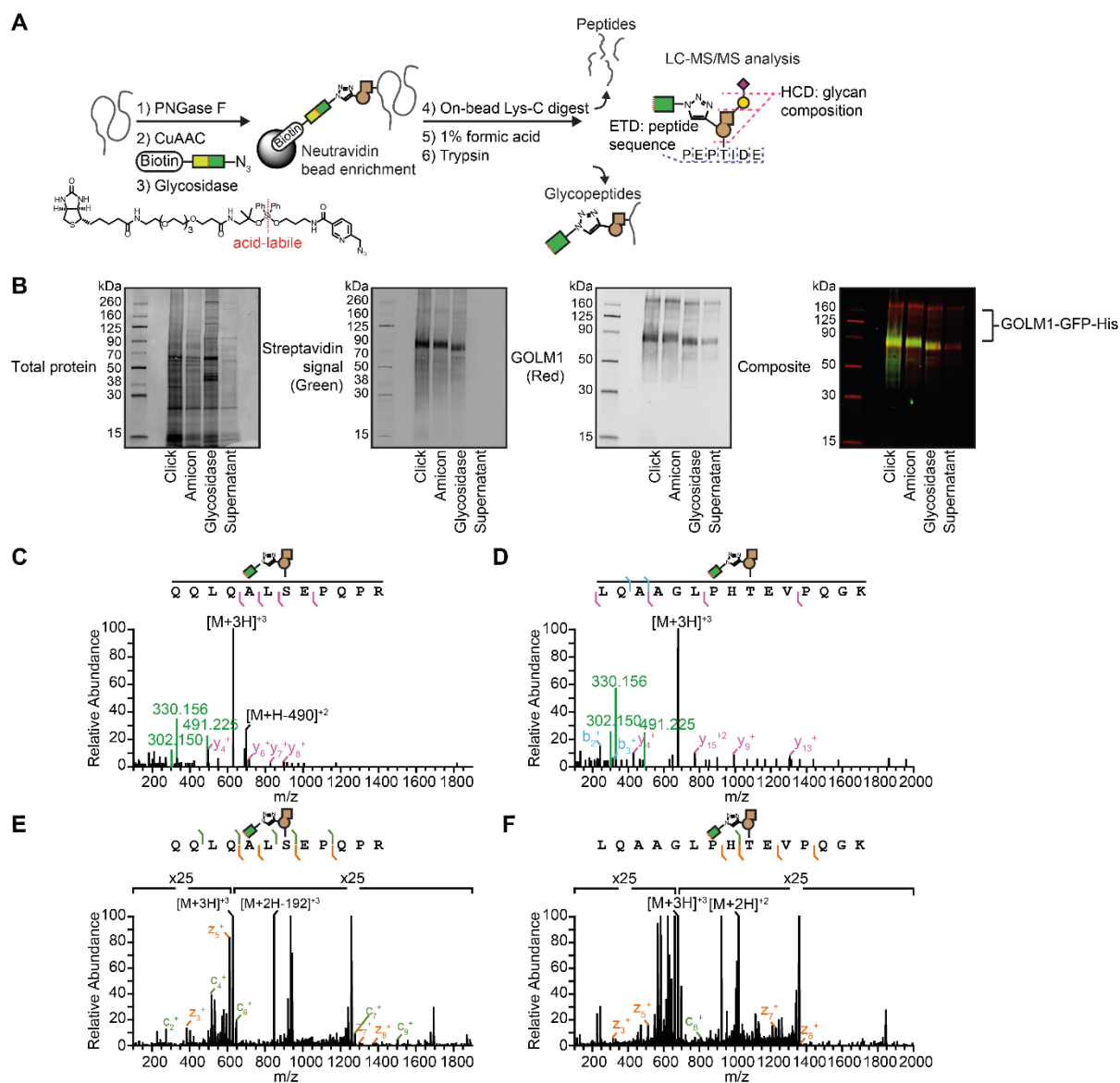


Fig. 45. Determining the GalNAc-T4-mediated glycosylation site(s) on recombinant GOLM1 by MS-glycoproteomics. **A**) Schematic representation of the workflow for the MS-glycoproteomics experiment on recombinant GOLM1. Glycosidase = SialEXO and OglyZOR. **B**) Tracking tagged recombinant GOLM1 during the MS-glycoproteomics experiment by Western blot. The stainings were performed on the same membrane in the order displayed. Red/Green refers to the imaging channel. The expected MW of the GOLM1 construct is 74 kDa (arrow pointing). Staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230) and an antibody against GOLM1 (1:500, HPA010638). HCD spectrum of the GOLM1 glycopeptide at **B**) Ser²⁰⁴ and **C**) Thr²¹⁸. B ions are represented in blue, y ions in pink, and unique HCD oxonium ions in green. ETD spectrum of the GOLM1 glycopeptide at **D**) Ser²⁰⁴ and **E**) Thr²¹⁸. C ions are represented in green, and z ions in orange. Data are from one experiment.

After neutravidin bead enrichment, the unmodified peptides were released from the beads by on-bead digestion with Lys-C. The tagged glycopeptides were eluted from the beads by acid treatment, digested with trypsin, desalted using UltraMicroSpin™ columns and vacuum-dried by Speedvac. The samples were reconstituted in 0.1% (v/v) formic acid in LC-MS grade water, and injected into an Orbitrap Eclipse™ Tribrid™ Mass Spectrometer coupled to an UltiMate™ 3000 RSLCnano System by Lucia Di Vagno. The raw MS data was processed using Byonic with the human and GOLM1 FASTA files for the search. Comparable results were obtained with both searches. We would have expected selective enrichment of recombinant GOLM1 since the strongest streptavidin signal overlapped with the GOLM1 bands (**Fig. 45B**) – instead 211 proteins were detected in the sample, suggesting that other proteins were binding non-specifically to the neutravidin beads. The output excel files were filtered for GOLM1 and for specific cleavage sites. Ser²⁰⁴ and Thr²¹⁸ were obtained as potential BH GalNAc-T4-modified glycosylation sites – one spectrum was found for the former and two for the latter corresponding to two different glycan compositions (HexNAc(1) and HexNAc(1)Hex(1)). Manual analysis of the MS spectra was performed with Xcalibur software by Lucia Di Vagno to validate the glycosite hits. The expected oxonium ions at 491.225 m/z, 330.156 m/z and 302.150 m/z^{182, 193} due to glycopeptides modified with GalN6yne and derivatised with biotin-DADPS-picolyl azide were observed in all the HCD spectra (indicated in green, **Fig. 45C-D**), confirming unambiguously that the peptides were glycosylated. The presence of the unique HCD oxonium ions triggered ETD on the glycopeptides and allowed GalNAc-T4-glycosylation site localisation in GOLM1 (**Fig. 45E-F**).

5.2.3. *In silico* evaluation of GalNAc-T4-mediated glycosylation of GOLM1

To explore a possible structural role of GalNAc-T4-mediated glycosylation on GOLM1, we performed *in silico* evaluation on the molecular structure of GOLM1. Since no crystal structure has been published for full-length GOLM1, we obtained the predicted molecular structure from the AlphaFold Protein Structure Database^{287, 288}(**Fig. 46**). However, the confidence score calculated by AlphaFold from Asn¹⁸² to Leu⁴⁰¹ was low, which made it challenging to evaluate a possible structural role of GalNAc-T4 using this molecular structure.

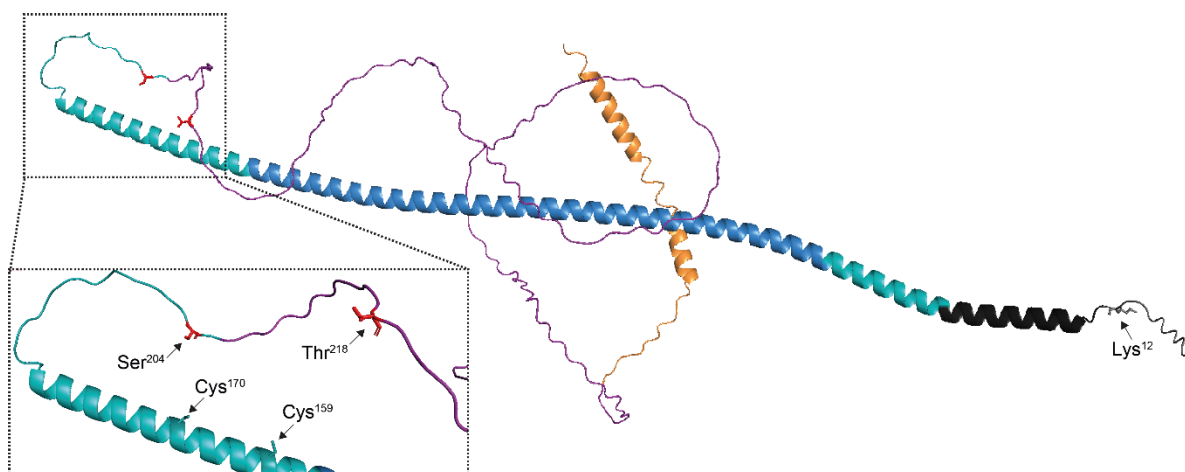


Fig. 46. *In silico* evaluation of GalNAc-T4-mediated glycosylation of GOLF1. Predicted molecular structure of GOLF1 obtained from AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk>)²⁸⁸. The N-terminal cytoplasmic domain is shown in grey, the transmembrane domain in black, the coiled-coil domain in teal (with the region mediating dimerisation in blue), the flexible loop in purple, the C-terminal luminal domain in orange and the two GalNAc-T4-mediated glycosylation sites in red (Ser²⁰⁴ and Thr²¹⁸). Lys¹² is positively charged residue at the cytoplasmic end of the transmembrane domain which is required for Golgi localisation. The GOLF1 dimer is stabilised by disulfide bonds at Cys¹⁵⁹ and Cys¹⁷⁰.

Overall, we successfully identified CD46 and GOLF1 as novel GalNAc-T4 substrates conserved in all the BCCs tested herein by SILAC-based MS-proteomics. Whilst we couldn't pinpoint the exact GalNAc-T4-mediated glycosylation site(s) in CD46, we established that the glycosylation site of GalNAc-T4 in CD46 must be located within the STP B domain. Ser²⁰⁴ and Thr²¹⁸ were successfully identified as the GalNAc-T4-mediated glycosylation sites in GOLF1 by MS-glycoproteomics and we detected elaboration of the Thr²¹⁸ site up to the disaccharide Hex-GalN6yne (where Hex is likely a Gal residue typical of core 1 O-GalNAc glycans). Additional work will be required to confirm the GalNAc-T4-mediated glycosylation site(s) in CD46 and to evaluate a possible structural role of GalNAc-T4-mediated glycosylation on GOLF1.

CHAPTER 6: Developing BH GalNAc-T systems *in vivo*

The experiments described up to this point were all performed in one-dimensional cellular model systems. However, these model systems cannot recapitulate the multifactorial nature of cancer involving the constant interplay between tumour and host cells²⁰⁴. We therefore endeavoured to extend the BH GalNAc-T4 system to an *in vivo* setting as a more biologically-relevant model system to investigate the role of GalNAc-T4 and its substrates in cancer initiation, progression, and metastasis. We first established the model system on a different isoenzyme, GalNAc-T2, as a well-characterised glycosyltransferase with a rich glycoprotein substrate repertoire^{193, 204}.

6.1. Establishing a BH GalNAc-T2 system *in vivo*

The work described in this section has been published in Cioce et al. (2022)²⁰⁴.

Dr Anna Cioce had previously generated GFP+ 4T1 murine metastatic BCCs expressing the two enzymes of the artificial biosynthetic pathway for UDP-sugar biosynthesis (AGX1^{F383A} and NahK) and BH GalNAc-T2 (henceforth referred to as “BH T2” cells)²⁰⁴. These cells were injected into the fat pad of female NSG mice by Dr Victoria Bridgeman, with mock-transfected GFP+ 4T1 cells injected into the other fat pad as an internal control (**Fig. 47A**)²⁰⁴. Tumour growth was monitored by Dr Victoria Bridgeman²⁰⁴. Eleven days after tumour cell injection, the mice harbouring the biggest tumours were treated intratumourally (i.t.) with Ac₄GalN6yne or the corresponding volume of vehicle once daily for 3 days by Dr Victoria Bridgeman²⁰⁴. Thirteen days after tumour cell injection, the remaining mice were treated intraperitoneally (i.p.) with Ac₄GalN6yne, Ac₄ManNAI or the corresponding volume of vehicle once daily for 5 days by Dr Victoria Bridgeman²⁰⁴. At the end of the sugar treatment, the tumours were harvested by Dr Victoria Bridgeman and Dr Tatiana Rizou²⁰⁴. Half of the tumour was digested, plated, and the cells cultured for protein expression control by Western blot by Dr Anna Cioce²⁰⁴. The other half was homogenised, incubated with neutravidin beads to remove endogenous biotinylated proteins, and reacted with biotin-picolyl azide²⁰⁴. Tagging of glycoproteins was assessed by streptavidin blot²⁰⁴.

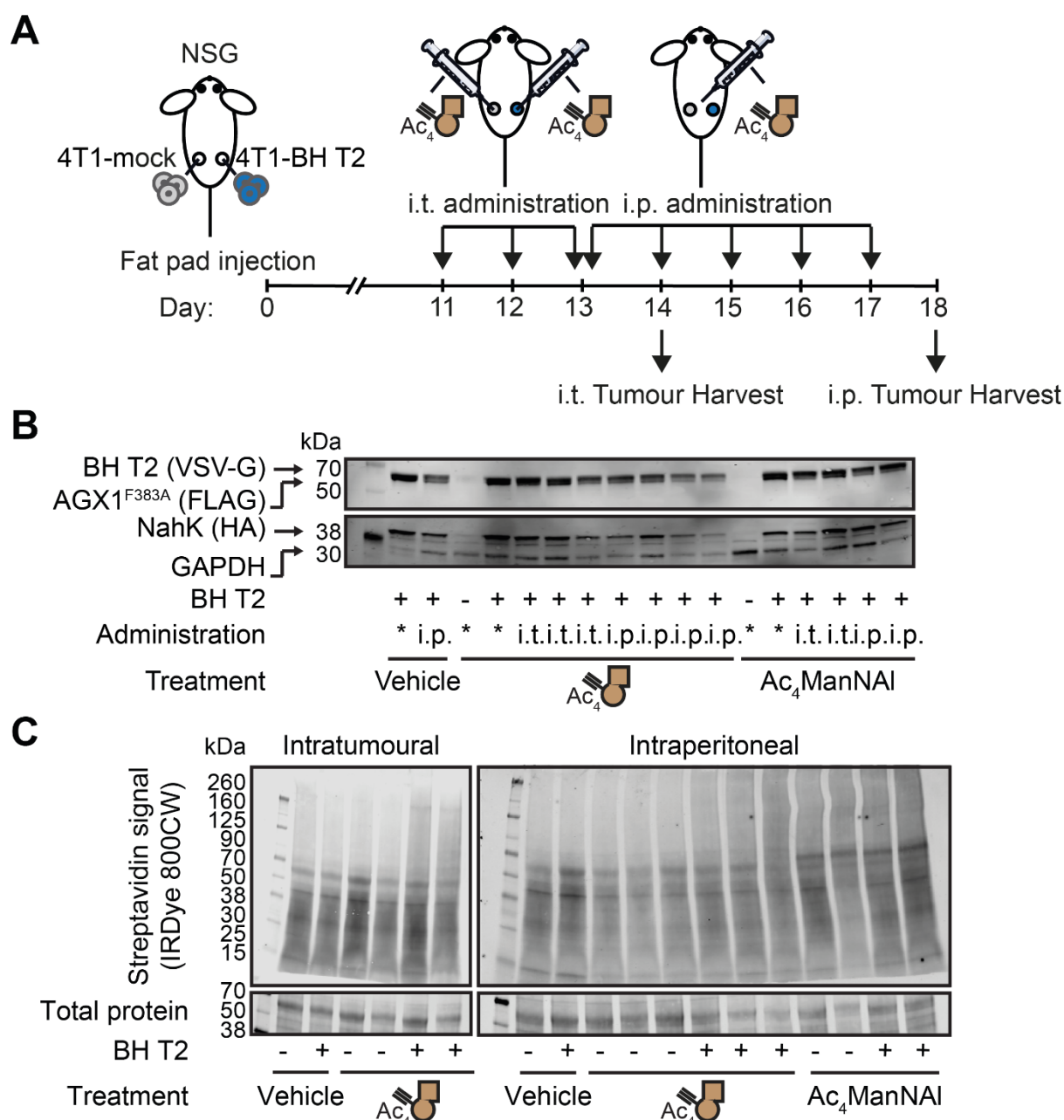


Fig. 47. Establishing a BH GalNAc-T2 system *in vivo*. **A**) Schematic representation of the workflow for the *in vivo* tagging of glycoproteins using BH GalNAc-T2 (BH T2). BH T2 tumours were grown in the fat pad of NSG mice. Mock tumours were grown in the opposite fat pad from the same mouse as an internal control. After allowing for the tumour to be established, mice were i.t. or i.p. injected with either vehicle, Ac₄GalNAc6S or Ac₄ManNAc by Dr Victoria Bridgeman. At the end of the sugar treatment, the tumours were harvested by Dr Victoria Bridgeman and Dr Tatiana Rizou. Figure adapted from Cioce et al. (2022)²⁰⁴. **B**) Protein expression analysis of the digested BH T2 tumours by Western blot (by Dr Anna Cioce). BH T2 or mock-transfected parental 4T1 cell lines were included as a control and are indicated by an asterisk (*). GAPDH was included as a loading control. The predicted MW of GalNAc-T2, AGX1, NahK and GAPDH is 65, 58, 41 and 36 kDa respectively. 15 µg of protein were used per lane. Antibody staining was performed using antibodies against VSV-G (1:2000, ab3861), FLAG (1:1000, PA1-984B), HA (1:1000, ab9110) and GAPDH (1:5000, ab181602). Figure adapted from Cioce et al. (2022)²⁰⁴. **C**) Tagging of glycoproteins in tumour homogenates as assessed by streptavidin blot after fat pad injection of 4T1-BH T2 or 4T1-mock transfected cells into NSG mice. 55 µg of protein were used per lane. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). 4T1-BH T2 (+) = 4T1-AGX1^{F383A}-NahK-GalNAc-T2^{BH} cells, 4T1-mock (-) = mock-transfected 4T1 cells.

Protein expression control performed on the digested tumours by Dr Anna Cioce confirmed the expression of the transgenes (**Fig. 47B**)²⁰⁴. The expression levels of the enzymes were comparable to those of parental cells (**Fig. 47B**). However, when assessing tagging of glycoproteins in the tumour homogenates by streptavidin blot, comparable signal was seen in all samples regardless of the treatment and the expression of BH GalNAc-T2 (**Fig. 47C**). Since the samples had been incubated with neutravidin beads prior to the CuAAC reaction, no endogenous biotinylated proteins should be present that would give rise to background signal. We wondered whether one of the components of the CuAAC reaction could be the source of the non-specific signal instead. Dr Anna Cioce therefore incubated lung homogenates of untreated control mice with the different components of the CuAAC reaction. A signal was observed in the samples incubated with an aqueous solution of biotin-picolyl azide even when all other components were absent (**SI Fig. 17C**), demonstrating that this probe was the source of the non-specific signal.

I next investigated the mechanism behind the background signal by biotin-picolyl azide. One hypothesis was that free thiols from cysteine-containing proteins could be interacting non-specifically with the probe. To test this, I incubated vehicle-treated (negative control) and Ac₄ManNAI-treated samples (positive control) with iodoacetamide before the CuAAC reaction to block any free thiols in the sample. To prevent the formation of free thiols after the reduction of disulfide bonds, I also excluded DTT from the loading buffer and skipped the boiling step. Minimal signal was seen in the vehicle samples whilst a strong signal was observed in the Ac₄ManNAI-treated samples when DTT and the boiling step were excluded (lanes 3 and 4; **Fig. 48A**). Comparable results were observed when the samples were incubated with iodoacetamide prior to the CuAAC reaction (lanes 1 and 2; **Fig. 48A**), as well as when DTT was excluded but the boiling step was maintained (lanes 5 and 6; **Fig. 48A**). Higher background signal was observed in the vehicle samples when DTT was included, and the boiling step excluded (lanes 7 and 8; **Fig. 48A**). Altogether, these results demonstrated that the addition of DTT and boiling of the samples were likely the cause of the background signal by biotin-picolyl azide, possibly due to a non-specific interaction between the probe and free thiols formed from reducing disulfide bonds.

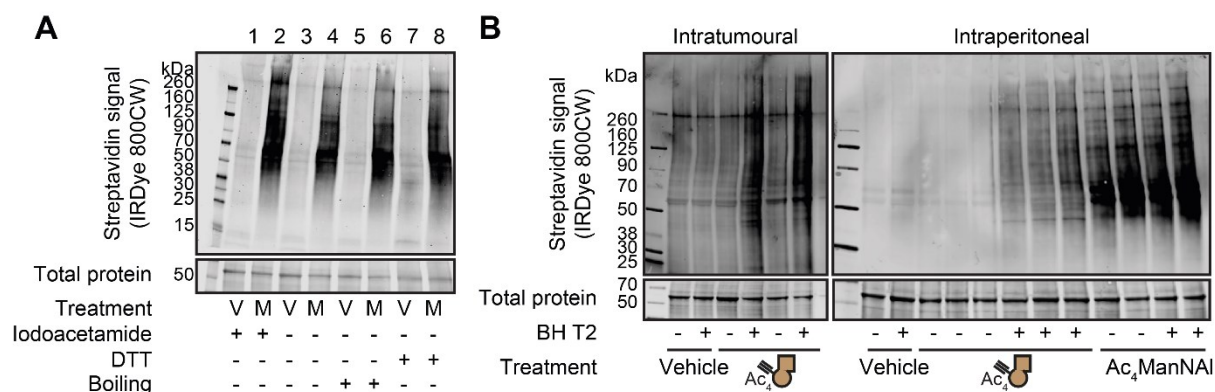


Fig. 48. Optimisation of *ex vivo* tagging analysis of BH GalNAc-T2 tumours by streptavidin blot. A) Identifying the source of background signal in BH GalNAc-T2 tumours. 40 µg of protein from tumours as depicted in **Fig. 47** were used per lane. V = vehicle; M = Ac₄ManNAI. **B)** Glycoprotein tagging analysis of tumour homogenates by streptavidin blot after fat pad injection of 4T1-BH T2 or 4T1-mock transfected cells into NSG mice (by Dr Anna Cioce). 15 µg of protein were used per lane. Figure adapted from Cioce et al. (2022)²⁰⁴. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). 4T1-BH T2 (+) = 4T1-AGX1^{F383A}-NahK-GalNAc-T2^{BH} cells, 4T1-mock (-) = mock-transfected 4T1 cells.

Having identified the source of the background signal by biotin-picolyl azide, Dr Anna Cioce repeated the streptavidin blot of the tumour homogenates excluding the boiling step and DTT. Minimal signal was seen in the vehicle-treated samples (**Fig. 48B**)²⁰⁴. In contrast, a strong signal was seen in the homogenates from mice treated with Ac₄ManNAI (**Fig. 48B**)²⁰⁴. In mice treated with Ac₄GalN6yne, signal was only observed when the tumours expressed BH GalNAc-T2 (**Fig. 48B**)²⁰⁴. The results from both methods of compound administration were comparable, although a higher background was seen in the intratumoural samples (**Fig. 48B**)²⁰⁴. Overall, the results demonstrated that we could successfully and selectively tag glycoproteins by BH GalNAc-T2 *in vivo*²⁰⁴.

6.2. Extending the BH GalNAc-T *in vivo* system to GalNAc-T4

The next objective was to extend the BH GalNAc-T *in vivo* approach to GalNAc-T4. T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells ("BH T4" cells) were injected into the fat pad of female NSG mice by Dr Victoria Bridgeman, with T47D^{WT}-AGX1^{F383A}-NahK cells injected into the other fat pad as an internal control ("control" cells)(**Fig. 49A**). When the tumours reached ~150 mm³, the mice were i.p. injected with either vehicle, Ac₄GalN6yne, or Ac₄ManNAI by Dr Victoria Bridgeman. In the experiment with BH GalNAc-T2, the dose of both sugars was 200 mg/kg²⁰⁴. In this experiment with BH GalNAc-T4 we performed a dose-response experiment and tested 3 doses of Ac₄GalN6yne: 200, 100 and 50 mg/kg. Since in the experiment with BH GalNAc-T2 a stronger signal was observed in the samples treated with Ac₄ManNAI compared to the samples treated with Ac₄GalN6yne (**Fig. 48B**)²⁰⁴, we chose an intermediate Ac₄ManNAI dose of 100 mg/kg for this experiment.

In the experiment with BH GalNAc-T2, we had to digest the tumours and culture the cells *ex vivo* to perform protein expression control since we struggled to perform protein expression control on the tumour homogenates²⁰⁴. In this experiment with BH GalNAc-T4, at the end of the sugar treatment, the tumours were harvested, digested, sorted through a MACS column for human CD298 and the positive selection collected and lysed by Dr Victoria Bridgeman (**Fig. 49A**). CD298 was chosen for this purpose since the T47D cells are human BCCs and CD298 was previously identified as a human-specific antigen, expressed in >99.9% of human BCCs with no background in mouse lines²⁸⁹. Isolating the tumour (CD298+) from the host (CD298-) cells and then analysing the CD298+ tumour digests should therefore facilitate protein expression control of the samples.

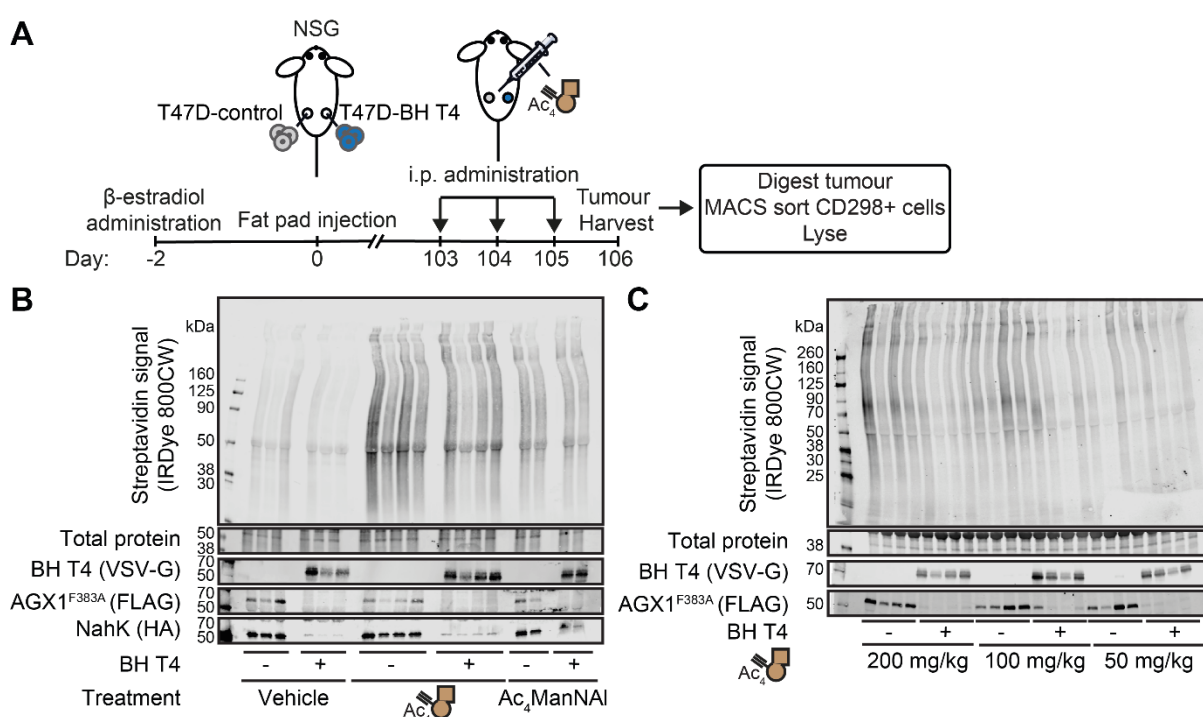


Fig. 49. Extending the BH GalNAc-T *in vivo* system to GalNAc-T4. **A**) BH-GalNAc-T4 (BH T4) tumours were grown in the fat pad of NSG mice. Control tumours (where BH T4 was absent) were grown in the opposite fat pad from the same mouse as an internal control. 103 days after tumour cell injection, the mice were i.p. injected with 200, 100 or 50 mg/kg of Ac₄GalN6yne; 100 mg/kg of Ac₄ManNAI; or the corresponding volume of vehicle once daily for 3 days by Dr Victoria Bridgeman. 24 h after the final injection, the tumours were harvested, digested, and sorted through a MACS column for human CD298 by Dr Victoria Bridgeman. The positive selection was collected and lysed by Dr Victoria Bridgeman. **B**) Tagging of glycoproteins in CD298+ tumour digests treated with vehicle, 200 mg/kg Ac₄GalN6yne or 100 mg/kg of Ac₄ManNAI as assessed by streptavidin blot after fat pad injection of T47D-BH T4 or T47D-control cells into NSG mice. 10 µg of protein were used per lane. Antibody staining was performed using antibodies against VSV-G (1:500, PA129903), FLAG (1:500, ab1257), and HA (1:2000, ab130275). Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). **C**) Tagging of glycoproteins in CD298+ tumour digests treated with different doses of Ac₄GalN6yne as assessed by streptavidin blot. 10 µg of protein were used per lane. Antibody staining was performed using antibodies against VSV-G (1:500, PA129903) and FLAG (1:500, PA1984B). Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). T47D-BH T4 (+) = T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells, T47D-control (-) = T47D^{WT}-AGX1^{F383A}-NahK cells.

Tagging of glycoproteins was assessed by streptavidin blot. I first assessed the tagging of glycoproteins in CD298+ tumour digests treated with vehicle, Ac₄ManNAI or the highest dose of Ac₄GalN6yne (200 mg/kg). Minimal signal was seen in the vehicle-treated samples (**Fig. 49B**). A boost in signal was seen in the sugar-treated samples compared to the vehicle-treated samples, however the intensity of all samples treated with Ac₄GalN6yne was comparable regardless of whether BH GalNAc-T4 was present or not (**Fig. 49B**).

We noticed that the expression levels of the enzymes from the artificial biosynthetic pathway were much lower in the BH GalNAc-T4 samples compared to the controls (**Fig. 49B**), which suggested silencing of the expression of AGX1^{F383A} and NahK by the BCCs. When Dr Anna Cioce tested the 4T1-BH T2 cells in immunocompetent BALB/c mice, she observed silencing of the expression of AGX1^{F383A} as well (**SI Fig. 17F**). Silencing of the artificial biosynthetic pathway may occur because the expression of the BH GalNAc-Ts was under an EF-1 α promoter whilst the expression of AGX1^{F383A} and NahK was under a synthetic RPBSA promoter instead (**SI Fig. 7**)²⁹⁰. Using NSG mice instead of BALB/c mice maintained the expression of AGX1^{F383A} and NahK in the 4T1-BH T2 cells as confirmed by Western blot (**Fig. 47B**). However, since tumour growth was slower with the T47D-BH T4 cells (106 days) compared to the 4T1-BH T2 cells (18 days), the competition between the two promoters could result in silencing of the artificial biosynthetic pathway even in the immunocompromised background.

Since a difference in signal was still observed between the vehicle- and sugar-treated samples (**Fig. 49B**) we wondered if residual expression of the artificial biosynthetic pathway could be enough to result in differential tagging of glycoproteins by BH GalNAc-T4. Since the samples used for **Fig. 49B** were treated with the highest dose of Ac₄GalN6yne tested (200 mg/kg), we wondered if we were observing saturation of the background signal in the control samples. I therefore repeated the Western blot using all the samples treated with Ac₄GalN6yne to see if lowering the dose 2-fold (100 mg/kg) or 4-fold (50 mg/kg) would reduce the background signal in the control samples. However, comparable signal was still observed between control and BH GalNAc-T4 samples at all sugar doses tested, suggesting that the reason we were observing a comparable signal between control and BH GalNAc-T4 samples was not saturation of the background signal due to the dose of the sugar (**Fig. 49C**).

All things considered, there was a boost in signal between sugar- and vehicle-treated samples, suggesting that the signal observed was dependent on sugar treatment. However, there was no difference between Ac₄GalN6yne-treated control and BH GalNAc-T4 samples. Two possible hypotheses could explain this observation. If residual expression of the artificial biosynthetic pathway was enough to biosynthesise UDP-GalN6yne, one hypothesis for the comparable signal observed between control and BH GalNAc-T4 samples could be that there were not many GalNAc-T4-specific substrates in the samples and that the signal observed in the samples was predominantly due to background tagging of glycoproteins. If residual expression of the artificial biosynthetic pathway was not enough to biosynthesise UDP-GalN6yne, an alternative explanation for the comparable results observed between control and BH GalNAc-T4 samples would be that the signal observed in the samples was not enzymatic glycosylation but instead a non-specific protein modification such as non-enzymatic S-glyco-modification between peracetylated sugars and protein cysteines^{291, 292}. To confirm these hypotheses, an additional control would be required consisting of mock-transfected cells treated with Ac₄GalN6yne. If the Ac₄GalN6yne-treated control and BH GalNAc-T4 samples showed a boost in signal compared to mock-transfected cells this would suggest that the signal is dependent on the biosynthesis of UDP-GalN6yne so the former hypothesis would be more likely. However, we decided that it was not feasible to pursue these experiments further due to time constraints.

Overall, preliminary work was performed to extend the BH GalNAc-T approach to an *in vivo* setting. We successfully tagged glycoproteins with BH GalNAc-T2 *in vivo*²⁰⁴, however further optimisation is required to extend the BH GalNAc-T *in vivo* system to BH GalNAc-T4.

DISCUSSION

The overall aim of this project was to use the BH GalNAc-T4 system to characterise the substrates and glycosylation sites of GalNAc-T4 in BCCs to better understand the role of GalNAc-T4 in BC.

The first objective of this project was to validate that GalNAc-T4 was implicated in BC by generating GalNAc-T4 KO T47D and MCF7 cells using CRISPR-Cas9 and comparing the growth of WT and GalNAc-T4 KO cells *in vitro* and *in vivo* using the same methodology for both cell lines. The difference between the WT and GalNAc-T4 KO cells was not statistically significant in the *in vitro* growth assays (**Fig. 12**), most likely due to the technical variability associated with the experiment. Nevertheless, there was a trend towards a greater number of apoptotic cells upon GalNAc-T4 KO with both cell lines (**Fig. 12D** and **Fig. 12G**), which could explain why a statistically significant growth defect was later observed with these cells *in vivo*. Herein a growth defect was observed with both T47D (**Fig. 13**) and MCF7 (**Fig. 14**) cells upon GalNAc-T4 KO whilst other groups had previously reported a growth defect upon GalNAc-T4 KD in T47D cells¹³² and a boost in growth upon GalNAc-T4 KO in MCF7 cells¹³³. Our results suggested that the contradicting phenotypes published previously were likely due to the different methodologies used by the two groups. Overall, the undisputable phenotype observed upon GalNAc-T4 KO *in vivo*, which is more representative of the growth and progression of BC tumours than one-dimensional cellular models, validated a role for GalNAc-T4 in BC.

Having confirmed that GalNAc-T4 is implicated in BC, the next objective was to evaluate the enzymatic activities of WT and BH GalNAc-T4 *in vitro* before extending the BH GalNAc-T4 system to BCCs. During my MSci project I had confirmed that BH GalNAc-T4 emulates the known glycopeptide substrate preference of WT GalNAc-T4 whilst displaying a UDP-sugar substrate switch from the native substrate to the bumped analogue UDP-GalN6yne²⁰⁸. Herein, I tested a greater panel of UDP-GalNAc analogues to see if there were any alternative UDP-sugars with more favourable glycopeptide formation and orthogonality than UDP-GalN6yne. Nevertheless, the most promising results were still observed with UDP-GalN6yne, where comparable glycopeptide turnover was observed between the BH GalNAc-T4/UDP-GalN6yne and WT GalNAc-T4/UDP-GalNAc pairs whilst reduced activity was observed with the opposite enzyme-substrate pair (**Fig. 15**), demonstrating the orthogonality of the BH GalNAc-T4/UDP-GalN6yne pair.

The next objective was to determine whether BH GalNAc-T4 retained the kinetic parameters of the native system. I therefore compared the Michaelis-Menten kinetics of the WT GalNAc-T4/UDP-GalNAc and BH GalNAc-T4/UDP-GalN6yne pairs. The kinetic parameters of both enzyme-substrate pairs were within the same order of magnitude (**Fig. 16**) and hence were considered comparable, demonstrating that the BH GalNAc-T4/UDP-GalN6yne pair retained the kinetic parameters of the native system. The K_M of WT GalNAc-T4 with glycopeptide **1** had been previously reported as $41 \pm 19 \mu\text{M}$ ⁷⁴ and $50.7 \pm 6.3 \mu\text{M}$ ⁸². These values are within the same order of magnitude as the value obtained for WT GalNAc-T4 herein ($200 \mu\text{M}$), validating the results obtained. In these reports, the glycosylation reaction mixtures contained recombinant WT GalNAc-T4, glycopeptide **1** and [³H]-radiolabelled UDP-GalNAc⁸². After quenching the reaction, free UDP and non-hydrolysed UDP-GalNAc were removed by passage over an anion exchange resin^{74, 82}. Total UDP-[³H]-GalNAc utilisation was determined by the difference after scintillation counting 1/20 of the sample before and after passage over the anion exchange resin⁸². The extent of [³H]-GalNAc transferred to the peptide and the extent of hydrolysis was determined by gel filtration analysis⁸². Even though the reported GalNAc-T4 activity accounted for UDP-[³H]-GalNAc hydrolysis^{74, 82}, they also stated that significant hydrolysis was observed with GalNAc-T4⁷⁴, which could explain the difference in the K_M values reported previously and those obtained herein.

