



The role of aberrant glycosylation of CD44 in gastrointestinal carcinoma and its application for cancer biomarker purposes

Álvaro M. Martins

Uminho | 2019



Universidade do Minho
Escola de Ciências

Álvaro Martinho de Sousa Martins

The role of aberrant glycosylation of CD44 in gastrointestinal carcinoma and its application for cancer biomarker purposes

Setembro de 2019

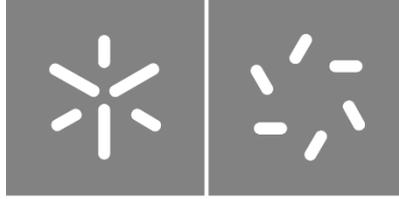


Universidade do Minho

Escola de Ciências



**INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE**
UNIVERSIDADE
DO PORTO



Universidade do Minho

Escola de Ciências

Álvaro Martinho de Sousa Martins

**The role of aberrant glycosylation of
CD44 in gastrointestinal carcinoma
and its application for cancer
biomarker purposes**

Dissertação de Mestrado

Bioquímica Aplicada

Biotechnologia

Trabalho efetuado sob a orientação do

Doutor Stefan Mereiter

Doutora Ana Preto

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos. Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada. Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.



Atribuição-Compartilha Igual
CC BY-SA

<https://creativecommons.org/licenses/by-sa/4.0/>

Agradecimentos

“A gratidão é a memória do coração” - Jean-Baptiste Massieu

Quero agradecer a todas as pessoas que fizeram parte desta aventura, pois elas estarão na minha memória para sempre.

Começo por agradecer ao meu orientador Doutor Stefan Mereiter por me ter ensinado muito e ajudado a crescer não só profissionalmente, mas também pessoalmente. Por me ter dado a confiança necessária para levar a cabo o trabalho científico e a paciência em orientar-me durante o mesmo.

Também quero agradecer ao Doutor Celso Reis pela oportunidade de trabalhar no seu grupo de investigação.

Quero agradecer aos restantes membros do grupo por me ajudarem durante esta jornada. À Meritxell Balmaña, pelos seus conselhos sábios e por me ter ajudado em várias ocasiões. À Juliana Poças, à Joana Rodrigues, à Francisca Diniz, ao Henrique Duarte, à Rita Matos, à Catarina Marquês, à Inês Moreira, à Rafaela Abrantes e à Ana Filipa por partilharem as suas experiências profissionais e perspetivas, e também por me proporcionarem momentos de lazer e risos.

Quero também agradecer à Sara Pereira, que me acompanhou durante esta aventura, apoiando-me e ouvindo sempre que precisava, e pelos momentos inesquecíveis.

Por último, mas não menos importante, estou muito grato aos meus pais e irmã por estarem sempre lá para me apoiar e reconfortar nos momentos difíceis quando mais precisava.

A todos, MUITO OBRIGADO!

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho

O papel da glicosilação aberrante da CD44 em carcinoma gastrointestinal e sua aplicação para fins como biomarcador para cancro

Cancro gástrico e colorretal estão entre as formas mais mortais de cancro mundialmente. Uma alteração frequente observada numa grande maioria de células cancerígenas gástricas e colorretais é a formação de glicanos aberrantes. Estes glicanos aberrantes podem conferir fenótipos agressivos às células cancerígenas, como aumento de invasão e metástase, afetando glicoproteínas e as suas funções. Uma proteína altamente glicosilada e com um papel importante na progressão do cancro gástrico e colorretal é a CD44. CD44 é descrita como sendo o recetor principal do ácido hialurónico e atuando como coreceptor de diferentes recetores de tirosina quinase, como por exemplo MET e “Recepteur d’origine Nantais” (RON). Já tem sido descrito que o perfil de glicosilação da CD44 está afetado no cancro gastrointestinal, mas o seu impacto funcional está ainda por apurar. Nesta tese de mestrado, tentamos desvendar os efeitos dos *O*-glicanos truncados da CD44 nas suas características moleculares e funções, usando um painel de modelos celulares de cancro modificados geneticamente e amostras clínicas. Os nossos resultados mostram que a CD44 é altamente *O*-glicosilada e o encurtamento dos *O*-glicanos afeta significativamente o seu peso molecular. Além do mais, encurtamento dos *O*-glicanos leva a um aumento na afinidade de ligação das células cancerígenas para o ácido hialurónico e a um aumento da ativação do recetor RON. Estes resultados foram validados em tecidos de cancro gástrico de pacientes usando uma abordagem inovadora de diferentes técnicas de imunofluorescência. Este trabalho fornece uma visão no impacto dos glicanos como moduladores da CD44 no cenário de cancro, além de dar novas perspetivas nos mecanismos de malignidade do cancro.

Palavras-chave: Glicosilação; Cancro gastrointestinal; CD44; ácido hialurónico; RON

The role of aberrant glycosylation of CD44 in gastrointestinal carcinoma and its application for cancer biomarker purposes

Gastric cancer and colorectal cancer are among the deadliest cancers worldwide. A frequent alteration observed in vast majority of gastric and colorectal cancer cells is the formation of aberrant glycan structures. These aberrant glycans may confer aggressive phenotypes to cancer cells, such as invasiveness and metastasis, by affecting specific glycoproteins and its functions. One highly glycosylated protein and key player in gastric and colorectal cancer progression is CD44. CD44 is described to be the main receptor for hyaluronic acid (HA) and acting as a co-receptor for different receptor tyrosine kinases (RTKs), such as MET and “Recepteur d’origine Nantais” (RON). CD44 has already been observed to have its glycosylation profile affected in gastrointestinal cancer, but the functional impact is yet to be addressed. In this master thesis, we try to uncover the effects of *O*-glycan truncation of CD44 in its molecular features and functions, using an array of glycoengineered cancer cell line models and clinical samples. Our results show that CD44 is heavily *O*-glycosylated and that truncation of *O*-glycans affect significantly its molecular weight. In addition, truncation of *O*-glycans leads to an increase of binding affinity of cancer cells towards hyaluronic acid while also leading to an increase in activation of RON. These results were validated using gastric tumor tissue samples using an innovative approach of different immunofluorescence techniques. This work provides new insights on the impact of glycans as key modulators of CD44 in the cancer setting, in addition to giving new perspectives on cancer malignancy mechanisms.

Keywords: Glycosylation; Gastrointestinal cancer; CD44; hyaluronic acid; RON

Content

1. Introduction	10
1.1 Glycosylation.....	10
1.2 OGalNAc glycosylation	11
1.3 Truncation of Oglycans.....	14
1.4 Hyaluronic acid.....	16
1.5 Gastrointestinal Carcinoma	18
1.5 Shortcomings of current serological assays and CD44 as a potential biomarker for gastrointestinal cancer detection.....	20
1.6 CD44: a signaling scaffold.....	22
1.7 The different CD44 isoforms	22
1.8 CD44 as a receptor for hyaluronic acid.....	23
1.9 CD44 as a co-receptor for receptor tyrosine kinases	26
1.10 CD44 as a cancer stem cell marker	28
2. Objectives.....	29
3. Materials and methods.....	30
3.1 Cell culture	30
3.2 Flow cytometry.....	30
3.3 Western Blot	30
3.4 O-glycosylation inhibition	31
3.5 Immunoprecipitation	31
3.6 Immunofluorescence (IF)	31
3.7 <i>in situ</i> Proximity ligation assay (PLA).....	31
3.8 Combined immunofluorescence proximity ligation assay (co-IF-PLA).....	32
3.9 Antibodies and Lectins	32
3.10 Microscopy and image analysis	32
4. Results	33
4.1 Truncated Oglycans as the main aberrant structures in CD44 affects its structure in the glycoengineered cell lines	33
4.2 Oglycosylation plays a role in the HA-binding affinity of CD44.....	36
4.3 Truncated Oglycans enhance RON activation in a CD44v6-dependent manner.....	38
4.4 Characterization of CD44 in the colorectal cancer cell line LS174T	42
5. Discussion	44
6. Conclusions and future perspectives.....	49
7. References.....	50
Annex	58

List of Abbreviations

C1GALT1 - Core 1 Synthase, Glycoprotein-N-Acetylgalactosamine 3-beta-Galactosyltransferase 1
C2GNT - β -1,3-Galactosyl-O-Glycosyl-Glycoprotein β -1,6-N-acetylglucosaminyltransferase
CIN - chromosomal instability
CSCs - cancer stem cells
DMSO - dimethyl sulfoxide
EBV - Epstein-Barr Virus
ECM - extracellular matrix
ER - endoplasmic reticulum
GAGs - glycosaminoglycans
Gal - Galactose
GalNAc - N-acetylgalactosamine
GALNT - GalNAc-transferase
GlcA - glucuronic acid
GlcNAc - N-acetylglucosamine
GNT - N-acetylglucosamine transferases
GS - genomically stable
HA - Hyaluronic acid
HGF - hepatocyte growth factor
IF – Immunofluorescence
LacNAc - N-acetyllactosamine
Le^a - Lewis a
Le^x - Lewis x
MMP-9 - matrix metalloproteinase-9
MMPs - matrix metalloproteinases
MSI - microsatellite instability
MSP - macrophage-stimulating protein
MT1-MMP - membrane-type 1 matrix metalloproteinase
RON - “Recepteur d’origine Nantais”
RTKs - receptor tyrosine kinases
Ser – Serine
SLe^a - sialic acidlyl-Lewis A
SLe^x - sialic acidlyl-Lewis X
SP-PLA - solid-phase proximity ligation assay
ST - Sialic acidlyl-T
ST3GAL - β -galactoside α -2,3-sialic acidlyltransferase
ST6GALNAC - α -N-acetylgalactosaminide α -2,6-sialic acidlyltransferase
STn – Sialic acidlyl-Tn
TCGA - The Cancer Genome Atlas
Thr – Threonine
TSG-6 - Tumor necrosis factor-simulated gene 6
VEGFR - vascular endothelial growth factor receptor
WHO - World Health Organization

Figures list

Figure 1: Schematic representation of the glycocalyx and transmission electron micrograph of a leukocyte glycocalyx (taken from Alberts, Bray et al. 1994).	10
Figure 2: First steps of O-glycosylation pathway and core structures.	12
Figure 3: Lewis type antigens Lea, Lex and their sialic acidylated counterparts SLea and SLex.	14
Figure 4: Schematic representation of the different truncated O-glycan structures (Tn, STn, T, ST and disialic acidylated core 2 structure) and their aberrant expression mechanisms.	16
Figure 5: Molecular structure of hyaluronic acid.	17
Figure 6: Schematic representation of gastric cancer progression stages from normal gastric mucosa towards gastric carcinoma. (adapted from Correa et al 2010 (57)).	20
Figure 7: Schematic representation of the general structure of the largest isoform (CD44v2-v10) and the smallest isoform (CD44s).	23
Figure 8: Representation of CD44v6 co-receptor function.	27
Figure 9: Western blot analysis of CD44 expression in cell lysates and flow cytometry analysis of the glycoengineered cell lines SC and ST6 compared to their controls WT and M6 respectively.	33
Figure 10: Western blot analysis of cell lysates from WT, SC, M6 and ST6 treated with benzyl- α -GalNAc.	34
Figure 11: Immunoprecipitation of CD44 in the glycoengineered cell line SC and its control WT.	35
Figure 12: Western blot analysis of cell lysates from glycoengineered cell line models (SC, ST6 and ST3) and their controls (WT, M6 and M3) stained with different lectins.	36
Figure 13: Immunofluorescence images of glycoengineered cells (SC, ST6 and ST3) and their respective controls (WT, M6 and M3) stained for CD44 and CD44v6.	37
Figure 14: Flow cytometry analysis of CD44 and CD44v6 in MKN45 glycoengineered cell lines and flow cytometry analysis of HA-Fluorescein (HA-FITC) binding. Comparative western blot analysis of CD44 present in the total cell lysates and in the secretome.	38
Figure 15: Immunofluorescence images of glycoengineered cells (SC, ST6 and ST3) and their respective controls (WT, M6 and M3) stained for RON and phospho RON (pRON).	39
Figure 16: Microscopy images of proximity ligation assay (PLA) of all cell lines.	40
Figure 17: Combinatory proximity ligation assay immunofluorescence (Co-PLA-IF) images of WT and SC cell lines and human gastric cancer tissue samples.	41
Figure 18: Comparative western blot analysis the gastric cancer MKN45 cell lines (WT and SC) and the colorectal cancer LS174T cell line. Immunofluorescence images of colorectal cancer cell line LS174T, stained for CD44, CD44v6, SLex, STn, RON and pRON.	43

1. Introduction

1.1 Glycosylation

In the beginning of the 20th century, carbohydrates were a main target of study in the metabolism and from a structural perspective (1). However, at the time other significant biological functions were not identified. It was not until the late 20th century that carbohydrates have also been getting attention as modulators of various cellular and biological processes, giving rise to a new rapidly growing field in molecular biology known as “Glycobiology” (2). All cells known in nature are surrounded by a layer of carbohydrate structures, referred to as the glycocalyx. The biosynthesis of carbohydrate structures of the glycocalyx occurs in an endoplasmic reticulum to Golgi complex directional axis using monosaccharides as building blocks that are added onto lipids and proteins affecting their biochemical properties both structurally and functionally. Additionally, glycans can also be secreted in the lipids of vesicles or in the form of secreted proteins. In recent years, the glycobiology field has made major advances thanks to the development of analytical methods, new antibodies and lectin-based assays, which are proteins that bind to specific carbohydrate structures (3).

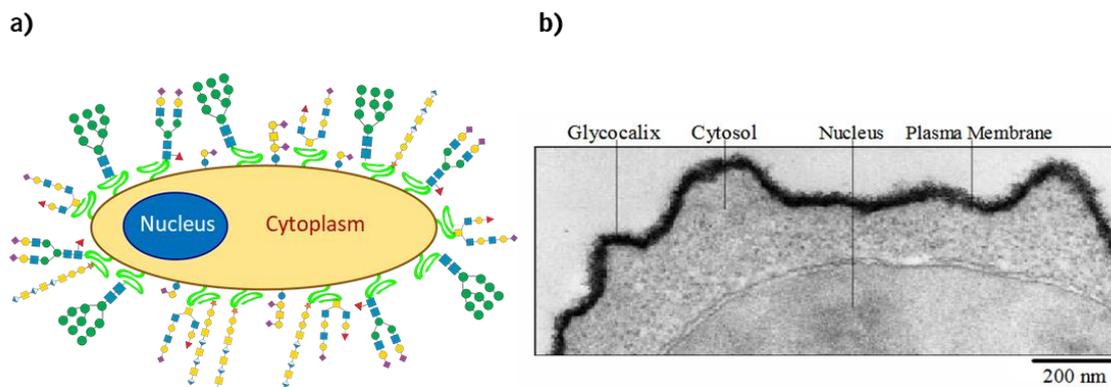


Figure 1: a) Schematic representation of the glycocalyx surrounding a eukaryotic cell; b) Transmission electron micrograph of a leukocyte glycocalyx, seen as a black layer surrounding the cell (taken from Alberts, Bray et al. 1994).

Protein glycosylation is one of the most common post-translational modification and arguably the most complex. It is characterized by the stepwise addition of monosaccharides to the protein backbone, giving rise to long carbohydrate structures termed glycans. Glycan structures play key roles in many cellular mechanisms, such as cellular adhesion, migration, cell signaling and immunomodulation, and are classified depending on their structure and synthesis pathway in *N*-glycans, *O*-glycans and glycosaminoglycans (4-6). Unlike DNA, RNA or proteins, the biosynthesis of glycans is not a template-based process (7, 8). Instead, it is a multistep process highly regulated by various enzymes named

glycosyltransferases. The biosynthesis pathway of the major glycan classes occurs mainly in the endoplasmic reticulum (ER) and the Golgi complex, which is catalyzed by different glycosyltransferases that add sugar residues provided by various nucleotide sugar transporters to the protein or glycan structure. Glycosyltransferases and nucleotide sugar transporters are the main key players in the glycosylation machinery and are responsible for its complex regulation.

1.2 *O*-GalNAc glycosylation

O-glycosylation is characterized by a glycan structure linked to the protein backbone through an oxygen atom. There are various types of *O*-glycans, but the most abundant form in humans is the *O*-GalNAc glycans or mucin-type *O*-glycans (henceforward referred to as *O*-glycans) (9).

The *O*-glycan biosynthesis pathway starts with the addition of an *N*-acetylgalactosamine (GalNAc) residue to a serine (Ser) or a threonine (Thr) of the protein backbone through an *O*-linkage. This reaction is catalyzed by a GalNAc-transferase (GALNT) (5, 10). More sugar residues are added to the GalNAc giving rise to different core structures that can be further elongated. The addition of a galactose (Gal) residue to the β 1-3 position of the GalNAc, generates a core 1 structure, catalyzed by the Core 1 Synthase, Glycoprotein-*N*-Acetylgalactosamine 3- β -Galactosyltransferase 1 (C1GALT1) glycosyltransferase. Further addition of a *N*-acetylglucosamine (GlcNAc) residue to the β 1-6 position of the GalNAc generates a core 2 structure, catalyzed by the β -1,3-Galactosyl-*O*-Glycosyl-Glycoprotein β -1,6-*N*-acetylglucosaminyltransferase (C2GNT) glycosyltransferase. Aside from core 1 and core 2 structures, there is also core 3 and core 4. These are generated by the addition of GlcNAc residues to the GalNAc in a β 1-3 position only (Core 3) or in both β 1-3 and β 1-6 position (Core 4).

Normally, core structures do not occur without extensions or modification. For example, a GlcNAc can be added to the terminal Gal residue in a β 1-4 position of the Core 2 structure, to which subsequently another Gal is added. This disaccharide [Gal β 1-4GlcNAc] is named *N*-acetylactosamine (LacNAc) and is

usually repeated, generating a long carbohydrate chain. These structures are not exclusive to *O*-glycans, as they can also occur in *N*-glycans and glycosphingolipids.

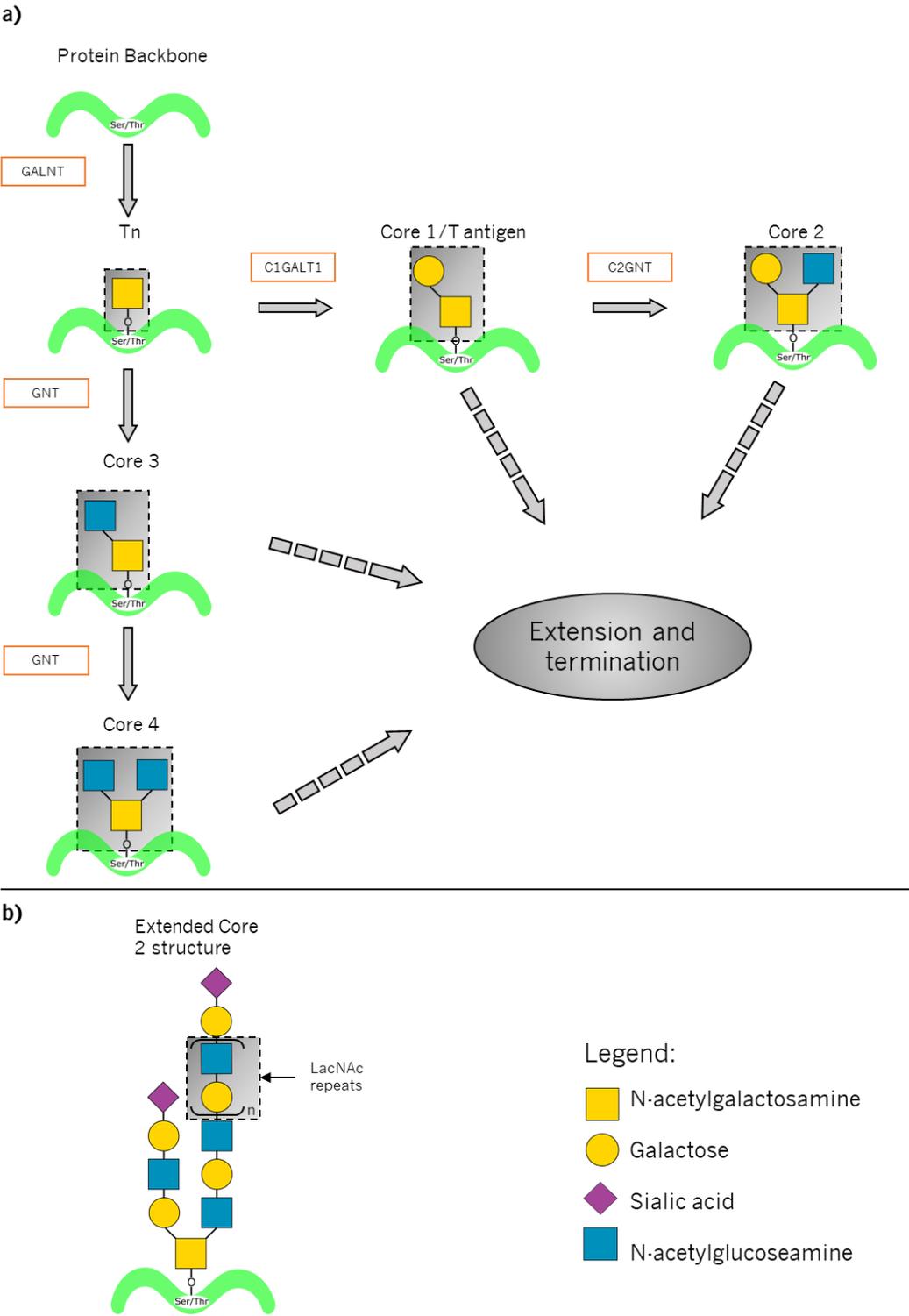


Figure 2: (a) First steps of *O*-glycosylation pathway and core structures. All glycotransferases responsible for each reaction are represented inside orange boxes. It is important to note that the synthesis of core 3 and core 4 structures are catalyzed by different *N*-acetylglucosamine transferases (GNT); (b) Example of an extended core 2 structure with LacNAc repeats.