The next objective was to compare the selectivity of WT and BH GalNAc-T4 in the presence of a 1:1 mixture of UDP-GalNAc and UDP-GalN6yne, which is more representative of the substrate exposure in the secretory pathway²³¹. The results demonstrated that WT GalNAc-T4 preferentially reacted with UDP-GalNAc whilst BH GalNAc-T4 exclusively reacted with UDP-GalN6yne in an environment that more closely resembles the secretory pathway (**Fig. 17**). This is important for application in biological systems as WT GalNAc-Ts should not accept the bumped UDP-GalNAc analogue in the presence of UDP-GalNAc to ensure selective characterisation of the tagged glycoproteins by the BH GalNAc-T in the presence of all other members of the GalNAc-T family⁶⁷. Choi et al. previously demonstrated that WT GalNAc-T2 exclusively reacted with UDP-GalNAc in the presence of a 1:1 mixture of UDP-GalNAc and UDP-GalN6yne⁶⁷. Future work should ensure that this selectivity towards UDP-GalNAc relative to UDP-GalN6yne is also observed with the other members of the GalNAc-T family. Altogether, the results of the *in vitro* glycosylation experiments demonstrated that the BH GalNAc-T4/UDP-GalN6yne pair is biochemically competent whilst orthogonal to the native enzyme-substrate pair.

With an orthogonal BH GalNAc-T4/UDP-sugar pair in hand, the next objective was to establish the BH GalNAc-T4 system in BCCs. Previous reports demonstrated that GalNAc-T4 is more highly expressed in ER+ luminal BC tumours and BCCs compared to healthy mammary epithelium^{129, 132}. Previous reports also demonstrated that the expression of GalNAc-T4 was lower in basal-like BCCs compared to luminal BCCs^{132, 133}, and that the expression of GalNAc-T4 was comparable between basal-like BC tumours and healthy breast tissue¹²⁹. Altogether, these results suggested that GalNAc-T4 appeared to be more implicated in luminal compared to basal-like BC. Nevertheless, for these experiments we decided to use a panel of BCCs encompassing both luminal and basal-like BC subtypes to reflect the heterogeneity in BC tumours and to see if determining the substrates of GalNAc-T4 in the different BCCs could help explain the preferential upregulation of GalNAc-T4 in luminal BC tumours. T47D, MCF7 and MDA-MB-231 cells were chosen for this work as these cell lines were previously used to investigate the role of GalNAc-T4 in BC^{129, 132, 133} and they account for more than two-thirds of BCCs used for BC modelling studies³⁶. T47D and MCF7 cells display a Luminal A phenotype whilst MDA-MB-231 cells display a TNBC phenotype³⁶.

T47D, MCF7 and MDA-MB-231 cells were stably transfected with plasmids encoding the two enzymes of the artificial biosynthetic pathway for UDP-sugar biosynthesis (AGX1^{F383A} and NahK) and WT/BH GalNAc-T4. After feeding the cells with Ac₄GalN6yne, the chemically-tagged glycoproteins were derivatised with a fluorophore by CuAAC and visualised by in-gel fluorescence. Multiple intense bands were selectively observed in the presence of BH GalNAc-T4 (**Fig. 18-20**), demonstrating that we could bioorthogonally tag the substrates of GalNAc-T4 in BCCs using BH GalNAc-T4. In the experiments with T47D cells we saw that BH GalNAc-T4 was more highly OE than WT GalNAc-T4 (**Fig. 18D**). Since this difference in expression levels would be a source of variation that could lead to artifacts, we decided not to use the cells where WT GalNAc-T4 was OE in future experiments.

The next objective was to characterise the chemically-tagged glycoproteins by MS-proteomics. SILAC-based MS-proteomics was performed on T47D, MCF7 and MDA-MB-231 cells transfected with plasmids encoding NahK, AGX1^{F383A} and BH GalNAc-T4. Since UDP-GalN6yne is epimerised to the corresponding UDP-GlcNAc analogue by GALE resulting in undesired tagging of GlcNAc-containing glycoconjugates¹⁹³, cells transfected with plasmids only expressing AGX1^{F383A} and NahK, (where UDP-GalN6yne can be biosynthesised but BH GalNAc-T4 is absent) were used as a control to dissect BH GalNAc-T4-specific tagging from any potential background tagging of glycoproteins. This allowed the successful characterisation of 42 (**Fig. 24** and **SI Table 3**), 114 (**Fig. 26** and **SI Table 6**) and 68 (**Fig. 26** and **SI Table 7**) proteins as substrates of GalNAc-T4 in the lysates of T47D, MCF7 and MDA-MB-231 cells, respectively, as well as 3 proteins as substrates of GalNAc-T4 in the secretome of T47D cells (**Fig. 25**). Most hits were cell-line specific, but a minority were conserved between the BCCs (**Table 4**). MUC1, the *bona fide* substrate of GalNAc-T4^{72, 156, 157, 234}, was conserved in the two luminal cell lines (T47D and MCF7 cells) but not in the basal-like MDA-MB-231 cells (**Table 4** and **Fig. 27B**), as anticipated since MUC1 is physiologically expressed in the apical surface of luminal epithelial cells²⁴⁰. CD46 and GOLM1 were identified as the two novel substrates of GalNAc-T4 which were conserved in the three BCC lines (**Fig. 27B**). Overall, the results demonstrated that SILAC-based MS-proteomics allowed the successful characterisation of the substrates of GalNAc-T4 in BCCs.

Since there was a statistically significant growth defect upon GalNAc-T4 KO *in vivo* (**Fig. 13** and **Fig. 14**), the next objective was to assess whether KO of GalNAc-T4 affected the expression levels of its substrates. Since CD46 and GOLM1 were conserved in the three cell lines, we decided to analyse the expression levels of these proteins as well as MUC1 as GalNAc-T4's *bona fide* substrate^{72, 156, 157, 234}. However, no differences were observed in the expression levels of CD46, GOLM1 and MUC1 between WT and KO GalNAc-T4 cells (**Fig. 29**). Overall, these results demonstrated that GalNAc-T4-mediated glycosylation must serve an alternative purpose other than affecting the expression levels of these proteins.

To help draw structural conclusions on the role of GalNAc-T4-mediated glycosylation in the identified substrates, the next objective was to determine the GalNAc-T4-mediated glycosylation site(s) by MS-glycoproteomics. However, there are multiple challenges when performing glycosylation site-localisation of O-GalNAc glycans¹⁸², as evidenced by the lack of hits obtained when analysing the glycopeptide fraction from the T47D lysates during the SILAC-based MS-(glyco)proteomics workflow (**Results Section 3.3**). Previous work in the Schumann lab had demonstrated that bioorthogonal reporter reagents can impart glycopeptides with beneficial MS properties such as unique HCD oxonium ions and ETD fingerprint ions which could be leveraged for confident glycopeptide assignment (**Fig. 31**)^{193, 256}. Encouraged by these results, we wondered if we could use a similar strategy to increase the charge density of the tagged glycopeptides and hence facilitate ETD fragmentation to improve glycosylation site-localisation by MS-glycoproteomics¹⁸². We therefore generated synthetic glycopeptides bearing azide- or alkyne-containing sugars which we derivatised with a panel of reporter reagents with increasingly basic or charged groups by CuAAC and then analysed by LC-MS/MS¹⁸². We saw an increase in peptide charge and detected unique HCD oxonium ions with all the reagents tested (**Fig. 32**)¹⁸². The unique HCD oxonium ions were used to successfully trigger HCD-pd-ETD in all samples (**Fig. 33**)¹⁸². The most promising results were observed in the samples with the iTag-containing reagents, where the increased charge density imparted by the permanent positive charge from the iTag increased the ETD fragmentation efficiency and hence the peptide sequence coverage¹⁸². An ETD fingerprint ion was also detected with one of the iTag-containing reagents (**Fig. 34**), facilitating confident glycopeptide assignment¹⁸². Altogether these experiments demonstrated the unique MS properties that modified sugars and bioorthogonal reporter reagents can impart on glycopeptides to facilitate MS-glycoproteomics analyses¹⁸².

Having validated biotin-DADPS-picolyl azide (for enrichment) and identified the iTag-containing reagents (without enrichment) as promising bioorthogonal reagents to facilitate MS-glycoproteomics analyses¹⁸², the next objective was to use these reagents to determine the GalNAc-T4-mediated glycosylation site(s) of CD46 and GOLM1 as the two hits conserved amongst the three BCCs investigated in the MS-proteomics experiments (**Fig. 27B**). Due to technical difficulties caused by the mucin-like nature of the STP domain of CD46, we couldn't pinpoint the exact GalNAc-T4-mediated glycosylation site(s) in CD46. Nevertheless, we established that the glycosylation site must be located within the STP B domain since the strongest streptavidin signal was observed in the CD46 BC samples containing BH GalNAc-T4 (**Fig. 40**).

C3b/C4b ligand binding to CD46 is mediated by the SCR domains (primarily SCR4), whereas the STP region was suggested to serve as a spacer between SCR4 and the cell membrane²⁴⁷. Liszewski et al. demonstrated that the CD46 BC isoforms display more cytoprotection and a greater cofactor activity than the C isoforms²⁷⁷. In a follow-up study, the same group demonstrated that STP deletion mutants had comparable cytoprotective and cofactor activity to the C isoforms, suggesting that the STP B domain is the key mediator for CD46's cytoprotective and cofactor activities²⁷⁵. Persson et al. later hypothesised that CD46 can exist in two extreme conformations: one conformation where the SCR4 and STP domains project vertically from the cell surface, allowing ligand binding to SCR4 (**Fig. 50A**); and a second conformation where the STP region is bent and the SCR4 domain lies parallel to the cell surface, hence preventing ligand binding (**Fig. 50B**)²⁴⁷. A possible explanation for the observations reported by Liszewski et al. could be that the heavily glycosylated nature of the STP B domain affects the rigidity of the STP domain and makes the former conformation more favourable, facilitating ligand binding and promoting CD46's cytoprotective and cofactor activities. GalNAc-T4-mediated glycosylation of the STP B domain may therefore contribute to the role of CD46 in protecting cells from complement-mediated damage and hence could explain why there was a trend towards a greater number of apoptotic cells upon GalNAc-T4 KO with both T47D and MCF7 cells (**Fig. 12D** and **Fig. 12G**). Additional work will be required to reveal the GalNAc-T4-mediated glycosylation site(s) in CD46 and validate the role of GalNAc-T4-mediated glycosylation on CD46.

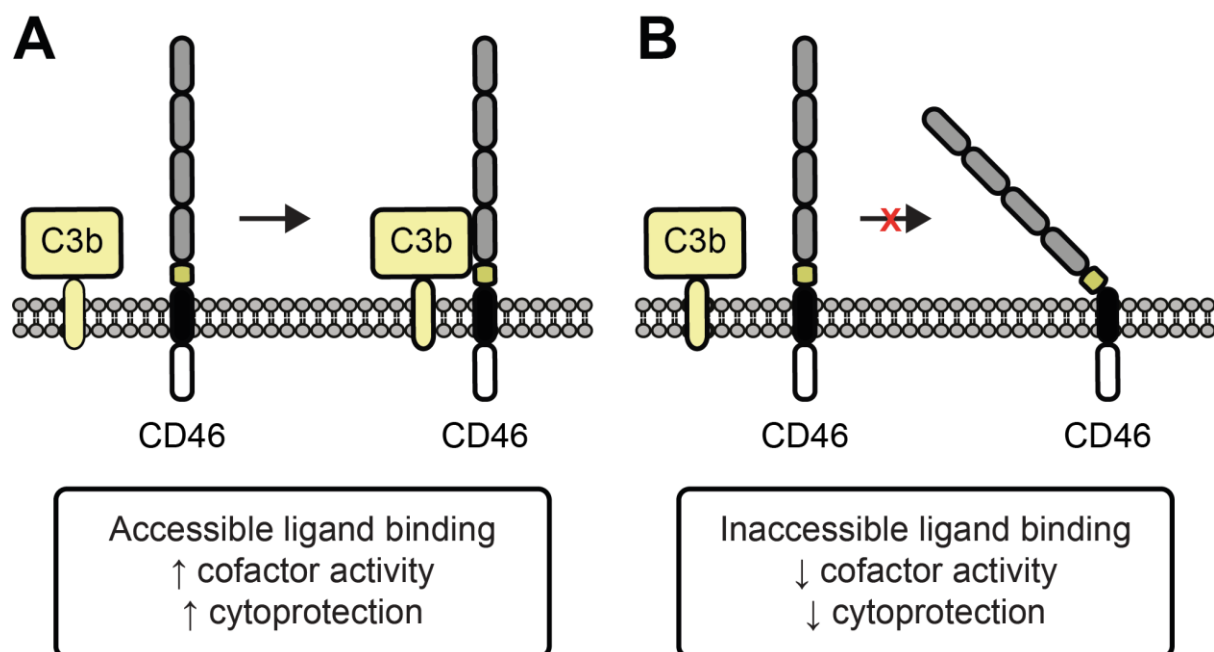


Fig. 50. Two possible extreme conformations of CD46 on the cell surface. A) In one conformation the SCR4 and STP domains project vertically from the cell surface, allowing ligand binding to SCR4. **B)** In another conformation the STP region is bent and the SCR4 domain lies parallel to the cell surface, preventing ligand binding. Figure adapted from Persson et al. (2010)²⁴⁷.

Furthermore, C3 (the pivotal complement component²⁹³) was obtained as a GalNAc-T4 substrate in the secretome of T47D cells (**Fig. 25**), providing further evidence for a possible role of GalNAc-T4 in regulating the complement system. Future investigation into the role of GalNAc-T4 in the complement system should be performed.

Ser²⁰⁴ and Thr²¹⁸ were successfully identified as the GalNAc-T4-mediated glycosylation sites in GOLM1 by MS-glycoproteomics. These two sites have been previously predicted to carry O-GalNAc glycans by NetOGlyc4.0 as reported in GlycoDomainViewer¹⁹². Nevertheless, as far as we are aware, this is the first time that these glycosites are confirmed by MS-glycoproteomics experiments and are established as GalNAc-T4 glycosylation sites. Furthermore, we detected elaboration of the Thr²¹⁸ site up to the disaccharide Hex-GalN6yne (where Hex is likely a Gal residue typical of core 1 O-GalNAc glycans). Whilst it is currently uncertain whether elaboration of these analogues recapitulates downstream glycosylation of the native GalNAc¹⁹³, the BH GalNAc-T approach is the only strategy that allows GalNAc-T-specific site and structural information beyond the initiating GalNAc.

Ser¹⁹¹ is the closest Ser/Thr residue (13 residues away) upstream of Ser²⁰⁴. If other GalNAc-Ts glycosylated this residue, it could promote lectin domain-mediated glycosylation of Ser²⁰⁴ by GalNAc-T4. However, since Ser¹⁹¹ is not predicted to carry O-GalNAc glycans by NetOGlyc4.0¹⁹² and GalNAc-T4 is thought to display poor activity with naked peptides⁸², it was surprising to obtain Ser²⁰⁴ as a hit. Nevertheless, GalNAc-T4 and -T7 are the only GalNAc-Ts from the 11 investigated to date which have an N-terminal neutral flanking charge preference⁷¹ like the one present in the QQLQALS²⁰⁴EPQPR peptide. In addition, this peptide contains a Pro-X-Pro motif, although in this case “X” corresponds to a Gln residue. The combination of these two factors may explain how Ser²⁰⁴ could be glycosylated by GalNAc-T4. Since Ser²⁰⁴ is located 14 residues upstream from Thr²¹⁸, once the former is glycosylated by GalNAc-T4 it could promote lectin domain-mediated glycosylation of the latter.

The glycosite at Thr²¹⁸ is at the beginning of the flexible loop of GOLM1. The function of the flexible loop remains poorly understood, although evidence suggests that its truncation leads to decreased survival rates and pathological changes to the kidney and liver in mice²⁹⁴. The glycosite at Ser²⁰⁴ is at the end of the coiled-coil domain. This domain has a protein-protein interaction function²⁹⁵, and it was shown to be the key structural motif mediating the interaction with the extracellular chaperone clusterin²⁹⁶ and the autophagy adaptor LC3²⁹⁷. Furthermore, the region from Ala⁵⁶ to Phe¹⁵² within the coiled-coil domain is sufficient to mediate the dimerisation of GOLM1, whilst disulfide bonds only serve to stabilise the dimer²⁴⁵. Whilst it has been previously reported that endogenous and secreted GOLM1 exit as disulfide-bonded dimers²⁴⁵, recent findings suggested that secreted GOLM1 may exist as a tetramer instead²⁸⁶. Bai et al. recently reported a crystal structure of the coiled-coil domain (from Thr¹¹¹ to Val¹⁷⁷) which demonstrates that secreted GOLM1 exists as two anti-parallel dimers which assemble to form a tetramer²⁸⁶. This group also assessed the oligomeric state of recombinantly expressed GOLM1 (from Arg⁵⁴ to Gln²⁰¹) in solution and confirmed that its tetrameric structure was not merely a result of crystallisation²⁸⁶. Based on their results, they hypothesised that membrane-bound full-length GOLM1 may be restricted to exist as a parallel dimer whilst secreted GOLM1 loses the restriction after furin cleavage and can form the tetrameric structure instead (**Fig. 51**)²⁸⁶.

To explore a possible structural role of GalNAc-T4-mediated glycosylation on GOLM1, we performed *in silico* evaluation on the molecular structure of GOLM1. Since no crystal structure has been published for full-length GOLM1, we obtained the predicted molecular structure from the AlphaFold Protein Structure Database (**Fig. 46**)²⁸⁸. However, the confidence score calculated by AlphaFold from Asn¹⁸² to Leu⁴⁰¹ was low, which made it challenging to evaluate a possible structural role of GalNAc-T4-mediated glycosylation using this molecular structure. Nevertheless, when tracking tagged recombinant GOLM1 by streptavidin blot during the MS-glycoproteomics workflow, the strongest streptavidin signal was observed in the monomer bands (**Fig. 45B**), suggesting that GalNAc-T4-mediated glycosylation may affect GOLM1's ability to oligomerise (**Fig. 51**). A similar mechanism had been previously suggested for TGFBR2 by Wu et al., where GalNAc-T4-mediated O-GalNAcylation of TGFBR2 prevented receptor dimerisation and inhibited downstream signalling¹²⁹. Nevertheless, we would have to confirm that this observation is also maintained using endogenous GalNAc-T4 and GalNAc to ensure that it is not an artifact of the chemically-tagged protein. Previous findings have suggested that GOLM1 may exhibit different roles intracellularly and extracellularly by dynamically changing its oligomeric state²⁸⁶, so it would be important to understand if GalNAc-T4-mediated glycosylation plays a role in the oligomeric state of GOLM1.

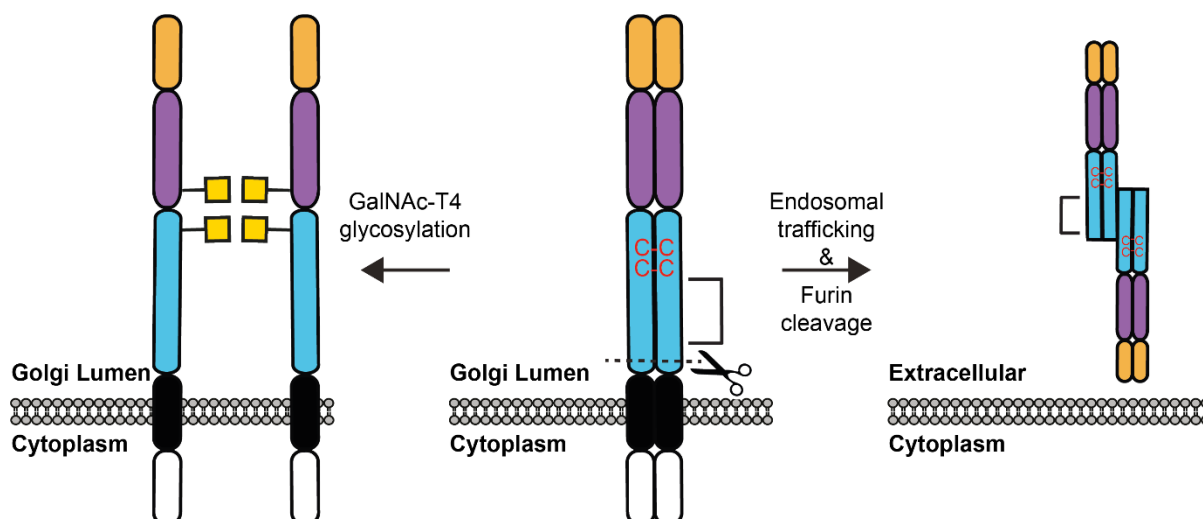


Fig. 51. Oligomerisation of GOLM1. Endogenous GOLM1 forms a dimer through the coiled-coil domain (the region mediating the interaction is indicated in the figure), which is stabilised by disulfide-bonds. GOLM1 undergoes endosomal trafficking, which results in GOLM1 being proteolytically cleaved by furin and the luminal portion secreted to the extracellular space. When GOLM1 is secreted two anti-parallel dimers can assemble into a tetramer. Glycosylation by GalNAc-T4 may inhibit the oligomerisation of GOLM1. The N-terminal cytoplasmic domain, the transmembrane domain, the coiled-coil domain, the flexible loop, and the acidic C-terminal luminal domain are represented in white, black, blue, purple, and orange respectively. The furin cleavage site is represented as a dashed line. Figure adapted from Bai et al. (2024)²⁸⁶.

Since the coiled-coiled domain has a protein-protein interaction function²⁹⁵, glycosylation of GOLM1 by GalNAc-T4 may affect the interaction between GOLM1 and its binding partners instead. To investigate this, we could harness the ability of UDP-sugar **15** to identify binding partners of tagged glycoproteins thanks to the diazirine moiety which allows photocrosslinking to neighbouring molecules upon ultraviolet irradiation^{205, 229}. Substantial turnover (>50%) was seen with the BH GalNAc-T4/UDP-sugar **15** pair whilst minimal activity was seen with the WT GalNAc-T4/UDP-sugar **15** pair in the *in vitro* glycosylation experiments (**Fig. 15**). Additional work will therefore be required to confirm the role of GalNAc-T4-mediated glycosylation on GOLM1.

FOXA1¹³², Cullin-1¹³³, and TGFBR2¹²⁹ had been previously proposed as substrates of GalNAc-T4 in BCCs, however none of these proteins were obtained as hits herein. These proteins were elucidated as GalNAc-T4 substrates after GalNAc-T4 KD/KO in BCCs, followed by immunoprecipitation using VVL and characterisation of differential VVL-binding proteins between KD/KO and control samples^{129, 132, 133}. Nevertheless, the complex interplay of the 20 GalNAc-Ts can lead to compensatory mechanisms taking place upon KD/KO of GalNAc-T4¹⁹³, as illustrated by findings that show an increase of VVL-binding proteins upon KO of GalNAc-T4¹³³.

In the cell surface tagging experiments performed herein some high MW bands were seen in the GalNAc-T4 KO T47D cells which were not seen in the corresponding WT cells (**Fig. 18C**), providing additional evidence of compensatory mechanisms taking place upon KO of GalNAc-T4. Taking these findings into account, FOXA1, Cullin-1, and TGFBR2 may thus not be substrates of GalNAc-T4 and instead may have been obtained as hits due to compensatory mechanisms taking place upon KD/KO of GalNAc-T4.

An alternative explanation as to why FOXA1, Cullin-1, and TGFBR2 weren't obtained as substrates herein could be due to limitations with our approach. For example, with our current methodology we may only be detecting the more highly expressed proteins, as evidenced with CD46 which is more highly expressed in T47D cells compared to MCF7 or MDA-MB-231 cells (**Fig. 28**). CD46 was only obtained as a hit in T47D cells but not in MCF7 or MDA-MB-231 cells in the MS-proteomics experiments, whilst it was subsequently confirmed as a hit in the three cell lines in the neutravidin bead pulldown experiments (**Fig. 27**). To confirm whether FOXA1, Cullin-1, and TGFBR2 substrates of GalNAc-T4 in BCCs, neutravidin-bead pulldown experiments could hence be performed.

Another limitation of our approach is that the BH GalNAc-T will be in competition for protein substrates and glycosylation sites with the corresponding WT enzyme. Similarly, when the substrate specificity of BH GalNAc-T1 and BH GalNAc-T2 was investigated *in vitro*, the results demonstrated that the BH GalNAc-Ts glycosylated fewer peptides than the corresponding WT enzymes¹⁹³. Both factors could suggest that we may be missing some hits with our methodology. Furthermore, the efficiency of GalN6yne incorporation or the chemical-tagging of glycoproteins may not be quantitative, as evidenced by the presence of untagged CD46 (**Fig. 42B**) and GOLM1 (**Fig. 45B**) in the supernatants after neutravidin bead enrichment. To overcome the former factor, higher sugar concentrations or additional days of sugar feeding could be performed. To overcome the latter, after an initial incubation of the CuAAC reaction mixtures, the supply of the CuAAC reactants could be replenished by adding more CuAAC solution to the samples and incubating the reaction for longer to maximise the chemical tagging of the glycoproteins. On top of this, BH GalNAc-Ts are OE compared to the endogenous enzymes, as reported previously¹⁹³, and as observed herein (**SI Fig. 8A**). Whilst previous findings demonstrated that BH GalNAc-Ts do not introduce new glycosylation sites in their protein substrates¹⁹³ so no false-positive hits should be seen even if BH GalNAc-Ts are OE, a perfect discovery tool would assess native glycosylation levels²⁰³. In future work, dox-inducible promoters could be used to fine-tune the expression levels of BH GalNAc-Ts to overcome this limitation²⁰³. Alternatively, homology-directed repair editing²²² could be performed to mutate the native allele to the corresponding BH mutant²⁰³.

Another limitation with our current approach is that when BH GalNAc-T2 was OE the sialylation of N- and O-glycans was reduced²⁰⁴, suggesting an alteration in the cellular glycome and that this approach doesn't fully replicate the native biological system²⁰³. Nevertheless, this should not affect the overall aim of identifying the substrates and glycosylation sites of a specific GalNAc-T in living cells²⁰³. All things considered, the BH GalNAc-T approach is the only method that allows the identification of the substrates and glycosylation sites of a specific GalNAc-T in the presence of all other GalNAc-Ts, which is crucial due to the interdependencies and substrate redundancies of the GalNAc-T family¹⁹³.

While O-GalNAc glycosylation is found in many types of glycoproteins, it is abundantly present in mucin domains²⁰³. As described in the “**Substrate identification of individual GalNAc-Ts**” section and as evidenced with CD46 herein, it is challenging to study mucin domains by MS^{182, 183, 187-190}. Nevertheless, the Schumann lab and others are investing efforts in addressing these challenges, such as the use of iTag-containing reagents to improve glycopeptide assignment^{182, 259-266} as described herein; using mucin-selective proteases to break down mucin-like glycoproteins into smaller fragments that are more amenable to MS analysis^{183, 187, 188, 190}; or the use of mucin-selective enrichment strategies^{183, 187-189}. Furthermore, Benito-Alifonso et al. demonstrated that iTag-modified glycoproteins could be isolated directly from cell lysates using a Hitrap™ Q HP affinity column and eluted using an imidazolium gradient²⁵⁹. Future experiments could therefore use a similar strategy to enrich the chemically-tagged glycoproteins derivatised with the iTag-containing reagents as an alternative to derivatising chemically-tagged glycoproteins with biotin-based reagents and enrichment with neutravidin beads, and hence still benefit from the unique MS properties imparted by the iTag-containing reagents. On top of this, we have demonstrated herein that O-GlyZOR can selectively remove unmodified O-glycans whilst leaving the modified sugar intact (**Fig. 42B** and **Fig. 45B**), which can also aid in mucin MS-glycoproteomics by facilitating access of proteases to the peptide backbone, reducing the hydrophilicity, and improving the ionizability of glycopeptides from mucin-like domains.

Since the canonical localisation of GalNAc-Ts is the secretory pathway⁶⁷, we would expect their substrates to be cell surface or extracellular proteins. However, some of the hits obtained in the MS-proteomics experiments were nucleocytoplasmic and mitochondrial proteins instead (**SI Table 3** and **SI Tables 6-7**). Even though some nucleocytoplasmic GalNAc-T substrates have been previously proposed (e.g. FOXA1 as a GalNAc-T2¹⁶² and GalNAc-T4¹³² substrate, Cullin-1 as a GalNAc-T4 substrate¹³³ or ER α as a GalNAc-T6 substrate¹⁷⁰), the existence of nucleocytoplasmic O-GalNAc glycosylation is still heavily debated^{298, 299}.

The genomic region of GalNAc-T4 may provide a possible mechanism for the nucleocytoplasmic activity of GalNAc-T4. *GALNT4* is located on chromosome 12q21.3 within an intron of the *POC1B* gene, which encodes a cytosolic protein (**Fig. 52A**)²²⁴. Fusion transcripts have been reported where the catalytic and lectin domains of GalNAc-T4 are intact, but the transmembrane domain is replaced with the N-terminus of *POC1B* (**Fig. 52B**)²²⁴. The fusion protein POC1B-GalNAc-T4 could be redirected to the nucleocytoplasm and hence glycosylate nuclear and cytosolic proteins. Nevertheless, such a fusion protein could not be formed with the BH GalNAc-T4 constructs used herein. Furthermore, since distinguishing between the isomeric O-GalNAc and O-GlcNAc glycans by MS is challenging²⁹⁸, and epimerisation of UDP-GalN6yne by GALE results in undesired tagging of O-GlcNAc glycans¹⁹³ (which are known to occur in the nucleocytoplasm and mitochondria⁶²), future work will be required to investigate the nucleocytoplasmic activity of GalNAc-T4 and to validate that the nucleocytoplasmic and mitochondrial hits obtained herein are indeed substrates of GalNAc-T4 and not O-GlcNAc glycans instead. Future investigation into alternative UDP-GalNAc analogues which are not epimerised by GALE to the corresponding UDP-GlcNAc analogue but show comparable glycopeptide formation and orthogonality to the BH GalNAc-T4/UDP-GalN6yne pair should also be performed.

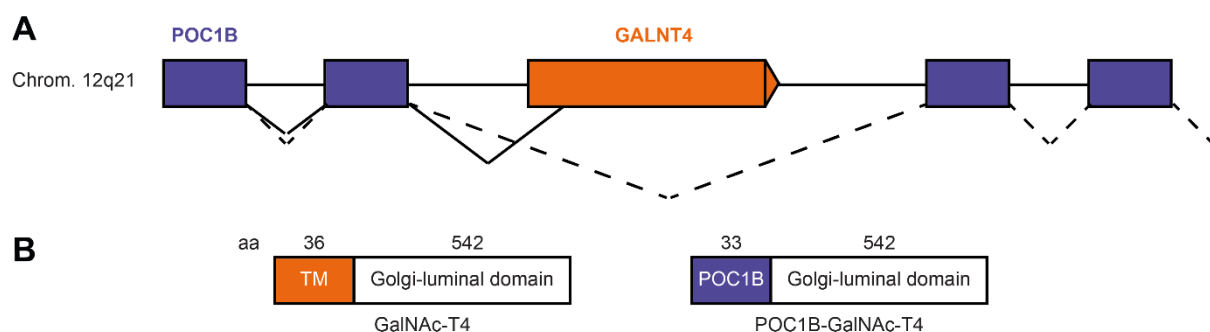


Fig. 52. Alternative splicing of the *POC1B-GALNT4* gene. A) *GALNT4* is located on chromosome 12q21.3 within an intron of the *POC1B* gene. **B)** Alternative splicing yields a fusion protein which retains the 542 amino acids (aa) from GalNAc-T4's glycosyltransferase domain, but the transmembrane (TM) domain is replaced with the N-terminus of POC1B.

Once the substrates of GalNAc-T4 are validated, pathway enrichment analysis could be performed to determine if there are certain pathways enriched in luminal BCCs but not basal-like BCCs that could explain the preferential upregulation of GalNAc-T4 in the former compared to the latter¹²⁹.

Since one-dimensional cellular model systems cannot recapitulate the multifactorial nature of cancer involving the constant interplay between tumour and host cells²⁰⁴, we endeavoured to extend the BH GalNAc-T system to an *in vivo* setting as a more biologically-relevant model system to investigate the role of GalNAc-T4 and its substrates in cancer initiation, progression, and metastasis. Whilst we successfully tagged glycoproteins with BH GalNAc-T2 *in vivo* (**Fig. 48**)²⁰⁴, additional work will be required to extend this approach to BH GalNAc-T4. Future work in the Schumann lab will endeavour to extend the BH GalNAc-T *in vivo* system to BH GalNAc-T4 using 4T1 cells (as the ones used with BH GalNAc-T2)²⁰⁴ instead of the T47D cells used herein. Since 4T1 cells grow faster than T47D cells, this should help overcome the silencing of the artificial biosynthetic pathway observed herein (**Fig. 49**).

Understanding the biological function of a single GalNAc-T is challenging due to substrate redundancies and overlapping glycosylation sites⁶⁷. For this reason, we hypothesised that to investigate the specific function of GalNAc-T4-mediated glycosylation, we would first identify the GalNAc-T4-specific glycosylation sites. These sites could then be mutated to prevent GalNAc-T4-mediated glycosylation and the phenotype of WT and mutant proteins could be subsequently compared using functional assays. However, time did not permit full investigation into the function of GalNAc-T4-mediated glycosylation in BC, so this should be the direction of future research. Furthermore, whilst we only focused on CD46 and GOLM1, the GalNAc-T4-mediated glycosylation site(s) of the other substrates of GalNAc-T4 should also be determined. Due to the complex interplay and interdependencies between GalNAc-Ts, it may be necessary to characterise the substrates and glycosylation sites of all the GalNAc-Ts that have been implicated in BC (as described in the section “**GalNAc-Ts implicated in breast cancer**”) to obtain a complete picture of the role of GalNAc-T-mediated glycosylation in BC.

Aberrantly glycosylated MUC1 is currently used to monitor cancer progression and response to treatment in the clinic (CA15-3 and CA 27.29)^{41, 42, 300}. A multi-centre first-in human phase I clinical trial has demonstrated that MUC1-chimeric antigen receptor T cells were expanded in all patients with MUC1-Tn positive solid tumours including TNBC and that there was no evidence of safety concerns or on-target/off-tumour toxicity (NCT04025216)³⁰¹. Similar strategies could therefore be performed with the newly identified GalNAc-T-specific substrates and glycosylation sites to help yield new diagnostic or therapeutic options towards BC.

CONCLUSION

Overall, we successfully used the BH GalNAc-T4 system to characterise the substrates of GalNAc-T4 in T47D, MCF7 and MDA-MB-231 BCCs. We also determined the GalNAc-T4-mediated glycosylation sites in the newly identified substrate GOLM1 and hypothesised a putative role of O-GalNAcylation of GOLM1 in its oligomerisation state or protein-protein interactions. We also established that glycosylation of another newly identified substrate CD46 by GalNAc-T4 occurs in the STP B domain and hypothesised a putative role of O-GalNAcylation of the STP B domain in regulating the conformation of the extracellular domain of CD46 required for C3b/C4b ligand binding, hence mediating the cytoprotective and cofactor activities of CD46. We also extended the BH GalNAc-T system to an *in vivo* setting, and successfully tagged glycoproteins with BH GalNAc-T2 in mouse models. Future work will focus on validating the GalNAc-T4 substrates identified herein, characterising the GalNAc-T4-mediated glycosylation sites, as well as investigating the role of GalNAc-T4-mediated glycosylation in BC using *in silico*, *in vitro* and *in vivo* methods. Once this has been completed for GalNAc-T4, the BH GalNAc-T approach can be extended to the other BC-associated GalNAc-Ts. Determining the GalNAc-T-specific substrates and the exact glycosylation sites that affect BC may help yield new diagnostic and therapeutic strategies. The BH GalNAc-T approach is the only strategy that provides GalNAc-T-specific site and structural information in the presence of all members of the GalNAc-T family, and hence is uniquely suited towards this end.

BILBIOGRAPHY

1. World Health Organization. (2024) *Breast cancer*. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> [Accessed 2nd September 2024]
2. Bayraktar, S., and Arun, B. K. (2019) Epidemiology, Risk Factors, and Prevention. In: Aydiner, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 39-61. https://doi.org/10.1007/978-3-030-04606-4_4
3. Admoun, C., and Mayrovitz, H. N. (2022) The Etiology of Breast Cancer. In: Mayrovitz, H. N. (ed.) *Breast Cancer*. Brisbane, Australia, Exon Publications, pp. 21-30. <https://doi.org/10.36255/exon-publications-breast-cancer-etiology>
4. McGuire, K.P. (2019) Breast Anatomy and Physiology. In: Aydiner, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 1-9. https://doi.org/10.1007/978-3-319-22843-3_1
5. De Silva, F., and Alcorn, J. (2022) A Tale of Two Cancers: A Current Concise Overview of Breast and Prostate Cancer. *Cancers*. 14 (12), 2954. <https://doi.org/10.3390/cancers14122954>
6. Gudjonsson, T., Adriance, M. C., Sternlicht, M. D., Petersen, O. W., and Bissell, M. J. (2005) Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. *J. Mammary Gland Biol. Neoplasia*. 10 (3), 261-272. <https://doi.org/10.1007/s10911-005-9586-4>
7. Scott, D. A., and Drake, R. R. (2019) Glycosylation and its implications in breast cancer. *Expert Rev. Proteomics*. 16 (8), 665-680. <https://doi.org/10.1080/14789450.2019.1645604>
8. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) Cancer. In: *Molecular biology of the cell*. 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 1091-1144
9. Weinberg, R. A. (2014) The Nature of Cancer. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 31-70.
10. Weinberg, R. A. (2014) Moving Out: Invasion and Metastasis. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 641-722.
11. Gedik, M. E., and Dogan, A. L. (2019) Biology and Genetics in Breast Cancer. In: Aydiner, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 125-142. https://doi.org/10.1007/978-3-030-04606-4_9
12. Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowitz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J. D., Tavaré, S., Caldas, C., and Aparicio, S. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 486, 346-352. <https://doi.org/10.1038/nature10983>
13. Russnes, H. G., Lingjærde, O. C., Børresen-Dale, A.-L., and Caldas, C. (2017) Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. *Am. J. Pathol.* 187 (10), 2152-2162. <https://doi.org/10.1016/j.ajpath.2017.04.022>
14. Weinberg, R. A. (2014) Maintenance of Genomic Integrity and the Development of Cancer. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 511-576.
15. Weinberg, R. A. (2014) p53 and Apoptosis: Master Guardian and Executioner. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 331-390.

16. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) The Cell Cycle. In: *Molecular biology of the cell*. 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 963-1020.
17. Weinberg, R. A. (2014) pRb and Control of the Cell Cycle Clock. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 275-330.
18. Weinberg, R. A. (2014) Growth Factors, Receptors, and Cancer. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 131-174.
19. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) Cell Signaling. In: *Molecular biology of the cell*. 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 813-888.
20. Albagoush, S. A., Zubair, M., and Limaiem, F. (2024) Tissue Evaluation for HER2 Tumor Marker. In: *StatPearls*. Treasure Island, FL, USA, StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK537134/>
21. Weinberg, R. A. (2014) Cellular Oncogenes. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 103-130.
22. Cilbir, E., and Yalcin, S. (2019) Clinical Aspects of Estrogen and Progesterone Receptors and ERBB2 Testing. In: Aydinler, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 143-161. https://doi.org/10.1007/978-3-030-04606-4_10
23. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) Cell Chemistry and Bioenergetics. In: *Molecular biology of the cell*. 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 43-108.
24. Pinho, S. S., and Reis, C. A. (2015) Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer*. 15, 540-555. <https://doi.org/10.1038/nrc3982>
25. Sotoca Covalada, A. M., van den Berg, H., Vervoort, J., van der Saag, P., Ström, A., Gustafsson, J. A., Rietjens, I., and Murk, A. J. (2008) Influence of Cellular ER α /ER β Ratio on the ER α -Agonist Induced Proliferation of Human T47D Breast Cancer Cells. *Toxicol. Sci.* 105 (2), 303-311. <https://doi.org/10.1093/toxsci/kfn141>
26. Kilickap, S., Aktas, B. Y., and Ozisik, Y. Y. (2019) Bone Marrow Micrometastases and Circulating Tumor Cells. In: Aydinler, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 191-199. https://doi.org/10.1007/978-3-030-04606-4_13
27. McAuliffe, P. (2019) Lobular Carcinoma In Situ. In: Aydinler, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 109-113. https://doi.org/10.1007/978-3-030-04606-4_7
28. Orrantia-Borunda, E., Anchondo-Núñez, P., Acuña-Aguilar, L. E., Gómez-Valles, F. O., and Ramírez-Valdespino, C. A. (2022) Subtypes of Breast Cancer. In: Mayrovitz, H. N. (ed.) *Breast Cancer*. Brisbane, Australia, Exon Publications, pp. 31-42. <https://doi.org/10.36255/exon-publications-breast-cancer-subtypes>
29. Rakha, E. A., Tse, G. M., and Quinn, C. M. (2023) An update on the pathological classification of breast cancer. *Histopathology*. 82 (1) 5-16. <https://doi.org/10.1111/his.14786>
30. Tuzlali, S. (2019) Pathology of Breast Cancer. In: Aydinler, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 201-220. https://doi.org/10.1007/978-3-030-04606-4_14
31. Cancer Research UK. (2023) *What is breast cancer?*. <https://www.cancerresearchuk.org/about-cancer/breast-cancer/about> [Accessed 2nd September 2024]
32. Cancer Research UK. (2023) *Types of breast cancer and related breast conditions*. <https://www.cancerresearchuk.org/about-cancer/breast-cancer/types> [Accessed 2nd September 2024]

33. Cancer Research UK. (2023) *Stages of cancer*. <https://www.cancerresearchuk.org/about-cancer/what-is-cancer/stages-of-cancer> [Accessed 2nd September 2024]
34. Manning, H. C., Buck, J. R., and Cook, R. S. (2016) Mouse Models of Breast Cancer: Platforms for Discovering Precision Imaging Diagnostics and Future Cancer Medicine. *J. Nucl. Med.* 57 (Suppl 1), 60S–68S. <https://doi.org/10.2967/jnumed.115.157917>
35. Miziak, P., Baran, M., Błaszczak, E., Przybyszewska-Podstawka, A., Kałafut, J., Smok-Kalwat, J., Dmoszyńska-Graniczka, M., Kielbus, M., and Stepulak, A. (2023) Estrogen Receptor Signaling in Breast Cancer. *Cancers*. 15 (19), 4689. <https://doi.org/10.3390/cancers15194689>
36. Dai, X., Cheng, H., Bai, Z., and Li, J. (2017) Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J. Cancer*. 8 (16), 3131-3141. <https://www.jcancer.org/v08p3131.htm>
37. Arzanova, E., and Mayrovitz, H. N. (2022) The Epidemiology of Breast Cancer. In: Mayrovitz, H. N. (ed.) *Breast Cancer*. Brisbane, Australia, Exon Publications, pp. 1-20. <https://doi.org/10.36255/exon-publications-breast-cancer-epidemiology>
38. Ganott, M., Griffith, B., and Rudzinski, S. M. (2019) Breast Imaging and Image-Guided Biopsy Techniques. In: Aydiner, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 63-94. https://doi.org/10.1007/978-3-030-04606-4_5
39. Onder, S., and Yavuz, E. (2019) Liquid Biopsy in Breast Carcinoma. In: Aydiner, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 247-252. https://doi.org/10.1007/978-3-030-04606-4_18
40. Duffy, M. J., Evoy, D., and McDermott, E. W. (2010) CA 15-3: uses and limitation as a biomarker for breast cancer. *Clin. Chim. Acta*. 411 (23-24), 1869-1874. <https://doi.org/10.1016/j.cca.2010.08.039>
41. Duffy, M. J., Shering, S., Sherry, F., McDermott, E., and O'Higgins, N. (2000) CA 15-3: a prognostic marker in breast cancer. *Int. J. Biol. Markers*. 15 (4), 330-333. <https://doi.org/10.1177/172460080001500410>
42. Gion, M., Mione, R., Leon, A. E., Lüftner, D., Molina, R., Possinger, K., and Robertson, J. F. (2001) CA27.29: a valuable marker for breast cancer management. A confirmatory multicentric study on 603 cases. *Eur. J. Cancer*. 37 (3), 355-363. [https://www.ejancer.com/article/S0959-8049\(00\)00396-8/](https://www.ejancer.com/article/S0959-8049(00)00396-8/)
43. Bon, G. G., von Mensdorff-Pouilly, S., Kenemans, P., van Kamp, G. J., Verstraeten, R. A., Hilgers, J., Meijer, S., and Vermorken, J. B. (1997) Clinical and technical evaluation of ACS BR serum assay of MUC1 gene-derived glycoprotein in breast cancer, and comparison with CA 15-3 assays. *Clin Chem*. 43 (4), 585-593. <https://doi.org/10.1093/clinchem/43.4.585>
44. Burchell, J. M., Beatson, R., Graham, R., Taylor-Papadimitriou, J., and Tajadura-Ortega, V. (2018) O-linked mucin-type glycosylation in breast cancer. *Biochem. Soc. Trans.* 46 (4), 779-788. <https://doi.org/10.1042/BST20170483>
45. Li, X., Xu, Y., and Zhang, L. (2019) Serum CA153 as biomarker for cancer and noncancer diseases. *Prog. Mol. Biol. Transl. Sci.* 162, 265-276. <https://doi.org/10.1016/bs.pmbts.2019.01.005>
46. Price, M. R., Rye, P. D., Petrakou, E., Murray, A., Brady, K., Imai, S., Haga, S., Kiyozuka, Y., Schol, D., Meulenbroek, M. F. A., Snijdwint, F. G. M., von Mensdorff-Pouilly, S., Verstraeten, R. A., Kenemans, K., Blockzijl, A., Nilsson, N., Nilsson, O., Reddish, R., Suresh, M. R., Koganty, K., Fortier, S., Baron, B., Berg, A., Longenecker, M. B., Hilken, H., Boer, M., Karanikas, K., McKenzie, I. F. C., Galanina, G., Simeoni, L. A., Ter-Grigoryan, A. G., Belyanchikov, I. M., Bovin, N. V., Cao, Y., Karsten, U., Dai, J., Allard, W. J., Davis, G., Yeung, K. K., Hanisch, F.-G., Lloyd, K. O., Kudryashov, V., Sikut, R., Sikut, A., Zhang, K., Baekström, D., Hansson, G. C., Reis, C. A., Hassan, H., Bennett, E. P., Claussen, H., Norum, L., Varaas, T., Kierulf, B., Nustad, K., Ciborowski, P., Konitzki, W. M., Magarian- Blander, J., Finn, O. J., and

- Hilgers, J. (1997) Summary report on the ISOBM TD-4 Workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin. *Tumor Biol.* 19 (Suppl. 1), 1-20. <https://doi.org/10.1159/000056500>
47. Takeuchi, H. , Kato, K. , Denda-Nagai, K. , Hanisch, F.-G., Clausen, H., and Irimura, T. (2002) The epitope recognized by the unique anti-MUC1 monoclonal antibody MY.1E12 involves sialyl alpha 2-3galactosyl beta 1-3N-acetylgalactosaminide linked to a distinct threonine residue in the MUC1 tandem repeat. *J. Immunol. Methods.* 270 (2), 199-209. [https://doi.org/10.1016/S0022-1759\(02\)00298-3](https://doi.org/10.1016/S0022-1759(02)00298-3)
 48. Zhou, D. , Xu, L., Huang, W., and Tonn, T. (2018) Epitopes of MUC1 Tandem Repeats in Cancer as Revealed by Antibody Crystallography: Toward Glycopeptide Signature-Guided Therapy. *Molecules.* 23 (6), 1326. <https://doi.org/10.3390/molecules23061326>
 49. Grinstead, J. S., Schuman, J. T., and Campbell, A. P. (2003) Epitope mapping of antigenic MUC1 peptides to breast cancer antibody fragment B27.29: a heteronuclear NMR study. *Biochemistry.* 42 (48), 14293-14305. <https://doi.org/10.1021/bi0301237>
 50. Crystal, J., Mella-Catinchi, J., Xu, K., and Weingrad, D. (2022) Current Surgical Innovations in the Treatment of Breast Cancer. In: Mayrovitz, H. N. (ed.) *Breast Cancer*. Brisbane, Australia, Exon Publications, pp. 43-62. <https://doi.org/10.36255/exon-publications-breast-cancer-surgical-innovation>
 51. McAuliffe, P. (2019) Ductal Carcinoma In Situ. In: Aydiner, A., İğci, A., and Soran, A. (eds) *Breast Disease: Diagnosis and Pathology*. 2nd ed. Cham, Switzerland, Springer Cham, pp. 115-123. https://doi.org/10.1007/978-3-030-04606-4_8
 52. Tecalco-Cruz, A. C., Ramírez-Jarquín, J. O., Macías-Silva, M., Sosa-Garrocho, M., and López-Camarillo, C. (2022) Novel Breast Cancer Treatment by Targeting Estrogen Receptor-Alpha Stability Using Proteolysis-Targeting Chimeras (PROTACs) Technology. In: Mayrovitz, H. N. (ed.) *Breast Cancer*. Brisbane, Australia, Exon Publications, pp. 179-194. <https://doi.org/10.36255/exon-publications-breast-cancer-protacs>
 53. Li, Z., Yang, D., Guo, T., and Lin, M. (2022) Advances in MUC1-Mediated Breast Cancer Immunotherapy. *Biomolecules.* 12 (7), 952. <https://doi.org/10.3390/biom12070952>
 54. Weinberg, R. A. (2014) The Rational Treatment of Cancer. In: *The Biology of Cancer*. 2nd ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 797-876.
 55. Varki, A., and Kornfeld, S. (2022) Historical Background and Overview. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 1. <https://www.ncbi.nlm.nih.gov/books/NBK579927/>
 56. Gagneux, P., Hennet, T., and Varki, A. (2022) Biological Functions of Glycans. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 7. <https://www.ncbi.nlm.nih.gov/books/NBK579984/>
 57. Seeberger, P. H. (2022) Monosaccharide Diversity. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 2. <https://www.ncbi.nlm.nih.gov/books/NBK579981/>
 58. Merry, C. L. R., Lindahl, U., Couchman, J., and Esko, J. D. (2022) Proteoglycans and Sulfated Glycosaminoglycans. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 17. <https://www.ncbi.nlm.nih.gov/books/NBK579925/>
 59. Stanley, P., Moremen, K. W., Lewis, N. E., Taniguchi, N., and Aebi, M. (2022) N-Glycans. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi,

- M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 9. <https://www.ncbi.nlm.nih.gov/books/NBK579964/>
60. Brockhausen, I., Wandall, H. H., Ten Hagen, K. G., and Stanley, P. (2022) O-GalNAc Glycans. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 10. <https://www.ncbi.nlm.nih.gov/books/NBK579921/>
 61. Gadi, M. R., Chen, C., Bao, S., Wang, S., Guo, Y., Han, J., Xiao, W., and Li, L. (2023) Convergent chemoenzymatic synthesis of O-GalNAc rare cores 5, 7, 8 and their sialylated forms. *Chem. Sci.* 14 (7),1837-1843. <https://doi.org/10.1039/D2SC06925C>
 62. Zachara, N. E., Akimoto, Y., Boyce, M., and Hart, G. W. (2022) The O-GlcNAc Modification. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 19. <https://www.ncbi.nlm.nih.gov/books/NBK579950/>
 63. Freeze, H. H., Boyce, M., Zachara, N. E., Hart, G. W., and Schnaar, R. L. (2022) Glycosylation Precursors. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 5. <https://www.ncbi.nlm.nih.gov/books/NBK579932/>
 64. Colley, K. J., Varki, A., Haltiwanger, R. S., and Kinoshita, T. (2022) Cellular Organization of Glycosylation. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 4. <https://www.ncbi.nlm.nih.gov/books/NBK579926/>
 65. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) Intracellular Compartments and Protein Sorting. In: *Molecular biology of the cell*. 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 641-694.
 66. Ten Hagen, K. G., Fritz, T. A., and Tabak, L. A. (2003) All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*. 13 (1), 1R-16R. <https://doi.org/10.1093/glycob/cwg007>
 67. Choi, J., Wagner, L. J. S., Timmermans, S. B. P. E., Malaker, S. A., Schumann, B., Gray, M. A., Debets, M. F., Takashima, M., Gehring, J., and Bertozzi, C. R. (2019) Engineering Orthogonal Polypeptide GalNAc-Transferase and UDP-Sugar Pairs. *J. Am. Chem. Soc.* 141 (34), 13442-13453. <https://doi.org/10.1021/jacs.9b04695>
 68. de Las Rivas, M., Lira-Navarrete, E., Gerken, T. A., and Hurtado-Guerrero, R. (2019) Polypeptide GalNAc-Ts: from redundancy to specificity. *Curr. Opin. Struct. Biol.* 56, 87-96. <https://doi.org/10.1016/j.sbi.2018.12.007>
 69. Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moinova, H., Ravi, L., Markowitz, S. D., Shen, W., Patel, H., and Tabak, L. A. (2011) Emerging paradigms for the initiation of mucin-type protein O-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J. Biol. Chem.* 286 (16),14493-507. [https://www.jbc.org/article/S0021-9258\(20\)51541-X/](https://www.jbc.org/article/S0021-9258(20)51541-X/)
 70. Lira-Navarrete, E., de Las Rivas, M., Compañón, I., Pallarés, M. C., Kong, Y., Iglesias-Fernández, J., Bernardes, G. J., Peregrina, J. M., Rovira, C., Bernadó, P., Bruscolini, P., Clausen, H., Lostao, A., Corzana, F., and Hurtado-Guerrero, R. (2015) Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation. *Nat. Commun.* 6, 6937. <https://doi.org/10.1038/ncomms7937>

71. Ballard, C. J., Paserba, M. R., Paul Daniel, E. J., Hurtado-Guerrero, R., and Gerken, T. A. (2023) Polypeptide N-acetylgalactosaminyltransferase (GalNAc-T) isozyme surface charge governs charge substrate preferences to modulate mucin type O-glycosylation. *Glycobiology*. 33 (10), 817-836. <https://doi.org/10.1093/glycob/cwad066>
72. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*. 22 (6), 736-756. <https://doi.org/10.1093/glycob/cwr182>
73. Li, X., Wang, J., Li, W., Xu, Y., Shao, D., Xie, Y., Xie, W., Kubota, T., Narimatsu, H., and Zhang, Y. (2012) Characterization of ppGalNAc-T18, a member of the vertebrate-specific Y subfamily of UDP-N-acetyl- α -D-galactosamine:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*. 22 (5), 602-615. <https://doi.org/10.1093/glycob/cwr179>
74. de Las Rivas, M., Lira-Navarrete, E., Paul Daniel, E. J., Compañón, I., Coelho, H., Diniz, A., Jiménez-Barbero, J., Peregrina, J. M., Clausen, H., Corzana, F., Marcelo, F., Jiménez-Osés, G., Gerken, T. A., and Hurtado-Guerrero, R. (2017) The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences. *Nat. Commun.* 8 (1), 1959. <https://doi.org/10.1038/s41467-017-02006-0>
75. Taujale, R., Venkat, A., Huang, L. C., Zhou, Z., Yeung, W., Rasheed, K. M., Li, S., Edison, A. S., Moremen, K. W., and Kannan, N. (2020) Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. *Elife*. 9, e54532. <https://doi.org/10.7554/eLife.54532>
76. Sanz-Martinez, I., Pereira, S., Merino, P., Corzana, F., and Hurtado-Guerrero, R. (2020) Molecular Recognition of GalNAc in Mucin-Type O-Glycosylation. *Acc. Chem. Res.* 56 (5), 548-560. <https://doi.org/10.1021/acs.accounts.2c00723>
77. Rini, J. M., Moremen, K. W., Davis, B. G., and Esko, J. D. (2022) Glycosyltransferases and Glycan-Processing Enzymes. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 6. <https://www.ncbi.nlm.nih.gov/books/NBK579908/>
78. Kubota, T., Shiba, T., Sugioka, S., Furukawa, S., Sawaki, H., Kato, R., Wakatsuki, S., and Narimatsu, H. (2006) Structural basis of carbohydrate transfer activity by human UDP-GalNAc: polypeptide α -N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J. Mol. Biol.* 359 (3), 708-727. <https://doi.org/10.1016/j.jmb.2006.03.061>
79. Fritz, T. A., Raman, J., and Tabak, L. A. (2006) Dynamic association between the catalytic and lectin domains of human UDP-GalNAc:polypeptide α -N-acetylgalactosaminyltransferase-2. *J. Biol. Chem.* 281 (13), 8613-8619. [https://www.jbc.org/article/S0021-9258\(19\)56563-2/](https://www.jbc.org/article/S0021-9258(19)56563-2/)
80. Lira-Navarrete, E., Iglesias-Fernández, J., Zandberg, W. F., Compañón, I., Kong, Y., Corzana, F., Pinto, B. M., Clausen, H., Peregrina, J. M., Voadlo, D. J., Rovira, C., and Hurtado-Guerrero, R. (2014) Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. *Angew. Chem. Int. Ed. Engl.* 53 (31), 8206-8210. <https://doi.org/10.1002/anie.201402781>
81. Revoredo, L., Wang, S., Bennett, E. P., Clausen, H., Moremen, K. W., Jarvis, D. L., Ten Hagen, K. G., Tabak, L. A., and Gerken, T. A. (2016) Mucin-type O-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. *Glycobiology*. 26 (4), 360-376. <https://doi.org/10.1093/glycob/cwv108>
82. de Las Rivas, M., Paul Daniel, E. J., Coelho, H., Lira-Navarrete, E., Raich, L., Compañón, I., Diniz, A., Lagartera, L., Jiménez-Barbero, J., Clausen, H., Rovira, C., Marcelo, F., Corzana, F., Gerken, T. A., and Hurtado-Guerrero, R. (2018) Structural and Mechanistic Insights into the Catalytic-Domain-Mediated Short-Range

- Glycosylation Preferences of GalNAc-T4. *ACS Cent. Sci.* 4 (9), 1274-1290. <https://doi.org/10.1021/acscentsci.8b00488>
83. Pedersen, J. W., Bennett, E. P., Schjoldager, K. T., Meldal, M., Holmér, A. P., Blixt, O., Cló, E., Levery, S. B., Clausen, H., and Wandall, H. H. (2011) Lectin domains of polypeptide GalNAc transferases exhibit glycopeptide binding specificity. *J. Biol. Chem.* 286 (37), 32684-32696. [https://www.jbc.org/article/S0021-9258\(20\)50977-0/](https://www.jbc.org/article/S0021-9258(20)50977-0/)
 84. Bagdonaite, I., Pallesen, E. M. H., Nielsen, M. I., Bennett, E. P., and Wandall, H. H. (2021). Mucin-Type O-GalNAc Glycosylation in Health and Disease. In: Lauc, G., and Trbojević-Akmačić, I. (eds) *The Role of Glycosylation in Health and Disease*. Cham, Switzerland, Springer Cham, pp. 25-60. https://doi.org/10.1007/978-3-030-70115-4_2
 85. Radhakrishnan, P., Dabelsteen, S., Madsen, F. B., Francavilla, C., Kopp, K. L., Steentoft, C., Vakhrushev, S. Y., Olsen, J. V., Hansen, L., Bennett, E. P., Woetmann, A., Yin, G., Chen, L., Song, H., Bak, M., Hlady, R. A., Peters, S. L., Opavsky, R., Thode, C., Qvortrup, K., Schjoldager, K. T., Clausen, H., Hollingsworth, M. A., and Wandall, H. H. (2014) Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc. Natl. Acad. Sci. USA.* 111 (39), E4066-E4075. <https://doi.org/10.1073/pnas.1406619111>
 86. Lavrsen, K., Dabelsteen, S., Vakhrushev, S. Y., Levann, A. M. R., Haue, A. D., Dylander, A., Mandel, U., Hansen, L., Frödin, M., Bennett, E. P., and Wandall, H. H. (2018) De novo expression of human polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) in colon adenocarcinoma inhibits the differentiation of colonic epithelium. *J. Biol. Chem.* 293 (4), 1298-1314. [https://www.jbc.org/article/S0021-9258\(20\)39131-6/](https://www.jbc.org/article/S0021-9258(20)39131-6/)
 87. Bagdonaite, I., Pallesen, E. M., Ye, Z., Vakhrushev, S. Y., Marinova, I. N., Nielsen, M. I., Kramer, S. H., Pedersen, S. F., Joshi, H. J., Bennett, E. P., Dabelsteen, S., and Wandall, H. H. (2020) O-glycan initiation directs distinct biological pathways and controls epithelial differentiation. *EMBO Rep.* 21 (6), e48885. <https://doi.org/10.15252/embr.201948885>
 88. Dabelsteen, S., Pallesen, E. M. H., Marinova, I. N., Nielsen, M. I., Adamopoulou, M., Rømer, T. B., Levann, A., Andersen, M. M., Ye, Z., Thein, D., Bennett, E. P., Büll, C., Moons, S. J., Boltje, T., Clausen, H., Vakhrushev, S. Y., Bagdonaite, I., and Wandall, H. H. (2020) Essential Functions of Glycans in Human Epithelia Dissected by a CRISPR-Cas9-Engineered Human Organotypic Skin Model. *Dev. Cell.* 54 (5), 669-684.e7. <https://doi.org/10.1016/j.devcel.2020.06.039>
 89. Goth, C. K., Halim, A., Khetarpal, S. A., Rader, D. J., Clausen, H., and Schjoldager, K. T. (2015) A systematic study of modulation of ADAM-mediated ectodomain shedding by site-specific O-glycosylation. *Proc. Natl. Acad. Sci. USA.* 112 (47), 14623-14628. <https://doi.org/10.1073/pnas.1511175112>
 90. Cummings, R. D., Chiffoleau, E., van Kooyk, Y., and McEver, R. P. (2022) C-Type Lectins. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 34. <https://www.ncbi.nlm.nih.gov/books/NBK579916/>
 91. Angata, T., von Gunten, S., Schnaar, R. L., and Varki, A. (2022) I-Type Lectins. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 35. <https://www.ncbi.nlm.nih.gov/books/NBK579966/>
 92. Macauley, M. S., Crocker, P. R., and Paulson, J. C. (2014) Siglec-mediated regulation of immune cell function in disease. *Nat. Rev. Immunol.* 14 (10), 653-666. <https://doi.org/10.1038/nri3737>
 93. Cummings, R. D., Liu, F. T., Rabinovich, G. A., Stowell, S. R., and Vasta, G. R. (2022) Galectins In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi,

- M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 36. <https://www.ncbi.nlm.nih.gov/books/NBK579987/>
94. Liu, F. T., and Stowell, S.R. (2023) The role of galectins in immunity and infection. *Nat. Rev. Immunol.* 23 (8), 479-494. <https://doi.org/10.1038/s41577-022-00829-7>
 95. Nielsen, M. I., Stegmayr, J., Grant, O. C., Yang, Z., Nilsson, U. J., Boos, I., Carlsson, M. C., Woods, R. J., Unverzagt, C., Leffler, H., and Wandall, H. H. (2018) Galectin binding to cells and glycoproteins with genetically modified glycosylation reveals galectin-glycan specificities in a natural context. *J. Biol. Chem.* 293 (52), 20249-20262. [https://www.jbc.org/article/S0021-9258\(20\)31897-4/](https://www.jbc.org/article/S0021-9258(20)31897-4/)
 96. Li, J., Li, X., and Guan, F. (2023) What are the diagnostic capabilities of glycans for breast cancer? *Expert Rev. Mol. Diagn.* 23 (1), 1-7. <https://doi.org/10.1080/14737159.2023.2173577>
 97. Lopes, N., Correia, V. G., Palma, A. S., and Brito, C. (2021) Cracking the Breast Cancer Glyco-Code through Glycan-Lectin Interactions: Targeting Immunosuppressive Macrophages. *Int. J. Mol. Sci.* 22 (4), 1972. <https://doi.org/10.3390/ijms22041972>
 98. Liang, D., Gao, Q., Meng, Z., Li, W., Song, J., and Xue, K. (2023) Glycosylation in breast cancer progression and mammary development: Molecular connections and malignant transformations. *Life Sci.* 326, 121781. <https://doi.org/10.1016/j.lfs.2023.121781>
 99. Chang, X., Obianwuna, U. E., Wang, J., Zhang, H., Qi, G., Qiu, K., and Wu, S. (2023) Glycosylated proteins with abnormal glycosylation changes are potential biomarkers for early diagnosis of breast cancer. *Int. J. Biol. Macromol.* 236, 123855. <https://doi.org/10.1016/j.ijbiomac.2023.123855>
 100. Tu, C. F., Wu, M. Y., Lin, Y. C., Kannagi, R., and Yang, R. B. (2017) FUT8 promotes breast cancer cell invasiveness by remodeling TGF- β receptor core fucosylation. *Breast Cancer Res.* 19 (1), 111. <https://doi.org/10.1186/s13058-017-0904-8>
 101. Gu, Y., Mi, W., Ge, Y., Liu, H., Fan, Q., Han, C., Yang, J., Han, F., Lu, X., and Yu, W. (2010) GlcNAcylation plays an essential role in breast cancer metastasis. *Cancer Res.* 70 (15), 6344-6351. <https://doi.org/10.1158/0008-5472.CAN-09-1887>
 102. Bellis, S. L., Reis, C. A., Varki, A., Kannagi, R., and Stanley, P. (2022) Glycosylation Changes in Cancer. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 47. <https://www.ncbi.nlm.nih.gov/books/NBK579911/>
 103. Magalhães, A., Duarte, H. O., and Reis, C. A. (2021) The role of O-glycosylation in human disease. *Mol. Aspects. Med.* 79, 100964. <https://doi.org/10.1016/j.mam.2021.100964>
 104. Cazet, A., Julien, S., Bobowski, M., Burchell, J., and Delannoy, P. (2010) Tumour-associated carbohydrate antigens in breast cancer. *Breast Cancer Res.* 12 (3), 204. <https://doi.org/10.1186/bcr2577>
 105. Lewis, A. L., Chen, X., Schnaar, R. L., and Varki, A. (2022) Sialic Acids and Other Nonulosonic Acids. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 15. <https://www.ncbi.nlm.nih.gov/books/NBK579976/>
 106. Liu, H., Ma, L., Lin, J., Cao, B., Qu, D., Luo, C., Huang, W., Han, L., Xu, H., Wu, Z., Xu, R., and Zhang, D. (2020) Advances in molecular mechanisms of drugs affecting abnormal glycosylation and metastasis of breast cancer. *Pharmacol Res.* 155, 104738. <https://doi.org/10.1016/j.phrs.2020.104738>

107. Beatson, R., Tajadura-Ortega, V., Achkova, D., Picco, G., Tsourouktsoglou, T. D., Klausning, S., Hillier, M., Maher, J., Noll, T., Crocker, P. R., Taylor-Papadimitriou, J., and Burchell, J. M. (2016) The mucin MUC1 modulates the tumor immunological microenvironment through engagement of the lectin Siglec-9. *Nat. Immunol.* 17 (11), 1273-1281. <https://doi.org/10.1038/ni.3552>
108. Tanida, S., Akita, K., Ishida, A., Mori, Y., Toda, M., Inoue, M., Ohta, M., Yashiro, M., Sawada, T., Hirakawa, K., and Nakada, H. (2013) Binding of the sialic acid-binding lectin, Siglec-9, to the membrane mucin, MUC1, induces recruitment of β -catenin and subsequent cell growth. *J. Biol. Chem.* 288 (44), 31842-31852. [https://www.jbc.org/article/S0021-9258\(20\)48655-7/](https://www.jbc.org/article/S0021-9258(20)48655-7/)
109. Hudak, J. E., Canham, S. M., and Bertozzi, C. R. (2014) Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. *Nat. Chem. Biol.* 10 (1), 69-75. <https://doi.org/10.1038/nchembio.1388>
110. van Vliet, S. J., van Liempt, E., Saeland, E., Aarnoudse, C. A., Appelmelk, B., Irimura, T., Geijtenbeek, T. B., Blixt, O., Alvarez, R., van Die, I., and van Kooyk, Y. (2005) Carbohydrate profiling reveals a distinctive role for the C-type lectin MGL in the recognition of helminth parasites and tumor antigens by dendritic cells. *Int. Immunol.* 17 (5), 661-669. <https://doi.org/10.1093/intimm/dxh246>
111. Beatson, R., Maurstad, G., Picco, G., Arulappu, A., Coleman, J., Wandell, H. H., Clausen, H., Mandel, U., Taylor-Papadimitriou, J., Sletmoen, M., and Burchell, J. M. (2015) The Breast Cancer-Associated Glycoforms of MUC1, MUC1-Tn and sialyl-Tn, Are Expressed in COSMC Wild-Type Cells and Bind the C-Type Lectin MGL. *PLoS One.* 10 (5), e0125994. <https://doi.org/10.1371/journal.pone.0125994>
112. van Vliet, S. J., Gringhuis, S. I., Geijtenbeek, T. B., and van Kooyk, Y. (2006) Regulation of effector T cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45. *Nat. Immunol.* 7 (11), 1200-1208. <https://doi.org/10.1038/ni1390>
113. Gay, L. J., and Felding-Habermann, B. (2011) Contribution of platelets to tumour metastasis. *Nat. Rev. Cancer.* 11 (2), 123-134. <https://doi.org/10.1038/nrc3004>
114. Girotti, M. R., Salatino, M., Dalotto-Moreno, T., and Rabinovich, G. A. (2020) Sweetening the hallmarks of cancer: Galectins as multifunctional mediators of tumor progression. *J. Exp. Med.* 217 (2), e20182041. <https://doi.org/10.1084/jem.20182041>
115. Santos, S. N., Junqueira, M. S., Francisco, G., Vilanova, M., Magalhães, A., Dias Baruffi, M., Chammas, R., Harris, A. L., Reis, C. A., and Bernardes, E. S. (2016) O-glycan sialylation alters galectin-3 subcellular localization and decreases chemotherapy sensitivity in gastric cancer. *Oncotarget.* 7 (50), 83570-83587. <https://doi.org/10.18632/oncotarget.13192>
116. Piyush, T., Chacko, A. R., Sindrewicz, P., Hilken, J., Rhodes, J. M., and Yu, L. G. (2017) Interaction of galectin-3 with MUC1 on cell surface promotes EGFR dimerization and activation in human epithelial cancer cells. *Cell Death Differ.* 24 (11), 1937-1947. <https://doi.org/10.1038/cdd.2017.119>
117. Zhao, Q., Guo, X., Nash, G. B., Stone, P. C., Hilken, J., Rhodes, J. M., and Yu, L. G. (2009) Circulating galectin-3 promotes metastasis by modifying MUC1 localization on cancer cell surface. *Cancer Res.* 69 (17), 6799-6806. <https://doi.org/10.1158/0008-5472.CAN-09-1096>
118. Zhao, Q., Barclay, M., Hilken, J., Guo, X., Barrow, H., Rhodes, J. M., and Yu, L. G. (2010) Interaction between circulating galectin-3 and cancer-associated MUC1 enhances tumour cell homotypic aggregation and prevents anoikis. *Mol. Cancer.* 9, 154. <https://doi.org/10.1186/1476-4598-9-154>
119. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) Cell Junctions and the Extracellular Matrix. In: *Molecular biology of the cell.* 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 1035-1091.
120. Paszek, M. J., DuFort, C. C., Rossier, O., Bainer, R., Mouw, J. K., Godula, K., Hudak, J. E., Lakins, J. N., Wijekoon, A. C., Cassereau, L., Rubashkin, M. G., Magbanua, M. J., Thorn, K. S., Davidson, M. W., Rugo, H. S., Park, J. W., Hammer, D. A., Giannone,

- G., Bertozzi, C. R., and Weaver, V. M. (2014) The cancer glycocalyx mechanically primes integrin-mediated growth and survival. *Nature*. 511 (7509), 319-325. <https://doi.org/10.1038/nature13535>
121. Simpson, M., Schaefer, L., Hascall, V., and Esko, J. D. (2022) Hyaluronan. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 16. <https://www.ncbi.nlm.nih.gov/books/NBK579913/>
 122. Rivinoja, A., Kokkonen, N., Kellokumpu, I., and Kellokumpu, S. (2006) Elevated Golgi pH in breast and colorectal cancer cells correlates with the expression of oncofetal carbohydrate T-antigen. *J. Cell Physiol.* 208 (1), 167-174. <https://doi.org/10.1002/jcp.20653>
 123. Rivinoja, A., Hassinen, A., Kokkonen, N., Kauppila, A., and Kellokumpu, S. (2009) Elevated Golgi pH impairs terminal N-glycosylation by inducing mislocalization of Golgi glycosyltransferases. *J. Cell Physiol.* 220 (1), 144-154. <https://doi.org/10.1002/jcp.21744>
 124. Gribbestad, I. S., Petersen, S. B., Fjøsne, H. E., Kvinnsland, S., and Krane, J. (1994) ¹H NMR spectroscopic characterization of perchloric acid extracts from breast carcinomas and non-involved breast tissue. *NMR Biomed.* 7 (4), 181-194. <https://doi.org/10.1002/nbm.1940070405>
 125. Nakajima, K., Kitazume, S., Angata, T., Fujinawa, R., Ohtsubo, K., Miyoshi, E., and Taniguchi, N. (2010) Simultaneous determination of nucleotide sugars with ion-pair reversed-phase HPLC. *Glycobiology*. 20 (7), 865-871. <https://doi.org/10.1093/glycob/cwq044>
 126. Burchell, J., Poulson, R., Hanby, A., Whitehouse, C., Cooper, L., Clausen, H., Miles, D., and Taylor-Papadimitriou, J. (1999) An alpha2,3 sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas. *Glycobiology*. 9 (12), 1307-1311. <https://doi.org/10.1093/glycob/9.12.1307>
 127. Fang, R., Xu, F., Shi, H., Wu, Y., Cao, C., Li, H., Ye, K., Zhang, Y., Liu, Q., Zhang, S., Zhang, W., and Ye, L. (2020) LAMTOR5 raises abnormal initiation of O-glycosylation in breast cancer metastasis via modulating GALNT1 activity. *Oncogene*. 39, (11), 2290-2304. <https://doi.org/10.1038/s41388-019-1146-2>
 128. Qiu, H., Xu, X., Liu, M., Wang, Z., Yuan, Y., Liu, C., Xu, L., and Wu, S. (2017) RNA interference-mediated silencing of ppGalNAc-T1 and ppGalNAc-T2 inhibits invasion and increases chemosensitivity potentially by reducing terminal α2,3 sialylation and MMP14 expression in triple-negative breast cancer cells. *Mol. Med. Rep.* 15 (6), 3724-3734. <https://doi.org/10.3892/mmr.2017.6449>
 129. Wu, Q., Zhang, C., Zhang, K., Chen, Q., Wu, S., Huang, H., Huang, T., Zhang, N., Wang, X., Li, W., Liu, Y., and Zhang, J. (2021) ppGalNAc-T4-catalyzed O-Glycosylation of TGF-β type II receptor regulates breast cancer cells metastasis potential. *J. Biol. Chem.* 296, 100119. <https://doi.org/10.1074/jbc.RA120.016345>
 130. Potapenko, I. O., Haakensen, V. D., Lüders, T., Helland, A., Bukholm, I., Sørli, T., Kristensen, V. N., Lingjaerde, O. C., and Børresen-Dale, A. L. (2010) Glycan gene expression signatures in normal and malignant breast tissue; possible role in diagnosis and progression. *Mol. Oncol.* 4 (2), 98-118. <https://doi.org/10.1016/j.molonc.2009.12.001>
 131. Nomoto, M., Izumi, H., Ise, T., Kato, K., Takano, H., Nagatani, G., Shibao, K., Ohta, R., Imamura, T., Kuwano, M., Matsuo, K., Yamada, Y., Itoh, H., and Kohno, K. (1999) Structural basis for the regulation of UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyl transferase-3 gene expression in adenocarcinoma cells. *Cancer Res.* 59 (24), 6214-6222. <https://aacrjournals.org/cancerres/article/59/24/6214/505754/>
 132. Niang, B., Jin, L., Chen, X., Guo, X., Zhang, H., Wu, Q., Padhiar, A. A., Xiao, M., Fang, D., and Zhang, J. (2016) GalNAc-T4 putatively modulates the estrogen regulatory