Sialic acidylation is another common modification of glycans. It consists in the addition of a sialic acid residue to the glycan chains which is catalyzed by different sialic acyltransferases. Usually Sialic acid residues are found in the terminal part of the glycan, acting as a capping molecule and preventing further chain elongation (11). There are various families of sialic acyltransferases that have different specific substrates and catalyze the addition in different positions. For instance, in humans, β -galactoside α -2,3-sialic acyltransferase (ST3GAL) is a type of sialic acyltransferases that catalyzes the addition of the Sialic acid residue to a Gal residue in a α 2-3 position, whilst α -N-acetylgalactosaminide α -2,6-sialic acyltransferase (ST6GALNAC), another family of sialic acyltransferases, catalyzes the addition of a Sialic acid residue to a GalNAc residue in a α 2-6 position (12). Overexpression of specific sialic acyltransferases can cause earlier sialic acidylation and thus truncation of *O*-glycan chains. This early sialic acidylation has been observed in many pathologies, especially in cancer, and will be discussed in more detail ahead (13-17).

O-glycans can also be fucosylated, process that is catalyzed by different fucosyltransferases, for example α 1-3 and α 1-4 fucosyltransferases that are required for the synthesis of Lewis type structures (18, 19). These are antigenic sugar structures which include Lewis A (Le^a) [$Gal\beta$ 1-3($Fuc\alpha$ 1-4) $GlcNAc$ -], Lewis X (Le^x) [$Gal\beta$ 1-4($Fuc\alpha$ 1-3) $GlcNAc$ -] and their sialic acidylated counterparts, sialic acidyl-Lewis A (SLe^a) [$Sialic\ acid\alpha$ 2-3 $Gal\beta$ 1-3($Fuc\alpha$ 1-4) $GlcNAc$ -] and sialic acidyl-Lewis X (SLe^x) [$Sialic\ acid\alpha$ 2-3 $Gal\beta$ 1-4($Fuc\alpha$ 1-3) $GlcNAc$ -]. Both SLe^x and SLe^a are major ligands for selectins, which are lectins important in the leukocyte adhesion to the endothelium and extravasation during inflammation processes (20). In cancer, the expression of SLe^x and SLe^a modulate metastatic behavior of cancer cells (21-23).

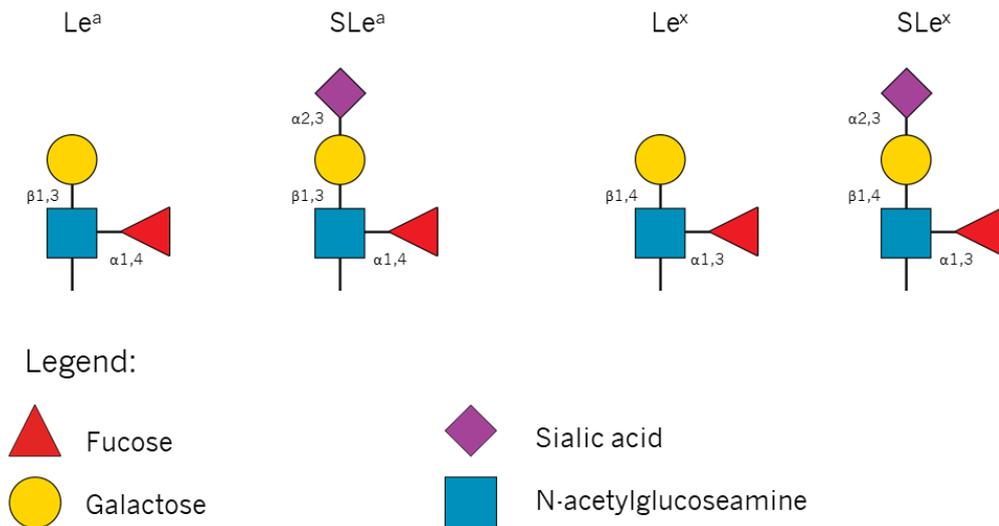


Figure 3: Lewis type antigens Le^a , Le^x and their sialic acidylated counterparts SLe^a and SLe^x . The type of linkage between sugars is represented close to its respective bond.

Considering that *O*-glycans are built in a stepwise manner, enzymatic localization is important for regulation of the biosynthesis pathway. Enzymes involved in the biosynthesis of *O*-glycans are separated in the Golgi apparatus, more specifically enzymes necessary for the first steps in the biosynthesis, for example C1GALT1, are enriched in the cis- and medial-Golgi, while enzymes required in the last steps, for example ST6GALNAC1, are mainly located in the trans-Golgi (24-27).

1.3 Truncation of *O*-glycans

One of the most prevalent alterations in protein *O*-glycosylation observed in cancer cells is the truncation of *O*-glycans (14, 15), and usually it is associated with the expression of specific tumor associated glycans, namely T, Tn, Sialic acidyl-T (ST) and Sialic acidyl-Tn (STn) antigens. The aberrant expression of these antigens was identified in many different types of carcinomas, for example in cancers of the stomach, pancreas and colorectum (28-30). The Tn antigen, being the simplest antigen among these, consist of a GalNAc residue α 1 linked to the Ser or Thr of the protein backbone. T antigen, also referred to as Thomsen–Friedenreich antigen, results in an extension of the Tn antigen by one Gal residue. These two antigens can also be sialic acidylated originating ST and STn antigens. It is described that sialic acidylation blocks further elongation and results in termination of the carbohydrate chain (15). One mechanism underlying these modifications is originated by upregulation or downregulation of important glycosyltransferases present in the biosynthetic pathway of *O*-linked glycans. For example, the overexpression of ST6GALNAC1 can cause premature sialic acidylation of glycans, blocking further elongation of the core 1 structure originating STn antigen (16, 17, 30). Another mechanism associated

with truncated *O*-glycans is the downregulation of C1GALT1, a galactosyltransferase responsible for the synthesis of core-1 structures from Tn antigen (31-33). Interestingly, the downregulation could be a result of mutations or hypermethylation on the *COSMC* gene, which encodes a C1GALT1 dedicated chaperone, thus the lack of a functional *COSMC* chaperone results in an improper folding of the C1GALT1 enzyme and consequently loss of its catalytic activity, resulting in an accumulation of STn and Tn structures (34). Yet another mechanism of expression of truncated *O*-glycans is through overexpression of ST3GAL4. This sialic acidyltransferase adds sialic acidic acids in a α 2,3-position to a Gal residue, and when overexpressed it results in an increase of another truncated *O*-glycan structure, the sialic acidylated core 2 structure. In addition, overexpression of ST3GAL4 also the general sialic acidylation profile, shifting the α 2,6 sialic acidylation to α 2,3 and leading to overexpression of SLe^x epitopes in both *N*-glycans and *O*-glycans (35, 36).

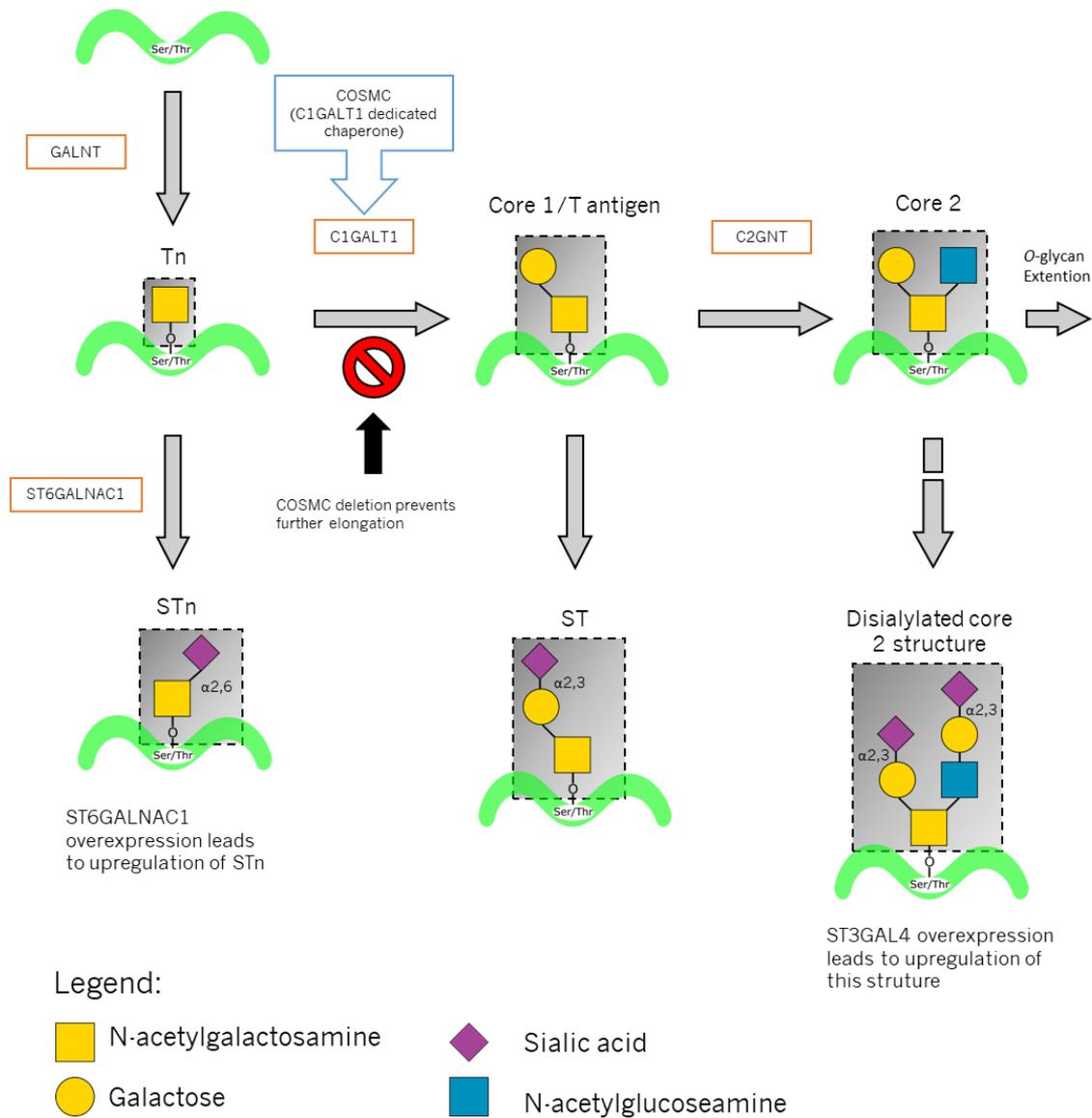


Figure 4: Schematic representation of the different truncated O-glycan structures (Tn, STn, T, ST and disialylated core 2 structure) and their aberrant expression mechanisms. Abrogation of COSMC expression prevents O-glycan elongation leading to an accumulation of Tn and STn structures. Overexpression of ST6GALNAC1 causes upregulation of STn structures. This enzyme will compete with C1GALT1 for the substrate. ST3GAL4 adds sialic acid units in a $\alpha 2,3$ position. Its overexpression leads to the upregulation of the disialylated core 2 structure and a shift from $\alpha 2,6$ to $\alpha 2,3$ sialic acidylation (36).

1.4 Hyaluronic acid

Another type of glycans are glycosaminoglycans (GAGs). GAGs are non-ramified long carbohydrate chains consisting of repeating disaccharide sugar units and play a variety of important roles in different cellular processes, for example providing a scaffold for epithelial cell proliferation, differentiation and migration (37). One of which is hyaluronic acid (HA), an extracellular matrix (ECM) component and consists of repeating disaccharide units of GlcNAc and glucuronic acid (GlcA).

Among all GAGs, HA is unique in the regard that it does not have sulfated groups and it is not synthesized in the Golgi complex, instead it is synthesized at the plasma membrane level by different hyaluronic acid synthases, being polymerized in the cytoplasmic surface and then extruded to the extracellular environment (38). HA is found conserved in different species, and in humans it is present in various connective and soft tissues, such as the skin, cartilage, blood vessels, vitreous humor of the eye and umbilical cord (39, 40). Due to its anionic nature and high hydrophilicity, HA is able to retain water molecules, maintaining tissue hydration, elasticity and permeation to different sized molecules (39). These biochemical properties are important in reducing tissue friction and compression resistance in the joints and in giving elasticity and youthful appearance to the skin (41). In addition, HA is able to bind specifically to several proteins called hyaladherins, for example LYVE-1, RHAMM, TSG-6, ICAM-1 and CD44 (42). In the majority of proteins of this superfamily it is present common sequence of 100 amino acids responsible for this binding activity often referred as “link module”. The ability of HA to bind to various proteins makes it not only a structural scaffold but also serves to modulate several biological processes such as cellular signaling and cell motility depending on which proteins are bound (43, 44). In cancer, HA has been observed to mediate several malignant phenotypes such as invasiveness, migration, tumor growth and serving as a barrier for compounds reducing the effectiveness of therapies (45-47). Furthermore, it is reported HA is strongly expressed in gastric cancer and that it correlated with tumor spread and short survival (48). It is hypothesized that the invasion phenotype of these cancer cells is due to increased production of type I collagenase, which digests collagen fibrils, and capability of stimulating HA synthesis in host fibroblasts promoting cell motility and invasion through the host tissues (48, 49).

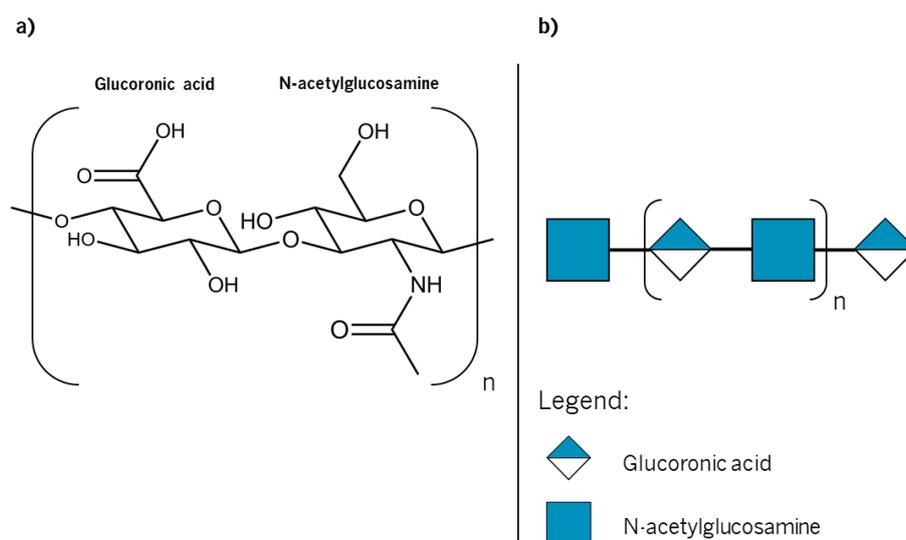


Figure 5: Molecular structure of hyaluronic acid with monosaccharides represented as (a) chair conformation and in (b) in its respective symbol nomenclature for glycans (48).

1.5 Gastrointestinal Carcinoma

Gastrointestinal cancer is among the most common types of carcinoma worldwide. These cancers include gastric, esophageal, pancreatic, liver and colorectal cancer, with gastric cancer being the fifth and colorectal cancer being the third most common form of carcinoma worldwide (50). In 2012, it was estimated 1.4 million cases of colorectal cancer and around 1 million of gastric cancer cases globally (50). That number increased to around 2 million for colorectal cancer and maintained for gastric cancer in 2018 (51). In Europe, colorectal cancer represents the second most common cause of cancer related death, after lung cancer, while gastric cancer represents the fourth most common (52). Survival rates of gastrointestinal cancers depend on a variety of factors such as lymph node metastases, molecular cancer subtype, geographic location, but above all, on the stage at which the cancer was diagnosed (53, 54). The 5-year-survival rate drastically increases if the cancer is diagnosed at early stages, as well a better prognostic for the patient, justifying the necessity to develop new diagnostic procedures with increased efficiency (55).

Factors that may contribute for development of gastric cancer high salt intake, smoking habit, alcohol consumption and infections caused by *Helicobacter pylori* or Epstein-Barr virus (56, 57). It has been shown that eradication of *H. pylori* infections decreases the risk of developing gastric cancers (58). According to Correa's hypothesis, gastric cancer progression occurs in a multistep process (57). High salt intake and *H. pylori* infections can lead to the first stages of gastric tissue inflammation, namely superficial gastritis and atrophic gastritis (57). At this stage, a variety of carcinogens may act progressing the lesion into intestinal metaplasialic acid, which is characterized by the stomach tissue presenting small intestine phenotypes. Intestinal metaplasialic acid can be subdivided complete and incomplete (59). When the stomach epithelium expresses a well-defined brush border, differentiated goblet cells and most of the digestive enzymes present in the small intestine it is classified as complete intestinal metaplasialic acid. This lesion can progress into incomplete intestinal metaplasialic acid, where a brush border is absent and presenting numerous irregular mucin droplets. The last stage before carcinoma is dysplasialic acid, defined as benign neoplastic lesions (60). Patients suffering from dysplasialic acid have a high risk of progressing into gastric cancer (60). Because gastric cancer showcases a high degree of heterogeneity it is subdivided in different groups depending on the tumor characteristics. One widely used classification method is the Laurén classification. The Laurén classification divides gastric cancer in two main groups depending on the tumor appearance and behavior: Diffuse-type and intestinal-type (61). Intestinal-type gastric cancer exhibits cell adhesion, forming organized structures, whereas diffuse-type seem to lack

cell-cell adhesion and usually these cancer cells infiltrate in the stroma leading to scattered and more homogeneous tumor cells (61). However, in a variety of cases, tumor characteristics do not fit in either group and are consequently classified as indeterminate type. Some studies suggest that the Laurén classification does not correlate with patient prognosis, sparking controversy regarding the clinical relevance of the Laurén classification (62). On the other hand, studies have also validated the prognostic relevance of this classification even as an independent prognostic factor (63, 64). In addition, presence of diffuse-type adenocarcinoma is correlated with worse prognosis for the patient when compared to the intestinal type (63, 64). In 2010, the World Health Organization (WHO) issued a more detailed classification method that includes not only the most common gastric adenocarcinomas but even rarer types. More recently, The Cancer Genome Atlas (TCGA) described four different gastric cancer molecular subtypes: Epstein-Barr Virus (EBV), microsatellite instability (MSI), genomically stable (GS) and chromosomal instability (CIN) (65). Resorting to this classification it was possible to discriminate patients with poor prognosis and predict which group would benefit more from adjuvant chemotherapy, for instance, patients with the MSI subtype were likely to have poor prognosis whilst patients with EBV subtype had better prognosis compared to other subtypes (65).

Colorectal cancer is yet another type of gastrointestinal carcinoma that is affecting 2 million people in 2018 around the world (51). Similar to gastric cancer, risk factors such as diet, sedentarism and alcohol consumption are associated with the development of colorectal cancer (66). In addition, hereditary diseases such as familial adenomatous polyposis , or inflammatory bowel diseases such as Chron's disease and ulcerative colitis are also included as risk factors (67-69). So far, 3 different molecular mechanisms have been identified that can lead to colorectal carcinoma *per se* or in combination. These are chromosomal instability, CpG island methylator phenotype and microsatellite instability (70). The colorectal cancer subtyping consortium characterized the existence of four consensus molecular subtypes based on transcriptomic data, gene expression, proteomic and other available data (71, 72). These are microsatellite instability immune, canonical, metabolic and mesenchymal subtype.