- network through FOXA1 glycosylation in human breast cancer cells. *Mol. Cell Biochem.* 411, 393-402. <https://doi.org/10.1007/s11010-015-2601-1>
133. Wu, Q., Xie, X., Zhang, K., Niang, B., Liu, Y., Zhang, C., Huang, T., Huang, H., Li, W., Zhang, J., and Liu, Y. (2021) Reduced expression of ppGalNAc-T4 promotes proliferation of human breast cancer cells. *Cell Biol. Int.* 45 (2), 320-333. <https://doi.org/10.1002/cbin.11488>
 134. Park, J. H., Nishidate, T., Kijima, K., Ohashi, T., Takegawa, K., Fujikane, T., Hirata, K., Nakamura, Y., and Katagiri, T. (2010) Critical roles of mucin 1 glycosylation by transactivated polypeptide N-acetylgalactosaminyltransferase 6 in mammary carcinogenesis. *Cancer Res.* 70 (7), 2759-2769. <https://doi.org/10.1158/0008-5472.CAN-09-3911>
 135. Berois, N., Mazal, D., Ubillos, L., Trajtenberg, F., Nicolas, A., Sastre-Garau, X., Magdelenat, H., and Osinaga, E. (2006) UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker. *J. Histochem. Cytochem.* 54 (3), 317-328. <https://doi.org/10.1369/jhc.5A6783.2005>
 136. Liu, C., Li, Z., Xu, L., Shi, Y., Zhang, X., Shi, S., Hou, K., Fan, Y., Li, C., Wang, X., Zhou, L., Liu, Y., Qu, X., and Che, X. (2020) GALNT6 promotes breast cancer metastasis by increasing mucin-type O-glycosylation of α 2M. *Aging.* 12 (12), 11794-11811. <https://doi.org/10.18632/aging.103349>
 137. Mao, Y., Zhang, Y., Fan, S., Chen, L., Tang, L., Chen, X., and Lyu, J. (2019) GALNT6 Promotes Tumorigenicity and Metastasis of Breast Cancer Cell via β -catenin/MUC1-C Signaling Pathway. *Int. J. Biol. Sci.* 15 (1), 169-182. <https://www.ijbs.com/v15p0169.htm>
 138. Freire, T., Berois, N., S  nora, C., Varangot, M., Barrios, E., and Osinaga, E. (2006) UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (ppGalNAc-T6) mRNA as a potential new marker for detection of bone marrow-disseminated breast cancer cells. *Int. J. Cancer.* 119 (6), 1383-1388. <https://doi.org/10.1002/ijc.21959>
 139. Patani, N., Jiang, W., and Mokbel, K. (2008) Prognostic utility of glycosyltransferase expression in breast cancer. *Cancer Genomics Proteomics.* 5 (6), 333-340. <https://cgp.iiarjournals.org/content/5/6/333>
 140. Kimura, R., Yoshimaru, T., Matsushita, Y., Matsuo, T., Ono, M., Park, J. H., Sasa, M., Miyoshi, Y., Nakamura, Y., and Katagiri, T. (2020) The GALNT6-LGALS3BP axis promotes breast cancer cell growth. *Int. J. Oncol.* 56 (2), 581-595. <https://doi.org/10.3892/ijo.2019.4941>
 141. Liesche, F., K  lbl, A. C., Ilmer, M., Hutter, S., Jeschke, U., and Andergassen, U. (2016) Role of N-acetylgalactosaminyltransferase 6 in early tumorigenesis and formation of metastasis. *Mol. Med. Rep.* 13 (5), 4309-4314. <https://doi.org/10.3892/mmr.2016.5044>
 142. Andergassen, U., Liesche, F., K  lbl, A. C., Ilmer, M., Hutter, S., Friese, K., and Jeschke, U. (2015) Glycosyltransferases as Markers for Early Tumorigenesis. *Biomed. Res. Int.* 2015, 792672. <https://doi.org/10.1155/2015/792672>
 143. Gadwal, A., Purohit, P., Khokhar, M., Vishnoi, J. R., Pareek, P., Choudhary, R., Elhence, P., Banerjee, M., and Sharma, P. (2024) GALNT6, GALNT14, and Gal-3 in association with GDF-15 promotes drug resistance and stemness of breast cancer via β -catenin axis. *Growth Factors.* 42 (2), 84-100. <https://doi.org/10.1080/08977194.2024.2368907>
 144. Xue, J. and Yu, Q. (2022) Knockdown of GALNT7 promotes cell apoptosis and autophagy of breast cancer cells by inactivation of STAT3. *Eur. J. Gynaecol. Oncol.* 43 (4), 79-85. <https://www.ejgo.net/articles/10.22514/ejgo.2022.035>
 145. Berkel, C., and Cacan, E. (2023) The expression of O-linked glycosyltransferase GALNT7 in breast cancer is dependent on estrogen-, progesterone-, and HER2-receptor status, with prognostic implications. *Glycoconj. J.* 40 (6), 631-644. <https://doi.org/10.1007/s10719-023-10137-4>

146. Li, J., Wan, X., Xie, D., Yuan, H., Pei, Q., Luo, Y., Chen, Y., Xian, J., and Ye, T. (2023) SPDEF enhances cancer stem cell-like properties and tumorigenesis through directly promoting GALNT7 transcription in luminal breast cancer. *Cell Death Dis.* 14 (8), 569. <https://doi.org/10.1038/s41419-023-06098-z>
147. Fernandez, E., Ubillos, L., Elgul, N., Festari, M. F., Mazal, D., Pritsch, O., Alonso, I., Osinaga, E., and Berois, N. (2021) Polypeptide-GalNAc-Transferase-13 Shows Prognostic Impact in Breast Cancer. *Cancers.* 13 (22), 5616. <https://doi.org/10.3390/cancers13225616>
148. Gadwal, A., Purohit, P., Khokhar, M., Vishnoi, J. R., Pareek, P., Choudhary, R., Elhence, P., Banerjee, M., and Sharma, P. (2024) GALNT14 in association with GDF-15 promotes stemness and drug resistance through β -catenin signalling pathway in breast cancer. *Mol. Biol. Rep.* 51 (1), 691. <https://doi.org/10.1007/s11033-024-09645-9>
149. Wu, J., Chen, X., Bao, Q., Duan, R., Jin, Y., Shui, Y., Yao, B., Lu, X., Wang, Y., Cui, H., Li, L., Yuan, H., and Ma, C. (2017) Osterix Decreases the Chemosensitivity of Breast Cancer Cells by Upregulating GALNT14. *Cell Physiol. Biochem.* 44 (3), 998-1010. <https://doi.org/10.1159/000485400>
150. Huang, T., Wu, Q., Huang, H., Zhang, C., Wang, L., Wang, L., Liu, Y., Li, W., Zhang, J., and Liu, Y. (2022) Expression of GALNT8 and O-glycosylation of BMP receptor 1A suppress breast cancer cell proliferation by upregulating ER α levels. *Biochim. Biophys. Acta. Gen. Subj.* 1866 (1), 130046. <https://doi.org/10.1016/j.bbagen.2021.130046>
151. Oshi, M., Ziazadeh, D., Wu, R., Chida, K., Yamada, A., Yamamoto, S., Narui, K., Yan, L., Ishikawa, T., Endo, I., and Takabe, K. (2023) GALNT1 Expression Is Associated with Angiogenesis and Is a Prognostic Biomarker for Breast Cancer in Adolescents and Young Adults (AYA). *Cancers.* 15 (13), 3489. <https://doi.org/10.3390/cancers15133489>
152. Milde-Langosch, K., Schütze, D., Oliveira-Ferrer, L., Wikman, H., Müller, V., Lebok, P., Pantel, K., Schröder, C., Witzel, I., Schumacher, U. (2015) Relevance of β Gal- β GalNAc-containing glycans and the enzymes involved in their synthesis for invasion and survival in breast cancer patients. *Breast Cancer Res. Treat.* 151 (3), 515-528. <https://doi.org/10.1007/s10549-015-3425-0>
153. Gill, D. J., Tham, K. M., Chia, J., Wang, S. C., Steentoft, C., Clausen, H., Bard-Chapeau, E. A., and Bard, F. A. (2013) Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. *Proc. Natl. Acad. Sci. USA.* 110 (34), E3152-E3161. <https://doi.org/10.1073/pnas.1305269110>
154. Brooks, S. A., Carter, T. M., Bennett, E. P., Clausen, H., and Mandel, U. (2007) 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 (4), 273-284. <https://doi.org/10.1016/j.acthis.2007.02.009>
155. Raghu, D., Mobley, R. J., Shendy, N. A. M., Perry, C. H., and Abell, A. N. (2019) GALNT3 Maintains the Epithelial State in Trophoblast Stem Cells. *Cell Rep.* 26 (13), 3684-3697.e7. [https://www.cell.com/cell-reports/fulltext/S2211-1247\(19\)30279-7](https://www.cell.com/cell-reports/fulltext/S2211-1247(19)30279-7)
156. Coelho, H., Rivas, M. L., Grosso, A. S., Diniz, A., Soares, C. O., Francisco, R. A., Dias, J. S., Compañón, I., Sun, L., Narimatsu, Y., Vakhrushev, S. Y., Clausen, H., Cabrita, E. J., Jiménez-Barbero, J., Corzana, F., Hurtado-Guerrero, R., and Marcelo, F. (2022) Atomic and Specificity Details of Mucin 1 O-Glycosylation Process by Multiple Polypeptide GalNAc-Transferase Isoforms Unveiled by NMR and Molecular Modeling. *JACS Au.* 2 (3), 631-645. <https://doi.org/10.1021/jacsau.1c00529>
157. Nason, R., Büll, C., Konstantinidi, A., Sun, L., Ye, Z., Halim, A., Du, W., Sørensen, D. M., Durbesson, F., Furukawa, S., Mandel, U., Joshi, H. J., Dworkin, L. A., Hansen, L., David, L., Iverson, T. M., Bensing, B. A., Sullam, P. M., Varki, A., Vries, E., de Haan, C. A. M., Vincentelli, R., Henrissat, B., Vakhrushev, S. Y., Clausen, H., and Narimatsu, Y. (2021) Display of the human mucinome with defined O-glycans by gene engineered cells. *Nat. Commun.* 12 (1), 4070. <https://doi.org/10.1038/s41467-021-24366-4>

158. Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merks, G., van Kessel, A. G., Eiberg, H., Steffensen, R., and Clausen, H. (1998) Cloning of a human UDP-N-acetyl-alpha-D-Galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. *J. Biol. Chem.* 273 (46), 30472-30481. [https://www.jbc.org/article/S0021-9258\(19\)59261-4/](https://www.jbc.org/article/S0021-9258(19)59261-4/)
159. Müller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A. M., and Hanisch, F. G. (1997) Localization of O-glycosylation sites on glycopeptide fragments from lactation-associated MUC1. All putative sites within the tandem repeat are glycosylation targets in vivo. *J. Biol. Chem.* 272 (40), 24780-24793. [https://www.jbc.org/article/S0021-9258\(19\)63472-1/](https://www.jbc.org/article/S0021-9258(19)63472-1/)
160. Müller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A. A., and Hanisch, F. G. (1999) High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells. *J. Biol. Chem.* 274 (26), 18165-18172. [https://www.jbc.org/article/S0021-9258\(19\)87141-7/](https://www.jbc.org/article/S0021-9258(19)87141-7/)
161. Wright, P. K., May, F. E., Darby, S., Saif, R., Lennard, T. W., and Westley, B. R. (2009) Estrogen regulates vesicle trafficking gene expression in EFF-3, EFM-19 and MCF-7 breast cancer cells. *Int. J. Clin. Exp. Pathol.* 2 (5), 463-475. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655148/>
162. Zhang, S., Bai, L., Chen, Q., Ren, Y., Zhang, K., Wu, Q., Huang, H., Li, W., Zhang, Y., Zhang, J., and Liu, Y. (2019) Identification of the O-GalNAcylation site(s) on FOXA1 catalyzed by ppGalNAc-T2 enzyme in vitro. *Biochem. Biophys. Res. Commun.* 514 (1), 157-165. <https://doi.org/10.1016/j.bbrc.2019.04.146>
163. Guo, L. W., Zhou, L., Wei, P., Li, S., and Li, D. (2024) GALNT4 promotes the endothelial cell inflammatory response via the NF-κB signaling pathway. *Research Square* [Preprint]. <https://doi.org/10.21203/rs.3.rs-3998175/v1> [Accessed 9th September 2024]
164. Ye, Z., Guo, H., Wang, L., Li, Y., Xu, M., Zhao, X., Song, X., Chen, Z., and Huang, R. (2022) GALNT4 primes monocytes adhesion and transmigration by regulating O-Glycosylation of PSGL-1 in atherosclerosis. *J. Mol. Cell. Cardiol.* 165, 54-63. <https://doi.org/10.1016/j.yjmcc.2021.12.012>
165. Liu, W., Ramachandran, V., Kang, J., Kishimoto, T. K., Cummings, R. D., and McEver, R. P. (1998) Identification of N-terminal residues on P-selectin glycoprotein ligand-1 required for binding to P-selectin. *J. Biol. Chem.* 273 (12), 7078-7087. [https://www.jbc.org/article/S0021-9258\(18\)63989-4/](https://www.jbc.org/article/S0021-9258(18)63989-4/)
166. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjöblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyansky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E., Vogelstein, B. (2007) The genomic landscapes of human breast and colorectal cancers. *Science*. 318 (5853), 1108-1113. <https://www.science.org/doi/10.1126/science.1145720>
167. Guda, K., Moinova, H., He, J., Jamison, O., Ravi, L., Natale, L., Lutterbaugh, J., Lawrence, E., Lewis, S., Willson, J. K., Lowe, J. B., Wiesner, G. L., Parmigiani, G., Barnholtz-Sloan, J., Dawson, D. W., Velculescu, V. E., Kinzler, K. W., Papadopoulos, N., Vogelstein, B., Willis, J., Gerken, T. A., and Markowitz, S. D. (2009) Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc. Natl. Acad. Sci. USA*. 106 (31), 12921-12925. <https://doi.org/10.1073/pnas.0901454106>
168. Xi, X., Wang, J., Qin, Y., Huang, W., You, Y., and Zhan, J. (2022) Glycosylated modification of MUC1 maybe a new target to promote drug sensitivity and efficacy for

- breast cancer chemotherapy. *Cell Death Dis.* 13 (8), 708. <https://doi.org/10.1038/s41419-022-05110-2>
169. Deng, B., Zhang, S., Zhou, Y., Sun, T., Zhu, Y., Fei, J., Li, A., and Miao, Y. (2024) CCDC88C, an O-GalNAc glycosylation substrate of GALNT6, drives breast cancer metastasis by promoting c-JUN-mediated CEMIP transcription. *Cancer Cell. Int.* 24 (1), 237. <https://doi.org/10.1186/s12935-024-03413-2>
 170. Deng, B., Tarhan, Y. E., Ueda, K., Ren, L., Katagiri, T., Park, J. H., and Nakamura, Y. (2018) Critical Role of Estrogen Receptor Alpha O-Glycosylation by N-Acetylgalactosaminyltransferase 6 (GALNT6) in Its Nuclear Localization in Breast Cancer Cells. *Neoplasia.* 20 (10), 1038-1044. <https://doi.org/10.1016/j.neo.2018.08.006>
 171. Park, J. H., Katagiri, T., Chung, S., Kijima, K., and Nakamura, Y. (2011) Polypeptide N-acetylgalactosaminyltransferase 6 disrupts mammary acinar morphogenesis through O-glycosylation of fibronectin. *Neoplasia.* 13 (4), 320-326. <https://doi.org/10.1593/neo.101440>
 172. Lin, J., Chung, S., Ueda, K., Matsuda, K., Nakamura, Y., and Park, J. H. (2017) GALNT6 Stabilizes GRP78 Protein by O-glycosylation and Enhances its Activity to Suppress Apoptosis Under Stress Condition. *Neoplasia.* 19 (1), 43-53. <https://doi.org/10.1016/j.neo.2016.11.007>
 173. Huang, T., Meng, F., Huang, H., Wang, L., Wang, L., Liu, Y., Liu, Y., Wang, J., Li, W., Zhang, J., and Liu, Y. (2022) GALNT8 suppresses breast cancer cell metastasis potential by regulating EGFR O-GalNAcylation. *Biochem. Biophys. Res. Commun.* 601, 16-23. <https://doi.org/10.1016/j.bbrc.2022.02.072>
 174. Pangen, R. P., Channathodiylil, P., Huen, D. S., Eagles, L. W., Johal, B. K., Pasha, D., Hadjistephanou, N., Nevell, O., Davies, C. L., Adewumi, A. I., Khanom, H., Samra, I. S., Buzatto, V. C., Chandrasekaran, P., Shinawi, T., Dawson, T. P., Ashton, K. M., Davis, C., Brodbelt, A. R., Jenkinson, M. D., Bièche, I., Latif, F., Darling, J. L., Warr, T. J., and Morris, M. R. (2015) The GALNT9, BNC1 and CCDC8 genes are frequently epigenetically dysregulated in breast tumours that metastasise to the brain. *Clin. Epigenetics.* 7 (1), 57. <https://doi.org/10.1186/s13148-015-0089-x>
 175. Zamani, A. R. N., Mashayekhi, M. R., Jadid, M. F. S., Faridvand, Y., Tajalli, H., and Rahbarghazi, R. (2018) Photo-modulation of zinc phthalocyanine-treated breast cancer cell line ZR-75-1 inhibited the normal tumor activity in vitro. *Lasers Med. Sci.* 33 (9), 1969-1978. <https://doi.org/10.1007/s10103-018-2563-0>
 176. Tao, C., Luo, R., Song, J., Zhang, W., and Ran, L. (2020) A seven-DNA methylation signature as a novel prognostic biomarker in breast cancer. *J. Cell Biochem.* 121 (3), 2385-2393. <https://doi.org/10.1002/jcb.29461>
 177. Wu, Z., Wang, Z., Wu, H., Zheng, N., Huang, D., Huang, Z., Han, H., Bao, J., Xu, H., Zhang, R., Du, Z., and Wu, D. (2023) The pan-cancer multi-omics landscape of key genes of sialylation combined with RNA-sequencing validation. *Comput. Biol. Med.* 166, 107556. <https://doi.org/10.1016/j.combiomed.2023.107556>
 178. Shan, J., Liu, Y., Wang, Y., Li, Y., Yu, X., and Wu, C. (2018) GALNT14 Involves the Regulation of Multidrug Resistance in Breast Cancer Cells. *Transl. Oncol.* 11 (3), 786-793. <https://doi.org/10.1016/j.tranon.2018.04.003>
 179. Zuo, T., Shan, J., Liu, Y., Xie, R., Yu, X., and Wu, C. (2018) EFEMP2 Mediates GALNT14-Dependent Breast Cancer Cell Invasion. *Transl. Oncol.* 11 (2), 346-352. <https://doi.org/10.1016/j.tranon.2018.01.021>
 180. Song, K. H., Park, M. S., Nandu, T. S., Gadad, S., Kim, S. C., and Kim, M. Y. (2016) GALNT14 promotes lung-specific breast cancer metastasis by modulating self-renewal and interaction with the lung microenvironment. *Nat. Commun.* 7, 13796. <https://doi.org/10.1038/ncomms13796>
 181. Wu, H., He, G., Song, T., Zhang, Y., Chen, X., Chen, H., Xiong, W., Sun, C., Zhao, C., and Chen, Y. (2019) Evaluation of GALNT16 polymorphisms to breast cancer risk in Chinese population. *Mol. Genet. Genomic Med.* 7 (8), e848. <https://doi.org/10.1002/mgg3.848>

182. Calle, B., Bineva-Todd, G., Marchesi, A., Flynn, H., Ghirardello, M., Tastan, O. Y., Roustan, C., Choi, J., Galan, M. C., Schumann, B., and Malaker, S. A. (2021) Benefits of Chemical Sugar Modifications Introduced by Click Chemistry for Glycoproteomic Analyses. *J. Am. Soc. Mass Spectrom.* 32 (9), 2366-2375. <https://doi.org/10.1021/jasms.1c00084>
183. Rangel-Angarita, V., and Malaker, S. A. (2021) Mucinomics as the Next Frontier of Mass Spectrometry. *ACS Chem. Biol.* 16 (10), 1866-1883. <https://doi.org/10.1021/acscchembio.1c00384>
184. Wandall, H. H., Nielsen, M. A. I., King-Smith, S., de Haan, N., and Bagdonaite, I. (2021) Global functions of O-glycosylation: promises and challenges in O-glycobiology. *FEBS J.* 288 (24), 7183-7212. <https://doi.org/10.1111/febs.16148>
185. Rudd, P. M., Karlsson, N. G., Khoo, K. H., Thaysen-Andersen, M., Wells, L., and Packer, N. H. (2022) Glycomics and Glycoproteomics. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 51. <https://www.ncbi.nlm.nih.gov/books/NBK579904>
186. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Lavery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat. Methods.* 8 (11), 977-982. <https://doi.org/10.1038/nmeth.1731>
187. Malaker, S. A., Pedram, K., Ferracane, M. J., Bensing, B. A., Krishnan, V., Pett, C., Yu, J., Woods, E. C., Kramer, J. R., Westerlind, U., Dorigo, O., and Bertozzi, C. R. (2019) The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. USA.* 116 (15), 7278-7287. <https://doi.org/10.1073/pnas.1813020116>
188. Shon, D. J., Malaker, S. A., Pedram, K., Yang, E., Krishnan, V., Dorigo, O., and Bertozzi, C. R. (2020) An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl. Acad. Sci. USA.* 117 (35), 21299-21307. <https://doi.org/10.1073/pnas.2012196117>
189. Malaker, S. A., Riley, N. M., Shon, D. J., Pedram, K., Krishnan, V., Dorigo, O., and Bertozzi, C. R. (2022) Revealing the human mucinome. *Nat. Commun.* 13 (1), 3542. <https://doi.org/10.1038/s41467-022-31062-4>
190. Chongsaritsinsuk, J., Steigmeyer, A. D., Mahoney, K. E., Rosenfeld, M. A., Lucas, T. M., Smith, C. M., Li, A., Ince, D., Kearns, F. L., Battison, A. S., Hollenhorst, M. A., Judy Shon, D., Tiemeyer, K. H., Attah, V., Kwon, C., Bertozzi, C. R., Ferracane, M. J., Lemmon, M. A., Amaro, R. E., and Malaker, S. A. (2023) Glycoproteomic landscape and structural dynamics of TIM family immune checkpoints enabled by mucinase SmE. *Nat Commun.* 14 (1), 6169. <https://doi.org/10.1038/s41467-023-41756-y>
191. Narimatsu, Y., Joshi, H. J., Schjoldager, K. T., Hintze, J., Halim, A., Steentoft, C., Nason, R., Mandel, U., Bennett, E. P., Clausen, H., and Vakhrushev, S. Y. (2019) Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics. *Mol. Cell. Proteomics.* 18 (7), 1396-1409. [https://www.mcponline.org/article/S1535-9476\(20\)31547-4/](https://www.mcponline.org/article/S1535-9476(20)31547-4/)
192. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Lavery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32 (10), 1478-1488. <https://doi.org/10.1038/emboj.2013.79>
193. Schumann, B., Malaker, S. A., Wisnovsky, S. P., Debets, M. F., Agbay, A. J., Fernandez, D., Wagner, L. J. S., Lin, L., Li, Z., Choi, J., Fox, D. M., Peh, J., Gray, M. A., Pedram, K., Kohler, J. J., Mrksich, M., and Bertozzi, C. R. (2020) Bump-and-Hole Engineering Identifies Specific Substrates of Glycosyltransferases in Living Cells. *Mol.*

- Cell.* 78 (5), 824-834.e15. [https://www.cell.com/molecular-cell/fulltext/S1097-2765\(20\)30198-2](https://www.cell.com/molecular-cell/fulltext/S1097-2765(20)30198-2)
194. Dalal, K., Yang, W., Tian, E., Chernish, A., McCluggage, P., Lara, A. J., Ten Hagen, K. G., and Tabak, L. A. (2024) In vivo mapping of the mouse Galnt3-specific O-glycoproteome. *J. Biol. Chem.* 300 (9), 107628. <https://doi.org/10.1016/j.jbc.2024.107628>
 195. Yang, W., Tian, E., Chernish, A., McCluggage, P., Dalal, K., Lara, A., Ten Hagen, K. G., and Tabak, L. A. (2023) Quantitative mapping of the in vivo O-GalNAc glycoproteome in mouse tissues identifies GalNAc-T2 O-glycosites in metabolic disorder. *Proc. Natl. Acad. Sci. USA.* 120 (43), e2303703120. <https://doi.org/10.1073/pnas.2303703120>
 196. Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhrushev, S. Y., and Clausen H. (2015) Deconstruction of O-glycosylation--GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep.* 16 (12), 1713-1722. <https://doi.org/10.15252/embr.201540796>
 197. Schjoldager, K. T., Vakhrushev, S. Y., Kong, Y., Steentoft, C., Nudelman, A. S., Pedersen, N. B., Wandall, H. H., Mandel, U., Bennett, E. P., Lavery, S. B., and Clausen, H. (2012) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc. Natl. Acad. Sci. USA.* 109 (25), 9893-9898. <https://doi.org/10.1073/pnas.1203563109>
 198. Islam, K. (2018) The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem. Biol.* 25 (10), 1171-1184. <https://doi.org/10.1016/j.chembiol.2018.07.001>
 199. Bishop, A. C., Buzko, O., and Shokat, K. M. (2001) Magic bullets for protein kinases. *Trends Cell Biol.* 11 (4), 167-172. [https://www.cell.com/trends/cell-biology/abstract/S0962-8924\(01\)01928-6](https://www.cell.com/trends/cell-biology/abstract/S0962-8924(01)01928-6)
 200. Li, J., Wei, H., and Zhou, M. M. (2011) Structure-guided design of a methyl donor cofactor that controls a viral histone H3 lysine 27 methyltransferase activity. *J. Med. Chem.* 54 (21), 7734-7738. <https://doi.org/10.1021/jm201000j>
 201. Carter-O'Connell, I., Jin, H., Morgan, R. K., David, L. L., and Cohen, M. S. (2014) Engineering the substrate specificity of ADP-ribosyltransferases for identifying direct protein targets. *J. Am. Chem. Soc.* 136 (14), 5201-5204. <https://doi.org/10.1021/ja412897a>
 202. Cioce, A., Malaker, S. A., and Schumann, B. (2021) Generating orthogonal glycosyltransferase and nucleotide sugar pairs as next-generation glycobiology tools. *Curr. Opin. Chem. Biol.* 60, 66-78. <https://doi.org/10.1016/j.cbpa.2020.09.001>
 203. Calle, B., Gonzalez-Rodriguez, E., Mahoney, K. E., Cioce, A., Bineva-Todd, G., Tastan, O. Y., Roustan, C., Flynn, H., Malaker, S. A., and Schumann, B. (2023) Bump-and-hole engineering of human polypeptide N-acetylgalactosamine transferases to dissect their protein substrates and glycosylation sites in cells. *STAR Protoc.* 4 (1), 101974. <https://doi.org/10.1016/j.xpro.2022.101974>
 204. Cioce, A., Calle, B., Rizou, T., Lowery, S. C., Bridgeman, V. L., Mahoney, K. E., Marchesi, A., Bineva-Todd, G., Flynn, H., Li, Z., Tastan, O. Y., Roustan, C., Soro-Barrio, P., Rafiee, M. R., Garza-Garcia, A., Antonopoulos, A., Wood, T. M., Keenan, T., Both, P., Huang, K., Parmeggian, F., Snijders, A. P., Skehel, M., Kjær, S., Fascione, M. A., Bertozzi, C. R., Haslam, S. M., Flitsch, S. L., Malaker, S. A., Malanchi, I., and Schumann, B. (2022) Cell-specific bioorthogonal tagging of glycoproteins. *Nat. Commun.* 13 (1), 6237. <https://doi.org/10.1038/s41467-022-33854-0>
 205. Yu, S. H., Boyce, M., Wands, A. M., Bond, M. R., Bertozzi, C. R., and Kohler, J. J. (2012) Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. *Proc. Natl. Acad. Sci. USA.* 109 (13), 4834-4839. <https://doi.org/10.1073/pnas.1114356109>
 206. Besanceney-Webler, C., Jiang, H., Zheng, T., Feng, L., Soriano del Amo, D., Wang, W., Klivansky, L. M., Marlow, F. L., Liu, Y., and Wu, P. (2011) Increasing the efficacy

- of bioorthogonal click reactions for bioconjugation: a comparative study. *Angew. Chem. Int. Ed. Engl.* 50 (35), 8051-8056. <https://doi.org/10.1002/anie.201101817>
207. Uttamapinant, C., Tangpeerachaikul, A., Grecian, S., Clarke, S., Singh, U., Slade, P., Gee, K. R., and Ting, A. Y. (2012) Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew. Chem. Int. Ed. Engl.* 51 (24), 5852-5856. <https://doi.org/10.1002/anie.201108181>
 208. Calle, B. (2020) *Characterisation of O-GalNAc Glycosylation by GalNAc-T4 Bump-and-Hole Engineering*. MSci thesis. Imperial College London.
 209. Boyce, M., Carrico, I. S., Ganguli, A. S., Yu, S. H., Hangauer, M. J., Hubbard, S. C., Kohler, J. J., and Bertozzi, C. R. (2011) Metabolic cross-talk allows labeling of O-linked beta-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway. *Proc. Natl. Acad. Sci. USA.* 108 (8), 3141-3146. <https://doi.org/10.1073/pnas.1010045108>
 210. Woo, C. M., Iavarone, A. T., Spicariich, D. R., Palaniappan, K. K., and Bertozzi, C. R. (2015) Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods.* 12 (6), 561-567. <https://doi.org/10.1038/nmeth.3366>
 211. Pouilly, S., Bourgeaux, V., Piller, F., and Piller, V. (2012) Evaluation of analogues of GalNAc as substrates for enzymes of the mammalian GalNAc salvage pathway. *ACS Chem. Biol.* 7 (4), 753-760. <https://doi.org/10.1021/cb200511t>
 212. Li, Y., Yu, H., Chen, Y., Lau, K., Cai, L., Cao, H., Tiwari, V. K., Qu, J., Thon, V., Wang, P. G., and Chen, X. (2011) Substrate promiscuity of N-acetylhexosamine 1-kinases. *Molecules.* 16 (8), 6396-6407. <https://doi.org/10.3390/molecules16086396>
 213. Gonzalez-Rodriguez, E., Zol-Hanlon, M., Bineva-Todd, G., Marchesi, A., Skehel, M., Mahoney, K. E., Roustan, C., Borg, A., Di Vagno, L., Kjær, S., Wrobel, A. G., Benton, D. J., Nawrath, P., Flitsch, S. L., Joshi, D., González-Ramírez, A. M., Wilkinson, K. A., Wilkinson, R. J., Wall, E. C., Hurtado-Guerrero, R., Malaker, S. A., and Schumann, B. (2023) O-Linked Sialoglycans Modulate the Proteolysis of SARS-CoV-2 Spike and Likely Contribute to the Mutational Trajectory in Variants of Concern. *ACS Cent. Sci.* 9 (3), 393-404. <https://doi.org/10.1021/acscentsci.2c01349>
 214. Hansen, S., Stüber, J. C., Ernst, P., Koch, A., Bojar, D., Batyuk, A., and Plückthun, A. (2017) Design and applications of a clamp for Green Fluorescent Protein with picomolar affinity. *Sci. Rep.* 7 (1), 16292. <https://doi.org/10.1038/s41598-017-15711-z>
 215. Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A., and Choi, S. Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One.* 6 (4), e18556. <https://doi.org/10.1371/journal.pone.0018556>
 216. Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., and Pinello, L. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 37 (3), 224-226. <https://doi.org/10.1038/s41587-019-0032-3>
 217. Faustino-Rocha, A., Oliveira, P. A., Pinho-Oliveira, J., Teixeira-Guedes, C., Soares-Maia, R., da Costa, R. G., Colaço, B., Pires, M. J., Colaço, J., Ferreira, R., and Ginja, M. (2013) Estimation of rat mammary tumor volume using caliper and ultrasonography measurements. *Lab Anim.* 42 (6), 217-224. <https://doi.org/10.1038/labon.254>
 218. Rafiee, M. R., Sigismondo, G., Kalxdorf, M., Förster, L., Brügger, B., Béthune, J., and Krijgsveld, J. (2020) Protease-resistant streptavidin for interaction proteomics. *Mol. Syst. Biol.* 16 (5), e9370. <https://doi.org/10.15252/msb.20199370>
 219. UniProt Consortium. (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* 51, D523-D531. <https://doi.org/10.1093/nar/gkac1052>
 220. Hughes, C. S., Moggridge, S., Müller, T., Sorensen, P. H., Morin, G. B., and Krijgsveld, J. (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* 14 (1), 68-85. <https://doi.org/10.1038/s41596-018-0082-x>

221. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11 (12), 2301-2319. <https://doi.org/10.1038/nprot.2016.136>
222. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8 (11), 2281-2308. <https://doi.org/10.1038/nprot.2013.143>
223. Synthego. (n.d.) *What Is CRISPR: The Ultimate Guide To CRISPR Mechanisms, Applications, Methods & More.* <https://www.synthego.com/learn/crispr> [Accessed 9th September 2024]
224. Prakash, T., Sharma, V. K., Adati, N., Ozawa, R., Kumar, N., Nishida, Y., Fujikake, T., Takeda, T., and Taylor, T. D. (2010) Expression of conjoined genes: another mechanism for gene regulation in eukaryotes. *PLoS One.* 5 (10), e13284. <https://doi.org/10.1371/journal.pone.0013284>
225. Narimatsu, Y., Joshi, H. J., Yang, Z., Gomes, C., Chen, Y. H., Lorenzetti, F. C., Furukawa, S., Schjoldager, K. T., Hansen, L., Clausen, H., Bennett, E. P., and Wandall, H. H. (2018) A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology.* 28 (5), 295-305. <https://doi.org/10.1093/glycob/cwx101>
226. Cen, H., Mao, F., Aronchik, I., Fuentes, R. J., and Firestone, G. L. (2008) DEVD-NucView488: a novel class of enzyme substrates for real-time detection of caspase-3 activity in live cells. *FASEB J.* 22 (7), 2243-2252. <https://doi.org/10.1096/fj.07-099234>
227. Özdemir, B. C., Sflomos, G., and Briskin, C. (2018) The challenges of modeling hormone receptor-positive breast cancer in mice. *Endocr. Relat. Cancer.* 25 (5), R319-R330. <https://doi.org/10.1530/ERC-18-0063>
228. Blackman, M. L., Royzen, M., Fox, J. M. (2008) Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc.* 130 (41), 13518-13519. <https://doi.org/10.1021/ja8053805>
229. Wu, H., Shajahan, A., Yang, J. Y., Capota, E., Wands, A. M., Arthur, C. M., Stowell, S. R., Moremen, K. W., Azadi, P., and Kohler, J. J. (2022) A photo-cross-linking GlcNAc analog enables covalent capture of N-linked glycoprotein-binding partners on the cell surface. *Cell Chem. Biol.* 29 (1), 84-97.e8. <https://doi.org/10.1016/j.chembiol.2021.07.007>
230. Puneekar, N. S. (2018) Henri–Michaelis–Menten Equation. In: *ENZYMES: Catalysis, Kinetics and Mechanisms.* Singapore, Singapore, Springer, pp. 155–176. https://doi.org/10.1007/978-981-13-0785-0_15
231. Liu, Y., Bineva-Todd, G., Meek, R., Mazo, L., Piniello, B., Moroz, O., Begum, N., Ohara, A., Roustan, C., Tomita, S., Kjaer, S., Polizzi, K., Rovira, C., Davies, G. J., and Schumann, B. (2024) A Bioorthogonal Precision Tool for human N-acetylglucosaminyltransferase V. *ChemRxiv* [Preprint]. <https://doi.org/10.26434/chemrxiv-2024-856gt> [Accessed 9th September 2024].
232. Hanisch, F. G., Stadie, T. R., Deutzmann, F., and Peter-Katalinic, J. (1996) MUC1 glycoforms in breast cancer. *Eur. J. Biochem.* 236 (1), 318-327. <https://doi.org/10.1111/j.1432-1033.1996.00318.x>
233. Chang, P. V., Chen, X., Smyrniotis, C., Xenakis, A., Hu, T., Bertozzi, C. R., and Wu, P. (2009) Metabolic labeling of sialic acids in living animals with alkynyl sugars. *Angew. Chem. Int. Ed. Engl.* 48 (22), 4030-4033. <https://doi.org/10.1002/anie.200806319>
234. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.* 275 (49), 38197-38205. [https://www.jbc.org/article/S0021-9258\(19\)55944-0/](https://www.jbc.org/article/S0021-9258(19)55944-0/)
235. Niers, J. M., Chen, J. W., Weissleder, R., and Tannous, B. A. (2011) Enhanced in vivo imaging of metabolically biotinylated cell surface reporters. *Anal. Chem.* 83 (3), 994-999. <https://doi.org/10.1021/ac102758m>