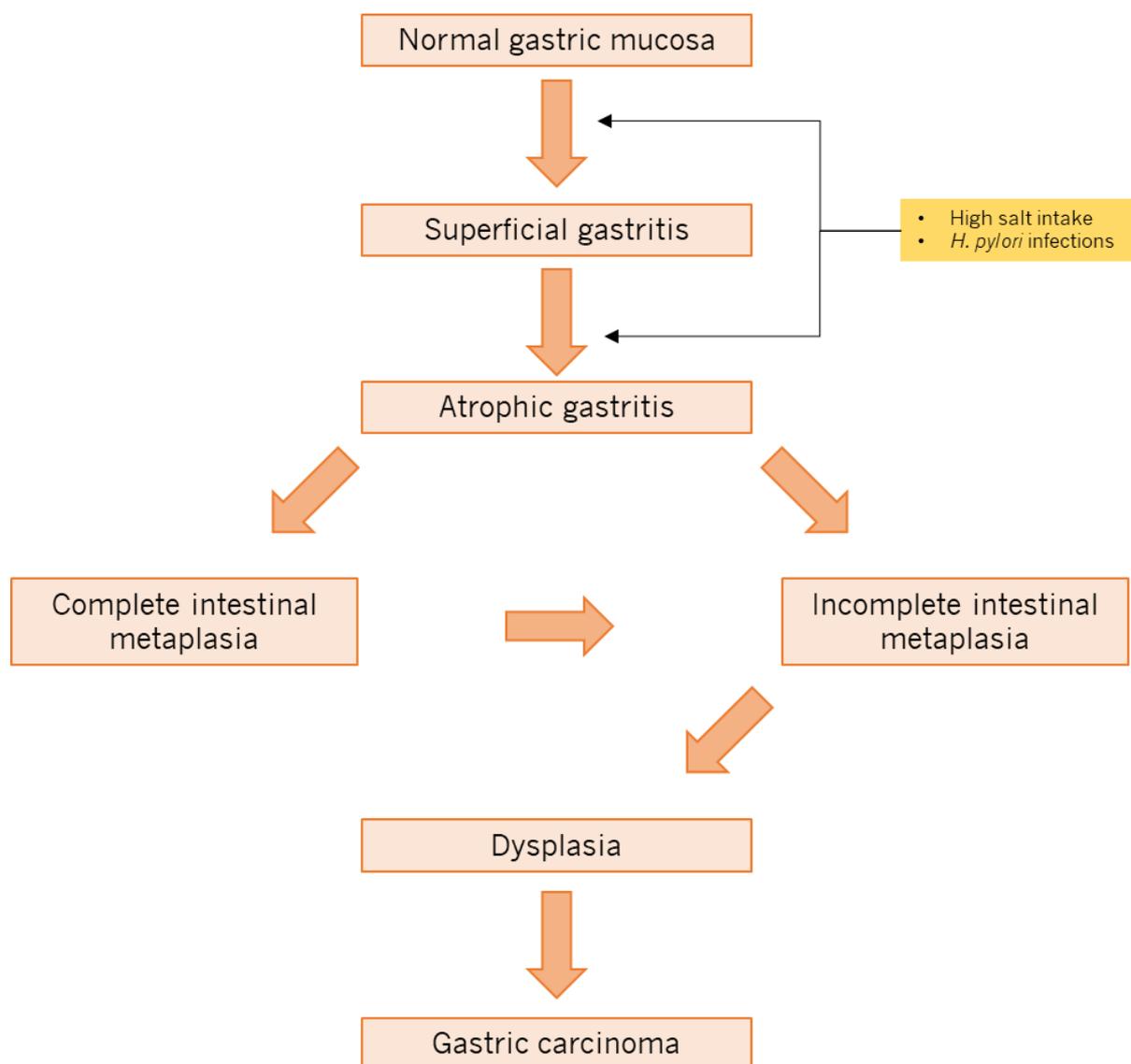


Figure 6: Schematic representation of gastric cancer progression stages from normal gastric mucosa towards gastric carcinoma. Risk factor for the development of superficial and atrophic gastritis include high salt intake and *H. pylori* infections. (adapted from Correa et al 2010 (57)).

1.5 Shortcomings of current serological assays and CD44 as a potential biomarker for gastrointestinal cancer detection

There is great potential of serological assays for gastrointestinal cancer diagnosis that are based on alteration of glycosylation profiles of secreted or transmembrane glycoproteins. Two commonly used assays that evaluate the serum quantity of glycoconjugates or glycans in gastrointestinal cancer patients are CEA and CA19-9. Even though these assays are not specific enough to serve as a tool for cancer diagnosis, they are used to monitor malignant progression and cancer recurrence. Studying the glycosylation profile of these biomarkers could prove to be advantageous by increasing their specificity. In order to quantify these biomarkers, mass spectrometry techniques and immune-affinity assays are

being implemented in the clinical setting as possible options for diagnosis (73). However, with these approaches entails the problem of low sensitivity, possibly due to the complexity of biological samples. Certainly, high sensitivity and specificity are extremely important in the clinical field for a clear diagnosis of the patient. Some studies have been focusing in addressing this problem by developing and discovering new techniques providing higher sensitivity for detection of specific cancer biomarkers in biological samples. For example, it was demonstrated that solid-phase proximity ligation assay (SP-PLA) is a possible option for detection of different post-translational modifications (74). As the name suggests, PLA allows proximate recognition of two different target molecules by using pairs of affinity probes, namely antibodies, each containing an oligonucleotide strand and if the target molecules are in proximity, these strands will hybridize allowing for a signal amplification (75). In conjunction with a solid phase, the SP-PLA could provide a high sensitivity detection technique suitable for clinical applications. Recent studies have shown the utility of SP-PLA in cancer biomarker detection. Using this technique, it was possible to detect a significant increase of CD44 containing STn antigen in the serum of gastric cancer patients (74). Other studies also disclose CD44 as a potential gastric cancer biomarker. With the intent to explore *O*-glycoproteins as potential gastric cancer biomarkers, previous studies were able to identify multiple circulating glycoproteins with truncated *O*-glycans and containing the STn antigen through characterization of the *O*-glycoproteome in different gastric cancer cell lines (76). One of the identified proteins was CD44, suggesting the importance of studying this protein as a potential biomarker for gastrointestinal carcinoma.

1.6 CD44: a signaling scaffold

CD44 is a highly heterogeneous transmembrane glycoprotein whose size ranges from 80 kDa to 200 kDa (77). CD44 consists of 3 major domains: A *N*-terminal globular domain and a stem structure in the cell surface, a transmembrane domain and an intracellular domain (77). All CD44 isoforms are encoded by one highly conserved gene, present on chromosome 11 in humans. It is responsible for modulating various cell mechanisms such as cell-cell and cell-matrix interaction, cell migration, proliferation and signaling, more specifically CD44 is the main receptor for HA and a co-receptor for many different receptor tyrosine kinases (RTKs). The heterogeneity of CD44 is due to three major factors: (i) high alternative splicing, (ii) its high number of glycosylation sites (iii) and as a target for proteolytic cleavage (78-80). It has also been discovered that some CD44 isoforms are overexpressed in cancer and play a key role in different cancer cell mechanisms, for example tumor growth and metastasis (81).

1.7 The different CD44 isoforms

The human gene encoding CD44 (NG_008937) located in the chromosome 11, has a total of around 100 kb of genetic information and is composed of 19 exons. It possesses a complex alternative splicing pattern. The first five and the last five exons are ubiquitously expressed, except for the second last exon in which contains an early 3'UTR site. The standard CD44 isoform, named CD44s, is the smallest isoform since it only expresses the 9 standard exons (82). The 9 other exons are called variant and are alternatively spliced giving rise to a variety of CD44 isoforms, designated CD44v(X), wherein "X" is referred to the number of the variant exon it includes (82). The stem structure of the CD44s isoform, has only 46 amino acids, but it is elongated when variant exons are expressed. The variant isoforms of CD44 are expressed mainly in high proliferating epithelial cells, however cancer cells also express large CD44 variants. For example, CD44v8-v10, often referred to as CD44E due to its expression in epithelial cells, has been observed to be overexpressed in gastric cancer (83). Another example is the variant isoform CD44v6, which has been observed overexpressed in gastric cancer cells, promoting metastatic behavior and growth stimuli (84, 85). Also, it is described as a potential biomarker to distinguish intestinal- and diffuse-type stomach carcinoma (86). Not only CD44v6, but also CD44s plays its role in cancer context. In previous studies, it has been demonstrated that CD44s is important for tumor initiation, conferring stem-like phenotypes, for example tumorsphere formation, and its downregulation decreases tumor initiation *in vitro* and *in vivo* (87).

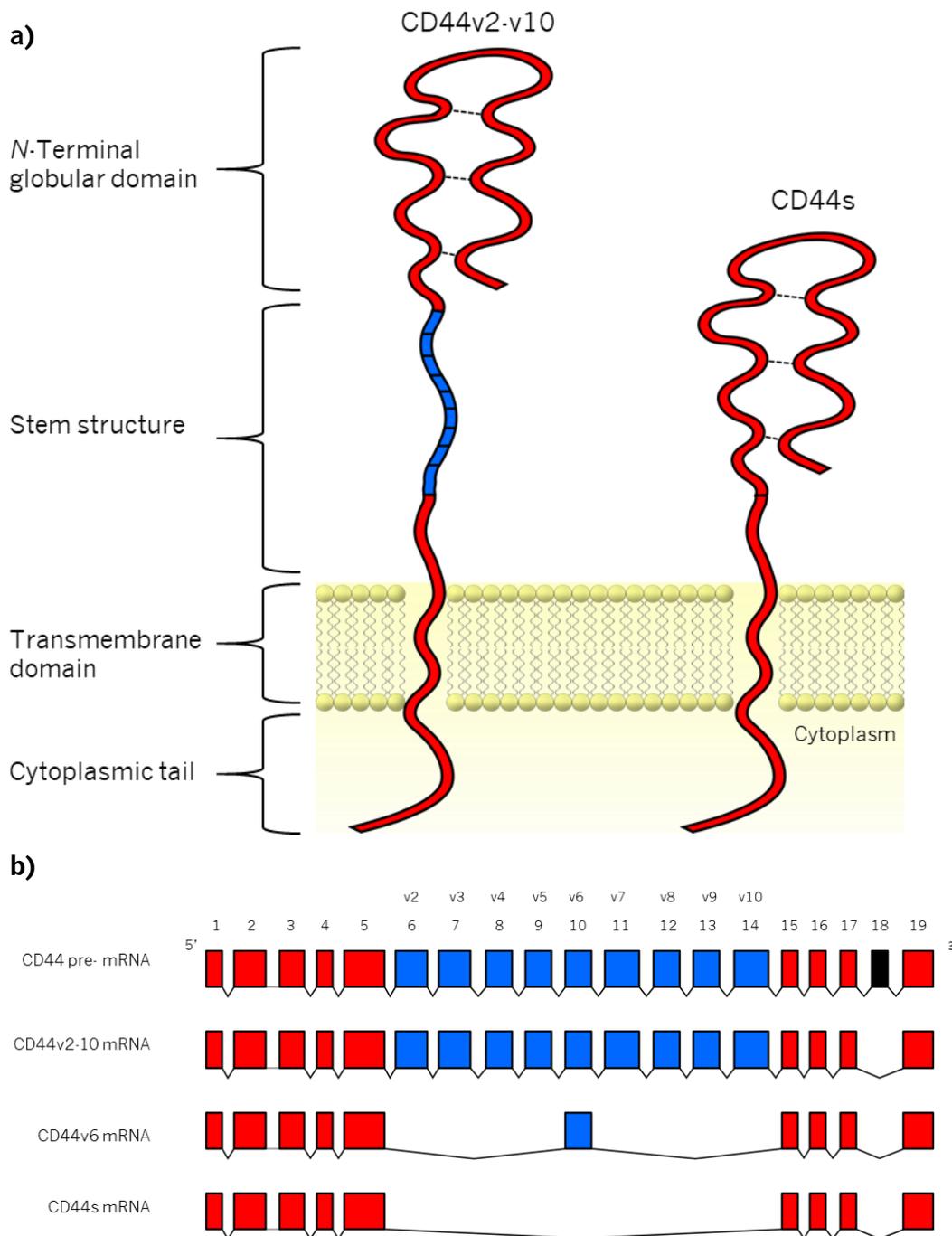


Figure 7: a) Schematic representation of the general structure of the largest isoform (CD44v2-v10) and the smallest isoform (CD44s). Red represents the ubiquitously expressed regions blue represents the variant region. Disulfide bonds are represented in dotted lines.

1.8 CD44 as a receptor for hyaluronic acid

CD44 is an important hyaladherin. The amino terminal globular domain is encoded by the first 5 exons of the gene, and it contains motifs that play a key role in the HA-binding mechanism. In this domain there is a region of around 100 amino acids, named “link module” (88), that is conserved in the hyaladherin superfamily (42). This family includes proteins such as brevican, aggrecan and the tumor necrosis factor-

simulated gene 6 (TSG-6) and in contrast to these proteins, CD44 binding affinity to HA is not only modulated by this “link module” but also by an adjacent region in the amino terminal globular domain (42). Structural analysis of CD44 N-globular domain revealed the existence of a disulfide bond between two cysteine residues (Cys28-Cys129 in human) responsible for stabilizing the region containing the “link module” and it is highly solvent accessible (89). Among this, other disulfide bonds in the N-terminal globular domain have been observed to be important for the HA-binding affinity, through structural stabilization of the protein and direct contact with the HA. It was also observed that cells expressing CD44 treated with reducing agents showed a significant decrease in adhesion to HA-coated plates (89).

There are also studies that report the effects of the cytoplasmic domain on the HA-binding mechanism. It has been observed that the cytoplasmic domain is not necessary but may contribute in the binding activity (90, 91). For example, a study conducted on cells expressing CD44 with and without truncated cytoplasmic domain reported that the cells with normal CD44 were capable of binding to HA in solution and immobilized in plastic culture wells while the cells with CD44 truncated in the cytoplasmic domain could only bind to immobilized HA (90). It is also known that the cytoplasmic domain of CD44 does indirectly interact with the actin cytoskeleton via two adaptors, ezrin/radixin/moesin (ERM) proteins and ankyrin (92). However, the role of the cytoplasmic domain for HA-binding mechanism is still controversial, considering there are other studies suggesting that this domain is not particularly involved in inducing this mechanism (91).

The stem structure also plays a part in modulating CD44 binding affinity to HA. This structure contains sites subject to proteolytic cleavage by enzymes (80), more specifically matrix metalloproteinases (MMPs), for example membrane-type 1 matrix metalloproteinase (MT1-MMP) and matrix metalloproteinase-9 (MMP-9) (93). It has been reported that CD44 undergoes proteolytic cleavage by MMP-9, producing an extracellular portion (CD44ECD) that increases cancer cell migration and invasion, and an intracellular portion (CD44ICD) that is translocated to the nucleus where it will initiate processes of transcriptional regulation of several genes promoting cancer cell adhesion (93). Alternative splicing has been reported to affect CD44-HA binding affinity, meaning that some CD44 isoforms have stronger binding affinities to HA than others (94, 95). In a previous study, transfecting rat pancreatic carcinoma cells with CD44v4-v7 constructs revealed that these transfectants had significantly higher affinity to HA in solution than their wild type counterparts expressing only CD44 standard isoform, possibly due to an increased clustering effect of CD44 variants compared to the standard isoform (94). Another interesting observation is that these CD44 variants do not require the cytoplasmic domain to bind to HA in solution (94).

Another factor that seems to heavily affect CD44 binding affinity to HA is post-translational modifications, namely glycosylation. For example, it has been described that CD44 *N*-glycosylation regulates the binding affinity to HA (96, 97). Early studies conducted on human cells that constitutively express CD44 reported that when these cells were treated with tunicamycin, an inhibitor of the enzyme GlcNAc phosphotransferase (GPT) necessary for early steps in the biosynthesis of *N*-linked glycans, they lose their ability to attach to HA-coated surfaces. It has also been hypothesized that tunicamycin treatment prevents clustering between CD44 variants (96). On the other hand, other studies showed that *N*-glycosylation is also able to negatively regulate the binding affinity of CD44 to HA (98). However, it is important to note that these studies were based on the induction of severe alterations, since tunicamycin has been reported to disturb protein folding and causing endoplasmic reticulum stress (99). Using molecular dynamics simulations to study *N*-glycosylation in the extracellular globular domain led to conclude that inhibition of the HA-binding mechanism could be due to terminal sialic acid residues in the *N*-linked glycans, and not to steric blockage of the binding site (100, 101). The disagreement between these studies suggest that biological context is important to understand how glycosylation regulates this function. The binding affinity of CD44 to HA also has its implications in the context of cancer. This interaction was studied in chronic lymphocytic leukemia cells, and it was concluded that the adhesion of these cells to HA-bearing primary mesenchymal stromal cells was potentiated by CD40L in a CD44-dependent manner, and that CD40L activation induced the expression of *N*-glycosylated CD44v6, facilitating the binding to HA by this, and possibly others, variant isoform (102). Similar to mucin-type proteins, CD44 shows high levels of *O*-glycosylation (103). Not only *N*-glycosylation but also *O*-glycosylation seems to modify the binding affinity of CD44 to HA and could therefore impact physiologic and pathologic processes (104, 105). Conducted research on colon carcinoma cell lines was done to assess the effects of CD44 *O*-glycosylation in the HA-binding function. It was found that colorectal cancer cells expressing high molecular weight CD44 isoforms when treated with phenyl- α -GalNAc, an inhibitor of *O*-glycosylation, show no significant difference in adhesion to HA, however cells expressing CD44s presented an enhancement of this binding mechanism when treated in the same conditions, suggesting that CD44 *O*-glycosylation possibly inhibits its binding affinity to HA (104). Furthermore, it was also observed that this enhancement was blocked when cells were pre-treated with BRIC 235, an anti-CD44 monoclonal antibody, demonstrating that this adhesion was CD44-mediated. Inhibition of *O*-glycosylation has been demonstrated to not affect CD44 cell surface expression. Moreover, studies also report that both CD44 *O*-linked glycans and *N*-linked glycans seem to be involved in the adhesion of endometrial cells to peritoneal mesothelial cells, which are known to secrete HA. Treating endometrial cells with tunicamycin and benzyl 2-acetamido-2-deoxy- α -

D-galactopyranoside (Benzyl- α -GalNAc), another *O*-glycosylation inhibitor, decreased attachment to peritoneal mesothelial cells (105).

1.9 CD44 as a co-receptor for receptor tyrosine kinases

RTKs are surface receptor important in many signaling cascades, forming docking sites for a variety of different molecules, for example growth factors, cytokines and hormones, by which they are activated (106). Activation of these receptors occurs typically through autophosphorylation activity and dimerization of two subunits which is induced or stabilized by the receptor ligand (106). It is well known that receptor activation is much more complex considering that there are many different proteins that possess co-receptor function, which are capable of modulating RTK activation or downstream signaling. It has been discovered that CD44 can act as a co-receptor for different RTKs such as “Recepteur d’origine Nantais” (RON), MET, vascular endothelial growth factor receptor (VEGFR), and many more (107). It is described that hepatocyte growth factor (HGF) acts as a ligand for MET receptor, activating it and inducing a multitude of cellular processes, for example cell proliferation, motility, survival, scattering, differentiation and morphogenesis (108). In cancer, MET receptor is overexpressed and aberrantly activated which confers more aggressive, metastatic and invasive phenotypes to those cells (108). Some CD44 variant isoforms are required for HGF-mediated of MET receptor activation. For example, studies found out that HGF was able to bind to Burkitt's lymphoma cells expressing CD44v3-v10, a heparan-sulfate (HS) modified variant isoform, promoting MET phosphorylation, and that treatment with heparinase decreased activation of MET, suggesting this signaling cascade is dependent on HS-modified CD44 (109). However, other studies have uncovered mechanisms of HGF-induced MET activation independent of HS-modified CD44 isoforms. For example, it has been demonstrated that CD44v6, a non-HS-modified isoform, is required for the activation of MET receptor in some cell lines and heparinase treatment does not affect HGF-induced MET activation (110). It is described that the extracellular domain with the v6 sequence is exclusively responsible for forming a ternary complex between MET and its ligand, HGF. Through this mechanism, CD44v6 is able to present the HGF to the MET receptor, facilitating its activation (110). Moreover, the requirement of CD44 for HGF-mediated MET receptor activation is not transversal and that some cell lines which do not express CD44 have their HGF-MET signaling cascade unaffected (111). MET and CD44v6 engage in a positive feed-back loop, in that MET can regulate CD44 alternative splicing promoting expression of CD44v6, which in turn is capable of interacting with MET and modulating its activation (112). Upregulation of CD44 and increased activation of MET receptors has also been observed in tumor tissue samples and they seem to be correlated with poor prognosis for the patient, mainly due

to more metastatic behavior of the tumor (113). Similar to MET, RON has been reported to be co-activated by CD44v6 (114). RON receptor is structurally and functionally related to MET and is activated by the HGF homologue macrophage-stimulating protein (MSP) (115). Both RON and MET activation through MSP and HGF, respectively, require CD44v6 to be anchored in the membrane in the human colorectal carcinoma cell lines (116). Furthermore, site directed mutagenesis studies discovered that a 3 amino acid sequence within the v6 region is crucial for this coreceptor function. In humans, this sequence is Arginine-Tryptophan-Histidine, and even small soluble peptides of 5 amino acids containing this sequence were able to compete and successfully block activation of both MET and RON receptors (116). Although it is known that glycosylation can modify CD44 binding affinity to HA, there is still much to be learned about the mechanisms of altered CD44 glycosylation on its function as a co-receptor of RTKs.