236. Chandler, C. S., and Ballard, F. J. (1986) Multiple biotin-containing proteins in 3T3-L1 cells. *Biochem. J.* 237 (1), 123-130. <https://doi.org/10.1042/bj2370123>
237. Chen, X., Wei, S., Ji, Y., Guo, X., and Yang, F. (2015) Quantitative proteomics using SILAC: Principles, applications, and developments. *Proteomics*. 15 (18), 3175-3192. <https://doi.org/10.1002/pmic.201500108>
238. Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell Proteomics*. 3 (6), 608-614. [https://www.mcponline.org/article/S1535-9476\(20\)33877-9/](https://www.mcponline.org/article/S1535-9476(20)33877-9/)
239. York, W. S., Mazumder, R., Ranzinger, R., Edwards, N., Kahsay, R., Aoki-Kinoshita, K. F., Campbell, M. P., Cummings, R. D., Feizi, T., Martin, M., Natale, D. A., Packer, N. H., Woods, R. J., Agarwal, G., Arpinar, S., Bhat, S., Blake, J., Castro, L. J. G., Fochtman, B., Gildersleeve, J., Goldman, R., Holmes, X., Jain, V., Kulkarni, S., Mahadik, R., Mehta, A., Mousavi, R., Nakarakommula, S., Navelkar, R., Pattabiraman, N., Pierce, M. J., Ross, K., Vasudev, P., Vora, J., Williamson, T., and Zhang, W. (2020) GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology*. 30 (2), 72-73. <https://doi.org/10.1093/glycob/cwz080>
240. Semba, R., Horimoto, Y., Sakata-Matsuzawa, M., Ishizuka, Y., Denda-Nagai, K., Fujihira, H., Noji, M., Onagi, H., Ichida, M., Miura, H., Watanabe, J., Saito, M., Saito, T., Arakawa, A., and Irimura, T. (2023) Possible correlation of apical localization of MUC1 glycoprotein with luminal A-like status of breast cancer. *Sci. Rep.* 13 (1), 5281. <https://doi.org/10.1038/s41598-023-32579-4>
241. Rahn, J. J., Dabbagh, L., Pasdar, M., and Hugh, J. C. (2001) The importance of MUC1 cellular localization in patients with breast carcinoma: an immunohistologic study of 71 patients and review of the literature. *Cancer*. 91 (11), 1973-1982. [https://doi.org/10.1002/1097-0142\(20010601\)91:11<1973::AID-CNCR1222>3.0.CO;2-A](https://doi.org/10.1002/1097-0142(20010601)91:11<1973::AID-CNCR1222>3.0.CO;2-A)
242. Rakha, E. A., Boyce, R. W., Abd El-Rehim, D., Kurien, T., Green, A. R., Paish, E. C., Robertson, J. F., and Ellis, I. O. (2005) Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer. *Mod. Pathol.* 18 (10), 1295-1304. [https://www.modernpathology.org/article/S0893-3952\(22\)04539-2/](https://www.modernpathology.org/article/S0893-3952(22)04539-2/)
243. Zhang, R., Zhu, Z., Shen, W., Li, X., Dhoomun, D. K., and Tian, Y. (2019) Golgi Membrane Protein 1 (GOLM1) Promotes Growth and Metastasis of Breast Cancer Cells via Regulating Matrix Metalloproteinase-13 (MMP13). *Med. Sci. Monit.* 25, 847-855. <https://www.medscimonit.com/abstract/index/idArt/911667>
244. Schork, K., Podwojski, K., Turewicz, M., Stephan, C., and Eisenacher, M. (2021) Important Issues in Planning a Proteomics Experiment: Statistical Considerations of Quantitative Proteomic Data. In: Marcus, K., Eisenacher, M., and Sitek, B. (eds.) *Quantitative Methods in Proteomics*. 2nd ed. New York, NY, USA, Humana, pp. 1-20. https://doi.org/10.1007/978-1-0716-1024-4_1
245. Hu, L., Li, L., Xie, H., Gu, Y., and Peng, T. (2011) The Golgi localization of GOLPH2 (GP73/GOLM1) is determined by the transmembrane and cytoplasmic sequences. *PLoS One*. 6 (11), e28207. <https://doi.org/10.1371/journal.pone.0028207>
246. Kim, M. J., Choi, J. R., Tae, N., Wi, T. M., Kim, K. M., Kim, D. H., and Lee, E. S. (2020) Novel Antibodies Targeting MUC1-C Showed Anti-Metastasis and Growth-Inhibitory Effects on Human Breast Cancer Cells. *Int. J. Mol. Sci.* 21 (9), 3258. <https://doi.org/10.3390/ijms21093258>
247. Persson, B. D., Schmitz, N. B., Santiago, C., Zocher, G., Larvie, M., Scheu, U., Casasnovas, J. M., and Stehle, T. (2010) Structure of the extracellular portion of CD46 provides insights into its interactions with complement proteins and pathogens. *PLoS Pathog.* 6 (9), e1001122. <https://doi.org/10.1371/journal.ppat.1001122>
248. Manchester, M., Liszewski, M. K., Atkinson, J. P., and Oldstone, M. B. (1994) Multiple isoforms of CD46 (membrane cofactor protein) serve as receptors for measles virus. *Proc. Natl. Acad. Sci. USA*. 91 (6), 2161-2165. <https://doi.org/10.1073/pnas.91.6.2161>

249. Christiansen, D., Milland, J., Thorley, B. R., McKenzie, I. F., Mottram, P. L., Purcell, L. J., and Loveland, B. E. (1996) Engineering of recombinant soluble CD46: an inhibitor of complement activation. *Immunology*. 87 (3), 348-354. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1384101/>
250. Uhlén, M., Björling, E., Agaton, C., Sziggyarto, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Sköllerö, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteomics*. 4 (12), 1920-1932. [https://www.mcponline.org/article/S1535-9476\(20\)30028-1/](https://www.mcponline.org/article/S1535-9476(20)30028-1/)
251. Wu, G., Kim, D., Kim, J. N., Park, S., Maharjan, S., Koh, H., Moon, K., Lee, Y., and Kwon, H. J. (2018) A Mucin1 C-terminal Subunit-directed Monoclonal Antibody Targets Overexpressed Mucin1 in Breast Cancer. *Theranostics*. 8 (1), 78-91. <https://www.thno.org/v08p0078.htm>
252. Banerjee, S., and Mazumdar, S. (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int. J. Anal. Chem.* 2012, 282574. <https://doi.org/10.1155/2012/282574>
253. Macias, L. A., Santos, I. C., and Brodbelt, J. S. (2020) Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* 92 (1), 227-251. <https://doi.org/10.1021/acs.analchem.9b04859>
254. Halim, A., Westerlind, U., Pett, C., Schorlemer, M., Rüetschi, U., Brinkmalm, G., Sihlbom, C., Lengqvist, J., Larson, G., and Nilsson, J. (2014) Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J. Proteome Res.* 13 (12), 6024-6032. <https://doi.org/10.1021/pr500898r>
255. Gerken, T. A., Raman, J., Fritz, T. A., and Jamison, O. (2006) Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *J. Biol. Chem.* 281 (43), 32403-32416. [https://www.jbc.org/article/S0021-9258\(20\)86815-X/](https://www.jbc.org/article/S0021-9258(20)86815-X/)
256. Debets, M. F., Tasthan, O. Y., Wisnovsky, S. P., Malaker, S. A., Angelis, N., Moeckl, L. K. R., Choi, J., Flynn, H., Wagner, L. J. S., Bineva-Todd, G., Antonopoulos, A., Cioce, A., Browne, W. M., Li, Z., Briggs, D. C., Douglas, H. L., Hess, G. T., Agbay, A. J., Roustan, C., Kjaer, S., Haslam, S. M., Snijders, A. P., Bassik, M. C., Moerner, W. E., Li, V. S. W., Bertozzi, C. R., and Schumann, B. (2020) Metabolic precision labeling enables selective probing of O-linked N-acetylgalactosamine glycosylation. *Proc. Natl. Acad. Sci. USA*. 117 (41), 25293-25301. <https://doi.org/10.1073/pnas.2007297117>
257. St Amant, A. H., Engbersm C., and Hudson, R. H. (2013) A solid-phase CuAAC strategy for the synthesis of PNA containing nucleobase surrogates. *Artif. DNA: PNA XNA*. 4 (1), 4-10. <https://doi.org/10.4161/adna.23982>
258. Bumpus, T. W., and Baskin, J. M. (2017) Clickable Substrate Mimics Enable Imaging of Phospholipase D Activity. *ACS Cent. Sci.* 3 (10), 1070-1077. <https://doi.org/10.1021/acscentsci.7b00222>
259. Benito-Alifonso, D., Tremell, S., Sadler, J. C., Berry, M., and Galan, M. C. (2016) Imidazolium-tagged glycan probes for non-covalent labeling of live cells. *Chem. Commun.* 52 (27), 4906-4909. <https://doi.org/10.1039/C5CC10040B>
260. Huang, K., Parmeggiani, F., Ledru, H., Hollingsworth, K., Mas Pons, J., Marchesi, A., Both, P., Matthey, A. P., Pallister, E., Bulmer, G. S., van Munster, J. M., Turnbull, W. B.,

- Galan, M. C., and Flitsch, S. L. (2019) Enzymatic synthesis of N-acetyllactosamine from lactose enabled by recombinant β 1,4-galactosyltransferases. *Org. Biomol. Chem.* 17 (24), 5920-5924. <https://doi.org/10.1039/C9OB01089K>
261. Sittel, I., and Galan, M. C. (2017) Imidazolium-labeled glycosides as probes to harness glycosyltransferase activity in human breast milk. *Org. Biomol. Chem.* 15 (17), 3575-3579. <https://doi.org/10.1039/C7OB00550D>
 262. Sittel, I., and Galan, M. C. (2015) Chemo-enzymatic synthesis of imidazolium-tagged sialyllactosamine probes. *Bioorg. Med. Chem. Lett.* 25 (19), 4329-4332. <https://doi.org/10.1016/j.bmcl.2015.07.049>
 263. Sittel, I., Tran, A. T., Benito-Alifonso, D., and Galan, M. C. (2013) Combinatorial ionic catch-and-release oligosaccharide synthesis (combi-ICROS). *Chem. Commun.* 49 (39), 4217-4219. <https://doi.org/10.1039/C2CC37164B>
 264. Huang, K., Marchesi, A., Hollingsworth, K., Both, P., Matthey, A. P., Pallister, E., Ledru, H., Charnock, S. J., Galan, M. C., Turnbull, W. B., Parmeggiani, F., and Flitsch, S. L. (2020) Biochemical characterisation of an α 1,4 galactosyltransferase from *Neisseria weaveri* for the synthesis of α 1,4-linked galactosides. *Org. Biomol. Chem.* 18 (16), 3142-3148. <https://doi.org/10.1039/D0OB00407C>
 265. Galan, M. C., Tran, A. T., Bromfield, K., Rabbani, S., and Ernst, B. (2012) Ionic-liquid-based MS probes for the chemo-enzymatic synthesis of oligosaccharides. *Org. Biomol. Chem.* 10 (35), 7091-7097. <https://doi.org/10.1039/C2OB25855B>
 266. Galan, M. C., Tran, A. T., and Bernard, C. (2010) Ionic-liquid-based catch and release mass spectroscopy tags for enzyme monitoring. *Chem. Commun.* 46 (47), 8968-8970. <https://doi.org/10.1039/C0CC04224B>
 267. Lřkov, M., Tshepelevitsh, S., Heering, A., Plieger, P.G., Vianello, R. and Leito, I. (2017), On the Basicity of Conjugated Nitrogen Heterocycles in Different Media. *Eur. J. Org. Chem.* 2017 (30), 4475-4489. <https://doi.org/10.1002/ejoc.201700749>
 268. Đorđević, M. G., and Premović, P. (2019) Iron, manganese, vanadium, copper and zinc of the Cretaceous-paleogene boundary fish clay at the Kirkevig site (Hřjerup, Stevns Klint, Denmark). *Bulgarian Chemical Communications.* 51 (1), 5-9. http://bcc.bas.bg/BCC_Volumes/Volume_51_Number_1_2019/BCC-51-1-2019-5-9-Premovic-4927.pdf
 269. Jiang, K., Li, W., Zhang, Q., Yan, G., Guo, K., Zhang, S., and Liu, Y. (2016) GP73 N-glycosylation at Asn144 reduces hepatocellular carcinoma cell motility and invasiveness. *Oncotarget.* 7 (17), 23530-23541. <https://doi.org/10.18632/oncotarget.8120>
 270. Hofman, P., His, B. L., Manie, S., Fenichel, P., Thyss, A., and Rossi, B. (1994) High expression of the antigen recognized by the monoclonal antibody GB24 on human breast carcinomas: a preventive mechanism of malignant tumor cells against complement attack? *Breast Cancer Res. Treat.* 32 (2), 213-219. <https://doi.org/10.1007/BF00665772>
 271. Maciejczyk, A., Szelachowska, J., Szynglarewicz, B., Szulc, R., Szulc, A., Wysocka, T., Jagoda, E., Lage, H., and Surowiak, P. (2011) CD46 Expression is an unfavorable prognostic factor in breast cancer cases. *Appl. Immunohistochem. Mol. Morphol.* 19 (6), 540-546. https://journals.lww.com/appliedimmunohist/abstract/2011/12000/cd46_expression_is_an_unfavorable_prognostic.12.aspx
 272. Liszewski, M. K., and Atkinson, J. P. (2021) Membrane cofactor protein (MCP; CD46): deficiency states and pathogen connections. *Curr. Opin. Immunol.* 72, 126-134. <https://doi.org/10.1016/j.coi.2021.04.005>
 273. Riley-Vargas, R. C., Gill, D. B., Kemper, C., Liszewski, M. K., and Atkinson, J. P. (2004) CD46: expanding beyond complement regulation. *Trends Immunol.* 25 (9), 496-503. [https://www.cell.com/trends/immunology/abstract/S1471-4906\(04\)00209-1](https://www.cell.com/trends/immunology/abstract/S1471-4906(04)00209-1)
 274. West, E. E., Kolevm M., and Kemper, C. (2018) Complement and the Regulation of T Cell Responses. *Annu. Rev. Immunol.* 36, 309-338. <https://doi.org/10.1146/annurev-immunol-042617-053245>

275. Liszewski, M. K., Leung, M. K., and Atkinson, J. P. (1998) Membrane cofactor protein: importance of N- and O-glycosylation for complement regulatory function. *J. Immunol.* 161 (7), 3711-3718. <https://doi.org/10.4049/jimmunol.161.7.3711>
276. Ni Choileain, S., Hay, J., Thomas, J., Williams, A., Vermeren, M. M., Benezech, C., Gomez-Salazar, M., Hugues, O. R., Vermeren, S., Howie, S. E. M., Dransfield, I., and Astier, A. L. (2017) TCR-stimulated changes in cell surface CD46 expression generate type 1 regulatory T cells. *Sci. Signal.* 10 (502), eaah6163. <https://doi.org/10.1126/scisignal.aah6163>
277. Liszewski, M. K., and Atkinson, J. P. (1996) Membrane cofactor protein (MCP; CD46). Isoforms differ in protection against the classical pathway of complement. *J. Immunol.* 156 (11), 4415-4421. <https://doi.org/10.4049/jimmunol.156.11.4415>
278. Christiansen, D., Milland, J., Thorley, B. R., McKenzie, I. F., and Loveland, B. E. (1996) A functional analysis of recombinant soluble CD46 in vivo and a comparison with recombinant soluble forms of CD55 and CD35 in vitro. *Eur. J. Immunol.* 26 (3), 578-585. <https://doi.org/10.1002/eji.1830260312>
279. Wang, Q., Chan, T. R., Hilgraf, R., Fokin, V. V., Sharpless, K. B., and Finn, M. G. (2003) Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* 125 (11), 3192-3193. <https://doi.org/10.1021/ja021381e>
280. Rostovtsev, V. V., Green, L. G., Fokin, V. V., and Sharpless, K. B. (2002) A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* 41 (14), 2596-2599. [https://doi.org/10.1002/1521-3773\(20020715\)41:14<2596::AID-ANIE2596>3.0.CO;2-4](https://doi.org/10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4)
281. Kim, H. J., Lv, D., Zhang, Y., Peng, T., and Ma, X. (2012) Golgi phosphoprotein 2 in physiology and in diseases. *Cell Biosci.* 2 (1), 31 <https://doi.org/10.1186/2045-3701-2-31>
282. Liu, Y., Hu, X., Liu, S., Zhou, S., Chen, Z., and Jin, H. (2021) Golgi Phosphoprotein 73: The Driver of Epithelial-Mesenchymal Transition in Cancer. *Front. Oncol.* 11, 783860. <https://doi.org/10.3389/fonc.2021.783860>
283. Bachert, C., Fimmel, C., and Linstedt, A. D. (2007) Endosomal trafficking and proprotein convertase cleavage of cis Golgi protein GP73 produces marker for hepatocellular carcinoma. *Traffic.* 8 (10), 1415-1423. <https://doi.org/10.1111/j.1600-0854.2007.00621.x>
284. Nagaraj, M., Höring, M., Ahonen, M. A., Nguyen, V. D., Zhou, Y., Vihinen, H., Jokitalo, E., Liebisch, G., Nidhina Haridas, P. A., and Olkkonen, V. M. (2022) GOLM1 depletion modifies cellular sphingolipid metabolism and adversely affects cell growth. *J. Lipid Res.* 63 (9), 100259. [https://www.jlr.org/article/S0022-2275\(22\)00092-X/](https://www.jlr.org/article/S0022-2275(22)00092-X/)
285. Frans, M. T., Kuipers, E. M., Bianchi, F., and van den Bogaart, G. (2023) Unveiling the impact of GOLM1/GP73 on cytokine production in cancer and infectious disease. *Immunol. Cell Biol.* 101 (8), 727-734. <https://doi.org/10.1111/imcb.12664>
286. Bai, W., Li, B., Wu, P., Li, X., Huang, X., Shi, N., Yang, C., Hu, F., and Xie, X. (2024) The first structure of human Golm1 coiled coil domain reveals an unexpected tetramer and highlights its structural diversity. *Int. J. Biol. Macromol.* 275, 133624. <https://doi.org/10.1016/j.ijbiomac.2024.133624>
287. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439-D444. <https://doi.org/10.1093/nar/gkab1061>
288. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Židek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., Steinegger, M., Hassabis, D., and Velankar, S. (2024)

- AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 52, D368-D375. <https://doi.org/10.1093/nar/gkad1011>
289. Lawson, D. A., Bhakta, N. R., Kessenbrock, K., Prummel, K. D., Yu, Y., Takai, K., Zhou, A., Eyob, H., Balakrishnan, S., Wang, C. Y., Yaswen, P., Goga, A., and Werb, Z. (2015) Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature.* 526 (7571), 131-135. <https://doi.org/10.1038/nature15260>
 290. Kowarz, E., Löscher, D., and Marschalek, R. (2015) Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol. J.* 10 (4), 647-653. <https://doi.org/10.1002/biot.201400821>
 291. Qin, W., Qin, K., Fan, X., Peng, L., Hong, W., Zhu, Y., Lv, P., Du, Y., Huang, R., Han, M., Cheng, B., Liu, Y., Zhou, W., Wang, C., and Chen, X. (2018) Artificial Cysteine S-Glycosylation Induced by Per-O-Acetylated Unnatural Monosaccharides during Metabolic Glycan Labeling. *Angew. Chem. Int. Ed. Engl.* 57 (7), 1817-1820. <https://doi.org/10.1002/anie.201711710>
 292. Qin, K., Zhang, H., Zhao, Z., and Chen, X. (2020) Protein S-Glyco-Modification through an Elimination-Addition Mechanism. *J. Am. Chem. Soc.* 142 (20), 9382-9388. <https://doi.org/10.1021/jacs.0c02110>
 293. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2015) The Innate and Adaptive Immune Systems. In: *Molecular biology of the cell*. 6th ed. New York, NY, USA, Garland Science, Taylor and Francis Group, pp. 1297-1342.
 294. Wright, L. M., Yong, S., Picken, M. M., Rockey, D., and Fimmel, C. J. (2009) Decreased survival and hepato-renal pathology in mice with C-terminally truncated GP73 (GOLPH2). *Int. J. Clin. Exp. Pathol.* 2 (1), 34-47. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491393/>
 295. Wang, Y., and Wan, Y. Y. (2020) Golgi protein 73, hepatocellular carcinoma and other types of cancers. *Liver Res.* 4 (4), 161-167. <https://doi.org/10.1016/j.livres.2020.09.003>
 296. Zhou, Y., Li, L., Hu, L., and Peng, T. (2011) Golgi phosphoprotein 2 (GOLPH2/GP73/GOLM1) interacts with secretory clusterin. *Mol. Biol. Rep.* 38 (3), 1457-1462. <https://doi.org/10.1007/s11033-010-0251-7>
 297. Shao, W. Q., Zhu, W. W., Luo, M. J., Fan, M. H., Li, Q., Wang, S. H., Lin, Z. F., Zhao, J., Zheng, Y., Dong, Q. Z., Lu, L., Jia, H. L., Zhang, J. B., Lu, M., Chen, J. H., and Qin, L. X. (2022) Cholesterol suppresses GOLM1-dependent selective autophagy of RTKs in hepatocellular carcinoma. *Cell Rep.* 39 (3), 110712. <https://doi.org/10.1016/j.celrep.2022.110712>
 298. Xu, Z., Ku, X., Tomioka, A., Xie, W., Liang, T., Zou, X., Cui, Y., Sato, T., Kaji, H., Narimatsu, H., Yan, W., and Zhang, Y. (2020) O-linked N-acetylgalactosamine modification is present on the tumor suppressor p53. *Biochim. Biophys. Acta Gen. Subj.* 1864 (8), 129635. <https://doi.org/10.1016/j.bbagen.2020.129635>
 299. West, C. M., Slawson, C., Zachara, N. E., and Hart, G. W. (2022) Nucleocytoplasmic Glycosylation. In: Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H. (eds.) *Essentials of Glycobiology*. (4th ed.) Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press, Chapter 18. <https://www.ncbi.nlm.nih.gov/books/NBK579979/>
 300. Cancer Research UK. (2023) *Tests for breast cancer*. <https://www.cancerresearchuk.org/about-cancer/breast-cancer/getting-diagnosed/tests> [Accessed 9th September 2024]
 301. Gutierrez, R., Shah, P. D., Hamid, O., Garfall, A. L., Posey, A. D., Bishop, M. R., Blumenschein, G. R., Johnson, M. L., Lee, S. M., Luke, J. J., Morgensztern, D., Fountaine, T. J., Dryer-Minnerly, R., Najmi, S., Hufner, P., and Chagin, K. D. (2021) Phase I experience with first in class TnMUC1 targeted chimeric antigen receptor T-cells in patients with advanced TnMUC1 positive solid tumors. *J. Clin. Oncol.* 39, e14513. https://doi.org/10.1200/JCO.2021.39.15_suppl.e14513

SUPPLEMENTARY INFORMATION

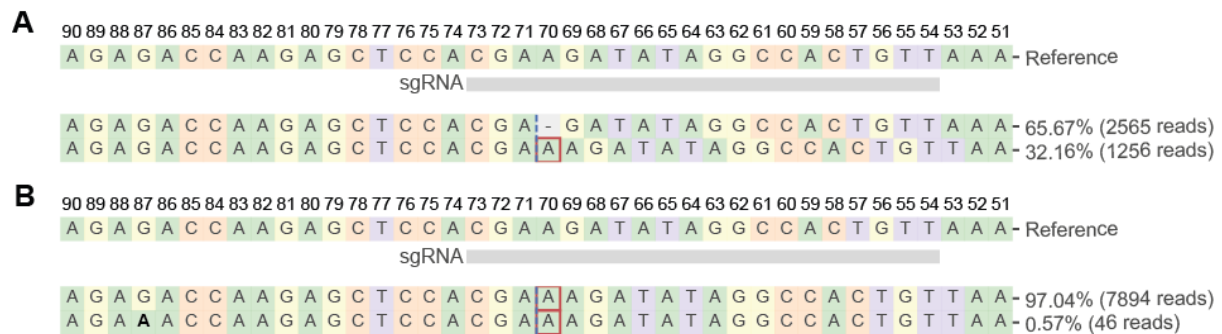
List of supplementary figures

SI Fig. 1. Establishing GalNAc-T4 KO T47D and MCF7 cells	203
SI Fig. 2. Premature stop codons generated in the GALNT4 sequence after deletion of A70 by CRISPR-Cas9	203
SI Fig. 3. Premature stop codons generated in the GALNT4 sequence after insertion of A70 by CRISPR-Cas9.	204
SI. Fig. 4. Premature stop codons generated in the GALNT4 sequence after insertion of A70 and substitution of G88A by CRISPR-Cas9	204
SI. Fig. 5. <i>In vitro</i> growth assays of WT and GalNAc-T4 KO BCCs.	205
SI Fig. 6. Protein expression control of recombinantly expressed WT and BH GalNAc-T4 by Western blot.....	205
SI Fig. 7. SnapGene map of the plasmid co-expressing BH GalNAc-T4 and the two enzymes of the artificial biosynthetic pathway (AGX1 ^{F383A} and NahK).	206
SI. Fig. 8. Characterisation of T47D ^{WT} and T47D ^{T4-KO} cells transfected with BH GalNAc-T4.	207
SI Fig. 9. Tracking tagged glycoproteins in lysates of transfected B) MCF7 ^{WT} and C) MDA-MB-231 ^{WT} cells during neutravidin bead enrichment by streptavidin blot.	208
SI Fig. 10. Testing different bioorthogonal reporter reagents to facilitate glycoproteomic analyses.....	209
SI Fig. 11. Expression construct of truncated CD46 BC tagged with GFP and a His-tag used in the MS-glycoproteomics experiments.....	209
SI Fig. 12. Expression construct of truncated CD46 C tagged with GFP and a His-tag used in the MS-glycoproteomics experiments.	210
SI Fig. 13. Assessing transduction success of the transfected T47D ^{WT} cells with viruses expressing CD46 tagged with GFP and a His-tag.	211
SI Fig. 14. Glycosidase treatment of recombinant CD46.....	212
SI Fig. 15. Expression construct of GOLM1 tagged with GFP and a His-tag used in the MS-glycoproteomics experiments.....	213
SI Fig. 16. Assessing transduction success of the transfected T47D ^{WT} cells with viruses expressing GOLM1 tagged with GFP and a His-tag	213
SI Fig. 17. Optimisation of the BH GalNAc-T2 <i>in vivo</i> system.	214

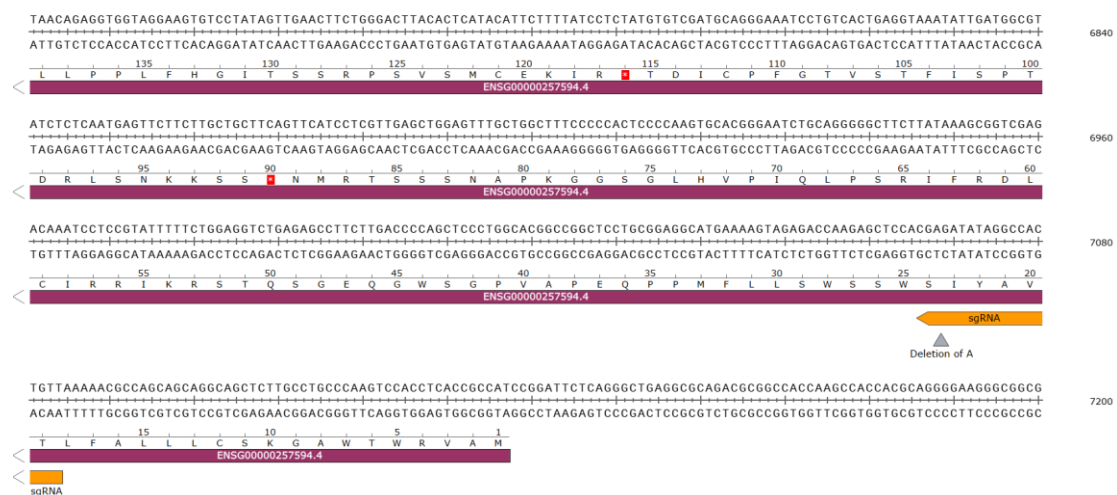
List of supplementary tables

SI Table 1. Supplier information for reagents and resources.....	215
SI Table 2. GalNAc-T4 substrates in lysates of T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH} cells	225
SI Table 3. GalNAc-T4 substrates in the secretome of T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH} cells.....	228
SI Table 4. Determining the incorporation efficiency of the heavy SILAC label on lysates of transfected MCF7 ^{WT} cells.....	229
SI Table 5. Determining the incorporation efficiency of the heavy SILAC label on lysates of transfected MDA-MB-231 ^{WT} cells.....	229
SI Table 6. GalNAc-T4 substrates in lysates of MCF7 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH} cells.....	229
SI Table 7. GalNAc-T4 substrates in lysates of MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc- T4 ^{BH} cells.....	237
SI Table 8. Protein concentration and amount from recombinant BH T4 CD46 BC and BH T4 GOLM1 after Ni-NTA purification	242

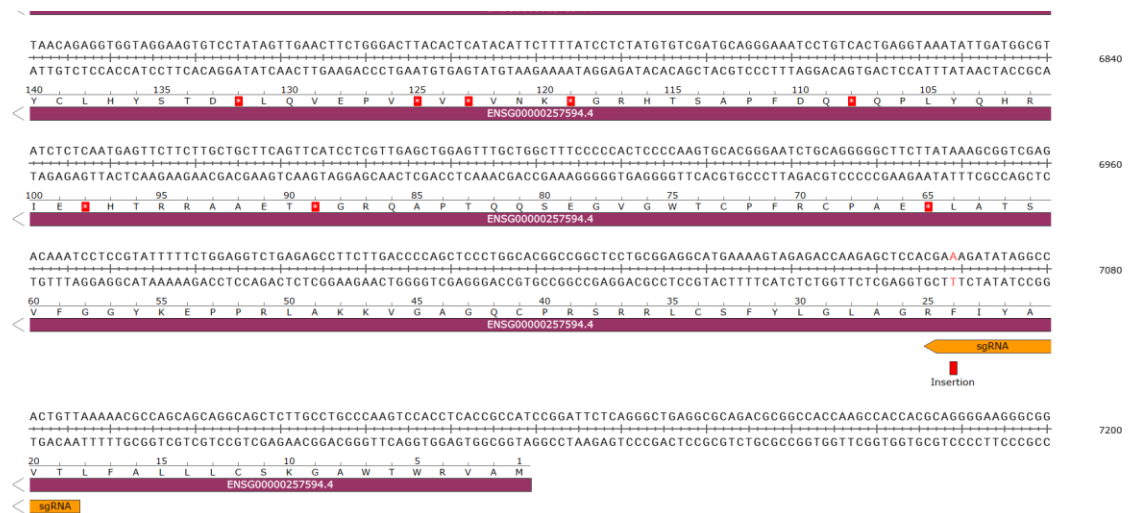
Supplementary Figures



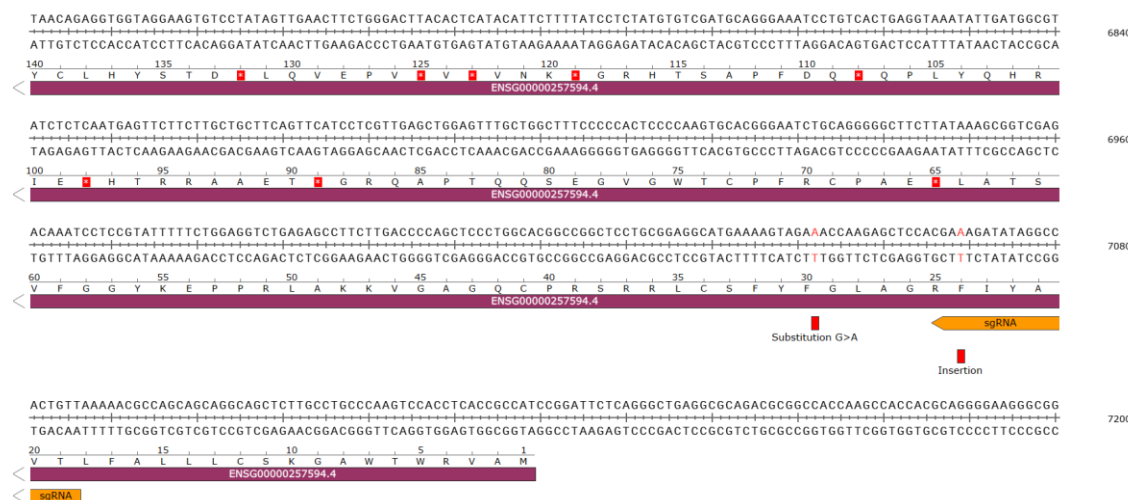
SI Fig. 1. Establishing GalNAc-T4 KO T47D and MCF7 cells. Representative allele frequency table of clones **A)** T47D^{T4-KO} ⁶ and **B)** MCF7^{T4-KO} ⁴ generated by CRISPResso2²¹⁶ after PCR-based amplicon sequencing. Nucleotides are indicated by specific colours (A = green; C = red; G = yellow; T = purple). The vertical dashed line indicates the predicted cleavage site. Red rectangles highlight inserted sequences. Horizontal dashed lines indicate deleted sequences. Substitutions are shown in bold.



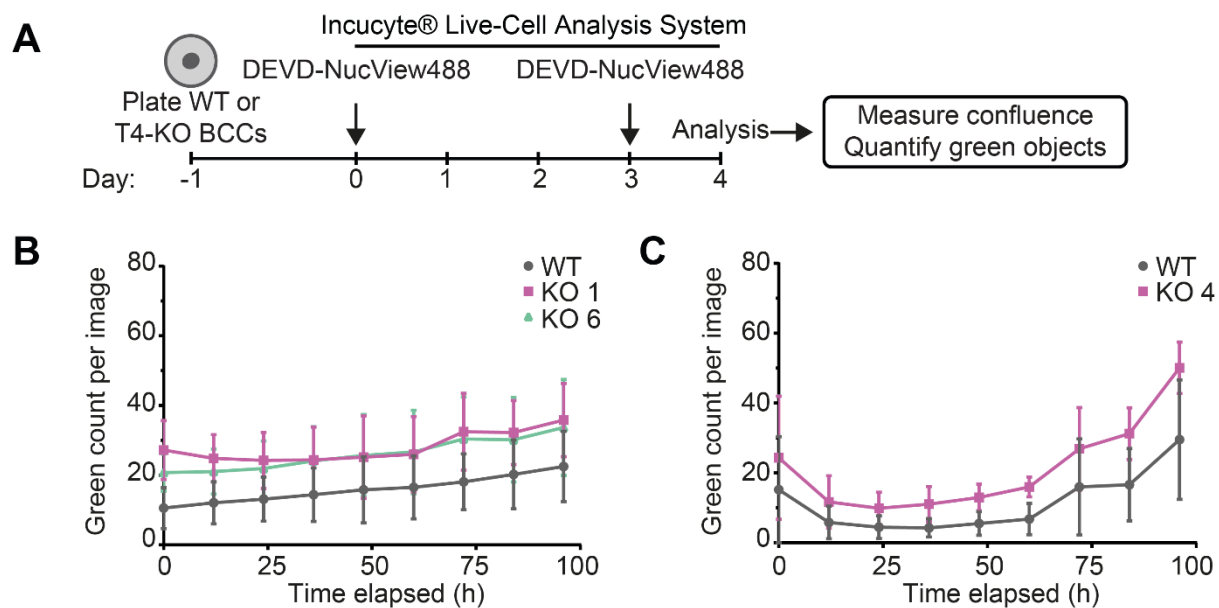
SI Fig. 2. Premature stop codons generated in the GALNT4 sequence after deletion of A70 by CRISPR-Cas9. Mutation observed in clones T47D^{T4-KO} ¹ and T47D^{T4-KO} ⁶.



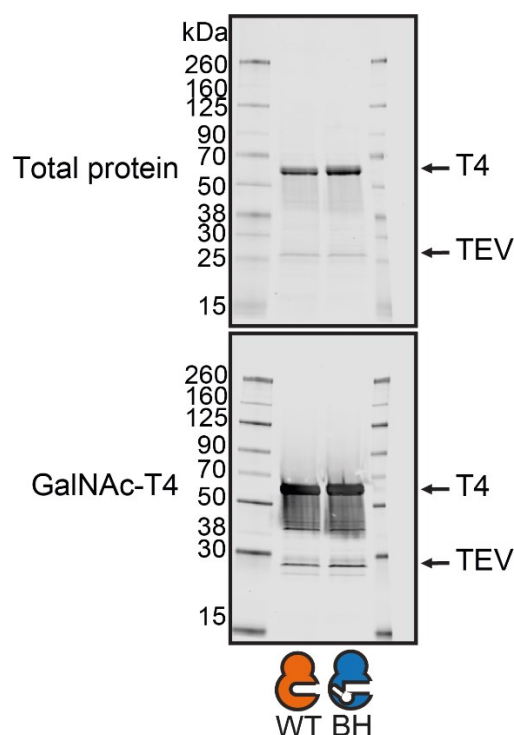
SI Fig. 3. Premature stop codons generated in the *GALNT4* sequence after insertion of A70 by CRISPR-Cas9. Mutation observed in clones T47D^{T4-KO 1}, T47D^{T4-KO 6} and MCF7^{T4-KO 4}.