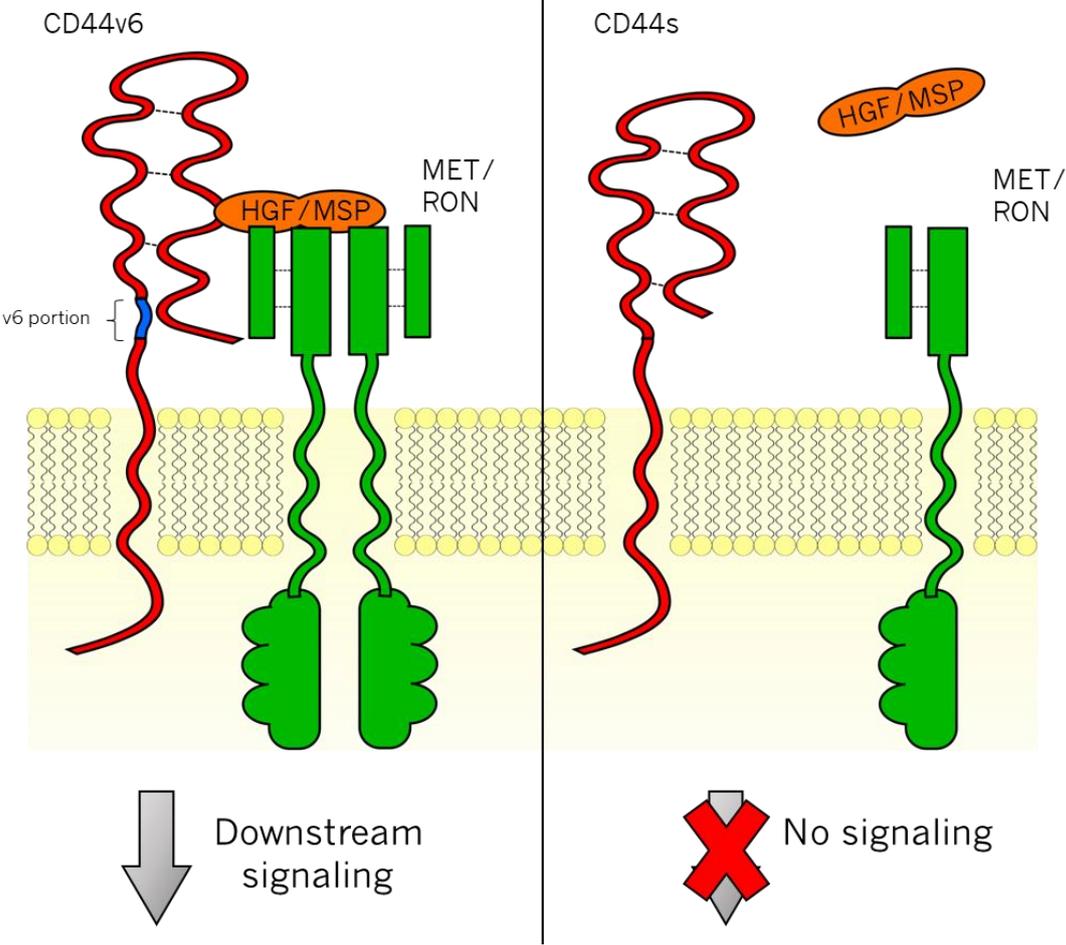


Figure 8: Representation of CD44v6 co-receptor function. When the v6 exon is expressed, CD44 is able to act as a co-receptor for MET/ROK, possibly by ligand presentation, allowing downstream signaling. However, when the v6 is not expressed, CD44 loses this co-receptor function.

1.10 CD44 as a cancer stem cell marker

It is known that cancer cells exhibit a high level of heterogeneity, even among cells within the same tumor mass. For some time, the existence of a subpopulation of cancer cells has been identified that are competent of self-renewal, differentiation, tumor-initiation and that exhibit increased resistance to therapies. These so called cancer stem cells (CSCs) (117) are believed to be responsible for tumor initiation and recurrence (117). Several markers have been identified specific for this subpopulation, for example CD24, Aldehyde Dehydrogenase 1-A1/ALDH1A1 and CD133. CD44 has also been identified as a cancer stem cell marker (118). Cancer stem cell markers are not expressed in all malignancies equally, for example CD44 is expressed in breast, gastric and colon carcinoma but not in skin or liver carcinomas (119). Remarkably, different CD44 isoforms are specific cancer stem cell markers for different tissues. For instance, CD44v6 is a cancer stem cell marker specific for colorectal carcinoma while CD44v8-10 is specific for gastric carcinoma (120, 121). The two major CD44 functions are important in tumor progression and metastatic capability, and this quality makes it a compelling therapeutic target (81).

2. Objectives

This master's thesis aims to test the hypothesis that the altered glycosylation of CD44 is a key process in cancer progression and seeks to further our understanding on the role of altered CD44 glycosylation in cancer. Various genetically engineered gastric cancer cell line models are used to study the impact of altered glycosylation on CD44 related functions, namely as a receptor for hyaluronic acid and as a co-receptor for RON. Additionally, the discovered functions of altered CD44 glycosylation will be validated in colorectal cancer cells likewise.

This thesis has following specific aims:

- (i) Describe CD44 glycoforms in genetically glycoengineered gastric cancer cell lines
- (ii) Unravel the effect of altered glycosylation on the HA-binding affinity of CD44
- (iii) Disclose the effect of altered glycosylation on the co-receptor function of CD44
- (iv) Corroborate the findings in gastric carcinoma tissue samples and colorectal cancer cell lines.

3. Materials and methods

3.1 Cell culture

The cell line MKN45, a gastric carcinoma cell line was obtained from the Japanese Cancer research Bank (Tsukuba, Japan). The MKN45 SimpleCells (SC) were obtained by targeting the *C1GALTC1* (COSMC) gene through zinc-finger nucleases. The MKN45 ST3 and ST6 were stably transfected with the full length human *ST3GAL4* gene and full length human *ST6GALNAC1* gene respectively, and each respective Mock control (M3 and M6) was transfected with the corresponding empty vector pcDNA3.1 (17, 18). The cell line LS174T was obtained from American Type Culture Collection (Virginia, USA), a colorectal cancer cell line reported to express STn and Tn antigens (31). All MKN45 cell lines were cultured in RPMI 1640 GlutaMAX, HEPES supplemented medium (Invitrogen, Waltham, MA) while LS174T was cultured in Gibco® DMEM, GlutaMAX, HEPES supplemented medium (Invitrogen, Waltham, MA). All cell lines were cultured in supplemented with 10% fetal bovine serum (FBS) at 37°C and in a 5% CO₂ atmosphere. Cells transfected with expression vectors had their media also supplemented with 0.5mg/ml G418 (Invitrogen, Waltham, MA). The cell culture medium was replaced between 2 to 3 days.

3.2 Flow cytometry

All cells were washed twice with PBS and detached using Gibco® versene solution (ThermoFisher, Waltham, MA). Cells were incubated with the primary antibody for 1 hour at 4°C and then with the secondary antibody for 30 minutes at 4°C protected from light. Hyaluronan-fluorescein (Merck Millipore, Burlington, MA) was added to the cells and incubated for 15 minutes at 4°C, also protected from light. In between each step, cells were washed with PBS. Finally, cells were labelled with propidium iodide and measured using BD FACSCanto™ II (BD Biosciences, San Jose, CA, USA). All flow cytometry data obtained was assessed through FlowJo v10.0.7.

3.3 Western Blot

Cells were washed twice with PBS and treated with lysis buffer 17 (R&D Systems, McKinley Place, MN) supplemented with 1 mM sodium orthovanadate, 1 mM phenylmethanesulfonyl fluoride (PMSF) and protease inhibitor cocktail (Roche, Basel, Switzerland). Protein concentration of each lysate was measured using DC protein assay kit (BioRad, Hercules, CA). Western blot (WB) was performed using the Mini-PROTEAN® Tetra Cell System (BioRad, Hercules, CA) and polyvinylidene difluoride (PVDF) membranes (GE Healthcare, Chicago, IL). Images were analysed using Image Lab 6.0.1.

3.4 *O*-glycosylation inhibition

The inhibitor used for *O*-glycan elongation inhibition experiments was Benzyl 2-acetamido-2-deoxy- α -D-galactopyranoside (Benzyl- α -GalNAc) (Sigma-Aldrich). All cells were grown in until reaching a 60% to 70% confluency. At that point cells were washed with PBS and treated with the inhibitor in medium without FBS. Finally, cell lysates were collected and run in a western blot. The treatment duration was for 1 and 3 days having obtained similar results.

3.5 Immunoprecipitation

For immunoprecipitation, 700 μ g of protein lysate were incubated with protein G fast flow sepharose beads (GE Healthcare, Little Chalfont, UK) preincubated with the CD44 antibody, overnight at 4°C. Next day, beads were washed with 1xPBS, boiled for 5 min and the samples were loaded in 7,5% polyacrylamide gel for Western blot.

3.6 Immunofluorescence (IF)

Cells were seeded in μ -Chamber 12 well glass slides (IBIDI, Martinsried, Germany) and fixed with 4% paraformaldehyde for 15 minutes for later use. Cells were permeabilized using 0.5% Triton in PBS for 10 minutes at 4°C, followed by blocking with 1:5 goat serum/10% bovine serum albumin in PBS for 30 minutes at room temperature and incubation with the primary antibody overnight at 4°C. Cells were incubated with a secondary antibody conjugated with a fluorophore and DAPI for 1 hour and for 5 minutes respectively, at room temperature. Finally, slides were mounted with vectashield mounting medium (Vector Laboratories, Burlingame, CA).

3.7 *in situ* Proximity ligation assay (PLA)

The *in situ* Proximity Ligation Assay (PLA) was performed using Duolink® PLA Technology (Sigma Aldrich, St.Louis, MO) according to the manufacture instructions. Cell slides were first washed with PBS and blocked according to Duolink® PLA Technology kit. To assess the PLA signal in gastric carcinoma samples, formalin-fixed paraffin embedded gastric carcinoma tissue slides were dewaxed and rehydrated prior to the blocking step. Samples were incubated overnight at 4°C with antibodies for CD44v6 and RON, each conjugated with positive or negative strand oligonucleotides for PLA by Duolink® In Situ Probemaker (Sigma Aldrich, St.Louis, MO). Ligation and amplification were performed according to the Duolink® PLA Technology kit. Finally, slides were incubated with DAPI for 5 minutes at room temperature and mounted with vectashield mounting medium (Vector Laboratories, Burlingame, CA).

3.8 Combined immunofluorescence proximity ligation assay (co-IF-PLA)

The combined immunofluorescence proximity ligation assay (co-IF-PLA) was established and optimized within the work of this thesis (results section). Cell slides washed with PBS and formalin-fixed paraffin embedded gastric carcinoma tissue slides were dewaxed and rehydrated prior to the blocking. Samples were incubated overnight at 4°C with antibodies for CD44v6 and RON, each conjugated with positive or negative strand oligonucleotides for PLA by Duolink® In Situ Probemaker (Sigma Aldrich, St.Louis, MO), and the previously biotinylated B72.3 antibody. Ligation and amplification were performed according to the Duolink® PLA Technology kit. For labelling, slides were incubated with Streptavidin-FITC and DAPI for 1 hour and for 5 min respectively, and mounted using vectashield mounting medium (Vector Laboratories, Burlingame, CA).

3.9 Antibodies and Lectins

Primary antibodies used for the aforementioned methods were as follows: CD44 (156/3C11; Cell Signaling Technology, Danvers, MA), CD44v6 (MA54; Invitrogen, Waltham, MA), RON (C-20; Santa Cruz Biotechnology, Dallas, TX), pRON (Y1238/Y1239; R&D Systems, McKinley Place, MN, USA), sialic acidyl Tn clone B72.3 described in (122), . The two lectins used were *Sambucus nigra* lectin (SNA) and *Aleuria aurantia* lectin (AAL) (Vector Laboratories).

3.10 Microscopy and image analysis

All IF, PLA and co-IF-PLA were visualized under a Zeiss Imager.Z1 Axio fluorescence microscope (Zeiss, Welwyn Garden City, UK). Images were acquired using a Zeiss Axio cam MRm and the AxioVision Rel. 4.8 software. For visualization purposes, PLA signal was enhanced using ImageJ 1.52g through the following established workflow: First, brightness and contrast of the red channel (PLA signal) was adjusted and a duplicate of the channel was generated. To one of the duplicates it was applied a “gaussialic acidn blur” filter and the difference between the two channels was calculated using “image calculator tool”. The brightness and contrast of the result channel was adjusted, and for visualization purposes a “maximum” filter was applied. Finally, all channels were merged and stacked to RGB. The PLA signal was quantified using Duolink® ImageTool (Sigma Aldrich, St.Louis, MO) and ImageJ 1.52g.

4. Results

In this thesis, several glycoengineered gastric cancer cell line models were used to study the impact of aberrant glycosylation on the oncogenic functions of CD44. The SC was conceived through a COSMC knockout, a C1GALT1 dedicated chaperone. The abrogation of this chaperone leads to the expression of a dysfunctional C1GALT1. This in turn results in the expression of Tn and STn structures. This mechanism of altered glycosylation as already been observed in various types of human cancer cell lines, either by hypermethylation, or deletion of this gene. ST6 also expresses STn structures, though through another mechanism. This cell line stably overexpresses the *ST6GALNAC1* gene, which leads to an upregulation of STn structures. This ST3 model stably overexpresses the sialic acidyltransferase ST3GAL4. This enzyme adds sialic acidlic acids in a a2,3 position giving rise to a different shortened O-glycan structure, the disialic acidylated core 2 structure. The aforementioned mechanisms of expressing truncated O-glycans are represented in figure 4.

4.1 Truncated O-glycans as the main aberrant structures in CD44 affects its structure in the glycoengineered cell lines

Western Blot analysis of the cell line models revealed a striking difference in the electrophoretic mobility of CD44 in the glycoengineered cells (Figure 9.a). MKN45 WT, M6 and M3 controls showcased CD44 as a smear which molecular weight ranged between 150 KDa and 250 KDa. This is supported by CD44 high heterogeneity due to glycosylation and alternative splicing (78, 79, 82). However, CD44 in SC, M6 and M3 shifted towards lighter molecular weights. This effect is evident in the SC and ST3 were we observe CD44 to be below 150 KDa, and an increase in the definition of the band. ST6 showcased a milder, but still considerable shift towards lighter molecular weights, specifically slightly below 150 KDa.

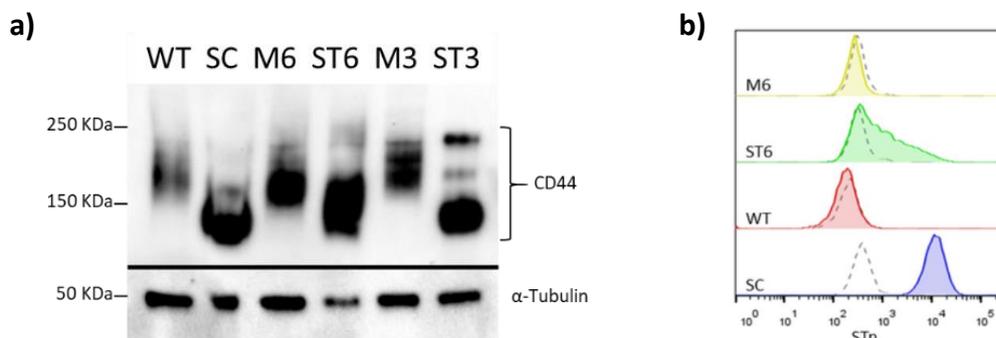


Figure 9: (a) Western blot analysis of CD44 expression in cell lysates. The glycoengineered cell line models (SC, ST6 and M3) express mainly CD44 below 150 KDa, while their respective controls (WT, M6 and ST3) express a smear between 150 and 250 KDa. (b) Flow cytometry analysis of the glycoengineered cell lines SC and ST6 compared to their controls WT and M6 respectively. The negative controls are represented in dotted lines.

Flow cytometry analysis of the SC and ST6 cell lines showed an upregulation of STn structures compared to their respective controls (Figure 9.b). In addition, WT, SC, M6 and ST6 cell lines were treated with Benzyl 2-acetamido-2-deoxy- α -D-galactopyranoside (Benzyl- α -GalNAc), an inhibitor of *O*-glycan elongation (Figure 10.a). WT cells seem to be affected by the inhibitor treatment and increased Benzyl- α -GalNAc concentration caused a shift of the CD44 band towards lighter molecular weights, almost mimicking the SC phenotype (Figure 10.b). However, SC cells did not show any shift in the CD44 band when treated with the inhibitor. This is according to expectation as Benzyl- α -GalNAc inhibits primarily the action of core 1 formation, a pathway that is dysfunctional in SC. Both M6 and ST6 presented slight shifts in the CD44 band when treated with the *O*-glycosylation inhibitor (Figure 10.c). Similar to WT, M6 does not express STn. STn staining was more evident when ST6 cells were treated with the inhibitor with concentration of 5 nM, demonstrating that this concentration was efficient in blocking core 1 formation, further boosting the STn formation. These findings confirm that *O*-glycosylation is highly prevalent in CD44 and contribute significantly to its molecular weight.

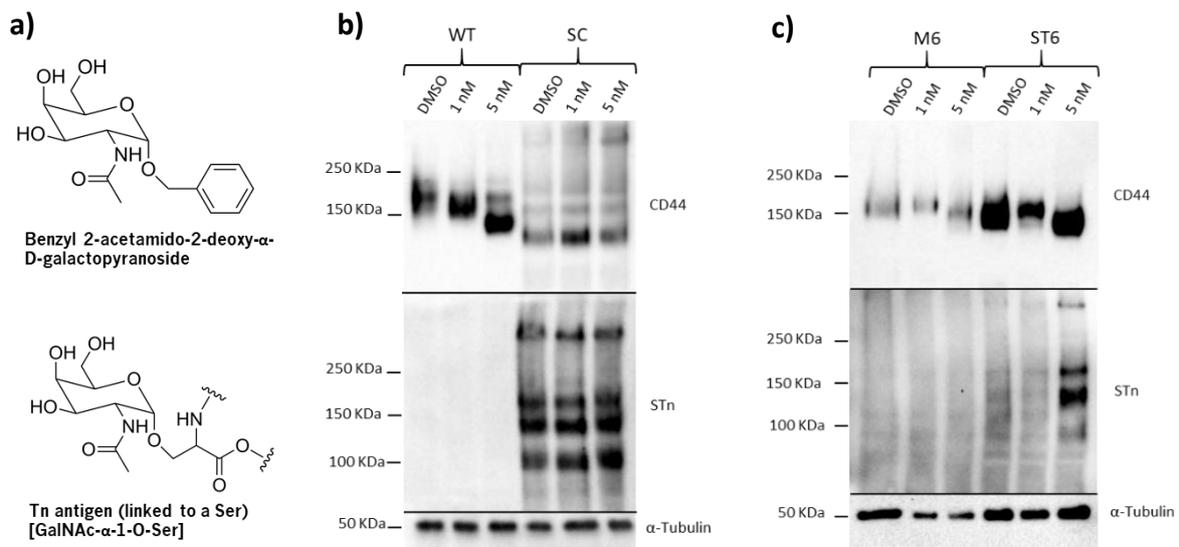


Figure 10: (a) Representation of the molecular structure of the Benzyl- α -GalNAc (Top) and Tn antigen when linked to a Ser residue (Bottom); (b-c) Western blot analysis of cell lysates from (b) WT and SC, (c) M6 and ST6 treated with benzyl- α -GalNAc. Concentrations used for the treatment were 1 nM and 5 nM for 3 days. For the negative control, cells were treated with dimethyl sulfoxide (DMSO), the solvent used to dissolve the inhibitor.

Immunoprecipitation of CD44 was performed on WT and SC lysate samples, and its efficiency was confirmed by western blot (Figure 11). Additionally, immunoprecipitation samples were stained for STn and Tn. As expected, WT does not have STn and Tn expression, seen in both the CD44 IP and the total lysate. On the other hand, SC showcased a strong expression of STn in the CD44 IP sample observed by the band in below the 150 KDa, which seems to match the band in the same molecular weight in the total lysate. The same observations were found when staining for Tn. These findings suggest that not only

CD44 carries a large quantity of *O*-glycosylation sites, but also is one of the major carriers of *O*-glycans in these cell line models.

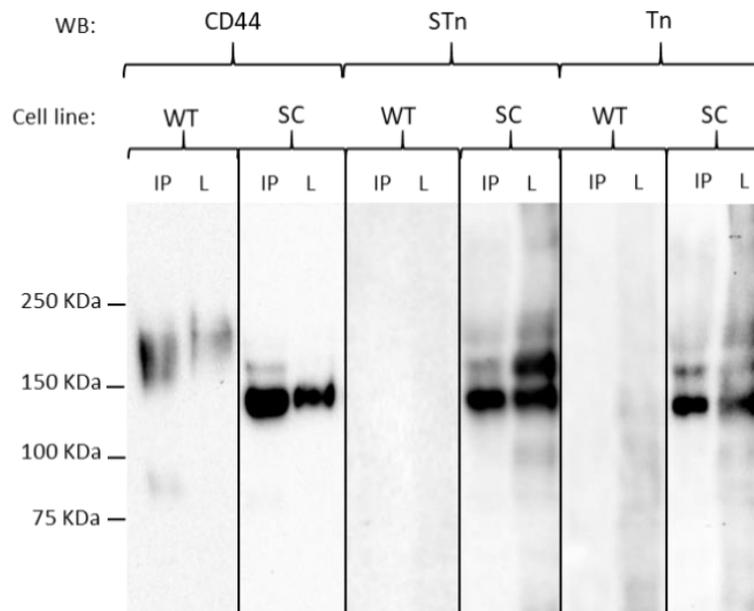


Figure 11: Immunoprecipitation of CD44 in the glycoengineered cell line SC and its control WT. the immunoprecipitated CD44 (IP) of each cell line is represented in the left lane and the total lysate used as an input is represented in the right lane. All samples were blotted for CD44, STn and Tn.

In addition, *Sambuccus nigra* Lectin (SNA) and *Aleuria aurantia* Lectin (AAL) were used to evaluate other possible glycosylation defects in the glycoengineered cell line (Figure 12). Lectins are proteins that have relatively high binding specificity towards different carbohydrate structures. SNA preferentially binds to α 2,6 sialic acidylation linked to a Gal or GalNAc which often found as a terminal glycan structure (123) in both *N* and *O*-glycans, and AAL mainly binds to fucosylation (124). Here, it was observed the SNA staining for ST3 cell line model to decrease relatively to its respective control (Figure 12.a), which expected since ST3GAL4 catalyzes the addition of sialic acidlic acids in a α 2,3 position, leading to a shift from α 2,6 to α 2,3 sialic acidylation (35, 36). Unexpectedly, the SNA staining also decreased in the SC and ST6, even though these cell line models have upregulation of STn structures which have α 2,6 sialic acidylation. Additionally, fucosylation seems also to be decreased in the glycoengineered cell line models when compared to each respective control (Figure 12.b), which could be due to the truncation of *O*-glycans.

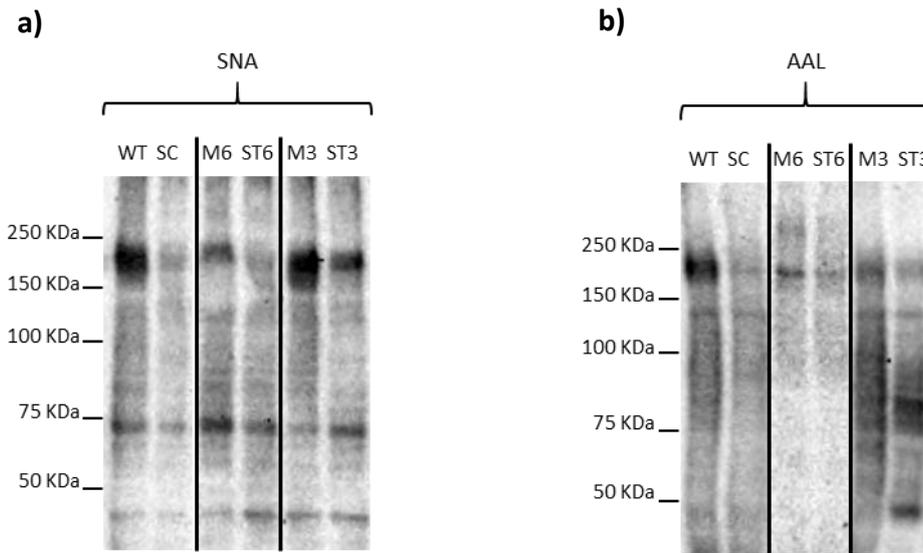


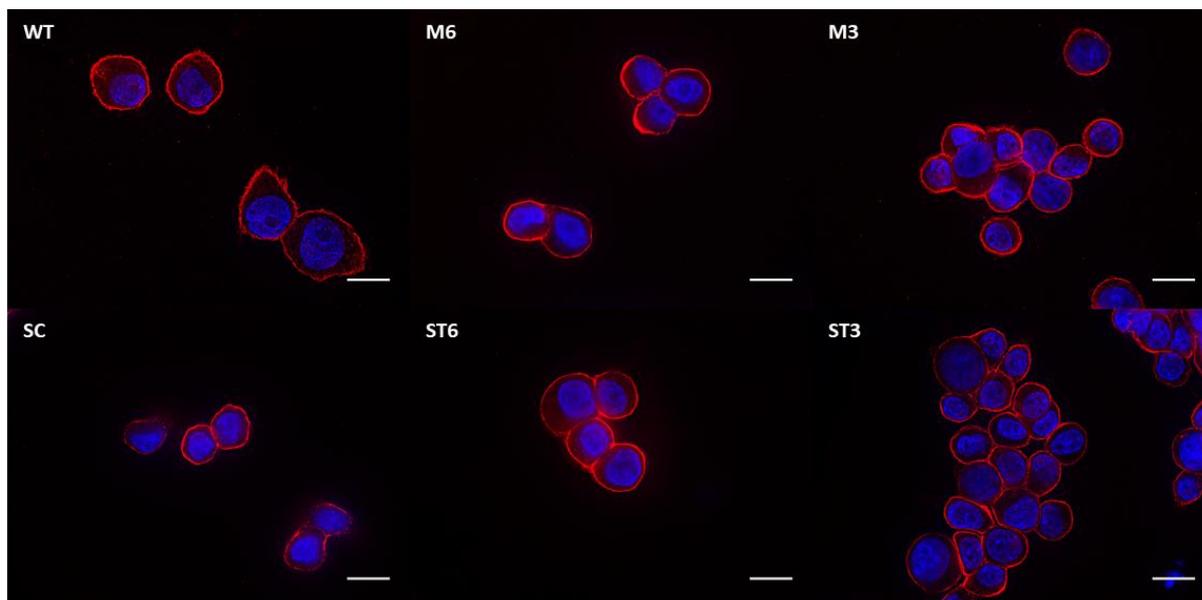
Figure 12: Western blot analysis of cell lysates from glycoengineered cell line models (SC, ST6 and ST3) and their controls (WT, M6 and M3) stained with different lectins. All samples were blotted with (a) *Sambuccus nigra* lectin (SNA), which binds mainly to α 2,6 sialic acidylation linked to a Gal or GalNAc, and (b) *Aleuria aurantia* lectin (AAL) which binds preferably to fucosylation.

4.2 *O*-glycosylation plays a role in the HA-binding affinity of CD44

To assess if altered glycosylation affects CD44 subcellular localization, an immunofluorescence protocol was performed (Figure 13). Both CD44 and CD44v6 seem to be mainly located in the transmembrane region on all cell line models, suggesting that the altered glycosylation does not affect its subcellular localization. In addition, cell line models were analyzed by Fluorescent-activated cell sorting (FACS). It was observed similar levels of CD44 and CD44v6 in the glycoengineered cell lines when compared to its respective control (Figure 14. a;b). However, when assessing the HA binding affinity of CD44 using HA-FITC labeled, it was observed that cell line models with truncated *O*-glycans showcased a striking increase in the HA binding capacity when compared to the WT and mock controls (Figure 14.c).

a)

Immunofluorescence: CD44



b)

Immunofluorescence: CD44v6

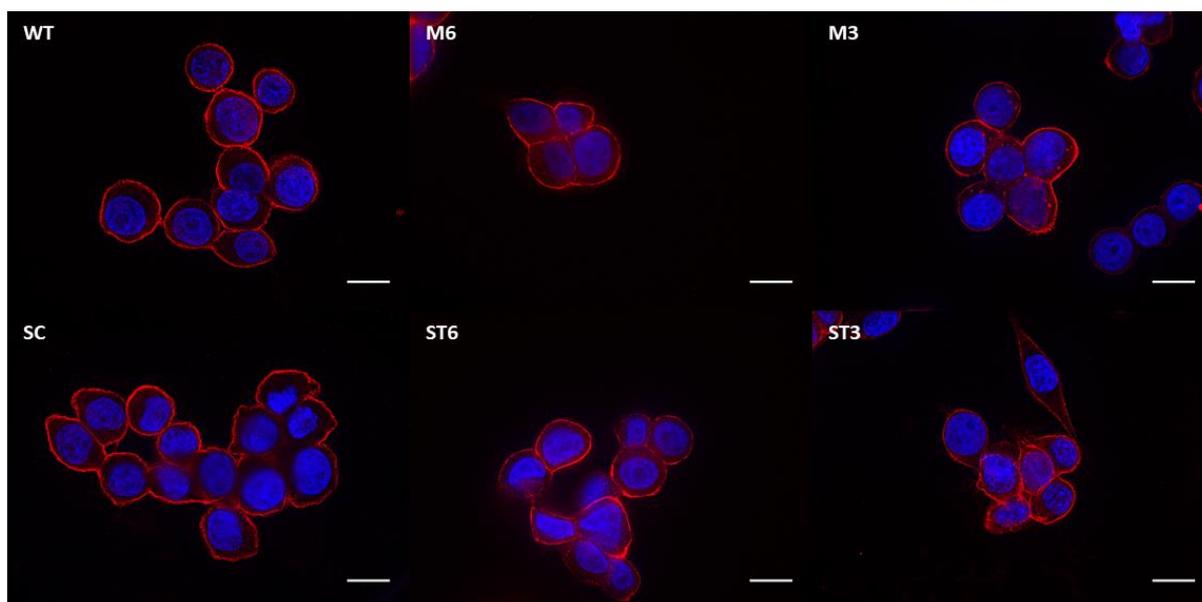


Figure 13: Immunofluorescence images of glycoengineered cells (SC, ST6 and ST3) and their respective controls (WT, M6 and M3) stained for CD44 and CD44v6. Red represents in (a) CD44 and in (b) CD44v6. Nuclear staining with DAPI is represented in blue. Scale bars indicate 15 μm .

One of the mechanisms that has been reported to modulate HA binding affinity of CD44 is proteolytic cleavage (80). CD44 is a common target for various metalloproteinases which can cleave the protein in the extracellular portion (80). To assess CD44 shedding, glycoengineered cell line models were grown in conditioned media for 72h. The medium was then collected, enriched and analyzed for CD44 in a western

blot (Figure 14.d). It was observed an effective CD44 shedding in the glycoengineered cell line models, especially in ST3. A shift of around 20KDa towards lighter molecular weights can be observed in the CD44. This finding is supported by the fact that CD44 is cleaved in the extracellular domain region, releasing the ectodomain and thus losing its transmembrane and cytoplasmic domains (80).

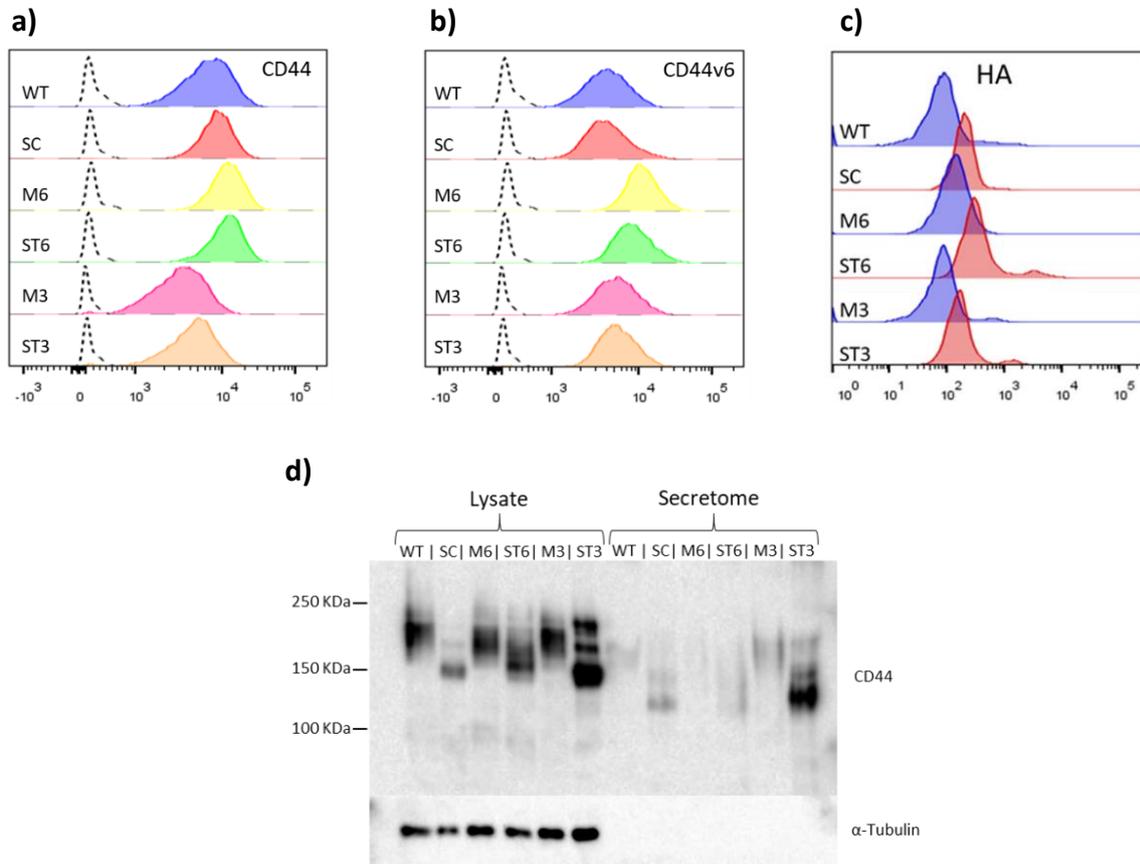


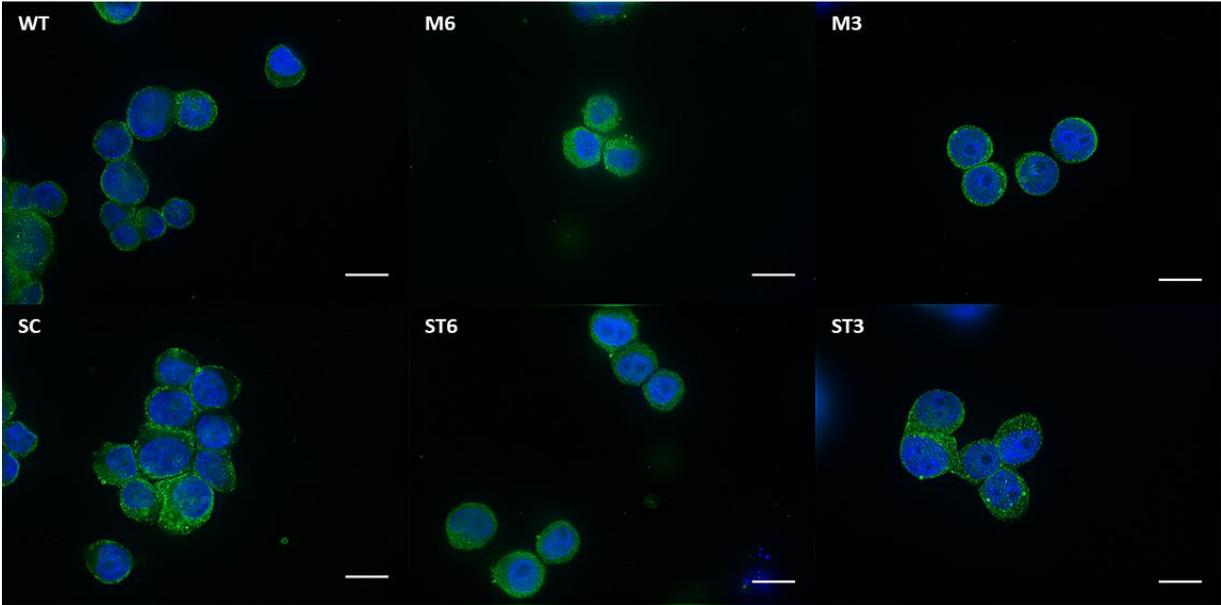
Figure 14: (a-b) Flow cytometry analysis of (a) CD44 and (b) CD44v6 in MKN45 glycoengineered cell lines (SC, ST6 and ST3) as compared to their control cell lines (WT, M6 and M3 respectively). The negative controls are shown in dotted lines. (c) Flow cytometry analysis of HA-Fluorescein (HA-FITC) binding to glycoengineered cell line models (SC, ST6 and ST3) compared to their respective controls (WT, M6 and M3). (d) Comparative western blot analysis of CD44 present in the total cell lysates and in the secretome of glycoengineered cell line models (SC, ST6 and ST3) and controls (WT, M6 and M3).

4.3 Truncated \mathcal{O} glycans enhance RON activation in a CD44v6-dependent manner

Several studies have reported that different CD44 isoforms are able to modulate receptor tyrosine kinase activation. For example, CD44 isoforms containing the variant 6 region are able to modulate activation of RON and MET (110, 116). Here, through immunofluorescence (IF), it was evaluated the phosphorylation status of RON in the glycoengineered cell line models and compared to their respective control (Figure 15). It was revealed that SC, ST6 and ST3 had increased activation of RON when compared to WT, M6 and M3 respectively. In addition, IF images for the phosphorylated RON showed that the signal is mainly

located in the membrane region (Figure 15.a). Interestingly, it was observed that the total RON is abundantly present in the cytoplasm suggesting the possible endocytosis and internalization mechanisms of the receptor (Figure 15.b).

a)
Immunofluorescence: RON



b)
Immunofluorescence: phospho RON

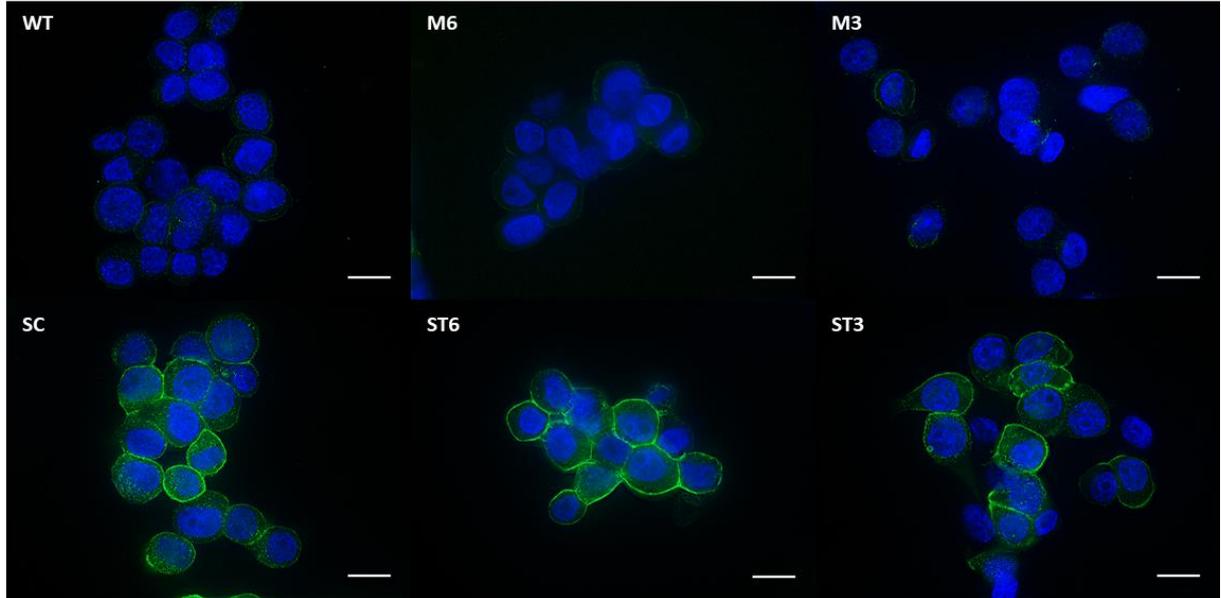


Figure 15: Immunofluorescence images of glycoengineered cells (SC, ST6 and ST3) and their respective controls (WT, M6 and M3) stained for RON and phospho RON (pRON). Green represents in (a) RON and in (b) pRON. Nuclear staining with DAPI is represented in blue. Scale bars indicate 15 μ m.

To further evaluate previous findings, a proximity ligation assay (PLA) was performed. PLA is a widely used technique to assess co-localization between two epitopes using a pair of nucleotide conjugated antibodies called PLA probes. Here, the PLA protocol was applied to assess the co-localization of CD44v6 and the RTK RON in the glycoengineered cell line models (Figure 16). Interestingly, it was observed a striking increase in the number of co-localization events on the cell line models with truncated *O*-glycans. These findings suggest that the increased phosphorylation of RON observed previously in these cell line models could be due to the truncation of *O*-glycans of CD44v6.

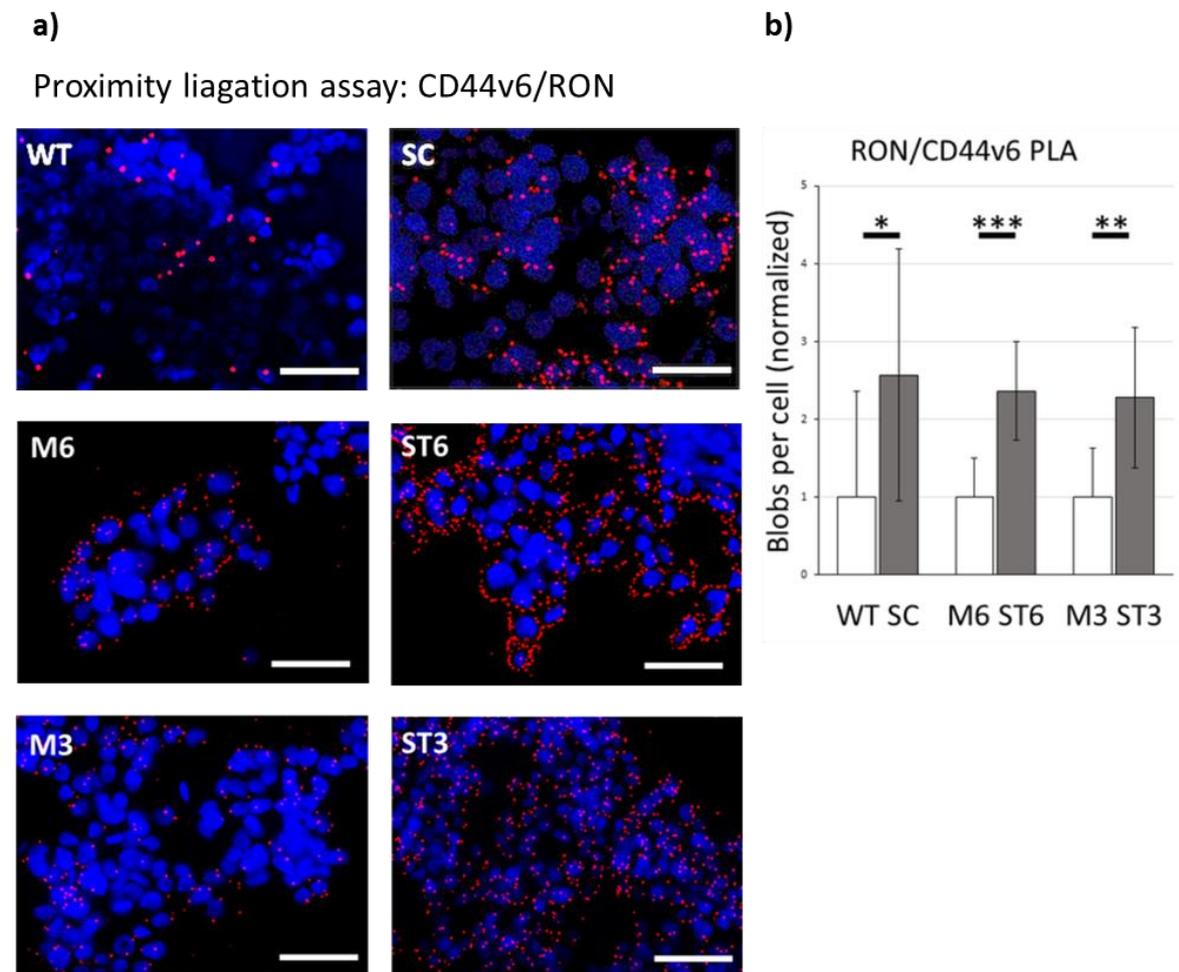


Figure 16: (a) Microscopy images of proximity ligation assay (PLA) of all cell lines. Red signal (blobs) represent colocalization events of the RTK RON and its coreceptor CD44v6 using a antibody for each epitope. Nuclear staining with DAPI is represented in blue. Scale bar indicates 50 μm . (b) Quantification of RON/CD44v6 PLA blobs per cell. Values have been normalized to the respective control cell lines (WT, M6 and M3) and are represented as average \pm SD. Statistical significance was determined by Student's t-test (P-value * < 0.05; ** < 0.01; *** < 0.001).