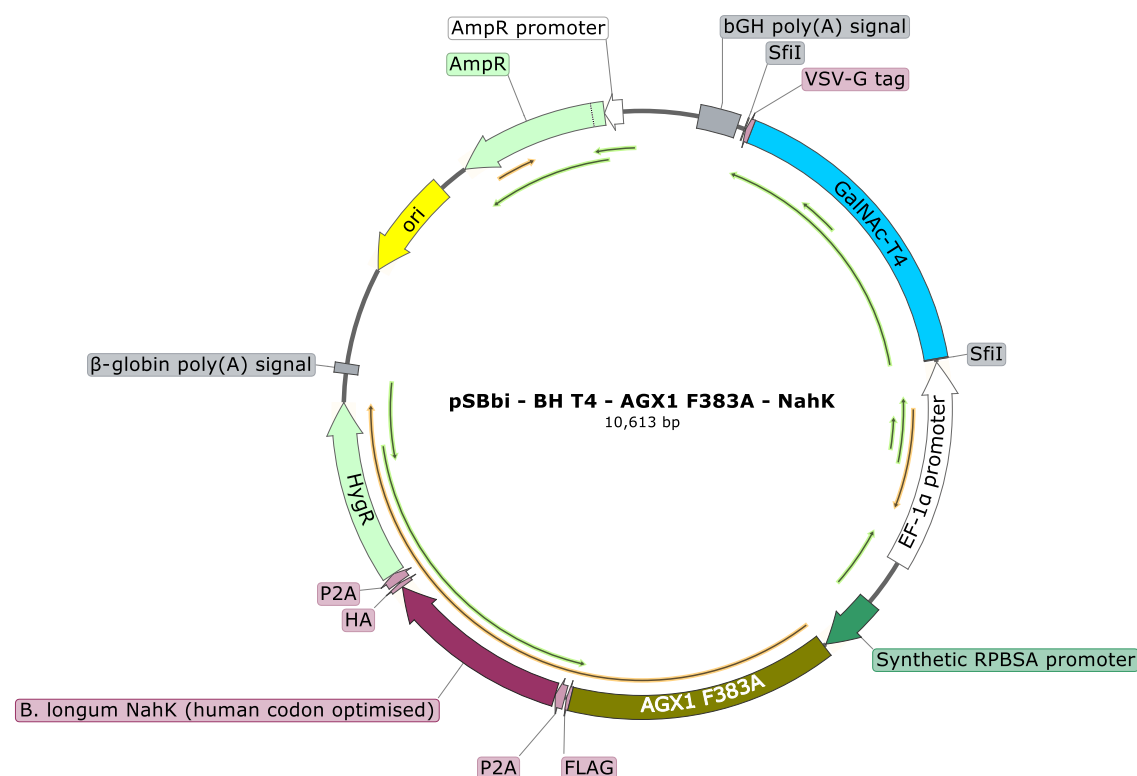
SI. Fig. 4. Premature stop codons generated in the *GALNT4* sequence after insertion of A70 and substitution of G87A by CRISPR-Cas9. Mutation observed in clone MCF7^{T4-KO 4}.



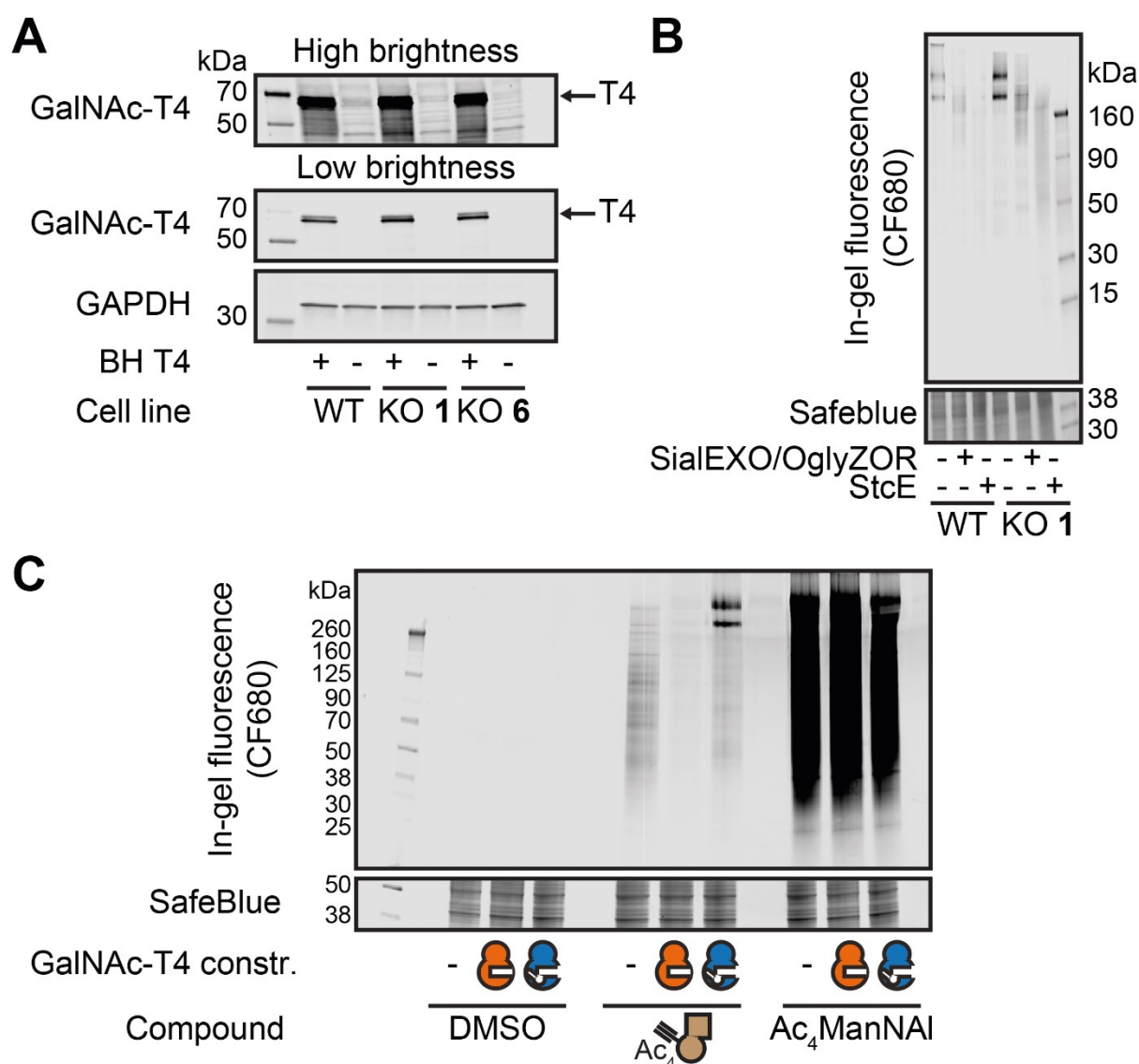
SI. Fig. 53. *In vitro* growth assays of WT and GalNAc-T4 KO BCCs. **A)** Schematic representation of the workflow to compare the growth of WT and GalNAc-T4 KO BCCs *in vitro*. The green count per image of WT and GalNAc-T4 KO **B)** T47D and **C)** MCF7 cells was determined for 4 days to assess differences in cell viability upon GalNAc-T4 KO.



SI Fig. 54. Protein expression control of recombinantly expressed WT and BH GalNAc-T4 by Western blot. 25 µg of protein were used per lane. Antibody staining was performed using an antibody against GalNAc-T4 (1:200, 12897-1-AP). The predicted molecular weight of GalNAc-T4 after TEV cleavage is 62 kDa. The predicted molecular weight of TEV is 27 kDa.



SI Fig. 55. SnapGene map of the plasmid co-expressing BH GalNAc-T4 and the two enzymes of the artificial biosynthetic pathway (AGX1^{F383A} and NahK). VSV-G-tagged BH GalNAc-T4 was cloned into the backbone of the pSBbi vector encoding FLAG-tagged AGX1^{F383A} and HA-tagged NahK using SfiI to digest the plasmid and the In-Fusion HD Cloning Kit to insert the GalNAc-T4 sequence. The pSBbi vector was designed based on the Sleeping Beauty transposable system to allow bidirectional constitutive expression and orientation-directed cloning of the genes of interest². The 2A self-cleaving peptides (P2A) allow efficient expression of multiple proteins from one promoter³. The synthetic RPBSA promoter consists of a fragment of the RPL13a promoter (chosen for its strong ubiquitous transcription in different human tissues) fused to a region of the RPL41 gene (chosen for improved expression of transcripts containing introns)². Co-transfecting cells with this plasmid and the SB100X transposase allows integration of the genes at the randomly distributed TA-dinucleotides sites in the host cell genome².

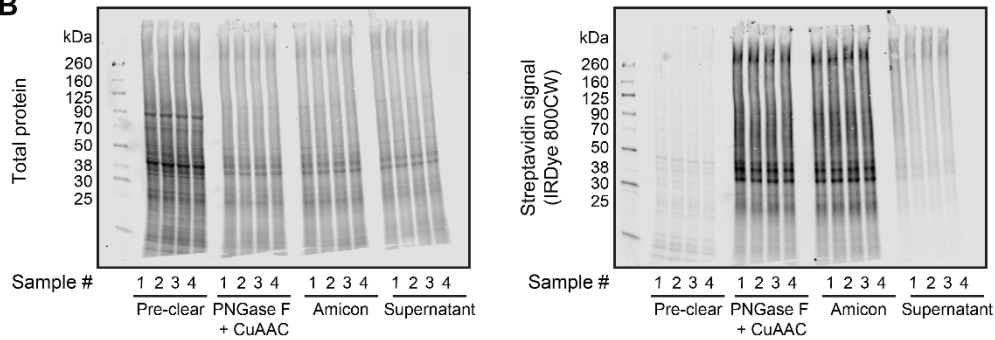


SI. Fig. 56. Characterisation of T47D^{WT} and T47D^{T4-KO} cells transfected with BH GalNAc-T4. **A)** Western blot against GalNAc-T4 to compare the expression levels of endogenous and BH GalNAc-T4 in the transfected T47D cells. The predicted MW of GalNAc-T4 and GAPDH is 67 (arrow pointing) and 36 kDa respectively. High (top) and low (bottom) brightness images included to better visualise the difference in the expression levels. GAPDH was included as a loading control. 25 µg of protein were used per lane. Antibody staining was performed using antibodies against GALNT4 (1:200, 12897-1-AP) and GAPDH (1:10000, ab181602). **B)** Labelling analysis by in-gel fluorescence after treating the metabolically labelled T47D cell lysates with the glycosidases SialEXO (digests sialic acids) and OglyZOR (digests core 1 and core 3 O-glycans in the presence of SialEXO) or StcE (digests mucin-type glycoproteins). 10 µg of protein were used per lane. **C)** Cell surface tagging of glycoproteins in transfected T47D^{WT} cell lysates without PNGase F treatment as assessed by in-gel fluorescence. 15 µg of protein were used per lane. Data are from one experiment.

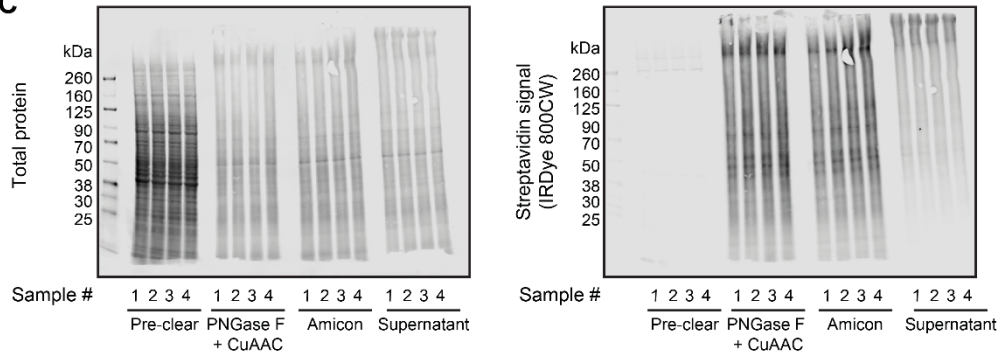
A

SILAC label	Heavy		Light	
Sample #	Cell line	Treatment	Cell line	Treatment
1	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	DMSO
2	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	DMSO	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne
3	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK	Ac ₄ GalN6yne
4	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK	Ac ₄ GalN6yne	MCF7 ^{WT} /MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne

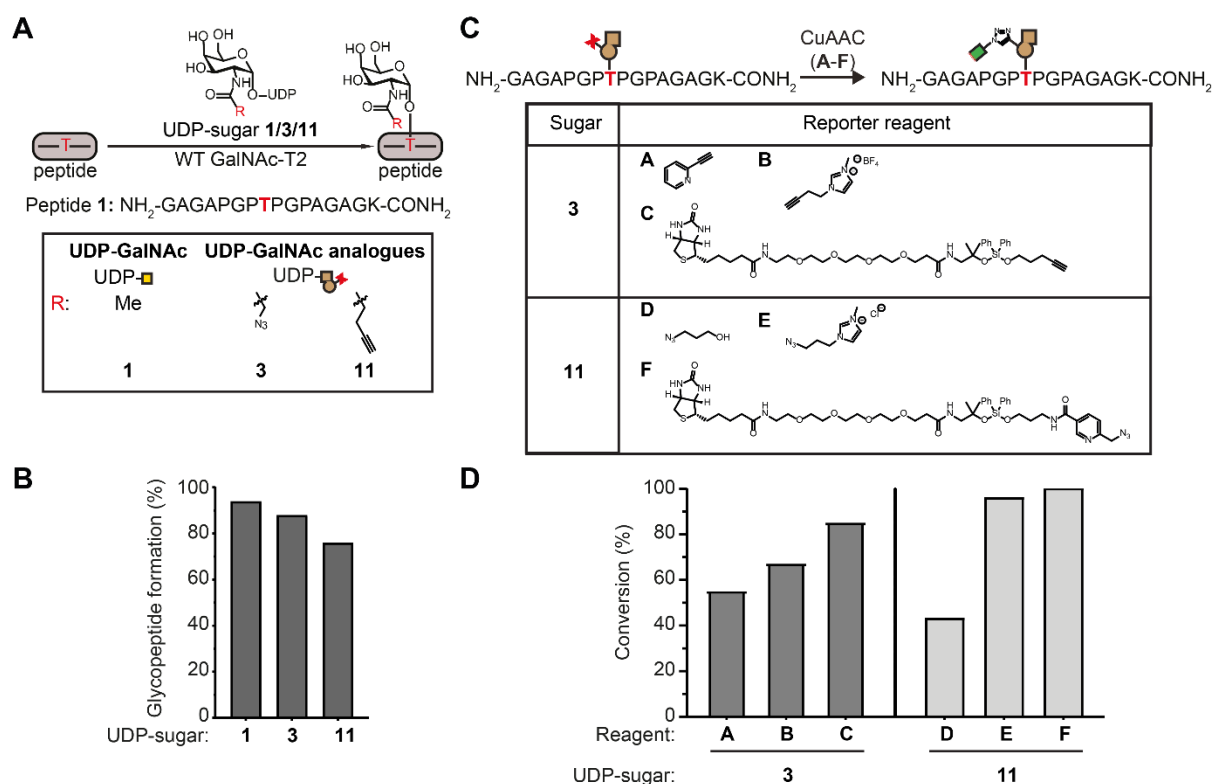
B



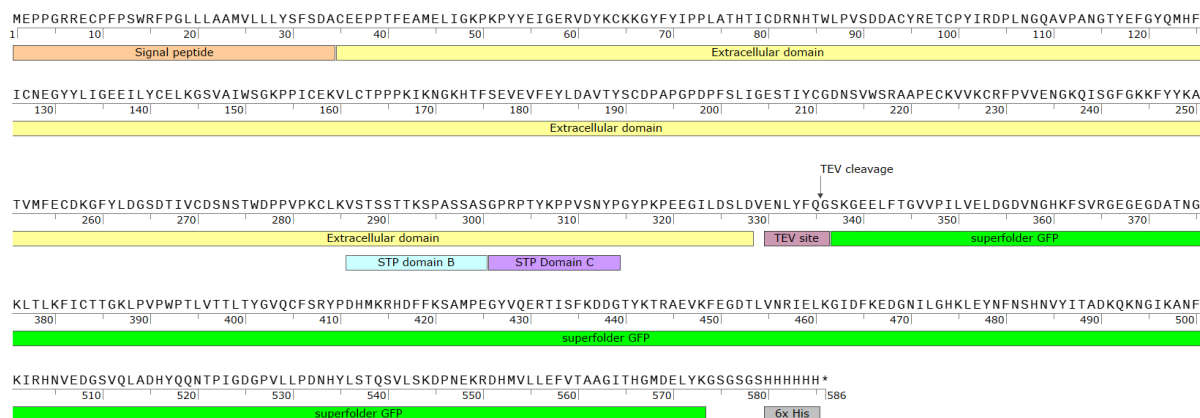
C



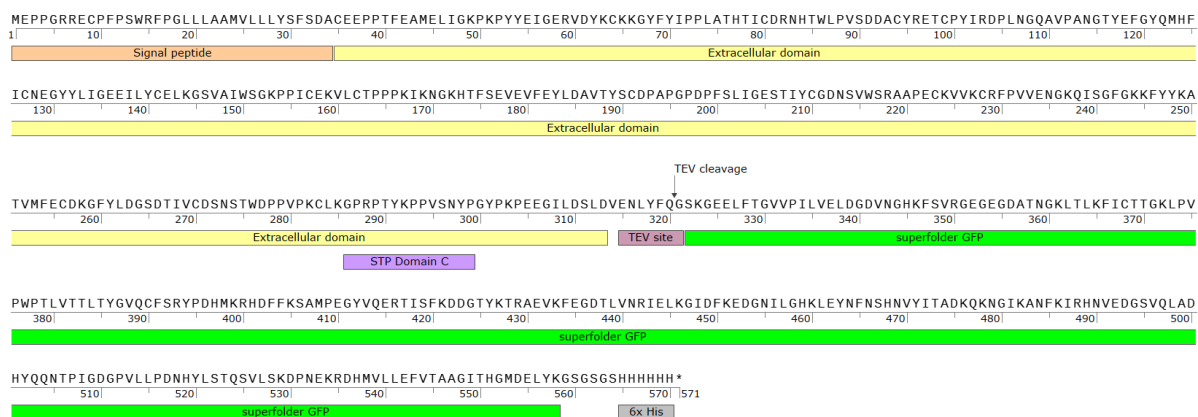
SI Fig. 57. Tracking tagged glycoproteins in lysates of transfected B) MCF7^{WT} and C) MDA-MB-231^{WT} cells during neutravidin bead enrichment by streptavidin blot. A) 1:1 mixtures of heavy and light MCF7 and MDA-MB-231 lysates. Data is representative of three independent replicates.



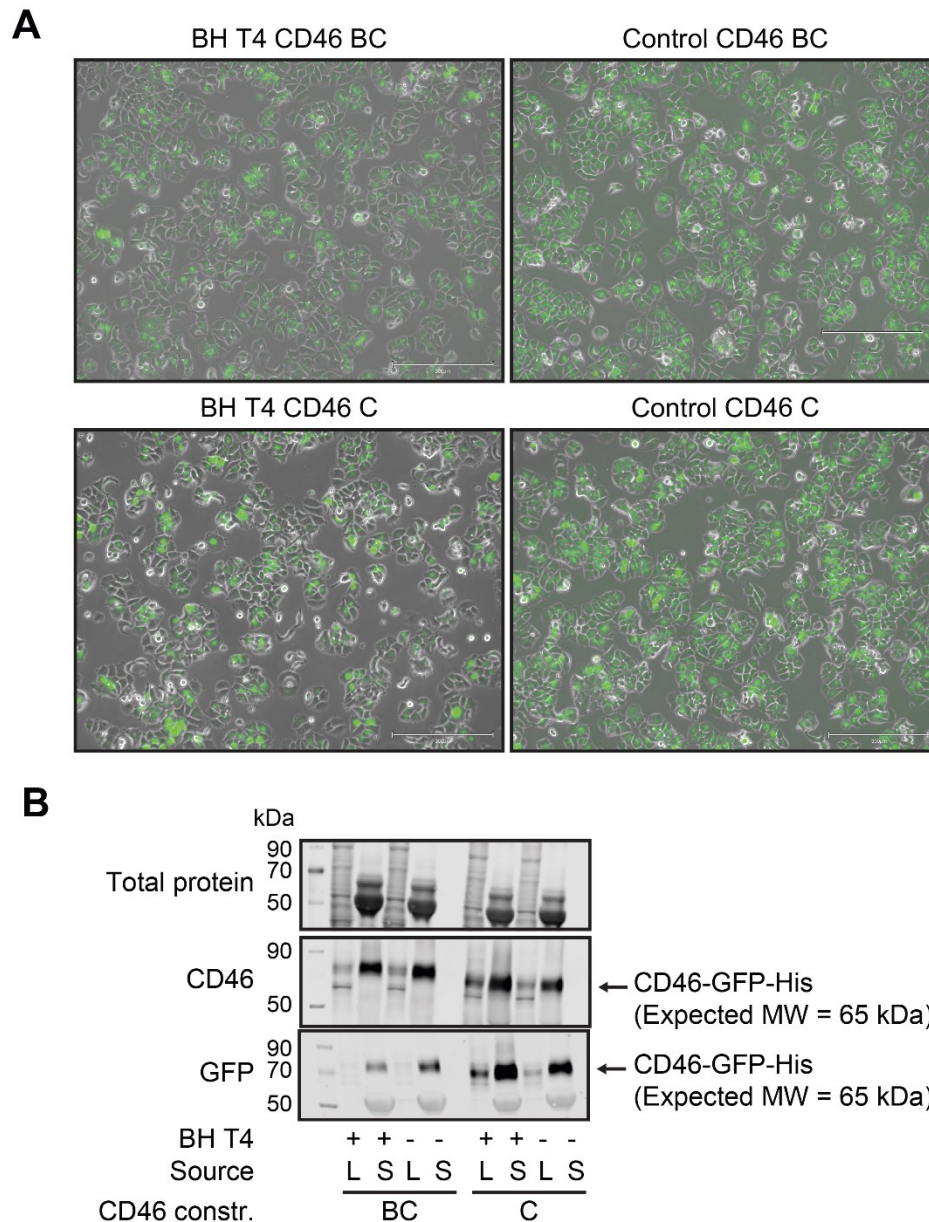
SI Fig. 58. Testing different bioorthogonal reporter reagents to facilitate glycoproteomic analyses. A) Schematic representation of *in vitro* glycosylation experiments performed with WT GalNAc-T2 and UDP-sugars **1**, **3** and **11**. . Figure adapted from Choi *et al.* (2019)⁴. **B)** Glycopeptide formation by WT GalNAc-T2 (100 nM) after incubation for 25 h at 37 °C with acceptor peptide (50 μM) and compounds **1**, **3** or **11** (51 μM). **C)** Schematic representation of the CuAAC reaction between the tagged glycopeptides and a panel of bioorthogonal reporter reagents. **D)** Conversion of the CuAAC reaction with the different reporter reagents. Data are from one experiment.



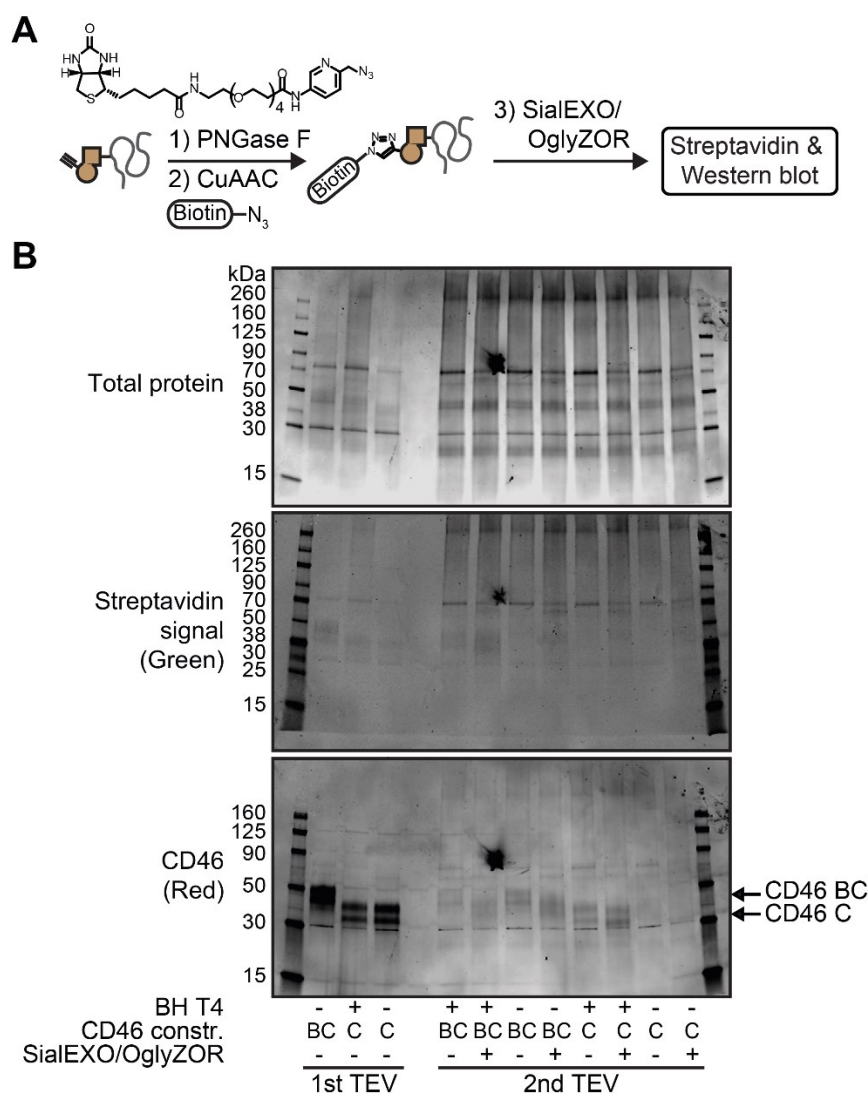
SI Fig. 59. Expression construct of truncated CD46 BC tagged with GFP and a His-tag used in the MS-glycoproteomics experiments.



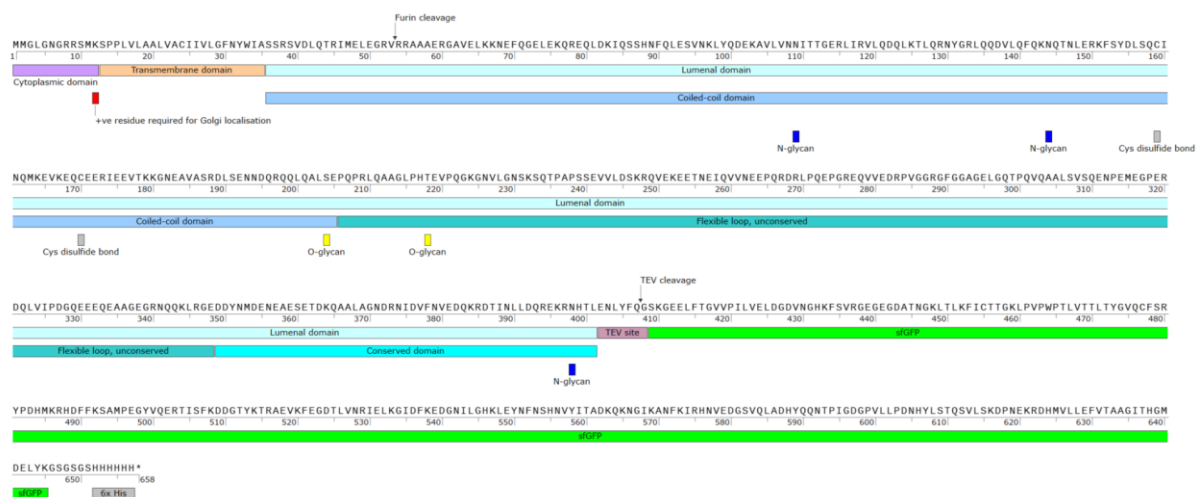
SI Fig. 60. Expression construct of truncated CD46 C tagged with GFP and a His-tag used in the MS-glycoproteomics experiments.



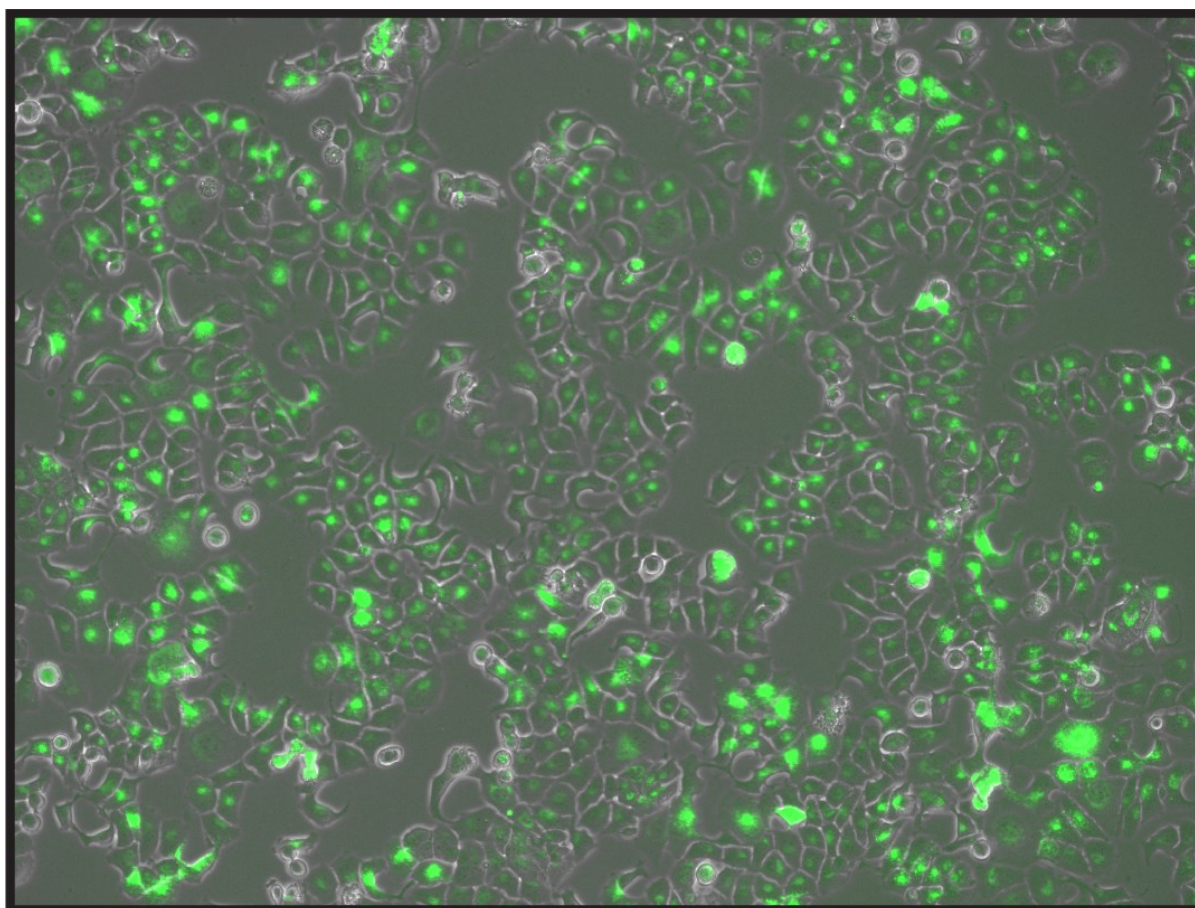
SI Fig. 61. Assessing transduction success of the transfected T47D^{WT} cells with viruses expressing CD46 tagged with GFP and a His-tag. A) Cells were imaged using an EVOS fluorescence imager and a 10X objective. **B)** Protein expression control by Western blot. 15 µg of protein were used per lane. Antibody staining was performed using antibodies against CD46 (1:500, 12494-1-AP) and GFP (1:2000, 66002-1-Ig). T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells ("BH T4" cells), and T47D^{WT}-AGX1^{F383A}-NahK cells as a control ("control" cells) were used for these experiments. L = lysate, S = secretome.



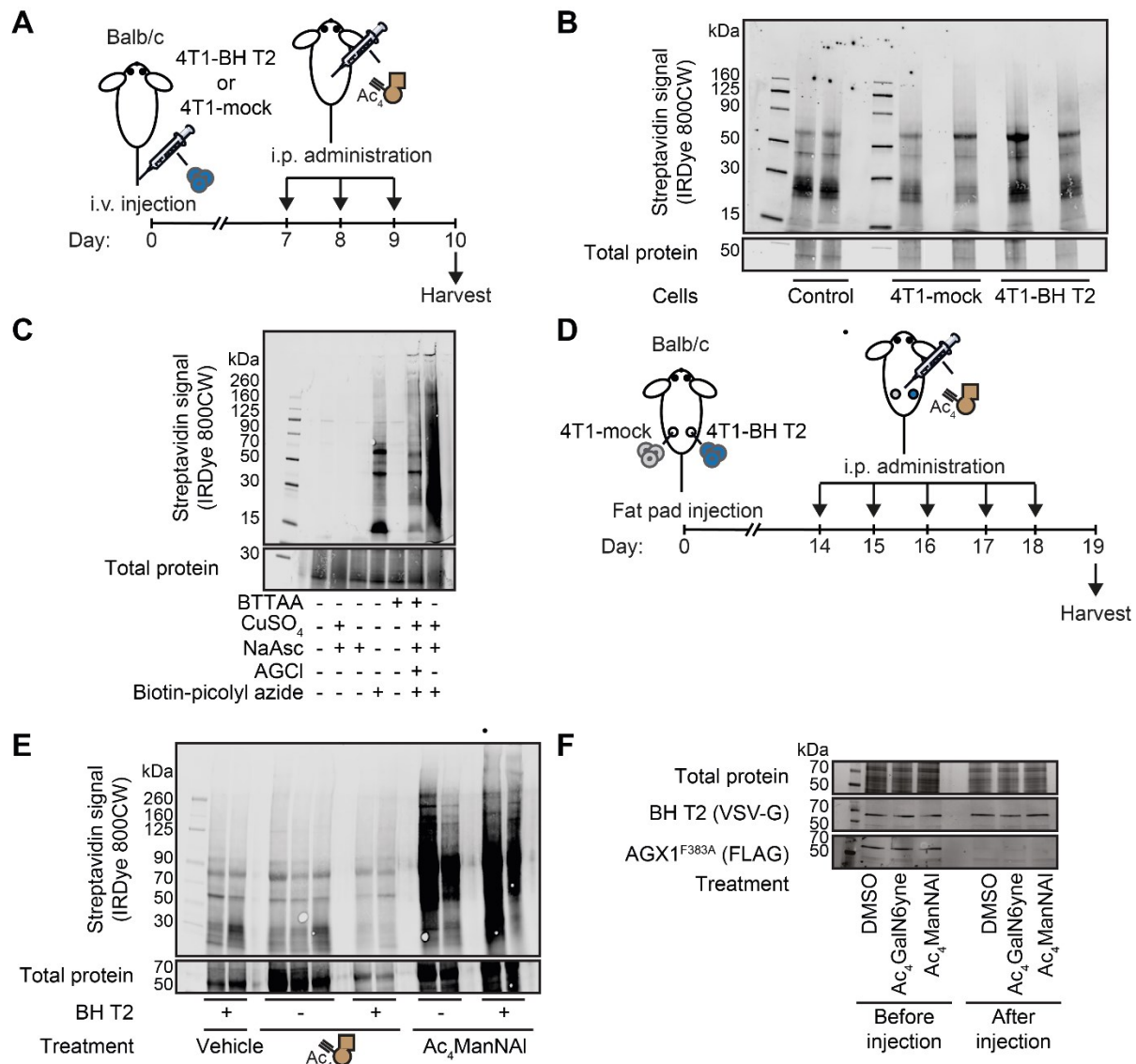
SI Fig. 62. Glycosidase treatment of recombinant CD46. **A)** Schematic representation of the experiment. The eluted fractions after 1st and 2nd TEV cleavage were subjected to CuAAC reaction with biotin-picolyl azide and glycosidase treatment. **B)** Tagging and protein expression control by Western blot after the eluted fractions were subjected to CuAAC reaction with biotin picolyl azide and glycosidase treatment. The stainings were performed on the same membrane in the order displayed. Red/Green refers to the imaging channel. The expected MW of the CD46 BC and C constructs after TEV cleavage is 38 and 36 kDa respectively (arrows pointing). 2 µg (1st TEV) and 5 µg (2nd TEV) of protein were used per lane. Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230). Antibody staining was performed using an antibody against CD46 (1:500, 12494-1-AP).



SI Fig. 63. Expression construct of GOLM1 tagged with GFP and a His-tag used in the MS-glycoproteomics experiments.



SI Fig. 64. Assessing transduction success of the transfected T47D^{WT} cells with viruses expressing GOLM1 tagged with GFP and a His-tag. Cells were imaged using an EVOS fluorescence imager and a 10X objective.



SI Fig. 65. Optimisation of the BH GalNAc-T2 *in vivo* system. **A)** Schematic representation of the workflow for the *in vivo* tagging of glycoproteins using BH GalNAc-T2 (BH T2) in a lung metastasis model. 4T1 cells transfected with the BH-GalNAc-T2 (4T1-BH T2) or 4T1-mock transfected cells (4T1-mock) were i.v. injected into the tail vein of BALB/c mice by Dr Victoria Bridgeman. 7 days after tumour cell injection, the mice were treated i.p. with Ac₄GalN6yne once daily for 3 days by Dr Victoria Bridgeman. 24 h after the final injection, the lungs were harvested and homogenised by Dr Anna Cioce. Lungs from untreated mice were also collected as a control. Endogenous biotinylated proteins were removed by incubation with neutravidin beads. The supernatant was collected and reacted with biotin-picolyl azide. **B)** Tagging of glycoproteins in lung homogenates as assessed by streptavidin blot after i.v. injection of 4T1-BH T2 or 4T1-mock transfected cells into BALB/c mice. 55 µg of protein were used per lane. **C)** Identifying which component of the CuAAC reaction is the source of the non-specific signal using control lung homogenates (by Dr Anna Cioce). 30 µg of protein were used per lane. **D)** Schematic representation of the workflow for the *in vivo* tagging of glycoproteins using BH T2 in a mammary fat pad model. BH T2 tumours were grown in the fat pad of BALB/c mice. Mock tumours were grown in the same mouse as an internal control. 14 days after tumour cell injection, the mice were treated i.p. with Ac₄GalN6yne, Ac₄ManNAI or the corresponding volume of vehicle once daily for 5 days by Dr Victoria Bridgeman and Tatiana Rizou. 24 h after the final injection, the tumours were harvested, homogenised, and reacted with biotin-picolyl azide. An extra piece of the BH T2 tumours was digested, and the cells plated for protein expression control by Western blot. **E)** Tagging of glycoproteins in tumour homogenates as assessed by streptavidin blot after fat pad injection of 4T1-BH T2 or 4T1-mock transfected cells into BALB/c mice. 30 µg of protein were used per lane. **F)** Protein expression analysis of the digested BH T2 tumours by Western blot (by Dr Anna Cioce). The predicted MW of GalNAc-T2 and AGX1 is 65 and 58 kDa respectively. 15 µg of protein were used per lane. Antibody staining was performed using antibodies against VSV-G (1:1000, ab386) and FLAG (1:1000, PA1-984B). Streptavidin staining was performed using IRDye® 800CW Streptavidin (1:5000, 926-32230).