Indeed, it is widely known CD44v6 to be overexpressed in gastric cancer. In addition, it has been reported that gastric carcinoma tissue can express a large amount of STn structures compared to the healthy adjacent mucosa that expresses none. To evaluate *in situ* the previous *in vitro* findings, an IF-PLA combinatory approach was developed (co-PLA-IF). The efficiency of the developed protocol was first tested

in MKN45 WT and SC cell line models (Figure 17.a). Co-PLA-IF was then applied to various gastric carcinoma tissue patients. As expected, normal tissue adjacent to tumors was negative for STn as well as for co-localization events of CD44v6/RON (Figure 17.b). Healthy gastric glands have been described to neither express CD44v6 nor STn structures (125, 126). Intestinal metaplasia, a premalignant lesion, presented a considerable number of co-localization events for CD44v6/RON and STn structures which are present in secretory vesicles of goblet cells (Figure 17.c). Interestingly, when comparing gastric carcinoma areas that are STn negative and STn positive, an increase in CD44v6/RON co-localization events was observed in the STn positive areas (Figure 17. d;e).

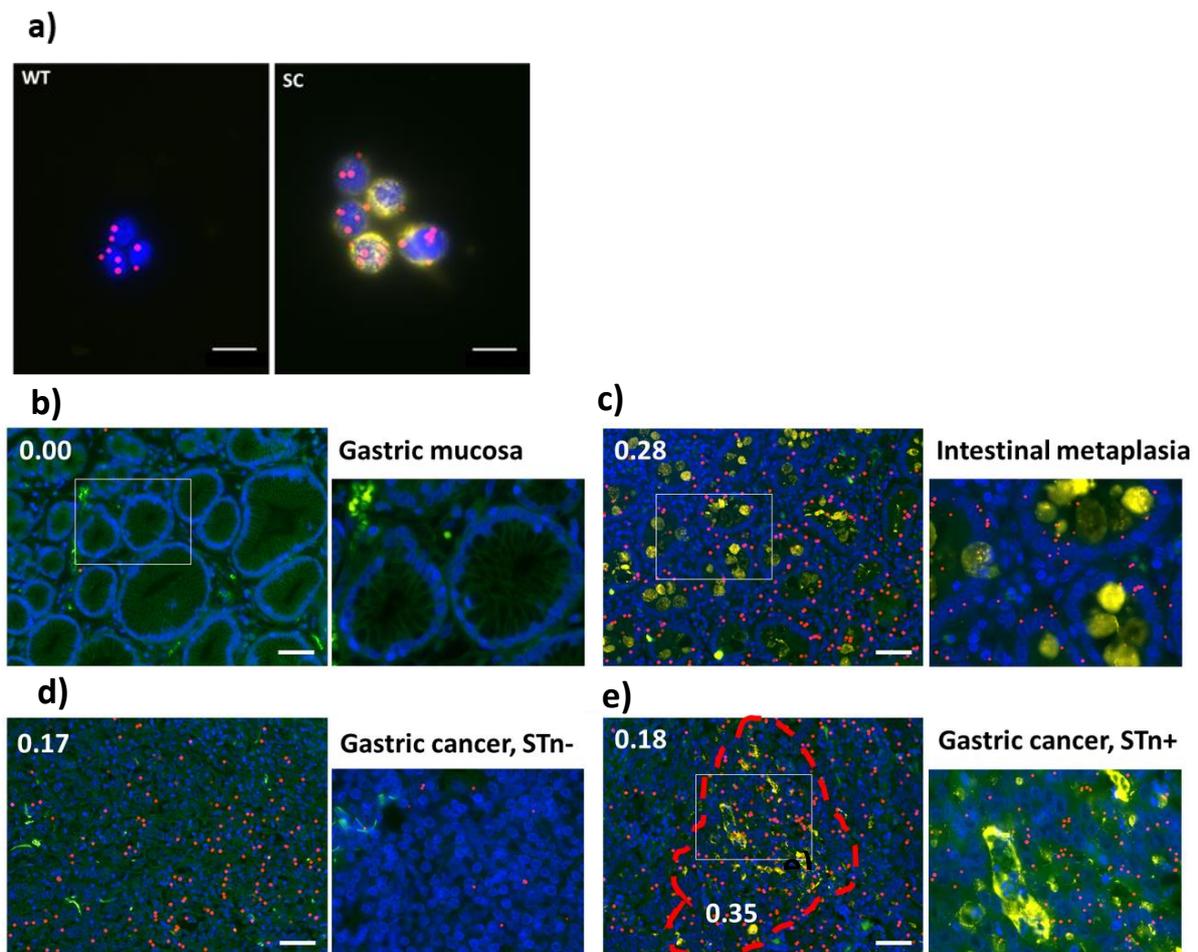


Figure 17: Combinatory proximity ligation assay immunofluorescence (Co-PLA-IF) images of WT and SC cell lines. Red signal (blobs) indicate colocalization events between RON and CD44v6. In yellow is represented STn staining. Nuclear staining with DAPI is represented in blue. Scale bar indicates 15 μ m. The number of PLA blobs per cell were quantified and are shown as numerical values within the image (b-e) Co-PLA-IF images of human gastric tissue sections in different areas. Red signal (blobs) indicate colocalization events between RON and CD44v6. In yellow is represented STn staining. Nuclear staining with DAPI is represented in blue. Autofluorescence is represented in green to visualize tissue architecture. Scale bar indicates 15 μ m. (b) Gastric mucosa adjacent to the gastric carcinoma areas. No colocalization events of RON/CD44v6 and STn staining was observed in these areas. (c) Intestinal metaplasia areas. It is visible the STn expression within the vacuoles of goblet cells which are full of highly O-glycosylated proteins, namely mucins. In addition, there is also a considerable amount of RON/CD44v6 colocalization events. (d) Gastric carcinoma area with no visible STn expression. Aside from the absence of

STn expression, there is also limited RON/CD44v6 colocalization events in these areas. (e) Gastric carcinoma area with STn expression. The area framed in dashed lines is considered to be STn positive and has around 2-fold increase of RON/CD44v6 colocalization events when compared to the adjacent STn negative area.

4.4 Characterization of CD44 in the colorectal cancer cell line LS174T

To corroborate these findings in colorectal cancer cells, the LS174T cell line was characterized (Figure.18). This colorectal cancer cell line has been previously reported to express CD44, STn and Tn structures (31, 127). Here, this cell line was characterized by Western blot and compared to the MKN45 WT and SC cell line models (Figure 18). Through western blot analysis, it was confirmed that LS174T cell line expressed CD44, as well as isoforms containing the variant 6 exon (Figure 18.a). This cell line has the characteristic of growing in big clusters, easily forming spheroid structures and it is possible to observe through immunofluorescence that CD44 and CD44v6 are mainly expressed in the periphery of these spheroid structures (Figure 18.d). SLe^x is also abundantly expressed in this cell line model. Western blot analysis shows that this structure is mainly present in high molecular weight proteins, being above 150KDa and even 250KDa (Figure 18.a). Interestingly, SLe^x coincides both in molecular weight and cellular localization with CD44 and CD44v6. STn structures were also detected in the LS174T cells, however not as abundant as in the MKN45 SC (Figure 18.b). Analyzing the immunofluorescence images, not all cells are positive for the STn, but those cells that express STn, carry it at the cellular membrane (Figure 18.d). Surprisingly LS174T also expresses a considerable amount of RON, however it is not endogenously activated as seen in the MKN45 cell line (Figure 18.c). The IF signal of RON in the LS174T seems to be cytoplasmic, suggesting that the internalization of the receptor also occurs in this cell line (Figure 18.d).

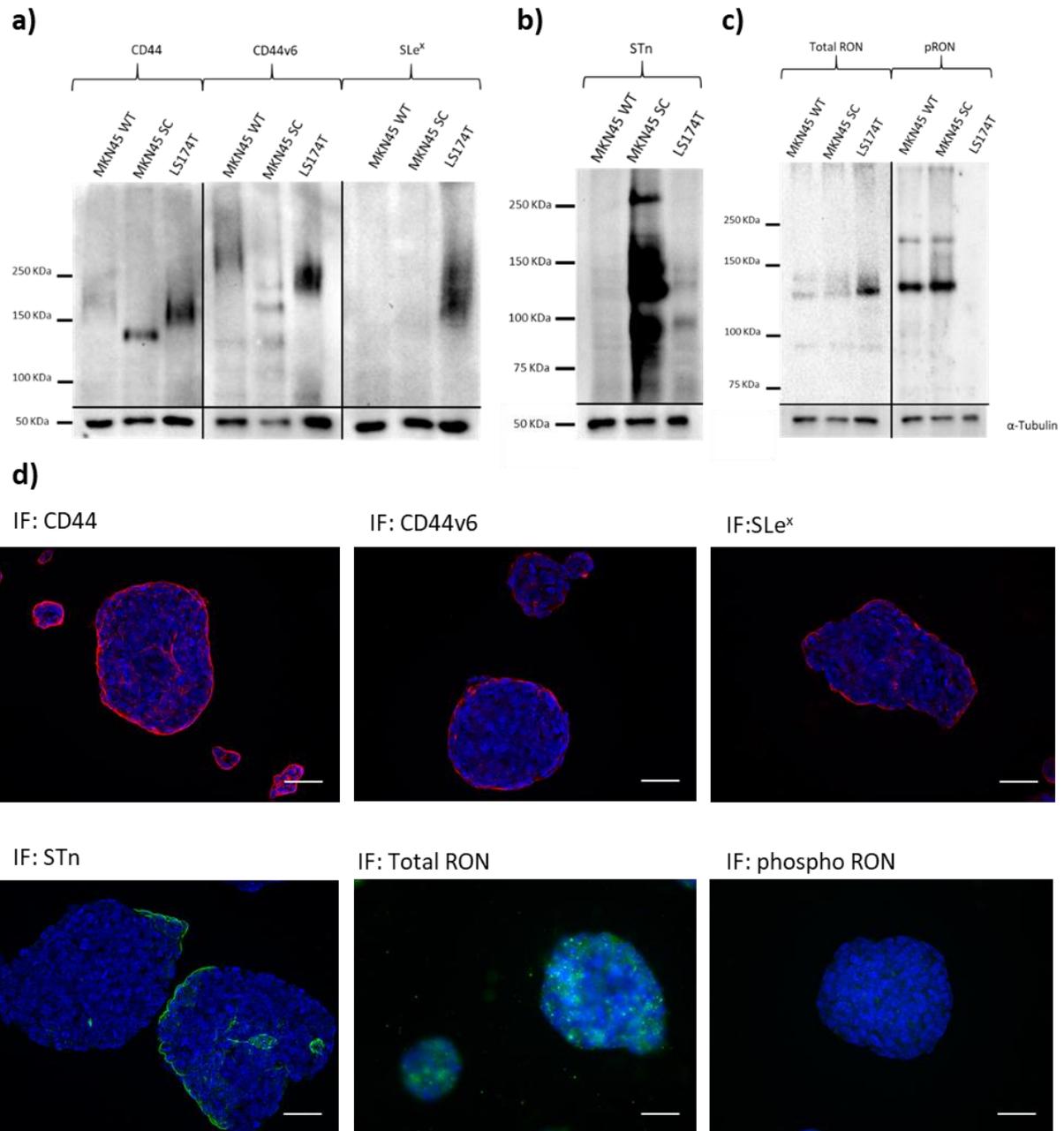


Figure 18: (a-c) Comparative western blot analysis the gastric cancer MKN45 cell lines (WT and SC) and the colorectal cancer LS174T cell line. All samples were blotted for (a) CD44, CD44v6, SLe^x; (b) STn and (c) Total RON and pRON. (d) Immunofluorescence images of colorectal cancer cell line LS174T, stained for CD44, CD44v6, SLe^x, STn, RON and pRON. Nuclear staining with DAPI is represented in blue. Scale bars indicate 15 μm.

In summary, our results showcase the alterations of CD44 molecular features in the glycoengineered cell line models expressing truncated *O*-glycans. These alterations included changes in molecular weight and the increased capacity to bind HA. We also observed that truncation of *O*-glycans led to an increase in CD44v6/RON colocalization and RON activation, which was validated in gastric carcinomas of patients.

5. Discussion

The glycosylation is a well-regulated mechanism that plays a key role in different biological processes crucial for homeostasis. In cancer, the glycosylation machinery is often disturbed leading to the expression of aberrant glycan structures (15). In fact, it has been identified in several studies an upregulation of aberrant *O*-glycan structures, namely Tn and STn in many types of cancer, especially gastric and colorectal cancer (128, 129). These are often referred to as tumor associated antigens, and in the cancer context they decorate several glycoproteins such as mucins (130). Upregulation of truncated *O*-glycans correlates with poor prognosis for the patient, as well as conferring aggressive phenotypes to cancer cells, including increased metastasis and migration (28). Various mechanisms can underpin the expression of truncated *O*-glycans. For instance, it has been shown that alterations in the glycosylation pathway, by overexpressing specific glycosyltransferases such as ST6GALNAC1 and ST3GAL4 or by deletion of crucial chaperones such as COSMC, can lead to the expression of truncated *O*-glycans (17, 28, 36).

Moreover, CD44 and some of its variant isoforms have also been reported to be overexpressed in different types of gastrointestinal cancer promoting oncogenic features such as metastasis, invasiveness, proliferation and drug resistance (131, 132). CD44 is a key component in cell adhesion, acting as a receptor for hyaluronic acid present in the extracellular matrix and serves as a scaffold for different cellular signaling processes being able to act as a co-receptor for different RTKs such as VEGFR, HER2, EGFR, MET and RON (103, 110, 114, 133) Our group has already identified CD44 to be highly present in the serum of gastric cancer patients as well as being heavily modified with truncated *O*-glycans, namely STn structures (74, 76).

In this work, several glycoengineered gastric cancer cell line models were used to study the impact of aberrant glycosylation in the oncogenic related functions of CD44, namely its binding affinity towards HA and its ability to act as a co-receptor for RON. These glycoengineered cell line models were established from the MKN45 parental cell line, a gastric cancer cell line that expresses mainly core 1 and 2 *O*-glycan structures elongated with LacNAc repeats as well as considerable amounts of CD44 and the variant isoform CD44v6 (35, 36, 134). Through different strategies, the glycoengineered cell lines were designed to express shortened *O*-glycan chains.

These different mechanisms to express truncated *O*-glycans showcased similar effects on the CD44 molecular weight in our western blot analysis. Previous studies have already identified several *O*-glycosylation sites present in CD44 (103). In addition, *in silico* analysis predicted a large number of

potential *O*-glycan sites (103). CD44 is described to be a heavily heterogeneous glycoprotein, to the extent of being sometimes addressed as a whole protein family (77). Underlined mechanisms for such are: (i) alternative splicing, (ii) glycosylation and (iii) proteolytic cleavage. Our group recently reported that the expression level of total CD44 and variant isoforms such as CD44v3 and CD44v6 were similar among these cell line models, excluding alternative splicing as one of the reasons for the observed shift (135). In addition, treating samples with chondroitinase and heparinase, enzymes that specifically digest chondroitin and heparan sulfate chains respectively, failed to induce the shift of CD44's electrophoretic mobility. This excludes GAGs as the cause of the molecular weight shift of CD44 in the glycoengineered cell line models. The CD44 shift observed in the ST3 is higher compared to the ST6, even though ST6 expresses shorter *O*-glycan chains. This can be explained by the fact that different glycosyltransferases could have different enzymatic activities. Thus, even though overexpression of ST6GALNAC1 produces shorter *O*-glycans than overexpression of ST3GAL4, the latter could be more efficient in adding sialic acid residues, generating more disialic acidylated core 2 structures than ST6GALNAC1 generates STn structures. It could also be possible that the induced overexpression of ST3GAL4 in the ST3 model is stronger than the overexpression of ST6GALNAC1 in the ST6 model.

The Benzyl- α -GalNAc is an inhibitor of the *O*-glycan elongation pathway. Due to its similar molecular structure to the Tn [GalNAc- α -1-*O*Ser/Thr] (Figure 10.a), this compound when in excess acts as a competitive inhibitor of the C1GALT1 enzyme, not letting it synthesize core 1 structures (136). This leads to accumulation of Tn structures. In our results, the inhibition of the *O*-glycan elongation pathway through Benzyl- α -GalNAc induced a similar shift in CD44 molecular weight of WT and M6 controls as the one observed in the SC and ST6 glycoengineered models. The shift from the treated controls did not completely match that of the untreated glycoengineered cell lines. This ought to be expected and suggests that the inhibition using Benzyl- α -GalNAc was not complete. In addition, we did not observe expression of STn in the WT and M6 controls when we treated them with Benzyl- α -GalNAc. Also, there were no variations in the STn expression of SC when treated with the inhibitor, since the *O*-glycan elongation pathway itself is disturbed in this cell model due to the COSMC knockout. However, the ST6 cell line, which does not have its *O*-glycan elongation pathway directly disturbed, has increased expression of STn when treated with higher concentrations of the inhibitor. The accumulation of Tn structures in this cell line, allows the ST6GALNAC1 enzyme to add sialic acid residues generating STn. Western blot analysis of the immunoprecipitated CD44 from WT and SC lysates further corroborate the heavy *O*-glycosylation profile of CD44.

Taking together the results of the glycoengineered cell lines and the *O*-glycan inhibition experiment, these results show the strong impact of *O*-glycosylation in the CD44 molecular weight. Importantly, the truncation of *O*-glycans not only induces the expression of truncated *O*-glycans but also decreases the expression of other glycan epitopes. This is shown by the decreased of AAL binding (Figure 12.b). The unexpected decrease in the SNA staining for SC and ST6 could be due to other unknown alterations in the glycosylation pathway, such as *N*-glycosylation, or possibly SNA is not able to effectively bind to STn structures in our glycoengineered cell line models for unknown reasons. Certainly, this should be further explored.

Indeed, glycosylation is an important post-translational modification affecting not only to the structure of the protein, but also its function. In several studies, glycosylation has been already described to affect CD44 binding affinity towards HA. Bartolazzi *et al* and English *et al* reported using site directed mutagenesis that several *N*-glycosylated sites on CD44 are important modulators of this binding affinity in different cell line models (96, 97). They also observed a reduced binding affinity towards HA when cells were treated with tunicamycin, an inhibitor of *N*-glycosylation. However, one could argue that this effect could be due to improper folding of CD44 since tunicamycin causes endoplasmic reticulum stress (99, 137). Here, gastric cancer cell line models that displayed truncated *O*-glycans were observed to have an increase in binding capacity towards hyaluronic acid. This was the first time that it was observed in gastric cancer cells the impact of CD44 glycosylation in the HA-binding function, suggesting that alterations in the glycosylation profile of CD44 can potentiate malignant phenotypes, in this case invasion and migration, by modulating its ability to bind to HA. A possible explanation of this observation could be due to an easier access of CD44 “link module” to hyaluronic acid in the glycoengineered cell line models. Faller *et al* described the terminal sialic acid residues in *N*-glycans negatively regulate the HA binding affinity of CD44 due to the HA chain and Sialic acid residues both have negative charge, causing repelling forces between them (101). Considering these observations, we can speculate that this effect also happens in sialic acid residues present in *O*-glycans. For instance, terminal sialic acid residues in elongated *O*-glycan are able to produce stronger repulsion forces than the ones present in truncated *O*-glycans. Another possibility would be the increased clustering of CD44 in the glycoengineered cell line models due to the truncated *O*-glycans, since it was already reported that CD44 clustering regulates HA binding affinity (94, 138). Though this is just speculation, further investigation should be done in order to fully understand the mechanism behind. CD44 is also a target of proteolytic cleavage by different metalloproteinases, releasing the extracellular domain in the stem region (80, 93). Here, it was observed that glycoengineered cell lines effectively shed CD44. Interestingly, the CD44 shedding also has been

observed to modulate the hyaluronic acid binding affinity, which can also explain the increase binding affinity (139). Besides, these results also exclude proteolytic cleavage in as a cause to the CD44 mass shift observed previously in the glycoengineered cell line models.

Several studies tried to elucidate the role of CD44 in potentiating malignant phenotypes. Indeed, different isoforms have been observed to be overexpressed in cancer, for example CD44v3 in head and neck cancer (140), CD44v9 in bile duct cancer (141) and CD44v6 in gastric and colorectal cancer (85, 142, 143). Variant isoform CD44v6 has been a very prominent subject of study to understand not only its biological importance in gastric cancer but also its potential as therapeutic target and biomarker. This variant isoform acts as a co-receptor for different RTKs, for example RON, which is also observed to be upregulated and hyperactivated in gastric cancer promoting proliferation and invasive growth (35, 144). Here it was evaluated the activation status of RON in these cell line models and observed the receptor to be hyperactivated in the models displaying truncated *O*-glycans. This phosphorylation status of RON as already been observed previously in the ST3 cell line model (35), but not in the SC or ST6. Furthermore, PLA results showed that RON and CD44v6 have increased co-localization events in the glycoengineered cell line models. This result suggest *O*-glycan truncation of CD44v6 is one of the key factors in promoting the phosphorylation status of RON and possibly potentiating malignant phenotypes. These findings were evaluated in gastric cancer tissue samples from patients using a combinatory approach of the PLA and immunofluorescence techniques, co-PLA-IF. It was observed an increase in the co-localization between CD44v6 and RON in STn positive tumor areas when compared to negative areas. Indeed, co-localization of CD44v6/RON is also observed in intestinal metaplasialic acidareas, which is a lesion of the gastric mucosa that possibly can lead to malignancy. This could suggest that RON activation dependent on CD44v6 is already present in premalignant lesions, however another key player is probably needed for the progression towards malignant stages of the gastric mucosa to occur. These observations prove the hypothesis that *O*-glycan chains are important structures in the cancer setting, either by promoting oncogenic functions of different proteins or through another mechanism yet to be elucidated. Moreover, further studies need to be performed in order to fully understand mechanism behind how truncated *O*-glycans increase the co-receptor function of CD44.

To evaluate previous finding in colorectal cancer, LS174T cell line was tested for the expression using different techniques. It was confirmed that this cell line expresses both CD44 and CD44v6, STn, SLe^x, and RON. The expression of STn in the LS174T cell line is significantly less than in the MKN45 SC. This could be explained by the fact that LS174T is reported to have mixed subpopulations, one that displays both STn and Tn antigens, and another that expresses extended *O*-glycans (31). Mechanisms that underlie

COSMC silencing in the LS174T cell line are deletions and insertions that cause shifts in the open-reading frame generating a premature stop codon (31). The SLe^x expression in the LS174T was already reported by Hanley, et al (127), and these results support the fact that LS174T expresses a SLe^x-rich CD44 glycoform, the HCELL (127, 145). This particular glycoform has been described to possess E-Selectin ligand activity, often used for the leukocyte rolling and extravasation. This mechanism could be hijacked during malignant transformation in order to promote metastatic spread *in vivo* (127, 146, 147). In addition, It was observed that the LS174T cell line does not possess endogenous activation of RON, contrary to the MKN45 (148). This suggests that the RON activation mechanism is not transversal, and it depends in the cellular context.

6. Conclusions and future perspectives

The aim of this master thesis was to address the effects of glycosylation, mainly *O*-glycans, in CD44 molecular features in the gastrointestinal cancer context.

In this work, we demonstrated that (i) *O*-glycans heavily impact CD44 molecular weight, (ii) truncation of *O*-glycans significantly impacts CD44 ability to bind to HA, in addition to affecting CD44 shedding, (iii) truncation of *O*-glycans significantly modulates interactions between CD44v6 and RON, possibly leading to an increase activation of the receptor (iv) and using an innovative combinatory approach of immunofluorescence-proximity ligation assay techniques we were able to measure the truncation of *O*-glycans and CD44v6/RON interactions simultaneously in patients clinical samples, corroborating our *in vitro* results. Although very preliminary, our results with the LS174T cell line showcase the high potential of applying this cell line models to corroborate our findings in colorectal cancer.

In conclusion, this work provides new insights on the impact of *O*-glycans as modulators of CD44 functions in the cancer setting, in addition to giving new perspectives on malignancy mechanism.

Furthermore, studies should be performed in order to fully elucidate the exact mechanism behind how *O*-glycans modulate CD44v6/RON interactions and the CD44 binding affinity towards HA. In addition, we would like to continue studying the effect of truncated *O*-glycans on CD44 using the LS174T cell line model and observe if our results apply also in the colorectal cancer context.

7. References

1. Varki A KS. Historical Background and Overview. *Essentials of Glycobiology* 3rd edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.
2. Rademacher TW, Parekh RB, Dwek RA. GLYCOBIOLOGY. *Annual Review of Biochemistry*. 1988;57(1):785-838.
3. Mereiter S, Balmaña M, Gomes J, Magalhães A, Reis CA. Glycomic Approaches for the Discovery of Targets in Gastrointestinal Cancer. *Frontiers in oncology*. 2016;6:55-.
4. Stanley P TN, Aebi M. N-glycans. *Essentials of Glycobiology* [Internet]. 3rd edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.
5. Brockhausen I SP. O-GalNAc Glycans. *Essentials of Glycobiology* [Internet]. 3rd edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
6. Lindahl U CJ, Kimata K, et al. Proteoglycans and Sulfated Glycosaminoglycans. *Essentials of Glycobiology* [Internet]. 3rd edition ed. Cold Spring Harbor (NY) Cold Spring Harbor Laboratory Press; 2015-2017.
7. Bieberich E. Synthesis, Processing, and Function of N-glycans in N-glycoproteins. *Adv Neurobiol*. 2014;9:47-70.
8. Wopereis S, Lefeber DJ, Morava É, Wevers RA. Mechanisms in Protein O-Glycan Biosynthesis and Clinical and Molecular Aspects of Protein O-Glycan Biosynthesis Defects: A Review. *Clinical Chemistry*. 2006;52(4):574.
9. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KTBG, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*. 2013;32(10):1478-88.
10. Eric Paul Bennett HH, and Henrik Clausen. cDNA Cloning and Expression of a Novel HumanUDP-N-acetyl-a-D-galactosamine. *The Journal of Biological Chemistry*. 1996;271(29):6.
11. Varki A. Sialic acidlic acids in human health and disease. *Trends Mol Med*. 2008;14(8):351-60.
12. Li Y, Chen X. Sialic acidlic acid metabolism and sialic acidlyltransferases: natural functions and applications. *Appl Microbiol Biotechnol*. 2012;94(4):887-905.
13. Schneider F, Kemmner W, Haensch W, Franke G, Gretschel S, Karsten U, et al. Overexpression of Sialic acidlyltransferase CMP-Sialic acidlic Acid:Gal β 1,3GalNAc-R α 6-Sialic acidlyltransferase Is Related to Poor Patient Survival in Human Colorectal Carcinomas. *Cancer Research*. 2001;61(11):4605-11.
14. Reis CA, Osorio H, Silva L, Gomes C, David L. Alterations in glycosylation as biomarkers for cancer detection. *Journal of Clinical Pathology*. 2010;63(4):322.
15. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*. 2015;15:540.
16. Marcos BE, Gomes J, Magalhaes A, Gomes C, David L, Dar I, Jeanneau C, DeFrees S, Krustrup D, Vogel LK, Kure EH, Burchell J, Taylor-Papadimitriou J, Clausen H, Mandel U, Reis CA. ST6GalNAc-I controls expression of sialic acidlyl-Tn antigen in gastrointestinal tissues. *Front Biosci (Elite Ed)*. 2011;1(3):1443.
17. Marcos NT, Pinho S, Grandela C, Cruz A, Samyn-Petit B, Harduin-Lepers A, et al. Role of the Human ST6GalNAc-I and ST6GalNAc-II in the Synthesis of the Cancer-Associated Sialic acidlyl-Tn Antigen. *Cancer Research*. 2004;64(19):7050.
18. Carvalho AS, Harduin-Lepers A, Magalhães A, Machado E, Mendes N, Costa LT, et al. Differential expression of α -2,3-sialic acidlyltransferases and α -1,3/4-fucosyltransferases regulates the levels of sialic acidlyl Lewis a and sialic acidlyl Lewis x in gastrointestinal carcinoma cells. *The International Journal of Biochemistry & Cell Biology*. 2010;42(1):80-9.

19. Kudelka MR, Ju T, Heimbürg-Molinaro J, Cummings RD. Chapter Three - Simple Sugars to Complex Disease—Mucin-Type O-Glycans in Cancer. In: Drake RR, Ball LE, editors. *Advances in Cancer Research*. 126: Academic Press; 2015. p. 53-135.
20. McEver RP. Selectins: initiators of leucocyte adhesion and signalling at the vascular wall. *Cardiovasc Res*. 2015;107(3):331-9.
21. Zhang B, Chen H, Yao X, Cong W, Wu M-C. E-selectin and its ligand-sLeX in the metastasis of hepatocellular carcinoma. *2000*. 534-6 p.
22. Nakamori S, Kameyama M, Imaoka S, Furukawa H, Ishikawa O, Sasaki Y, et al. Increased expression of Sialic acidyl Lewis x antigen correlates with poor survival in patients with colorectal carcinoma: Clinicopathological and Immunohistochemical Study. *1993*. 3632-7 p.
23. Trinchera M, Aronica A, Dall'Olio F. Selectin Ligands Sialic acidyl-Lewis a and Sialic acidyl-Lewis x in Gastrointestinal Cancers. *Biology (Basel)*. 2017;6(1):16.
24. Colley KJ VA, Kinoshita T. Cellular Organization of Glycosylation. *Essentials of Glycobiology* [Internet]. 3rd edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.
25. Sewell R, Bäckström M, Dalziel M, Gschmeissner S, Karlsson H, Noll T, et al. The ST6GalNAc-I Sialic acidyltransferase Localizes throughout the Golgi and Is Responsible for the Synthesis of the Tumor-associated Sialic acidyl-Tn O-Glycan in Human Breast Cancer. *Journal of Biological Chemistry*. 2006;281(6):3586-94.
26. Tu L, Banfield DK. Localization of Golgi-resident glycosyltransferases. *Cellular and Molecular Life Sciences*. 2010;67(1):29-41.
27. Narimatsu Y, Ikehara Y, Iwasaki H, Nonomura C, Sato T, Nakanishi H, et al. Immunocytochemical analysis for intracellular dynamics of C1GalT associated with molecular chaperone, Cosmc. *Biochemical and Biophysical Research Communications*. 2008;366(1):199-205.
28. Fu C, Zhao H, Wang Y, Cai H, Xiao Y, Zeng Y, et al. Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen. *HLA*. 2016;88(6):275-86.
29. Itzkowitz S, Kjeldsen T, Frieri A, Hakomori S-I, Yang U-S, Kim YS. Expression of Tn, sialic acidlosyl Tn, and T antigens in human pancreas. *Gastroenterology*. 1991;100(6):1691-700.
30. Pinho S, Marcos NT, Ferreira B, Carvalho AS, Oliveira MJ, Santos-Silva F, et al. Biological significance of cancer-associated sialic acidyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. *Cancer Letters*. 2007;249(2):157-70.
31. Ju T, Lanneau GS, Gautam T, Wang Y, Xia B, Stowell SR, et al. Human Tumor Antigens Tn and Sialic acidyl Tn Arise from Mutations in Cosmc. *Cancer Research*. 2008;68(6):1636.
32. Ju T, Cummings RD. A unique molecular chaperone Cosmc required for activity of the mammalian core 1 β 3-galactosyltransferase. *Proceedings of the National Academy of Sciences*. 2002;99(26):16613-8.
33. Lin M-C, Chien P-H, Wu H-Y, Chen S-T, Juan H-F, Lou P-J, et al. C1GALT1 predicts poor prognosis and is a potential therapeutic target in head and neck cancer. *Oncogene*. 2018;37(43):5780-93.
34. Radhakrishnan P, Dabelsteen S, Madsen FB, Francavilla C, Kopp KL, Steentoft C, et al. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(39):E4066-E75.
35. Mereiter S, Magalhães A, Adamczyk B, Jin C, Almeida A, Drici L, et al. Glycomic analysis of gastric carcinoma cells discloses glycans as modulators of RON receptor tyrosine kinase activation in cancer. *2015*.
36. Mereiter S, Magalhães A, Adamczyk B, Jin C, Almeida A, Drici L, et al. Glycomic and sialic acidloproteomic data of gastric carcinoma cells overexpressing ST3GAL4. *Data Brief*. 2016;7:814-33.
37. Yamada S, Sugahara K, Özbek S. Evolution of glycosaminoglycans: Comparative biochemical study. *Communicative & Integrative Biology*. 2011;4(2):150-8.

38. Hascall V EJ. Hyaluronan. *Essentials of Glycobiology* [Internet]. 16th edition ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.
39. Necas J, Bartosikova L, Brauner P, Kolář J. Hyaluronic acid (Hyaluronan): A review 2008.
40. Price RD, Berry MG, Navsaria HA. Hyaluronic acid: the scientific and clinical evidence. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2007;60(10):1110-9.
41. Papakonstantinou E, Roth M, Karakiulakis G. Hyaluronic acid: A key molecule in skin aging. *Dermatoendocrinol*. 2012;4(3):253-8.
42. Day AJ, Prestwich GD. Hyaluronan-binding Proteins: Tying Up the Giant. *Journal of Biological Chemistry*. 2002;277(7):4585-8.
43. Wu R-L, Sedlmeier G, Kyjacova L, Schmaus A, Philipp J, Thiele W, et al. Hyaluronic acid-CD44 interactions promote BMP4/7-dependent Id1/3 expression in melanoma cells. *Scientific Reports*. 2018;8(1):14913.
44. Pedron S, Hanselman JS, Schroeder MA, Sarkaria JN, Harley BAC. Extracellular Hyaluronic Acid Influences the Efficacy of EGFR Tyrosine Kinase Inhibitors in a Biomaterial Model of Glioblastoma. *Adv Healthc Mater*. 2017;6(21):10.1002/adhm.201700529.
45. Bharadwaj AG, Kovar JL, Loughman E, Elowsky C, Oakley GG, Simpson MA. Spontaneous metastasis of prostate cancer is promoted by excess hyaluronan synthesis and processing. *The American journal of pathology*. 2009;174(3):1027-36.
46. Li Y, Li L, Brown TJ, Heldin P. Silencing of hyaluronan synthase 2 suppresses the malignant phenotype of invasive breast cancer cells. *International Journal of Cancer*. 2007;120(12):2557-67.
47. Elahi-Gedwillo KY, Carlson M, Zettervall J, Provenzano PP. Antifibrotic Therapy Disrupts Stromal Barriers and Modulates the Immune Landscape in Pancreatic Ductal Adenocarcinoma. *Cancer Research*. 2019;79(2):372.
48. Setälä LP, Tammi MI, Tammi RH, Eskelinen MJ, Lipponen PK, Ågren UM, et al. Hyaluronan expression in gastric cancer cells is associated with local and nodal spread and reduced survival rate. *British Journal Of Cancer*. 1999;79:1133.
49. Knudson W, Biswas C, Toole BP. Interactions between human tumor cells and fibroblasts stimulate hyaluronate synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. 1984;81(21):6767-71.
50. Torre Lindsey A, Bray F, Siegel Rebecca L, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*. 2015;65(2):87-108.
51. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6):394-424.
52. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *European Journal of Cancer*. 2013;49(6):1374-403.
53. Orman S, Cayci HM. Gastric cancer: factors affecting survival. *Acta Chirurgica Belgica*. 2019;119(1):24-30.
54. Moghimi-Dehkordi B, Safaee A. An overview of colorectal cancer survival rates and prognosis in Asialic acid. *World J Gastrointest Oncol*. 2012;4(4):71-5.
55. Griffin-Sobel JP. Gastrointestinal Cancers: Screening and Early Detection. *Seminars in Oncology Nursing*. 2017;33(2):165-71.
56. Singh S, Jha HC. Status of Epstein-Barr Virus Coinfection with *Helicobacter pylori* in Gastric Cancer. *Journal of Oncology*. 2017;2017:3456264.
57. Correa P. Human Gastric Carcinogenesis: A Multistep and Multifactorial Process—First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Research*. 1992;52(24):6735.

58. Tsukamoto T, Nakagawa M, Kiriya Y, Toyoda T, Cao X. Prevention of Gastric Cancer: Eradication of *Helicobacter pylori* and Beyond. *International Journal of Molecular Sciences*. 2017;18(8):1699.
59. Correa P, Piazuelo MB, Wilson KT. Pathology of gastric intestinal metaplasia: clinical implications. *The American journal of gastroenterology*. 2010;105(3):493-8.
60. Sung JK. Diagnosis and management of gastric dysplasia. *The Korean journal of internal medicine*. 2016;31(2):201-9.
61. Laurén P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. *Acta Pathologica Microbiologica Scandinavica*. 1965;64(1):31-49.
62. Demirkan NÇ, Tunçyürek M, Ertan EU, Alkanat MB, İçöz I. Correlation of histological classifications of gastric carcinomas with location and prognosis. *Gastroenterol Clin Biol*. 2002;1220(6):5.
63. Qiu M-z, Cai M-y, Zhang D-s, Wang Z-q, Wang D-s, Li Y-h, et al. Clinicopathological characteristics and prognostic analysis of Lauren classification in gastric adenocarcinoma in China. *J Transl Med*. 2013;11:58-.
64. Guglielmi A, de Manzoni G, Romezzoli A, Ricci F, Pelosi G, Laterza E, et al. Prognostic value of histologic classifications of advanced stomach cancer: comparative study of Lauren's and Goseki's classifications. *Chir Ital*. 1997(49):4.
65. Sohn BH, Hwang J-E, Jang H-J, Lee H-S, Oh SC, Shim J-J, et al. Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2017;10.1158/078-0432.CCR-16-2211.
66. Marley AR, Nan H. Epidemiology of colorectal cancer. *Int J Mol Epidemiol Genet*. 2016;7(3):105-14.
67. Freeman H-J. Colorectal cancer risk in Crohn's disease. *World J Gastroenterol*. 2008;14(12):1810-1.
68. Bojuwoye MO, Olokoba AB, Ogunlaja OA, Agodirin SO, Ibrahim OK, Okonkwo KC, et al. Familial adenomatous polyposis syndrome with colorectal cancer in two Nigerians: a report of two cases and review of literature. *Pan Afr Med J*. 2018;30:6-.
69. Plawski A, Banasiewicz T, Borun P, Kubaszewski L, Krokowicz P, Skrzypczak-Zielinska M, et al. Familial adenomatous polyposis of the colon. *Hered Cancer Clin Pract*. 2013;11(1):15-.
70. Granados-Romero J, Valderrama-Treviño A, Contreras Flores E, Barrera-Mera B, Herrera M, Uriarte-Ruiz K, et al. Colorectal cancer: a review 2017. 4667 p.
71. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med*. 2015;21(11):1350-6.
72. Fontana E, Eason K, Cervantes A, Salazar R, Sadanandam A. Context matters-consensus molecular subtypes of colorectal cancer as biomarkers for clinical trials. *Ann Oncol*. 2019;30(4):520-7.
73. Strathmann FG, Hoofnagle AN. Current and Future Applications of Mass Spectrometry to the Clinical Laboratory. *American Journal of Clinical Pathology*. 2011;136(4):609-16.
74. Oliveira FMSd, Mereiter S, Lönn P, Sialic acid B, Shen Q, Heldin J, et al. Detection of post-translational modifications using solid-phase proximity ligation assay. *New Biotechnology*. 2017.
75. Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gústafsdóttir SM, et al. Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*. 2002;20:473.
76. Campos D, Freitas D, Gomes J, Magalhães A, Steentoft C, Gomes C, et al. Probing the O-Glycoproteome of Gastric Cancer Cell Lines for Biomarker Discovery. *Molecular & Cellular Proteomics : MCP*. 2015;14(6):1616-29.
77. Ponta H, Sherman L, Herrlich PA. CD44: From adhesion molecules to signalling regulators. *Nature Reviews Molecular Cell Biology*. 2003;4:33.