Supplementary Tables

SI Table 3. Supplier information for reagents and resources.

REAGENT/RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CD46 antibody (rabbit, polyclonal)	Proteintech (Rosemont, USA)	Proteintech Cat# 12494-1-AP; RRID: AB_2260042
Anti-FLAG antibody (rabbit, polyclonal)	Thermo Fisher Scientific (Waltham, USA)	Thermo Fisher Scientific Cat# PA1-984B; RRID: AB_347227
Anti-FLAG antibody (goat, polyclonal)	Abcam (Cambridge, UK)	Abcam Cat# ab1257; RRID:AB_299216
anti-GALNT4 antibody (rabbit, polyclonal)	Proteintech	Proteintech Cat# 12897-1-AP; RRID: AB_10643389
anti-GAPDH antibody (rabbit, monoclonal)	Abcam	Abcam Cat# ab181602; RRID: AB_2630358
Anti-GFP antibody (mouse, monoclonal)	Proteintech	Proteintech Cat# 66002-1-Ig; RRID: AB_11182611
Anti-GOLM1 antibody (rabbit, polyclonal)	Atlas Antibodies (Stockholm, Sweden)	Atlas Antibodies Cat# HPA010638, RRID: AB_1849886
Anti-HA antibody (mouse, monoclonal)	Abcam	Abcam Cat# ab130275; RRID: AB_11156884
Anti-HA antibody (mouse, monoclonal)	Abcam	Abcam Cat# ab18181; RRID: AB_444303
Anti-HA antibody (rabbit, polyclonal)	Abcam	Abcam Cat# ab9110; RRID: AB_307019
Anti-MUC1 antibody (mouse, monoclonal)	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# MA1-06503; RRID: AB_558126
Anti-VSV-G antibody (goat, polyclonal)	Abcam	Abcam Cat# ab3861; RRID: AB_304118
Anti-VSV-G antibody (rabbit, polyclonal)	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# PA1-29903; RRID: AB_1961363
Anti-VSV-G antibody (mouse, monoclonal)	Abcam	Abcam Cat# ab50549; RRID: AB_883494
IRDye® 680RD Donkey anti-goat IgG	LI-COR Biosciences (Lincoln, USA)	LI-COR Biosciences Cat# 926-68074; RRID AB_10956736
IRDye® 680RD Donkey anti-mouse IgG	LI-COR Biosciences	LI-COR Biosciences Cat# 926-68072; RRID AB_10953628
IRDye® 680RD Donkey anti-rabbit IgG	LI-COR Biosciences	LI-COR Biosciences Cat# 926-68073; RRID AB_10954442
IRDye® 680RD Streptavidin	LI-COR Biosciences	LI-COR Biosciences Cat# 926-68079
IRDye® 800CW Donkey anti-goat IgG	LI-COR Biosciences	LI-COR Biosciences Cat# 926-32214; RRID AB_621846

IRDye® 800CW Donkey anti-mouse IgG	LI-COR Biosciences	LI-COR Biosciences Cat# 926-32212; RRID AB_621847
IRDye® 800CW Donkey anti-rabbit IgG	LI-COR Biosciences (Lincoln, USA)	LI-COR Biosciences Cat# 926-32213; RRID AB_621848
IRDye® 800CW Streptavidin	LI-COR Biosciences	LI-COR Biosciences Cat# 926-32230
Chemicals, peptides, and recombinant proteins		
0.05% Trypsin-EDTA (1X)	Gibco (Grand Island, USA)	Gibco Cat# 25300
1 Kb Plus DNA Ladder	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 10787018
4X Protein Loading Buffer	LI-COR Biosciences	LI-COR Biosciences Cat# 928-40004
10X CutSmart Buffer	New England Biolabs (Ipswich, USA)	New England Biolabs Cat# B6004S
10X Tris/Glycine/SDS Buffer	Bio-Rad Laboratories (Hercules, USA)	Bio-Rad Cat# 1610772
50X TAE (solution)	Media Preparation Facility (Francis Crick Institute)	N/A
100 bp DNA ladder	Invitrogen (Waltham, USA)	Invitrogen Cat# 15628019
Ac ₄ ManNAI	Click Chemistry Tools (Scottsdale, USA)	Click Chemistry Tools Cat# 1154-100; CAS No. 935658-93-8
Acetic acid glacial	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# A-0400-PB17
Ac ₄ GalN6yne	Synthesised by Ganka Bineva-Todd (Schumann Lab)	N/A
Aminoguanidine hydrochloride	Cayman Chemical Company (Ann Arbor, USA)	Cayman Chemical Company Cat# 81530; CAS No. 1937-19-5
Ammonium bicarbonate (solid)	Sigma Aldrich (St. Louis, USA)	Sigma Aldrich Cat# A6141; CAS No. 1066-33-7
β-oestradiol	Sigma Aldrich	Sigma Aldrich Cat# E8875; CAS No. 50-28-2
Bathocuproinedisulfonic acid	Sigma Aldrich	Sigma Aldrich Cat# B1125; CAS No. 52698-84-7
BbsI Restriction Enzyme	New England Biolabs	New England Biolabs Cat# R0539S
Benzonase® Nuclease	Merck (Darmstadt, Germany)	Merck Cat# E1014-5KU; CAS No. 9025-65-4
Biotin-picolyl azide	Sigma Aldrich	Sigma Aldrich Cat# 900912
BTAA	Jena Bioscience (Jena, Germany)	Jena Bioscience Cat# CLK-067-25; CAS No. 1334179-85-9
Calf Intestinal Alkaline Phosphatase	Invitrogen	Invitrogen Cat# 18009-019
Chameleon® Duo Pre-stained Protein Ladder	LI-COR Biosciences	LI-COR Biosciences Cat# 928-60000
Chymotrypsin, Sequencing Grade	Promega (Madison, USA)	Promega Cat# V1062
CF680 picolyl azide	Biotium (Fremont, USA)	Biotium Cat# 96003

CloneAmp™ HiFi PCR Premix	Takara Bio (Kusatsu, Japan)	Takara Cat# 639298
Compound 1 (UDP-GalNAc)	Sigma Aldrich	Sigma Aldrich Cat# U5252; CAS No. 108320-87-2
Compound 2 (UDP-GalN6yne)	Synthesised by Ganka Bineva-Todd (Schumann Lab); Choi <i>et al.</i> (2019) ⁴	N/A
Compound 3 (UDP-GalNAz)	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 4 (UDP-GalNAzMe (R))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 5 (UDP-GalNAzMe (S))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 6 (UDP-GalNAzEt (S))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 7 (UDP-NAzMeOMe (R))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 8 (UDP-NAzMeOMe (S))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 9 (UDP-PrAzMe (R))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 10 (UDP-PrAzMe (S))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 11 (UDP-GalNAIk)	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 12 (UDP-GalNAIkMe (R))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 13 (UDP-GalNAIkMe (S))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 14 (UDP-GalN6yneMe (S))	Choi <i>et al.</i> (2019) ⁴	N/A
Compound 15 (UDP-GalNDAz)	Gift from Jennifer Kohler (UT Southwestern Medical Center)	N/A
Compound 16 (UDP-GalNCyoc)	Synthesised by Ganka Bineva-Todd (Schumann Lab)	N/A
Copper (II) sulfate pentahydrate	Sigma Aldrich	Sigma Aldrich Cat# C8027; CAS No. 7758-99-8
DEVD-NucView488 peptide	Chemical Biology STP (Francis Crick Institute); Cen <i>et al.</i> (2008) ⁵	N/A
DNase I	Sigma Aldrich	Sigma Aldrich Cat# DN25; CAS No. 9003-98-9
DMEM	Gibco	Gibco Cat# 41966
DMEM for SILAC	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 88364
DMSO	Sigma Aldrich	Sigma Aldrich Cat# D2650; CAS No.
DMSO Cell Culture reagent	Santa Cruz Biotechnology (Dallas, USA)	Santa Cruz Biotechnology Cat# SC-358801; CAS No. 67-68-5
DTT, molecular biology grade (solid)	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# R0862

DTT (solution)	Sigma Aldrich	Sigma Aldrich Cat# 646563; CAS No. 3463-12-3
EDTA pH 8 (solution, 0.5M)	Media Preparation Facility	N/A
Ethanol	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# E-0555DF-17
FBS	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# A3840402
FBS (dialysed)	Labtech (Heathfield, UK)	Labtech Cat# FB-1001D/500
GalNAc-T2 peptide substrate: NH ₂ -GAGAPGPTPGPAGAGK-CONH ₂	Chemical Biology STP (Francis Crick Institute); Calle <i>et al.</i> (2021) ⁶	N/A
GalNAc-T4 glycopeptide substrate: NH ₂ -KGGGGAT*GAGAGAGTTP GPG-CONH ₂	Chemical Biology STP; Calle (2020) ⁷	N/A
GelGreen Nucleic Acid Stain	Biotium	Biotium Cat# 41005
Gel Loading Dye Purple (6X)	New England Biolabs	New England Biolabs Cat# B7024A
GFR Matrigel	Corning (New York, USA)	Corning Cat# 356231
GFP-clamps	Structural Biology STP (Francis Crick Institute)	N/A
Glycerol	Sigma Aldrich	Sigma Aldrich Cat# G2025; CAS No. 56-81-5
Halt™ Protease Inhibitor Cocktail (100X)	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 1861279
HBSS without CaCl ₂ or MgCl ₂	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 14175-053
HEPES pH 7.5 (solution, 1M)	Media Preparation Facility	N/A
Hygromycin B	Thermo Fisher Scientific	Thermo Fischer Scientific Cat# 10687010
Imidazole	Sigma Aldrich	Sigma Aldrich Cat# 56750; CAS No. 288-32-4
Intercept Blocking Buffer	LI-COR Biosciences	LI-COR Biosciences Cat# 927-60001
Intercept Protein-Free Blocking Buffer	LI-COR Biosciences	LI-COR Biosciences Cat# 927-80001
Iodoacetamide (solid)	Merck	Merck Cat# I5161, CAS No. 144-48-9
K0 L-Lysine	Cambridge Isotope Laboratories (Tewksbury, USA)	Cambridge Isotope Laboratories Cat# ULM-8766
K8 L-Lysine	Cambridge Isotope Laboratories	Cambridge Isotope Laboratories Cat# CNLM-291-H
L-ascorbic acid sodium salt (sodium ascorbate)	ACROS Organics (Geel, Belgium)	ACROS Organics Cat# 352685000; CAS No. 134-03-2
LC-MS grade acetonitrile	Thermo Fisher Scientific	Thermo Fischer Scientific Cat# 51101, CAS No. 75-05-8

LC-MS grade formic acid	Honeywell Research Chemicals (Morris Plains, USA)	Honeywell Cat# 14265, CAS No. 64-18-6
LC-MS grade methanol	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# A456-1, CAS No. 67-56-1
LC-MS grade water	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# W6-212, CAS No. 7732-18-5
Liberase TH	Roche (Basel, Switzerland)	Roche Cat# 05401151001
Liberase TM	Roche	Roche Cat# 05401127001
Lys-C, Mass Spec Grade	Promega (Madison, USA)	Promega Cat# VA1170
Methanol	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# M-3900-17
MgCl ₂ (solution)	Sigma Aldrich	Sigma Cat# 63069; CAS No. 7786-30-3
MnCl ₂ (solid)	Sigma Aldrich	Sigma Cat# M5005; CAS No. 13446-34-9
NaCl (solution, 5M)	Media Preparation Facility	N/A
Ni-NTA® Agarose beads	Qiagen (Hilden, Germany)	Qiagen Cat# 30210
NiSO ₄ (solid)	Sigma Aldrich	Sigma Cat# 227676; CAS No. 10101-97-0
Nuclease Free Water	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# AM9937
OglyZOR	Genovis (Lund, Sweden)	Genovis Cat# G2-OG1-020
Opti-MEM® I Reduced Serum Medium	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 31985062
PBS (with CaCl ₂ and MgCl ₂)	Sigma Aldrich	Sigma Cat# D8662
PBS (without CaCl ₂ and MgCl ₂)	Sigma Aldrich	Sigma Cat# D8537
P0 L-Proline	Cambridge Isotope Laboratories	Cambridge Isotope Laboratories Cat# ULM-8333-PK
Penicillin Streptomycin (PenStrep)	Gibco	Gibco Cat# 15140-122
Phosphoric acid	Sigma Aldrich	Sigma Cat# 49685
Polyethylene Glycol 400 (PEG-400)	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# P-3676-08
PNGase F	Promega	Promega Cat# V4831A
PUGNAc	Sigma Aldrich	Sigma Cat# A7229; CAS No. 132489-69-1
Puromycin dihydrochloride	Santa Cruz Biotechnology	Santa Cruz Biotechnology Cat# SC-108071, CAS No. 58-58-2
Q5® High-Fidelity 2X Master Mix	New England Biolabs	New England Biolabs Cat# M0494S
SafeBlue Protein Stain	NBS Biologicals (Huntingdon, UK)	NBS Biologicals Cat# NBS-SB1L

SDS (solid)	Sigma Aldrich	Sigma Cat# 75746; CAS No. 151-21-3
SDS (solution, 20% (v/v))	Media Preparation Facility	N/A
Sera-Mag™ SpeedBead Carboxylate-Modified [E7] Magnetic Particles	Cytiva (Marlborough, USA)	Cytiva Cat# 45152105050250
Sera-Mag SpeedBead Carboxylate-Modified [E3] magnetic particles	Cytiva	Cytiva Cat# 65152105050250
Sera-Mag SpeedBeads Neutravidin-Coated Magnetic Particles	Cytiva	Cytiva Cat# 78152104011150
Sodium deoxycholate (solid)	Sigma Aldrich	Sigma Cat# D6750; CAS No. 302-95-4
R0 L-Arginine	Cambridge Isotope Laboratories	Cambridge Isotope Laboratories Cat# ULM-8347-PK
R10 L-Arginine	Cambridge Isotope Laboratories	Cambridge Isotope Laboratories Cat# CNLM-539-H-PK
RapiGest SF	Waters (Milford, USA)	Waters Cat# 186002122
Recombinant full-length GOLM1	This paper	N/A
Recombinant soluble BH GalNAc-T4	Structural Biology STP; Calle (2020) ⁷	N/A
Recombinant soluble CD46 BC	This paper	N/A
Recombinant soluble CD46 BC	This paper	N/A
Recombinant soluble CD46 C	This paper	N/A
Recombinant soluble CD46 C	This paper	N/A
Recombinant soluble WT GalNAc-T2	Structural Biology STP; Calle <i>et al.</i> (2021) ⁶	N/A
Recombinant soluble WT GalNAc-T4	Structural Biology STP; Calle (2020) ⁷	N/A
Reporter reagent A (2-Ethynylpyridine)	Sigma Aldrich	Sigma Aldrich Cat# 469920; CAS No. 1945-84-2
Reporter reagent B (iTag-alkyne)	Gift from Carmen Galan (University of Bristol)	N/A
Reporter reagent C (Biotin-DADPS-alkyne)	Click Chemistry Tools	Click Chemistry Tools Cat# CCT-1331
Clickable reagent D (3-Azido-1-propanol)	Sigma Aldrich	Sigma Aldrich Cat# 776130; CAS No. 72320-38-8
Reporter reagent E (iTag-azide)	Synthesised by Andrea Marchesi (Schumann Lab)	N/A
Reporter reagent F (Biotin-DADPS-picolyl azide)	Schumann <i>et al.</i> (2020) ⁸ ; subsequently custom	N/A

	synthesized from Sussex Research Laboratories (Ontario, USA)	
RIPA buffer	Sigma Aldrich	Sigma Aldrich Cat# R0278
RPMI-1640 Medium	Merck	Merck Cat# R8758
RPMI 1640 Media for SILAC	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 88365
SialEXO	Genovis	Genovis Cat# G1-SM1-020
SmE	Gift from Stacy Malaker (Yale University)	N/A
Sodium deoxycholate (solid)	Sigma Aldrich	Sigma Cat# D6750; CAS No. 302-95-4
StcE	Gift from Stacy Malaker	N/A
T4 DNA Ligase Reaction Buffer	New England Biolabs	New England Biolabs Cat# B0202S
T4 Polynucleotide Kinase	New England Biolabs	New England Biolabs Cat# M0201
TEV protease	Structural Biology STP	N/A
TransducelT Transduction Reagent	Mirus Bio (Madison, USA)	Mirus Bio Cat# MIR 6620
TransIT-BrCa Transfection Kit	Mirus Bio	Mirus Bio Cat# MIR 5500
Trypsin Gold, Mass Spectrometry Grade	Promega	Promega Cat# V528A
Tris-HCl pH 6.5 (solution, 1M)	Media Preparation Facility	N/A
Tris-HCl pH 7.5 (solution, 1M)	Media Preparation Facility	N/A
Tris-HCl pH 8 (solution, 1M)	Media Preparation Facility	N/A
Triton X-100	Merck	Merck Cat# 108643; CAS No. 9035-19-5
Tween-20 Ultra-Pure	Rockland (Limerick, USA)	Rockland Cat#TW0020
UltraPure™ Agarose	Invitrogen	Invitrogen Cat# 16500
Urea	Sigma Aldrich	Sigma Cat# 51456; CAS No. 57-13-6
Commercial assays		
NucleoSpin® Gel and PCR Clean-up Kit	Macherey-Nagel (Düren, Germany)	Macherey-Nagel Cat# 740609.50
Pierce™ BCA Protein Assay Kit	Thermo Fischer Scientific	Thermo Fischer Scientific Cat# 23227
QIAquick Gel Extraction Kit	Qiagen	Qiagen Cat# 28704
QIAprep Spin Miniprep Kit	Qiagen	Qiagen Cat# 27106
Quick Ligation™ Kit	New England Biolabs	New England Biolabs Cat# M2200
PureLink™ Genomic DNA Mini Kit	Invitrogen	Invitrogen Cat# K1820-01
Revert™ 700 Total Protein Stain Kit	LI-COR Bioscience	LI-COR Bioscience Cat# 926-11016

Trans-Blot Turbo RTA Midi 0.2 µm Nitrocellulose Transfer Kit	Bio-Rad Laboratories	Bio-Rad Laboratories Cat# 1704271
ZymoPURE™ II Plasmid Maxiprep Kit	Zymo Research (Irvine, USA)	Zymo Research Cat# D4203
Experimental models		
4T1-AGX1 ^{F383A} -NahK- GalNAc-T2 ^{BH} (“ 4T1-BH T2 ”) cells	Cioce <i>et al.</i> (2022) ⁹	N/A
4T1 mock transfected (“ 4T1- mock ”) cells	Cioce <i>et al.</i> (2022) ⁹	N/A
BALB/c mice	Biological Research Facility (Francis Crick Institute)	N/A
MCF7 ^{T4-KO 4} cells	This paper	N/A
MCF7 ^{WT} cells	Cell Services (Francis Crick Institute)	N/A
MCF7 ^{WT} -AGX1 ^{F383A} -NahK cells	This paper	N/A
MCF7 ^{WT} -AGX1 ^{F383A} -NahK- GalNAc-T4 ^{BH} cells	This paper	N/A
MCF7 ^{WT} -AGX1 ^{F383A} -NahK- GalNAc-T4 ^{WT} cells	This paper	N/A
MDA-MB-231 ^{WT} cells	Cell Services	N/A
MDA-MB-231 ^{WT} -AGX1 ^{F383A} - NahK cells	This paper	N/A
MDA-MB-231 ^{WT} -AGX1 ^{F383A} - NahK-GalNAc-T4 ^{BH} cells	This paper	N/A
MDA-MB-231 ^{WT} -AGX1 ^{F383A} - NahK-GalNAc-T4 ^{WT} cells	This paper	N/A
NSG mice	Biological Research Facility	N/A
T47D ^{T4-KO 1} cells	This paper	N/A
T47D ^{T4-KO 6} cells	This paper	N/A
T47D ^{WT} cells	Cell Services	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK cells	This paper	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK- CD46 ^{BC} (“ Control CD46 BC ”) cells	This paper	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK- CD46 ^C (“ Control CD46 C ”) cells	This paper	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK- GalNAc-T4 ^{BH} cells	This paper	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK- GalNAc-T4 ^{BH} -CD46 ^{BC} (“ BH T4 CD46 BC ”) cells	This paper	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK- GalNAc-T4 ^{BH} -CD46 ^C (“ BH T4 CD46 C ”) cells	This paper	N/A
T47D ^{WT} -AGX1 ^{F383A} -NahK- GalNAc-T4 ^{BH} -GOLM1 (“ BH T4 GOLM1 ”) cells	This paper	N/A

T47D ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{WT} cells	This paper	N/A
Stellar TM Competent Cells	Takara Bio	Takara Bio Cat# 636763
Oligonucleotides		
sgRNA oligonucleotide (forward): 5'-CACCGAACAGTGGCCTATATCTTCG-3'	Calle (2020) ⁷	N/A
sgRNA oligonucleotide (reverse): 5'-AAACCGAAGATATAGGCCACTGTTC-3'	Calle (2020) ⁷	N/A
Sanger sequencing primer for GalNAc-T4 sgRNA-coding plasmid: 5'-GAGGGCCTATTTCCCATGATT-3'	This paper	N/A
Amplicon sequencing primer to validate GalNAc-T4 KO (forward): 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAATCTGCAGGGGGCTTCTT-3'	This paper	N/A
Amplicon sequencing primer to validate GalNAc-T4 KO (reverse): 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAGGGAGAGAGGAGTTGGGC-3'	This paper	N/A
Recombinant DNA		
pCMV(CAT)T7-SB100	Addgene (Watertown, USA)	Addgene Cat# 34879; RRID: Addgene_34879
pSBbi-AGX1 ^{F383A} -NahK	Cioce <i>et al.</i> (2022) ⁹	N/A
pSBbi-AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Calle (2020) ⁷	N/A
pSBbi-AGX1 ^{F383A} -NahK-GalNAc-T4 ^{WT}	Calle (2020) ⁷	N/A
pSpCas9(BB)-2A-GFP plasmid	Addgene	Addgene Cat# 48138; RRID: Addgene_48138
GalNAc-T4 sgRNA-coding plasmid	This paper	N/A
Software and algorithms		
AlphaFold Protein Structure Database	Varadi <i>et al.</i> (2021) ¹⁰ ; Varadi <i>et al.</i> (2023) ¹¹	N/A
Byonic	Protein Metrics (Cupertino, USA)	N/A
CRISPResso2	Clement <i>et al.</i> (2019) ¹	N/A
GraphPad Prism (non-linear Michaelis-Menten and kcat fitting programs)	Dotmatics (Boston, USA)	N/A
Perseus	Tyanova <i>et al.</i> (2016) ¹²	N/A

PyMOL	Schrödinger, Inc. (New York, USA)	N/A
MaxQuant	Cox <i>et al.</i> (2008) ¹³	N/A
SnapGene	Dotmatics	N/A
Other		
4–20% Criterion™ TGX™ Precast Midi Protein Gel, 12+2 well, 45 µl	Bio-Rad Laboratories	Bio-Rad Laboratories Cat# 5671093
4–20% Criterion™ TGX™ Precast Midi Protein Gel, 18 well, 30 µl	Bio-Rad Laboratories	Bio-Rad Laboratories Cat# 5671094
4–20% Criterion™ TGX™ Precast Midi Protein Gel, 26 well, 15 µl	Bio-Rad Laboratories	Bio-Rad Laboratories Cat# 5671095
Amicon Ultra-0.5 Centrifugal Filters (3 kDa MWCO)	Merck	Merck Cat# UFC500396
Amicon Ultra-0.5 Centrifugal Filters (10 kDa MWCO)	Merck	Merck Cat# UFC501024
Amicon Ultra-0.5 Centrifugal Filters (30 kDa MWCO)	Merck	Merck Cat# UFC503024
Amicon Ultra-15 Centrifugal Filters (3 kDa MWCO)	Merck	Merck Cat# UFC900308
Amicon Ultra-15 Centrifugal Filters (10 kDa MWCO)	Merck	Merck Cat# UFC901024
Amicon Ultra-15 Centrifugal Filters (30 kDa MWCO)	Merck	Merck Cat# UFC903024
ACQUITY™ QDa™ Detector	Waters	Waters Cat# 176003208
ACQUITY™ UPLC BEH C18 column	Waters	Waters Cat# 186002350
ACQUITY™ H-Class PLUS (CH-A) Core System	Waters	Waters Cat# 176015101
Orbitrap Eclipse™ Tribrid™ Mass Spectrometer	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# FSN04-10000
Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# IQLAAEGAAPFADBMBHQ
Acclaim™ PepMap™ 100 C18 HPLC Columns	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 164535
Protein LoBind® Tubes, 0.5 mL	Eppendorf (Hamburg, Germany)	Eppendorf Cat# 022431064
Protein LoBind® Tubes, 1.5 mL	Eppendorf	Eppendorf Cat# 0030108116
Protein LoBind® Tubes, 2 mL	Eppendorf	Eppendorf Cat# 0030108132
Nalgene™ 0.2 µM PES syringe filter	Thermo Fisher Scientific	Thermo Fisher Scientific Cat# 720-1320
Odyssey CLx	LI-COR Bioscience	LI-COR Bioscience Model # 9140
Precellys Lysing Kit	Bertin Technologies (Montigny-le-bretonneux, France)	Bertin Technologies Cat# P000916-LYSK0-A
S-Trap™ micro spin columns	ProtiFi (Fairport, USA)	ProtiFi Cat# CO2-micro

Stericup® Quick Release Vacuum Filtration System	Merck	Merck Cat# S2GPU05RE
Strata-X columns	Phenomenex	Phenomenex Cat# 8B-S100-AAK
Ultrafree-MC Centrifugal Filter	Merck	Merck Cat# UFC30GVNB
UltraMicroSpin™ columns	The Nest group, Inc. (Ipswich, USA)	The Nest group, Inc. Cat# SUM SS18V
UltiMate™ 3000 RSLCnano System	Thermo Fisher Scientific	Thermo Fischer Scientific Cat# ULTIM3000RSLCNANO

SI Table 4. GalNAc-T4 substrates in lysates of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. Cellular localisation was first checked in Uniprot¹⁴ and the Human Protein Atlas¹⁵ (enhanced, approved & supported locations included; uncertain locations excluded). If there was a disagreement between both databases regarding the localisation of the proteins, literature was also investigated. Previous reports on the glycosylation status of the proteins were checked in the GlyGen database¹⁶. N/A = no entries were detected for the protein of interest in the database.

Gene name	Protein name	Cellular localisation			Glycosylation status
		Uniprot	HPA	Other	
CAPZB	F-actin-capping protein subunit beta	Cytoplasm	Nucleoplasm / cytosol / vesicles	-	O-linked
CD46	Membrane cofactor protein	Plasma membrane / extracellular space	Plasma membrane	-	N- and O-linked
CPD	Carboxypeptidase D	Cell membrane	Cytosol	Golgi apparatus ¹⁷	N- and O-linked
CS	Citrate synthase, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
DDX1	ATP-dependent RNA helicase DDX1	Nucleus / cytoplasm / mitochondria	Nucleoplasm / cytosol	-	N- and O-linked
DECR1	2,4-dienoyl-CoA reductase [(3E)-enoyl-CoA-producing], mitochondrial	Mitochondria	Mitochondria	-	N/A
FAM49B	CYFIP-related Rac1 interactor B	Mitochondria	Nucleoplasm / cytosol	Nucleus / cytoplasm / mitochondria ¹⁸⁻²⁰	O-linked
GALNT2	Polypeptide N-acetylgalactosaminyltransferase 2	Golgi apparatus / secreted	Golgi apparatus	-	O-linked
GART	Trifunctional purine biosynthetic protein adenosine-3	Cytosol	Cytosol / mitochondria	-	O-linked

GLG1	Golgi apparatus protein 1	Golgi apparatus / cytoplasm	Golgi apparatus	-	N- and O-linked
GOLGA4	Golgin subfamily A member 4	Golgi apparatus / cytoplasm	Golgi apparatus / nucleoplasm	-	N- and O-linked
GOLM1	Golgi membrane protein 1	Golgi apparatus	Golgi apparatus	-	N- and O-linked
GTF2I	General transcription factor II-I	Nucleus/ cytoplasm	Nucleoplasm	-	O-linked
H2AFY2	Core histone macro-H2A.2	Nucleus	Nucleoplasm	-	O-linked
IQGAP1	Ras GTPase-activating-like protein IQGAP1	Cell membrane/ nucleus/ cytoplasm	Plasma membrane / cell junctions	-	N- and O-linked
JTB	Protein JTB	Mitochondria / cytoplasm	Vesicles	Cell membrane/ mitochondria/ cytoplasm ^{21, 22}	O-linked
KRI1	Protein KRI1 homolog	Nucleolus	Nucleoli	-	O-linked
LAMP2	Lysosome-associated membrane glycoprotein 2	Cell membrane / lysosomes / endosomes	Vesicles	-	N- and O-linked
LGALS1	Galectin-1	Secreted/ cytoplasm	Cytosol / secreted	-	N- and O-linked
LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial	Mitochondria / nucleus	Mitochondria	-	N- and O-linked
MTA1	Metastasis-associated protein MTA1	Nucleus / cytoplasm	Nucleoplasm / cytosol	-	O-linked
MUC1	Mucin-1	Cell membrane / secreted / cytoplasm / nucleus	Plasma membrane / secreted	-	N- and O-linked
MYOF	Myoferlin	Cell membrane / nucleus membrane / cytoplasmic vesicle membrane	Vesicles / plasma membrane	-	N- and O-linked

NCAPD2	Condensin complex subunit 1	Nucleus / cytoplasm	Nucleoplasm / cytosol / nucleoli	-	O-linked
PARN	Poly(A)-specific ribonuclease PARN	Nucleus / cytoplasm/ nucleolus	Nuclear speckles	-	N- and O-linked
PDCD11	Protein RRP5 homolog	Nucleus / nucleolus	Nucleoli	-	O-linked
PLEC	Plectin	Cytoplasm / cytoskeleton / cell junction	Intermediate filaments / cytosol / focal adhesion sites	-	N- and O-linked
PMPCA	Mitochondrial-processing peptidase subunit alpha	Mitochondria	Mitochondria	-	O-linked
POLR2A	DNA-directed RNA polymerase II subunit RPB1	Nucleus / cytoplasm	Nucleoplasm / cytosol	-	N- and O-linked
PSMD1	26S proteasome non-ATPase regulatory subunit 1	Nucleus/ cytosol / extracellular region / proteasome	Nucleoplasm / actin filaments	-	N- and O-linked
RAD50	DNA repair protein RAD50	Nucleus	Nucleoplasm / nuclear bodies	-	O-linked
RANBP2	E3 SUMO-protein ligase RanBP2	Nucleus/ nucleus membrane	Nuclear membrane/ nucleoplasm / vesicles	-	N- and O-linked
SLC5A6	Sodium-dependent multivitamin transporter	Cell membrane	Plasma membrane	-	N- and O-linked
SNRNP200	U5 small nuclear ribonucleoprotein 200 kDa helicase	Nucleus	Nucleoplasm	-	O-linked
SORL1	Sortilin-related receptor	Golgi apparatus/ endosomes/ cell membrane/ cytoplasmic vesicle/ secreted	N/A (predicted to be secreted)	-	N- and O-linked

SPTAN1	Spectrin alpha chain, non-erythrocytic 1	Cytoplasm / cytoskeleton / cell cortex	Microtubules	-	N- and O-linked
TLN1	Talin-1	Cell projection/ cytoplasm / cytoskeleton / cell surface/ cell junction/ focal adhesion	Focal adhesion sites / cytosol/ plasma membrane/ centriolar satellite	-	O-linked
TRPS1	Zinc finger transcription factor Trps1	Nucleus	Nucleoplasm	-	O-linked
UBR4	E3 ubiquitin-protein ligase UBR4	Cytoplasm/ cytoskeleton / nucleus	Nucleoplasm	-	N- and O-linked
VPS35	Vacuolar protein sorting-associated protein 35	Cytoplasm/ endosomes	Endosomes/ lysosomes	-	O-linked
XRCC5	X-ray repair cross-complementing protein 5	Nucleus/ nucleolus	Nucleoplasm	-	O-linked

SI Table 5. GalNAc-T4 substrates in the secretome of T47D^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. Previous reports on the glycosylation status of the proteins were checked in the GlyGen database¹⁶.

Gene name	Protein name	Glycosylation status
AGRN	Agrin	N- and O-linked
C3	Complement C3	N- and O-linked
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1	O-linked

SI Table 6. Determining the incorporation efficiency of the heavy SILAC label on lysates of transfected MCF7^{WT} cells.

Cell line	Treatment	Incorporation efficiency (%)			
		R1	R2	R3	Average
MCF7 ^{WT} -AGX1 ^{F383A} -NahK	Ac ₄ GalN6yne	98	98	98	98
MCF7 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	DMSO	98	98	97	98
MCF7 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne	98	98	97	98

SI Table 7. Determining the incorporation efficiency of the heavy SILAC label on lysates of transfected MDA-MB-231^{WT} cells.

Cell line	Treatment	Incorporation efficiency (%)			
		R1	R2	R3	Average
MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK	Ac ₄ GalN6yne	98	97	98	98
MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	DMSO	97	97	98	97
MDA-MB-231 ^{WT} -AGX1 ^{F383A} -NahK-GalNAc-T4 ^{BH}	Ac ₄ GalN6yne	97	97	98	97

SI Table 8. GalNAc-T4 substrates in lysates of MCF7^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. Cellular localisation was first checked in Uniprot¹⁴ and the Human Protein Atlas¹⁵ (enhanced, approved & supported locations included; uncertain locations excluded). If there was a disagreement between both databases regarding the localisation of the proteins, literature was also investigated. Previous reports on the glycosylation status of the proteins were checked in the GlyGen database¹⁶. N/A = no entries were detected for the protein of interest in the database.

Gene name	Protein name	Cellular localisation			Glycosylation status
		Uniprot	HPA	Other	
ABHD10	Palmitoyl-protein thioesterase ABHD10, mitochondrial	Mitochondria	Mitochondria	-	O-linked
ACO1	Cytoplasmic aconitate hydratase	Cytoplasm/ cytosol	Mitochondria / cytosol	-	N- and O-linked
ACO2	Aconitate hydratase, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
AIFM1	Apoptosis-inducing factor 1, mitochondrial	Mitochondria / cytoplasm/ nucleus	Mitochondria	-	N- and O-linked
ALDH18A1	Delta-1-pyrroline-5-carboxylate synthase	Mitochondria	Mitochondria	-	N- and O-linked
ALDH7A1	Alpha-aminoadipic semialdehyde dehydrogenase	Mitochondria / cytoplasm/ nucleus	Mitochondria / cytosol	-	O-linked
ALDH9A1	4-trimethylaminobutyraldehyde dehydrogenase	Cytoplasm/ cytosol	Cytosol	-	O-linked

ASH2L	Set1/Ash2 histone methyltransferase complex subunit ASH2	Nucleus	Nucleoplasm / plasma membrane	-	N- and O-linked
ATP5A1	ATP synthase subunit alpha, mitochondrial	Mitochondria / cell membrane	Mitochondria	-	O-linked
ATP5B	ATP synthase subunit beta, mitochondrial	Mitochondria	Mitochondria	-	O-linked
CALR	Calreticulin	Endoplasmic reticulum/ cytoplasm/ secreted/ cytoplasmic vesicle/ cell surface	Endoplasmic reticulum	-	N- and O-linked
CAPN1	Calpain-1 catalytic subunit	Cytoplasm/ cell membrane	Cytosol	-	N- and O-linked
CCDC93	Coiled-coil domain-containing protein 93	Endosome	Vesicles/ plasma membrane	-	O-linked
DCTN1	Dynactin subunit 1	Cytoplasm/ cytoskeleton/ cell cortex	Microtubules / cytokinetic bridge/ mitotic spindle	-	O-linked
DDX23	Probable ATP-dependent RNA helicase DDX23	Nucleus	Nucleoplasm / nucleoli	-	O-linked
DDX39B	Spliceosome RNA helicase DDX39B	Nucleus/ cytoplasm	Nuclear speckles	-	O-linked
DHX9	ATP-dependent RNA helicase A	Nucleus/ nucleolus/ cytoplasm/ cytoskeleton	Nucleoplasm / nucleoli	-	N- and O-linked
DLD	Dihydrolipoyl dehydrogenase, mitochondrial	Mitochondria / nucleus/ cytoplasmic vesicle	Mitochondria / nucleoplasm	-	N- and O-linked
DLST	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	Mitochondria / nucleus	Nucleoplasm	-	O-linked

DNPEP	Aspartyl aminopeptidase	Cytoplasm	Cytosol	-	O-linked
EFTUD2	116 kDa U5 small nuclear ribonucleoprotein component	Nucleus	Nucleoplasm / vesicles/ cytosol	-	N- and O-linked
EPPK1	Epiplakin	Cytoplasm/ cytoskeleton/ cell junction/ cell projection	Intermediate filaments	-	O-linked
FH	Fumarate hydratase, mitochondrial	Mitochondria / nucleus/ cytoplasm	Mitochondria	-	O-linked
FLNB	Filamin-B	Cytoplasm/ cytoskeleton/ cell cortex	Plasma membrane/ Golgi apparatus/ actin filaments/ cytosol	-	N- and O-linked
GANAB	Neutral alpha-glucosidase AB	Endoplasmic reticulum / Golgi apparatus/ melanosome	Endoplasmic reticulum	-	N- and O-linked
GFRA1	GDNF family receptor alpha-1	Cell membrane/ Golgi apparatus/ endosomes	Golgi apparatus/ nucleoplasm	-	N- and O-linked
GOLM1	Golgi membrane protein 1	Golgi apparatus	Golgi apparatus	-	N- and O-linked
GOT2	Aspartate aminotransferase, mitochondrial	Mitochondria / cell membrane	N/A	Mitochondria ² ₃	O-linked
HADH	Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	Mitochondria	Mitochondria	-	O-linked
HADHA	Trifunctional enzyme subunit alpha, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
HADHB	Trifunctional enzyme subunit beta, mitochondrial	Mitochondria / Endoplasmic reticulum	Mitochondria	-	O-linked

HDGF	Hepatoma-derived growth factor	Nucleus/ cytoplasm/ secreted	Nucleoplasm / secreted	-	O-linked
HDLBP	Vigilin	Nucleus/ cytoplasm	Endoplasmic reticulum / nuclear bodies	-	N- and O- linked
HNRNPA0	Heterogeneous nuclear ribonucleoprotein A0	Nucleus	Nucleoplasm	-	O-linked
HNRNPA2B 1	Heterogeneous nuclear ribonucleoproteins A2/B1	Nucleus/ cytoplasm/ secreted	Nucleoplasm	-	O-linked
HNRNPA3	Heterogeneous nuclear ribonucleoprotein A3	Nucleus	Nucleoplasm	-	N- and O- linked
HNRNPH1	Heterogeneous nuclear ribonucleoprotein H	Nucleus	Nucleoplasm	-	N- and O- linked
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	Cytoplasm/ nucleus/ cell projection	Nucleoplasm	-	N- and O- linked
HNRNPR	Heterogeneous nuclear ribonucleoprotein R	Nucleus/ cytoplasm	Nucleoplasm	-	O-linked
HNRNPU	Heterogeneous nuclear ribonucleoprotein U	Nucleus/ cytoplasm/ cell surface	Nucleoplasm	-	N- and O- linked
HSD17B10	3-hydroxyacyl-CoA dehydrogenase type-2	Mitochondria	N/A	Mitochondria ² 4	O-linked
HUWE1	E3 ubiquitin-protein ligase HUWE1	Cytoplasm/ nucleus/ mitochondria	Nucleoplasm / cytosol	-	N- and O- linked
HYOU1	Hypoxia up-regulated protein 1	Endoplasmic reticulum	N/A	Endoplasmic reticulum ²⁵	N- and O- linked
IDH2	Isocitrate dehydrogenase [NADP], mitochondrial	Mitochondria	Mitochondria	-	O-linked
INTS3	Integrator complex subunit 3	Nucleus/ cytoplasm	Nucleoli/ nucleoplasm	-	O-linked
KDM1A	Lysine-specific histone demethylase 1A	Nucleus	Nucleoplasm	-	O-linked
KHSRP	Far upstream element- binding protein 2	Nucleus/ cytoplasm	Nucleoplasm	-	N- and O- linked
LAMP1	Lysosome-associated membrane glycoprotein 1	Cell membrane / lysosomes / endosomes	N/A	Lysosomes / endosomes ²⁶	N- and O- linked

LAMP2	Lysosome-associated membrane glycoprotein 2	Cell membrane / lysosomes / endosomes	Vesicles	-	N- and O-linked
LMNA	Prelamin-A/C	Nucleus	Nuclear speckles	-	O-linked
LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial	Mitochondria / nucleus	Mitochondria	-	N- and O-linked
LRRC59	Leucine-rich repeat-containing protein 59	Endoplasmic reticulum	Endoplasmic reticulum	-	O-linked
MATR3	Matrin-3	Nucleus	Nucleoplasm	-	O-linked
MCCC1	Methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
MCCC2	Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	Mitochondria	Mitochondria	-	O-linked
MDH1	Malate dehydrogenase, cytoplasmic	Cytoplasm	Centrosome/ cytosol	-	N- and O-linked
MRPL1	Large ribosomal subunit protein uL1m	Mitochondria	Mitochondria	-	O-linked
MRPL28	Large ribosomal subunit protein bL28m	Mitochondria	Mitochondria	-	O-linked
MRPS7	Small ribosomal subunit protein uS7m	Mitochondria	Mitochondria	-	O-linked
MSH2	DNA mismatch repair protein Msh2	Nucleus	Nucleoplasm	-	N- and O-linked
MUC1	Mucin-1	Cell membrane / secreted / cytoplasm / nucleus	Plasma membrane / secreted	-	N- and O-linked
NAT10	RNA cytidine acetyltransferase	Nucleus/ nucleolus/ midbody	Nucleoli/ midbody	-	N- and O-linked
NDUFA6	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 6	Mitochondria	N/A	Mitochondria ² ₇	O-linked
NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	Nucleus	Nucleoplasm / nucleoli	-	O-linked
NUMA1	Nuclear mitotic apparatus protein 1	Nucleus/ cytoplasm/	Nucleoplasm / cytosol	-	O-linked

		cytoskeleton/ cell membrane			
NUP160	Nuclear pore complex protein Nup160	Nucleus	N/A	Nucleus/ cytoplasm ²⁸	O-linked
NUP93	Nuclear pore complex protein Nup93	Nucleus	N/A	Nucleus ²⁹	N- and O- linked
OPA1	Dynamin-like GTPase OPA1, mitochondrial	Mitochondria	Nucleoplasm / mitochondria	-	O-linked
PCYOX1	Prenylcysteine oxidase 1	Lysosome	N/A	Lysosome ³⁰	N- and O- linked
PDHB	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	Mitochondria	Mitochondria	-	O-linked
PLEC	Plectin	Cytoplasm / cytoskeleton / cell junction	Intermediate filaments / cytosol / focal adhesion sites	-	N- and O- linked
POLR2A	DNA-directed RNA polymerase II subunit RPB1	Nucleus / cytoplasm	Nucleoplasm / cytosol	-	N- and O- linked
POP1	Ribonucleases P/MRP protein subunit POP1	Nucleus/ nucleolus	Nucleoli	-	N- and O- linked
PRKDC	DNA-dependent protein kinase catalytic subunit	Nucleus/ nucleolus	Nucleoplasm	-	N- and O- linked
PRPF6	Pre-mRNA-processing factor 6	Nucleus/ nucleus speckle	Nuclear speckles	-	O-linked
PRPF8	Pre-mRNA-processing- splicing factor 8	Nucleus/ nucleus speckle	Nucleoplasm	-	O-linked
PSMC6	26S proteasome regulatory subunit 10B	Nucleus/ cytoplasm	Nucleoplasm / cytosol/ plasma membrane	-	O-linked
PSME1	Proteasome activator complex subunit 1	Cytoplasm/ nucleoplasm	Cytosol/ nuclear bodies	-	O-linked
RAB12	Ras-related protein Rab-12	Endosome/ Lysosome/ Golgi apparatus/	Vesicles	-	N/A

		cytoplasmic vesicles			
RALY	RNA-binding protein Raly	Nucleus	Nucleoplasm / vesicles	-	O-linked
RANBP2	E3 SUMO-protein ligase RanBP2	Nucleus/ nucleus membrane	Nuclear membrane/ nucleoplasm/ vesicles	-	N- and O-linked
RPL27A	Large ribosomal subunit protein uL15	Cytoplasm	Endoplasmic reticulum / cytosol	-	N- and O-linked
RPN2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	Endoplasmic reticulum	Endoplasmic reticulum	-	N- and O-linked
RRBP1	Ribosome-binding protein 1	Endoplasmic reticulum	Endoplasmic reticulum	-	N- and O-linked
RRP12	RRP12-like protein	Nucleus membrane/ nucleolus	Nucleoli/ vesicles/ plasma membrane/ cytosol	-	O-linked
SCD	Stearoyl-CoA desaturase	Endoplasmic reticulum	Endoplasmic reticulum	-	O-linked
SF3B2	Splicing factor 3B subunit 2	Nucleus/ nucleus speckle	Nuclear speckles	-	O-linked
SFXN1	Sideroflexin-1	Mitochondria	Mitochondria	-	O-linked
SLC25A5	ADP/ATP translocase 2	Mitochondria	Mitochondria	-	O-linked
SLC5A6	Sodium-dependent multivitamin transporter	Cell membrane	Plasma membrane	-	N- and O-linked
SMARCA5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	Nucleus	Nucleoplasm / nucleoli fibrillar centre	-	N- and O-linked
SND1	Staphylococcal nuclease domain-containing protein 1	Cytoplasm/ nucleus/ melanosome	Cytosol	-	O-linked
SNRNP200	U5 small nuclear ribonucleoprotein 200 kDa helicase	Nucleus	Nucleoplasm	-	O-linked
SQRDL	Sulfide:quinone oxidoreductase, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked

SRPRB	Signal recognition particle receptor subunit beta	Endoplasmic reticulum	Plasma membrane	Cytoplasm ³¹	O-linked
STXBP3	Syntaxin-binding protein 3	Cytoplasm/ cell membrane	Nucleoplasm / cytosol	-	N/A
SUCLA2	Succinate--CoA ligase [ADP-forming] subunit beta, mitochondrial	Mitochondria	Mitochondria	-	O-linked
TFRC	Transferrin receptor protein 1	Cell membrane/ melanosome / secreted	Endosomes/ lysosomes/ secreted	-	N- and O-linked
TMED7	Transmembrane emp24 domain-containing protein 7	Endoplasmic reticulum / Golgi apparatus/ cytoplasmic vesicle	N/A	Golgi apparatus ³²	N- and O-linked
TNPO2	Transportin-2	Nucleus/ cytoplasm	Nucleoplasm / nucleoli	-	O-linked
TOP1	DNA topoisomerase 1	Nucleus/ nucleolus	Nucleoplasm / Nucleoli fibrillar centre	-	N- and O-linked
TOP2B	DNA topoisomerase 2-beta	Nucleus/ nucleolus	Nucleoplasm	-	O-linked
TRAP1	Heat shock protein 75 kDa, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
TRIM28	Transcription intermediary factor 1-beta	Nucleus	Nucleoplasm	-	N- and O-linked
UBTF	Nucleolar transcription factor 1	Nucleus/ nucleolus	Nucleoli fibrillar centre	-	N- and O-linked
UGDH	UDP-glucose 6-dehydrogenase	Cytosol/ nucleus	Nucleoplasm	-	O-linked
UQCRC1	Cytochrome b-c1 complex subunit 1, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
UQCRC2	Cytochrome b-c1 complex subunit 2, mitochondrial	Mitochondria	Mitochondria	-	N- and O-linked
USO1	General vesicular transport factor p115	Cytoplasm/ Golgi apparatus	Golgi apparatus/ Nucleoli fibrillar centre	-	O-linked

VDAC3	Voltage-dependent anion-selective channel protein 3	Mitochondria	Mitochondria	-	N- and O-linked
VPS35	Vacuolar protein sorting-associated protein 35	Cytoplasm/endosomes	Endosomes/lysosomes	-	O-linked
VRK1	Serine/threonine-protein kinase VRK1	Nucleus/cytoplasm	Nucleoplasm / cytosol	-	O-linked
VWA8	von Willebrand factor A domain-containing protein 8	Mitochondria / peroxisome	Vesicles/lipid droplets	Mitochondria ³ 3	N- and O-linked
XRCC6	X-ray repair cross-complementing protein 6	Nucleus	Nucleoplasm	-	N- and O-linked

SI Table 9. GalNAc-T4 substrates in lysates of MDA-MB-231^{WT}-AGX1^{F383A}-NahK-GalNAc-T4^{BH} cells. Cellular localisation was first checked in Uniprot¹⁴ and the Human Protein Atlas¹⁵ (enhanced, approved & supported locations included; uncertain locations excluded). If there was disagreement between both databases regarding the localisation of the proteins, literature was also investigated. Previous reports on the glycosylation status of the proteins were checked in the GlyGen database¹⁶. N/A = no entries were detected for the protein of interest in the database.

Gene name	Protein name	Cellular localisation			Glycosylation status
		Uniprot	HPA	Other	
AARS	Alanine--tRNA ligase, cytoplasmic	Cytoplasm	Cytosol	-	O-linked
ABCE1	ATP-binding cassette sub-family E member 1	Cytoplasm/mitochondria	Cytosol	-	N- and O-linked
ACTR3	Actin-related protein 3	Cytoplasm/cytoskeleton/cell projection/nucleus	Cytoplasm	-	O-linked
AHNAK	Neuroblast differentiation-associated protein AHNAK	Nucleus	Plasma membrane/cytosol	Plasma membrane/cytosol/vesicles ³⁴	N- and O-linked
ARHGAP18	Rho GTPase-activating protein 18	Cytoplasm	Cytosol	-	O-linked
ARPC4	Actin-related protein 2/3 complex subunit 4	Cytoplasm/cytoskeleton/cell projection/nucleus	N/A	Cytoskeleton/nucleus ³⁵	O-linked
CAPZA2	F-actin-capping protein subunit alpha-2	Cytosol/cytoskeleton/extracellular region	Cytosol	-	N/A
CARS	Cysteine--tRNA ligase, cytoplasmic	Cytoplasm	Cytosol	-	N- and O-linked
CCT2	T-complex protein 1 subunit beta	Cytoplasm	Cytosol	-	N- and O-linked

CMAS	N-acylneuraminate cytidyltransferase	Nucleus	Nucleoli	-	O-linked
CNDP2	Cytosolic non-specific dipeptidase	Cytoplasm	Nucleoplasm/ cytosol	-	O-linked
COPB1	Coatomer subunit beta	Cytoplasm/ Golgi apparatus/ cytoplasmic vesicle/ cell membrane	Golgi apparatus/ vesicles/ cytosol	-	O-linked
CPD	Carboxypeptidase D	Cell membrane	Cytosol	Golgi apparatus ¹⁷	N- and O- linked
CTNNA1	Catenin alpha-1	Nucleus/ cell membrane/ cell junction	Plasma membrane/ cell junctions	-	N- and O- linked
CYR61	CCN family member 1	Secreted	Secreted	-	O-linked
DNMBP	Dynamin-binding protein	Cytoplasm/ cell junction/ Golgi apparatus	Cytosol/ Golgi apparatus/ nucleoli/ nuclear bodies	-	O-linked
EDC4	Enhancer of mRNA- decapping protein 4	Cytoplasm/ P- body/ nucleus	Cytoplasmic bodies/ nucleoplasm	-	O-linked
EEF1D	Elongation factor 1-delta	Nucleus	Nucleoplasm/ nucleoli fibrillar centre	-	O-linked
EEF1G	Elongation factor 1-gamma	Cytoplasm/ nucleus	Cytosol	-	N- and O- linked
EIF2S1	Eukaryotic translation initiation factor 2 subunit 1	Cytoplasm	Cytosol	-	N- and O- linked
EIF4G1	Eukaryotic translation initiation factor 4 gamma 1	Cytoplasm/ nucleus	Cytosol	-	N- and O- linked
ERC1	ELKS/Rab6- interacting/CAST family member 1	Cytoplasm/ cytoskeleton	Cytosol/ vesicles/ plasma membrane	-	N- and O- linked
ERCC6L	DNA excision repair protein ERCC-6-like	Chromosome/ centromere/ kinetochore	N/A	Cytoplasm/ nucleus ³⁶	O-linked
EXOC4	Exocyst complex component 4	Cytoplasm/ cytoskeleton/ midbody	Nucleoli	Cytoplasmic vesicles/	N- and O- linked

				plasma membrane ³⁶	
FAM129B	Protein Niban 2	Cytoplasm/ cell junction	Plasma membrane/ cytosol/ nucleoplasm	-	O-linked
FARSB	Phenylalanine--tRNA ligase beta subunit	Cytoplasm	Cytosol/ nucleoplasm	-	O-linked
FEN1	Flap endonuclease 1	Nucleus/ mitochondria	Nucleoplasm/ nucleoli	-	O-linked
FLNA	Filamin-A	Cytoplasm	Plasma membrane/ actin filaments/ cytosol	-	N- and O- linked
FLNB	Filamin-B	Cytoplasm/ cytoskeleton/ cell cortex	Plasma membrane/ Golgi apparatus/ actin filaments/ cytosol	-	N- and O- linked
GBE1	1,4-alpha-glucan-branching enzyme	Cytoplasm	Cytosol	-	O-linked
GOLM1	Golgi membrane protein 1	Golgi apparatus	Golgi apparatus	-	N- and O- linked
GPKOW	G-patch domain and KOW motifs-containing protein	Nucleus	Nucleoplasm	-	O-linked
GSTP1	Glutathione S-transferase P	Cytoplasm/ mitochondria/ nucleus	Cytosol/ mitochondria	-	O-linked
HDLBP	Vigilin	Nucleus/ cytoplasm	Endoplasmic reticulum / nuclear bodies	-	N- and O- linked
HSD17B4	Peroxisomal multifunctional enzyme type 2	Peroxisomes	Peroxisomes	-	N- and O- linked
KIF5B	Kinesin-1 heavy chain	Cytoplasm/ cytoskeleton/ lysosomes	Cytosol/ centriolar satellite	-	N- and O- linked
KPNB1	Importin subunit beta-1	Cytoplasm/ nucleus envelope	Nuclear membrane/ nucleoplasm/ cytosol	-	O-linked

LDLR	Low-density lipoprotein receptor	Cell membrane/ Golgi apparatus/ endosomes/ lysosomes	N/A	Cell membrane/ endosomes/ Golgi apparatus ³⁷	N- and O-linked
MAP4	Microtubule-associated protein 4	Cytoplasm/ cytoskeleton	Microtubules	-	N- and O-linked
MRE11A	Double-strand break repair protein MRE11	Nucleus	Nucleoplasm	-	O-linked
NANS	Sialic acid synthase	Cytoplasm	Nucleoplasm/ cytoplasm	-	O-linked
NAP1L1	Nucleosome assembly protein 1-like 1	Nucleus/ melanosome/ cytoplasm	N/A	Cytoplasm/ nucleus ³⁸	O-linked
NMT1	Glycylpeptide N-tetradecanoyltransferase 1	Cytoplasm	Cytosol/ plasma membrane	-	O-linked
NQO2	Ribosyldihydronicotinamide dehydrogenase [quinone]	Cytoplasm	Cytosol/ nucleoplasm	-	N/A
NT5DC1	5'-nucleotidase domain-containing protein 1	N/A	Nucleoplasm/ cytosol	Cytosol ³⁹	O-linked
NT5E	5'-nucleotidase	Cell membrane	Plasma membrane	-	N- and O-linked
NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	Nucleus	Nucleoplasm/ nucleoli	-	O-linked
NUFIP2	FMR1-interacting protein NUFIP2	Nucleus/ cytoplasm	Nucleoplasm/ nuclear bodies/ cytosol	-	N- and O-linked
NUMA1	Nuclear mitotic apparatus protein 1	Nucleus/ cytoplasm/ cytoskeleton/ cell membrane	Nucleoplasm / cytosol	-	O-linked
PGM3	Phosphoacetylglucosamine mutase	Cytosol	Nucleoplasm / cytosol	-	O-linked
PHLDB1	Pleckstrin homology-like domain family B member 1	Cytosol/ plasma membrane	Cytosol/ plasma membrane/ vesicles	-	N- and O-linked
PLS3	Phospholipid scramblase 3	Mitochondria	Mitochondria	-	N/A

PRKACA	cAMP-dependent protein kinase catalytic subunit alpha	Cytoplasm/ cell membrane/ nucleus	Cytosol	-	O-linked
RPA1	Replication protein A 70 kDa DNA-binding subunit	Nucleus	Nucleoplasm	-	O-linked
RPL4	Large ribosomal subunit protein uL4	Cytoplasm	Endoplasmic reticulum / cytosol/ nucleoli, nuclear bodies	-	N- and O-linked
SEPT2	Septin-2	Cytoplasm / cytoskeleton/ cell projection	Actin filaments/ cytokinetic bridge	-	O-linked
SLK	STE20-like serine/threonine-protein kinase	Cytoplasm	Cytosol	-	O-linked
SPTBN1	Spectrin beta chain, non-erythrocytic 1	Cytoplasm/ cell membrane	Cytosol	-	N- and O-linked
TDRD7	Tudor domain-containing protein 7	Cytoplasm	Cytoplasmic bodies/ nucleoplasm	-	O-linked
TKT	Transketolase	Cytosol/ nucleoplasm/ Endoplasmic reticulum / peroxisome	Nucleoplasm	-	N- and O-linked
TLN1	Talin-1	Cell projection/ cytoplasm / cytoskeleton/ cell surface/ cell junction/ focal adhesion	Focal adhesion sites / cytosol/ plasma membrane/ centriolar satellite	-	O-linked
TXNL1	Thioredoxin-like protein 1	Cytoplasm/ nucleus	Cytosol	-	O-linked
UBA2	SUMO-activating enzyme subunit 2	Cytoplasm/ nucleus	Nucleoplasm	-	O-linked
UGDH	UDP-glucose 6-dehydrogenase	Cytosol/ nucleus	Nucleoplasm	-	O-linked

USP9X	Ubiquitin carboxyl-terminal hydrolase 9X	Cytoplasm/ cell projection/ cytoskeleton	Cytosol	-	O-linked
UTRN	Utrophin	Cell membrane/ cytoplasm/ cytoskeleton	Plasma membrane	-	N- and O- linked
ZC3HAV1	Zinc finger CCCH-type antiviral protein 1	Cytosol/ nucleus	Cytosol/ Golgi apparatus	-	N- and O- linked
ZHX3	Zinc fingers and homeoboxes protein 3	Nucleus	Nucleoplasm	-	O-linked

SI Table 10. Protein concentration and amount from recombinant BH T4 CD46 BC and BH T4 GOLM1 after Ni-NTA purification.

Sample	Concentration (mg/mL)	Amount of protein (µg)
BH T4 CD46 BC	3.25	325
BH T4 GOLM1	2.15	215

Supplementary Bibliography

1. Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., and Pinello, L. (2019) CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 37 (3), 224-226. <https://doi.org/10.1038/s41587-019-0032-3>
2. Kowarz, E., Löscher, D., and Marschalek, R. (2015) Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnol. J.* 10 (4), 647-653. <https://doi.org/10.1002/biot.201400821>
3. Kim, J. H., Lee, S. R., Li, L. H., Park, H. J., Park, J. H., Lee, K. Y., Kim, M. K., Shin, B. A., and Choi, S. Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One.* 6 (4), e18556. <https://doi.org/10.1371/journal.pone.0018556>
4. Calle, B. (2020) Characterisation of O-GalNAc Glycosylation by GalNAc-T4 Bump-and-Hole Engineering. MSci thesis. Imperial College London
5. Choi, J., Wagner, L. J. S., Timmermans, S. B. P. E., Malaker, S. A., Schumann, B., Gray, M. A., Debets, M. F., Takashima, M., Gehring, J., and Bertozzi, C. R. (2019) Engineering Orthogonal Polypeptide GalNAc-Transferase and UDP-Sugar Pairs. *J. Am. Chem. Soc.* 141 (34), 13442-13453. <https://doi.org/10.1021/jacs.9b04695>
6. Cen, H., Mao, F., Aronchik, I., Fuentes, R. J., and Firestone, G. L. (2008) DEVD-NucView488: a novel class of enzyme substrates for real-time detection of caspase-3 activity in live cells. *FASEB J.* 22 (7), 2243-2252. <https://doi.org/10.1096/fj.07-099234>
7. Calle, B., Bineva-Todd, G., Marchesi, A., Flynn, H., Ghirardello, M., Tastan, O. Y., Roustan, C., Choi, J., Galan, M. C., Schumann, B., and Malaker, S. A. (2021) Benefits of Chemical Sugar Modifications Introduced by Click Chemistry for Glycoproteomic Analyses. *J. Am. Soc. Mass Spectrom.* 32 (9), 2366-2375. <https://doi.org/10.1021/jasms.1c00084>
8. Schumann, B., Malaker, S. A., Wisnovsky, S. P., Debets, M. F., Agbay, A. J., Fernandez, D., Wagner, L. J. S., Lin, L., Li, Z., Choi, J., Fox, D. M., Peh, J., Gray, M. A., Pedram, K., Kohler, J. J., Mrksich, M., and Bertozzi, C. R. (2020) Bump-and-Hole Engineering Identifies Specific Substrates of Glycosyltransferases in Living Cells. *Mol. Cell.* 78 (5), 824-834.e15. [https://www.cell.com/molecular-cell/fulltext/S1097-2765\(20\)30198-2](https://www.cell.com/molecular-cell/fulltext/S1097-2765(20)30198-2)
9. Cioce, A., Calle, B., Rizou, T., Lowery, S. C., Bridgeman, V. L., Mahoney, K. E., Marchesi, A., Bineva-Todd, G., Flynn, H., Li, Z., Tastan, O. Y., Roustan, C., Soro-Barrio, P., Rafiee, M. R., Garza-Garcia, A., Antonopoulos, A., Wood, T. M., Keenan, T., Both, P., Huang, K., Parmeggiani, F., Snijders, A. P., Skehel, M., Kjær, S., Fascione, M. A., Bertozzi, C. R., Haslam, S. M., Flitsch, S. L., Malaker, S. A., Malanchi, I., and Schumann, B. (2022) Cell-specific bioorthogonal tagging of glycoproteins. *Nat. Commun.* 13 (1), 6237. <https://doi.org/10.1038/s41467-022-33854-0>
10. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439-D444. <https://doi.org/10.1093/nar/gkab1061>
11. Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., Tsenkov, M., Nair, S., Mirdita, M., Yeo, J., Kovalevskiy, O., Tunyasuvunakool, K., Laydon, A., Židek, A., Tomlinson, H., Hariharan, D., Abrahamson, J., Green, T., Jumper, J., Birney, E., Steinegger, M., Hassabis, D., and Velankar, S. (2024) AlphaFold Protein Structure Database in 2024: providing structure coverage for over

- 214 million protein sequences. *Nucleic Acids Res.* 52, D368-D375. <https://doi.org/10.1093/nar/gkad1011>
12. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26 (12), 1367-1372. <https://doi.org/10.1038/nbt.1511>
 13. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., Cox, J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods.* 13 (9), 731-740. <https://doi.org/10.1038/nmeth.3901>
 14. UniProt Consortium. (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* 51, D523-D531. <https://doi.org/10.1093/nar/gkac1052>
 15. Uhlén, M., Björling, E., Agaton, C., Szigartyo, C. A., Amini, B., Andersen, E., Andersson, A. C., Angelidou, P., Asplund, A., Asplund, C., Berglund, L., Bergström, K., Brumer, H., Cerjan, D., Ekström, M., Elobeid, A., Eriksson, C., Fagerberg, L., Falk, R., Fall, J., Forsberg, M., Björklund, M. G., Gumbel, K., Halimi, A., Hallin, I., Hamsten, C., Hansson, M., Hedhammar, M., Hercules, G., Kampf, C., Larsson, K., Lindskog, M., Lodewyckx, W., Lund, J., Lundberg, J., Magnusson, K., Malm, E., Nilsson, P., Odling, J., Oksvold, P., Olsson, I., Oster, E., Ottosson, J., Paavilainen, L., Persson, A., Rimini, R., Rockberg, J., Runeson, M., Sivertsson, A., Sköllerö, A., Steen, J., Stenvall, M., Sterky, F., Strömberg, S., Sundberg, M., Tegel, H., Tourle, S., Wahlund, E., Waldén, A., Wan, J., Wernérus, H., Westberg, J., Wester, K., Wrethagen, U., Xu, L. L., Hober, S., and Pontén, F. (2005) A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteomics.* 4 (12), 1920-1932. [https://www.mcponline.org/article/S1535-9476\(20\)30028-1/](https://www.mcponline.org/article/S1535-9476(20)30028-1/)
 16. York, W. S., Mazumder, R., Ranzinger, R., Edwards, N., Kahsay, R., Aoki-Kinoshita, K. F., Campbell, M. P., Cummings, R. D., Feizi, T., Martin, M., Natale, D. A., Packer, N. H., Woods, R. J., Agarwal, G., Arpinar, S., Bhat, S., Blake, J., Castro, L. J. G., Fochtman, B., Gildersleeve, J., Goldman, R., Holmes, X., Jain, V., Kulkarni, S., Mahadik, R., Mehta, A., Mousavi, R., Nakarakomula, S., Navelkar, R., Pattabiraman, N., Pierce, M. J., Ross, K., Vasudev, P., Vora, J., Williamson, T., and Zhang, W. (2020) GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology.* 30 (2), 72-73. <https://doi.org/10.1093/glycob/cwz080>
 17. Varlamov, O., Eng, F. J., Novikova, E. G., and Fricker, L. D. (1999) Localization of metallopeptidase D in AtT-20 cells. Potential role in prohormone processing. *J. Biol. Chem.* 274 (21), 14759-14767. [https://www.jbc.org/article/S0021-9258\(19\)73153-6/](https://www.jbc.org/article/S0021-9258(19)73153-6/)
 18. Chattaragada, M. S., Riganti, C., Sassoe, M., Principe, M., Santamorenna, M. M., Roux, C., Curcio, C., Evangelista, A., Allavena, P., Salvia, R., Rusev, B., Scarpa, A., Cappello, P., and Novelli, F. (2018) FAM49B, a novel regulator of mitochondrial function and integrity that suppresses tumor metastasis. *Oncogene.* 37 (6), 697-709. <https://doi.org/10.1038/onc.2017.358>
 19. Chen, Y., Jiang, Y., Lao, J., Zhou, Y., Su, L., and Huang, X. (2022) Characterization and Functional Study of FAM49B Reveals Its Effect on Cell Proliferation in HEK293T Cells. *Genes.* 13 (2), 388. <https://doi.org/10.3390/genes13020388>
 20. Li, Y., Xiong, Y., Wang, Z., Han, J., Shi, S., He, J., Shen, N., Wu, W., Wang, R., Lv, W., Deng, Y., and Liu, W. (2021) FAM49B promotes breast cancer proliferation, metastasis, and chemoresistance by stabilizing ELAVL1 protein and regulating downstream Rab10/TLR4 pathway. *Cancer Cell Int.* 21 (1), 534. <https://doi.org/10.1186/s12935-021-02244-9>
 21. Rousseau, F., Pan, B., Fairbrother, W. J., Bazan, J. F., and Lingel, A. (2012) The structure of the extracellular domain of the jumping translocation breakpoint protein reveals a variation of the midkine fold. *J. Mol. Biol.* 415 (1), 22-28. <https://doi.org/10.1016/j.jmb.2011.10.048>
 22. Kanome, T., Itoh, N., Ishikawa, F., Mori, K., Kim-Kaneyama, J. R., Nose, K., and Shibamura, M. (2007) Characterization of Jumping translocation breakpoint (JTB) gene product isolated as a TGF-beta1-inducible clone involved in regulation of

- mitochondrial function, cell growth and cell death. *Oncogene*. 26 (41), 5991-6001. <https://doi.org/10.1038/sj.onc.1210423>
23. Kerk, S. A., Garcia-Bermudez, J., Birsoy, K., Sherman, M. H., Shah, Y. M., and Lyssiotis, C. A. (2023) Spotlight on GOT2 in Cancer Metabolism. *OncoTargets Ther.* 16, 695-702. <https://doi.org/10.2147/OTT.S382161>
 24. Liu, L., Chen, S., Yu, M., Ge, C., Ren, M., Liu, B., Yang, X., Christian, T. W. Jr, Hou, Y. M., Zou, J., Zhu, W. G., and Luo, J. (2020) Deacetylation of HSD17B10 by SIRT3 regulates cell growth and cell resistance under oxidative and starvation stresses. *Cell Death Dis.* 11 (7), 563. <https://doi.org/10.1038/s41419-020-02763-9>
 25. Rao, S., Oyang, L., Liang, J., Yi, P., Han, Y., Luo, X., Xia, L., Lin, J., Tan, S., Hu, J., Wang, H., Tang, L., Pan, Q., Tang, Y., Zhou, Y., and Liao, Q. (2021) Biological Function of HYOU1 in Tumors and Other Diseases. *OncoTargets Ther.* 14, 1727-1735. <https://doi.org/10.2147/OTT.S297332>
 26. Baba, K., Kuwada, S., Nakao, A., Li, X., Okuda, N., Nishida, A., Mitsuda, S., Fukuoka, N., Kakeya, H., and Kataoka, T. (2020) Different localization of lysosomal-associated membrane protein 1 (LAMP1) in mammalian cultured cell lines. *Histochem. Cell Biol.* 153 (4), 199-213. <https://doi.org/10.1007/s00418-019-01842-z>
 27. Talla, V., Koilkonda, R., Porciatti, V., Chiodo, V., Boye, S. L., Hauswirth, W. W., and Guy, J. (2015) Complex I subunit gene therapy with NDUFA6 ameliorates neurodegeneration in EAE. *Invest. Ophthalmol. Vis. Sci.* 56 (2), 1129-1140. <https://doi.org/10.1167/iovs.14-15950>
 28. Chatel, G., and Fahrenkrog, B. (2011) Nucleoporins: leaving the nuclear pore complex for a successful mitosis. *Cell Signal.* 23 (10), 1555-1562. <https://doi.org/10.1016/j.cellsig.2011.05.023>
 29. Braun, D. A., Sadowski, C. E., Kohl, S., Lovric, S., Astrinidis, S. A., Pabst, W. L., Gee, H. Y., Ashraf, S., Lawson, J. A., Shril, S., Airik, M., Tan, W., Schapiro, D., Rao, J., Choi, W. I., Hermle, T., Kemper, M. J., Pohl, M., Ozaltin, F., Konrad, M., Bogdanovic, R., Büscher, R., Helmchen, U., Serdaroglu, E., Lifton, R. P., Antonin, W., and Hildebrandt, F. (2016) Mutations in nuclear pore genes NUP93, NUP205 and XPO5 cause steroid-resistant nephrotic syndrome. *Nat. Genet.* 48 (4), 457-465. <https://doi.org/10.1038/ng.3512>
 30. Banfi, C., Amadio, P., Zarà, M., Brioschi, M., Sandrini, L., and Barbieri, S. S. (2022) Prenylcysteine Oxidase 1 (PCYOX1), a New Player in Thrombosis. *Int. J. Mol. Sci.* 23 (5), 2831. <https://doi.org/10.3390/ijms23052831>
 31. Ma, Q., Wu, X., Wu, J., Liang, Z., and Liu, T. (2017) SERP1 is a novel marker of poor prognosis in pancreatic ductal adenocarcinoma patients via anti-apoptosis and regulating SRPRB/NF-κB axis. *Int. J. Oncol.* 51 (4), 1104-1114. <https://doi.org/10.3892/ijo.2017.4111>
 32. Liaunardy-Jopeace, A., Bryant, C. E., and Gay, N. J. (2014) The COP II adaptor protein TMED7 is required to initiate and mediate the delivery of TLR4 to the plasma membrane. *Sci. Signal.* 7 (336), ra70. <https://doi.org/10.1126/scisignal.2005275>
 33. Luo, M., Ma, W., Sand, Z., Finlayson, J., Wang, T., Brinton, R. D., Willis, W. T., and Mandarino, L. J. (2020) Von Willebrand factor A domain-containing protein 8 (VWA8) localizes to the matrix side of the inner mitochondrial membrane. *Biochem. Biophys. Res. Commun.* 521 (1), 158-163. <https://doi.org/10.1016/j.bbrc.2019.10.095>
 34. Silva, T. A., Smuczek, B., Valadão, I. C., Dzik, L. M., Iglesía, R. P., Cruz, M. C., Zelanis, A., de Siqueira, A. S., Serrano, S. M., Goldberg, G. S., Jaeger, R. G., and Freitas, V. M. (2016) AHNK enables mammary carcinoma cells to produce extracellular vesicles that increase neighboring fibroblast cell motility. *Oncotarget.* 7 (31), 49998-50016. <https://doi.org/10.18632/oncotarget.10307>
 35. Zhang, C., Mallery, E. L., and Szymanski, D. (2013) B. ARP2/3 localization in Arabidopsis leaf pavement cells: a diversity of intracellular pools and cytoskeletal interactions. *Front. Plant Sci.* 4, 238. <https://doi.org/10.3389/fpls.2013.00238>
 36. Tummala, H., Kirwan, M., Walne, A. J., Hossain, U., Jackson, N., Ponderre, C., Plagnol, V., Vulliamy, T., and Dokal, I. (2014) ERCC6L2 mutations link a distinct bone-

- marrow-failure syndrome to DNA repair and mitochondrial function. *Am. J. Hum. Genet.* 94 (2), 246-256. [https://www.cell.com/ajhg/fulltext/S0002-9297\(14\)00008-1](https://www.cell.com/ajhg/fulltext/S0002-9297(14)00008-1)
37. Nassoury, N., Blasiole, D. A., Tebon Oler, A., Benjannet, S., Hamelin, J., Poupon, V., McPherson, P. S., Attie, A. D., Prat, A., and Seidah, N. G. (2007) The cellular trafficking of the secretory proprotein convertase PCSK9 and its dependence on the LDLR. *Traffic*. 8 (6), 718-732. <https://doi.org/10.1111/j.1600-0854.2007.00562.x>
38. Okuwaki, M., Kato, K., and Nagata, K. (2010) Functional characterization of human nucleosome assembly protein 1-like proteins as histone chaperones. *Genes Cells*. 15 (1), 13-27. <https://doi.org/10.1111/j.1365-2443.2009.01361.x>
39. Lyu, P., Jiang, K., Zhou, Y., Hu, J., Chang, Y., Zhang, Z., Huang, M., Zhang, Z. M., Ding, K., Hao, P., Lin, L., and Li, Z. (2022) Proteome-wide Identification of Off-Targets of a Potent EGFR^{L858R/T790M} Mutant Inhibitor. *ACS Med. Chem. Lett.* 13 (2), 292-297. <https://doi.org/10.1021/acsmedchemlett.1c00651>