78. Marhaba R, Zöller M. CD44 in Cancer Progression: Adhesion, Migration and Growth Regulation. *Journal of Molecular Histology*. 2004;35(3):211-31.
79. Prochazka L, Tesarik R, Turanek J. Regulation of alternative splicing of CD44 in cancer. *Cellular Signalling*. 2014;26(10):2234-9.
80. Stamenkovic I, Yu Q. Shedding Light on Proteolytic Cleavage of CD44: The Responsible Sheddase and Functional Significance of Shedding. *The Journal of investigative dermatology*. 2009;129(6):1321-4.
81. Sottnik JL, Theodorescu D. CD44: A metastasis driver and therapeutic target. *Oncoscience*. 2016;3(11-12):320-1.
82. Cheng C, Sharp PA. Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. *Molecular and cellular biology*. 2006;26(1):362-70.
83. Zhu S, Chen Z, Katsha A, Hong J, Belkhiri A, El-Rifai W. Regulation of CD44E by DARPP-32-dependent activation of SRp20 splicing factor in gastric tumorigenesis. *Oncogene*. 2016;35(14):1847-56.
84. Zavros Y. Initiation and Maintenance of Gastric Cancer: A Focus on CD44 Variant Isoforms and Cancer Stem Cells. *Cellular and Molecular Gastroenterology and Hepatology*. 2017;4(1):55-63.
85. Xin Y, Grace A, Gallagher MM, Curran BT, Leader MB, Kay EW. CD44V6 in Gastric Carcinoma: A Marker of Tumor Progression. *Applied Immunohistochemistry & Molecular Morphology*. 2001;9(2).
86. Heider K-H, Dämmrich J, Skroch-Angel P, Müller-Hermelink H-K, Vollmers HP, Herrlich P, et al. Differential Expression of CD44 Splice Variants in Intestinal- and Diffuse-Type Human Gastric Carcinomas and Normal Gastric Mucosa. *Cancer Research*. 1993;53(18):4197.
87. Dang H, Steinway SN, Ding W, Rountree CB. Induction of tumor initiation is dependent on CD44s in c-Met⁺ hepatocellular carcinoma. *BMC Cancer*. 2015;15(1):161.
88. Banerji S, Day AJ, Kahmann JD, Jackson DG. Characterization of a Functional Hyaluronan-Binding Domain from the Human CD44 Molecule Expressed in *Escherichia coli*. *Protein Expression and Purification*. 1998;14(3):371-81.
89. Kellett-Clarke H, Stegmann M, Barclay AN, Metcalfe C. CD44 Binding to Hyaluronic Acid Is Redox Regulated by a Labile Disulfide Bond in the Hyaluronic Acid Binding Site. *PLOS ONE*. 2015;10(9):e0138137.
90. Requirements for hyaluronic acid binding by CD44: a role for the cytoplasmic domain and activation by antibody. *The Journal of Experimental Medicine*. 1992;175(1):257-66.
91. Lesley J, English N, Charles C, Hyman R. The role of the CD44 cytoplasmic and transmembrane domains in constitutive and inducible hyaluronan binding. *European Journal of Immunology*. 2000;30(1):245-53.
92. Wang Y, Yago T, Zhang N, Abdisalaam S, Alexandrakis G, Rodgers W, et al. Cytoskeletal Regulation of CD44 Membrane Organization and Interactions with E-selectin. *The Journal of Biological Chemistry*. 2014;289(51):35159-71.
93. Chetty C, Vanamala SK, Gondi CS, Dinh DH, Gujrati M, Rao JS. MMP-9 Induces CD44 Cleavage and CD44 mediated Cell Migration in Glioblastoma Xenograft Cells. *Cellular signalling*. 2012;24(2):549-59.
94. Regulated clustering of variant CD44 proteins increases their hyaluronate binding capacity. *The Journal of Cell Biology*. 1996;135(4):1139-50.
95. Seiter S TW, Herrmann K, Schadendorf D, Patzelt E, Möller P, Zöller M. Expression of CD44 splice variants in human skin and epidermal tumours. *Virchows Archiv*. 1996;428(3):141-9.
96. Bartolazzi A, Nocks A, Aruffo A, Spring F, Stamenkovic I. Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *The Journal of Cell Biology*. 1996;132(6):1199.
97. English NM, Lesley JF, Hyman R. Site-specific De-N-glycosylation of CD44 Can Activate Hyaluronan Binding, and CD44 Activation States Show Distinct Threshold Densities for Hyaluronan Binding. *Cancer Research*. 1998;58(16):3736.

98. Glycosylation of CD44 negatively regulates its recognition of hyaluronan. *The Journal of Experimental Medicine*. 1995;182(2):419-29.
99. Osowski CM, Urano F. Measuring ER stress and the unfolded protein response using mammalian tissue culture system. *Methods in enzymology*. 2011;490:71-92.
100. Guvench O. Revealing the Mechanisms of Protein Disorder and N-Glycosylation in CD44-Hyaluronan Binding Using Molecular Simulation. *Frontiers in Immunology*. 2015;6:305.
101. Faller CE, Guvench O. Terminal sialic acidic acids on CD44 N-glycans can block hyaluronan binding by forming competing intramolecular contacts with arginine sidechains. *Proteins*. 2014;82(11):3079-89.
102. Girbl T, Hinterseer E, Grössinger EM, Aslhaber D, Oberascher K, Weiss L, et al. CD40-Mediated Activation of Chronic Lymphocytic Leukemia Cells Promotes Their CD44-Dependent Adhesion to Hyaluronan and Restricts CCL21-Induced Motility. *Cancer Research*. 2013;73(2):561.
103. Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL, et al. CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. *Clin Proteomics*. 2018;15(1):22.
104. Dasgupta A, Takahashi K, Cutler M, Tanabe KK. O-linked Glycosylation Modifies CD44 Adhesion to Hyaluronate in Colon Carcinoma Cells. *Biochemical and Biophysical Research Communications*. 1996;227(1):110-7.
105. Rodgers AK, Nair A, Binkley PA, Tekmal R, Schenken RS. Inhibition of CD44 N- and O-linked Glycosylation Decreases Endometrial Cell Lines Attachment to Peritoneal Mesothelial Cells. *Fertility and sterility*. 2011;95(2):823-5.
106. Paul MD, Hristova K. The RTK Interactome: Overview and Perspective on RTK Heterointeractions. *Chemical Reviews*. 2019;119(9):5881-921.
107. Orian-Rousseau V, Sleeman J. Chapter Nine - CD44 is a Multidomain Signaling Platform that Integrates Extracellular Matrix Cues with Growth Factor and Cytokine Signals. In: Simpson MA, Heldin P, editors. *Advances in Cancer Research*. 123: Academic Press; 2014. p. 231-54.
108. Organ SL, Tsao M-S. An overview of the c-MET signaling pathway. *Therapeutic Advances in Medical Oncology*. 2011;3(1 Suppl):S7-S19.
109. van der Voort R, Taher TEI, Wielenga VJM, Spaargaren M, Prevo R, Smit L, et al. Heparan Sulfate-modified CD44 Promotes Hepatocyte Growth Factor/Scatter Factor-induced Signal Transduction through the Receptor Tyrosine Kinase c-Met. *Journal of Biological Chemistry*. 1999;274(10):6499-506.
110. Orian-Rousseau V, Chen L, Sleeman JP, Herrlich P, Ponta H. CD44 is required for two consecutive steps in HGF/c-Met signaling. *Genes & Development*. 2002;16(23):3074-86.
111. Dortet L, Veiga E, Bonazzi M, Cossart P. CD44-independent activation of the Met signaling pathway by HGF and In1B2010. 919-27 p.
112. Cheng C, Yaffe MB, Sharp PA. A positive feedback loop couples Ras activation and CD44 alternative splicing. *Genes & Development*. 2006;20(13):1715-20.
113. Elliott VA, Rychahou P, Zaytseva YY, Evers BM. Activation of c-Met and Upregulation of CD44 Expression Are Associated with the Metastatic Phenotype in the Colorectal Cancer Liver Metastasis Model. *PLoS ONE*. 2014;9(5):e97432.
114. Orian-Rousseau V. CD44 Acts as a Signaling Platform Controlling Tumor Progression and Metastasis. *Frontiers in Immunology*. 2015;6:154.
115. Ronsin C, Muscatelli F, G Mattei M, Breathnach R. A novel putative receptor protein tyrosine kinase of the MET family1993. 1195-202 p.
116. Matzke A, Herrlich P, Ponta H, Orian-Rousseau V. A Five-Amino-Acid Peptide Blocks Met- and Ron-Dependent Cell Migration. *Cancer Research*. 2005;65(14):6105.
117. Yu Z, Pestell TG, Lisanti MP, Pestell RG. Cancer stem cells. *The international journal of biochemistry & cell biology*. 2012;44(12):2144-51.

118. Takaishi S, Okumura T, Tu S, Wang SSW, Shibata W, Vigneshwaran R, et al. Identification of gastric cancer stem cells using the cell surface marker CD44. *Stem cells* (Dayton, Ohio). 2009;27(5):1006-20.
119. Najafi M, Farhood B, Mortezaee K. Cancer stem cells (CSCs) in cancer progression and therapy. *Journal of Cellular Physiology*. 2019;234(6):8381-95.
120. Todaro M, Gaggianesi M, Catalano V, Benfante A, Iovino F, Biffoni M, et al. CD44v6 Is a Marker of Constitutive and Reprogrammed Cancer Stem Cells Driving Colon Cancer Metastasis. *Cell Stem Cell*. 2014;14(3):342-56.
121. Lau WM, Teng E, Chong HS, Lopez KAP, Tay AYL, Salto-Tellez M, et al. CD44v8-10 Is a Cancer-Specific Marker for Gastric Cancer Stem Cells. *Cancer Research*. 2014;74(9):2630.
122. Thor A, Ohuchi N, Szpak CA, Johnston WW, Schlom J. Distribution of Oncofetal Antigen Tumor-associated Glycoprotein-72 Defined by Monoclonal Antibody B72.3. *Cancer Research*. 1986;46(6):3118.
123. Shibuya N, Goldstein IJ, Broekaert WF, Nsimba-Lubaki M, Peeters B, Peumans WJ. The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *Journal of Biological Chemistry*. 1987;262(4):1596-601.
124. Kochibe N, Furukawa K. Purification and properties of a novel fucose-specific hemagglutinin of *Aleuria aurantia*. *Biochemistry*. 1980;19(13):2841-6.
125. David L, Nesland JM, Clausen H, Carneiro F, Sobrinho-Simões M. Simple Mucin-type carbohydrate antigens (Tn, sialic acidlosyl-Tn and T) in gastric mucosa, carcinomas and metastases. *APMIS Supplementum*. 1992;27:162-72.
126. Huang L, Xu A-M, Chen B. Expression of CD44v6 in gastric cancer and its significance. *Int J Clin Exp Pathol* 2017;10(2):1970-6.
127. Hanley WD, Burdick MM, Konstantopoulos K, Sackstein R. CD44 on LS174T Colon Carcinoma Cells Possesses E-Selectin Ligand Activity. *Cancer Research*. 2005;65(13):5812-7.
128. T Marcos N, Bennett E, Gomes J, Magalhaes A, Gomes C, David L, et al. ST6GalNAc-I controls expression of sialic acidlyl-Tn antigen in gastrointestinal tissues2011. 1443-55 p.
129. Zhao Q, Zhan T, Deng Z, Li Q, Liu Y, Yang S, et al. Glycan analysis of colorectal cancer samples reveals stage-dependent changes in CEA glycosylation patterns. *Clin Proteomics*. 2018;15:9-.
130. Conze T, Carvalho AS, Landegren U, Almeida R, Reis CA, David L, et al. MUC2 mucin is a major carrier of the cancer-associated sialic acidlyl-Tn antigen in intestinal metaplasialic acidand gastric carcinomas. *Glycobiology*. 2009;20(2):199-206.
131. Senbanjo LT, Chellaiah MA. CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. *Front Cell Dev Biol*. 2017;5:18-.
132. Cain JW, Hauptschein RS, Stewart JK, Bagci T, Sahagian GG, Jay DG. Identification of CD44 as a surface biomarker for drug resistance by surface proteome signature technology. *Mol Cancer Res*. 2011;9(5):637-47.
133. Matzke-Ogi A, Jannasch K, Shatirishvili M, Fuchs B, Chiblak S, Morton J, et al. Inhibition of Tumor Growth and Metastasis in Pancreatic Cancer Models by Interference With CD44v6 Signaling. *Gastroenterology*. 2016;150(2):513-25.e10.
134. da Cunha CB, Oliveira C, Wen X, Gomes B, Sousa S, Suriano G, et al. De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Laboratory Investigation*. 2010;90:1604.
135. Mereiter S, Martins ÁM, Gomes C, Balmaña M, Macedo JA, Polom K, et al. O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer. *FEBS Letters*. 2019;593(13):1675-89.
136. Delacour D, Gouyer V, Leteurtre E, Ait-Slimane T, Drobecq H, Lenoir C, et al. 1-Benzyl-2-acetamido-2-deoxy- α -D-galactopyranoside Blocks the Apical Biosynthetic Pathway in Polarized HT-29 Cells. *Journal of Biological Chemistry*. 2003;278(39):37799-809.

137. Wu J, Chen S, Liu H, Zhang Z, Ni Z, Chen J, et al. Tunicamycin specifically aggravates ER stress and overcomes chemoresistance in multidrug-resistant gastric cancer cells by inhibiting N-glycosylation. *J Exp Clin Cancer Res.* 2018;37(1):272-.
138. Yang C, Cao M, Liu H, He Y, Xu J, Du Y, et al. The high and low molecular weight forms of hyaluronan have distinct effects on CD44 clustering. *The Journal of biological chemistry.* 2012;287(51):43094-107.
139. Okamoto I, Kawano Y, Tsuiki H, Sasaki J-i, Nakao M, Matsumoto M, et al. CD44 cleavage induced by a membrane-associated metalloprotease plays a critical role in tumor cell migration. *Oncogene.* 1999;18(7):1435-46.
140. Reategui EP, Mayolo AAd, Das PM, Astor FC, Singal R, Hamilton KL, et al. Characterization of CD44v3 containing isoforms in head and neck cancer. *Cancer Biology & Therapy.* 2006;5(9):1163-8.
141. Suwannakul N, Ma N, Thanan R, Pinlaor S, Ungarreevittaya P, Midorikawa K, et al. Overexpression of CD44 Variant 9: A Novel Cancer Stem Cell Marker in Human Cholangiocarcinoma in Relation to Inflammation. *Mediators Inflamm.* 2018;2018:4867234.
142. Lu L, Huang F, Zhao Z, Li C, Liu T, Li W, et al. CD44v6: A metastasis-associated biomarker in patients with gastric cancer?: A comprehensive meta-analysis with heterogeneity analysis. *Medicine (Baltimore).* 2016;95(50):e5603-e.
143. Ma L, Dong L, Chang P. CD44v6 engages in colorectal cancer progression. *Cell Death & Disease.* 2019;10(1):30.
144. Zhou D, Huang L, Zhou Y, Wei T, Yang L, Li C. RON and RONΔ160 promote gastric cancer cell proliferation, migration, and adaption to hypoxia via interaction with β-catenin. *Aging (Albany NY).* 2019;11(9):2735-48.
145. Trinchera M, Malagolini N, Chiricolo M, Santini D, Minni F, Caretti A, et al. The biosynthesis of the selectin-ligand sialic acidyl Lewis x in colorectal cancer tissues is regulated by fucosyltransferase VI and can be inhibited by an RNA interference-based approach. *The International Journal of Biochemistry & Cell Biology.* 2011;43(1):130-9.
146. Jacobs PP, Sackstein R. CD44 and HCELL: preventing hematogenous metastasis at step 1. *FEBS letters.* 2011;585(20):3148-58.
147. Strell C, Entschladen F. Extravasation of leukocytes in comparison to tumor cells. *Cell Commun Signal.* 2008;6:10-.
148. Benvenuti S, Lazzari L, Arnesano A, Li Chiavi G, Gentile A, Comoglio PM. Ron Kinase Transphosphorylation Sustains MET Oncogene Addiction. *Cancer Research.* 2011;71(5):1945.

Annex

O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer

Stefan Mereiter^{1,2,†} , Álvaro M. Martins^{1,2}, Catarina Gomes^{1,2}, Meritxell Balmaña^{1,2} , Joana A. Macedo^{1,2,‡} , Karol Polom^{3,4}, Franco Roviello⁴, Ana Magalhães^{1,2}  and Celso A. Reis^{1,2,5,6} 

- 1 I3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
- 2 IPATIMUP – Institute of Molecular Pathology and Immunology, University of Porto, Portugal
- 3 Department of Surgical Oncology, Medical University of Gdansk, Poland
- 4 General Surgery and Surgical Oncology Department, University of Siena, Italy
- 5 Faculty of Medicine, University of Porto, Portugal
- 6 Instituto de Ciências Biomédicas Abel Salazar, University of Porto, Portugal

Correspondence

C. A. Reis, I3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal
Tel: +351 22 040 88 00 (ext. 6068)
E-mail: celsor@ipatimup.pt

Present addresses

[†]IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, VBC – Vienna BioCenter Campus, Vienna, Austria
[‡]3B's Research Group, I3Bs – Research Institute on Biomaterials, Biodegradables and Biomimetics, University of Minho, Braga, Portugal

(Received 4 February 2019, revised 4 May 2019, accepted 6 May 2019)

doi:10.1002/1873-3468.13432

Edited by Maria Papatriantafyllou

CD44 isoforms are often upregulated in gastric cancer and have been associated with increased metastatic potential and poor survival. To evaluate the functional impact of O-glycan truncation on CD44 we have analysed glyco-engineered cancer cell models displaying shortened O-glycans. Here, we demonstrate that induction of aberrant O-glycan termination through various molecular mechanisms affects CD44 molecular features. We show that CD44 is a major carrier of truncated O-glycans and that this truncation is accompanied by an increased hyaluronan binding capacity and affects extracellular shedding. In addition, short O-glycans promoted the colocalization of CD44v6 with the receptor tyrosine kinase RON and concomitantly increased activation. Our *in vitro* findings were validated in gastric cancer clinical samples.

Keywords: CD44; gastric cancer; glycosylation; hyaluronic acid; proximity ligation assay; sialylation

CD44 is an exceptionally versatile glycoprotein and consequently often described as a whole-protein family encoded by a single gene [1,2]. The molecular weight of CD44 largely varies among human cells, and ranges from 80 kDa to more than 250 kDa [3,4]. This heterogeneity has been mainly attributed to three mechanisms, namely (a) alternative splicing at the RNA level [2,5,6], (b) extensive glycosylation of the protein backbone

[7–10] and (c) the possible proteolytic cleavage through metalloproteases at the cell surface [4,11,12].

The process of alternative splicing leads to the expression of variable exon products in the extracellular, membrane proximal region. Due to the nine variable exons (v), numbered from v2 to v10, and through a complex underlying regulatory machinery that enables countless combinations, hundreds of

Abbreviations

CS, chondroitin sulfate; HA, hyaluronan; HS, heparan sulfate; IP, immunoprecipitation; M3, mock-transfected MKN45 cells and control of ST3; M6, mock-transfected MKN45 cells and control of ST6; PLA, proximity ligation assay; RTK, receptor tyrosine kinase; SC, MKN45 SimpleCells; ST3, ST3GAL4-transfected MKN45 cells; ST6, ST6GALNAC1-transfected MKN45 cells; STn, sialyl Thomsen-nouvelle antigen; T, Thomsen-Friedenreich antigen; Tn, Thomsen-nouvelle antigen; WB, western blot; WT, wild-type MKN45 cells.

functionally distinct splice variants of CD44 exist [2,13]. A single cell can express several CD44 isoforms simultaneously and many intra- and extracellular cues have been shown to dictate this splicing signature [5,14–16], making CD44 a sentinel for cellular and environmental alterations.

Functionally, CD44 is involved in cellular adhesion and signal transduction through several mechanisms: (a) CD44 is the main receptor of the extracellular matrix component hyaluronan (HA) [17,18]; (b) it may carry heparan sulfate (HS) and chondroitin sulfate (CS) chains, which are docking sites for various growth factors and proteases [8,19,20]; (c) it is a coreceptor to cell surface receptor tyrosine kinases (RTKs), such as MET, RON and HER2 [16,21–23]; and (d) it can be a potent ligand for selectins in the extravasation process [10]. However, these biological functions can vary significantly among the various CD44 isoforms. For instance, the efficiency of CD44 to bind HA is controlled by both the alternatively spliced variable (v) chains and the glycosylation status [7,24]. Furthermore, the single heparan sulfate chain of CD44 is linked to the variant chain v3, whereas the peptide sequence that acts as coreceptor of MET and RON is encoded in the v6 variable exon [21–23]. In gastric cancer, CD44 is commonly upregulated and the CD44v6 isoform is frequently expressed *de novo* [15,25,26]. CD44 has been implicated in cancer stem cell properties, and cancer cells that overexpress CD44 display an increased metastatic potential and are associated with poor survival of the patients [15,27]. Particularly the expression of the v6 variant form of CD44 has been associated with distant metastasis, poor prognosis, but also susceptibility to chemotherapy in gastric cancer [15,28,29]. Functionally, CD44v6 has been shown to play a crucial role in activation of RTKs, such as MET and RON, and thereby to contribute to oncogenic features [21,23]. Thus, CD44, in general, and CD44v6, in particular, have been shown to contribute to various cancer hallmarks through the regulation of cellular adhesion and RTK activation [16].

CD44 and its variable isoforms are extensively glycosylated and may carry N-glycans, O-glycans and glycosaminoglycans [8,9,30]. In fact, a major part of the molecular mass of CD44 is derived from these numerous glycan modifications [8,9,31]. Alterations in the glycosylation machinery is a key event in carcinogenesis and cancer progression [32]. In gastric cancer, the abnormal biosynthesis of O-glycans is a frequent event leading to the expression of short, truncated glycan epitopes such as T, Tn or sialyl Tn (STn) [33–37]. Recently, we demonstrated that CD44v6 carries the cancer-associated glycan epitope STn in gastric cancer

[38] and that gastric cancer patients present elevated serum levels of CD44v6 with STn when compared to healthy individuals [39]. Yet, despite the indications that glycosylation may have a strong modulatory effect on CD44, potentially driving malignancy, the functional impact of O-glycan truncation on CD44 has not been addressed in cancer.

In the present work, we use a panel of genetically engineered gastric cancer cell models to demonstrate that alterations in the glycosylation machinery have a significant impact on molecular features of CD44, leading to differential molecular weight and affecting the antibody recognition of the protein backbone. Moreover, we show that CD44 is a major carrier of truncated O-glycans driving the CD44-mediated activation of the RTK RON and binding of HA.

Materials and methods

Cell lines

The gastric carcinoma cell line MKN45 was obtained from the Japanese Cancer Research Bank (Tsukuba, Japan) and was stably transfected with the full-length human *ST3GAL4* gene, the full-length human *ST6GALNAC1* gene or the corresponding empty vector pcDNA3.1 (Mock) as previously described [40,41]. The MKN45 SimpleCells (SC) were obtained by targeting the *CIGALTC1* (COSMC) gene by zinc-finger nuclease as previously described [38,42]. All cells were grown in monolayer in uncoated cell culture flasks. Cells were maintained at 37 °C in an atmosphere of 5% CO₂, in RPMI 1640 GlutaMAX, HEPES medium supplemented with 10% FBS. The media of cells transfected with expression vectors were supplemented by 0.5 mg·mL⁻¹ G418 (all from Invitrogen, Waltham, MA, USA). Cell culture medium was replaced every 2–3 days. Cultured cell lines were routinely tested for mycoplasma contamination by PCR amplification for mycoplasma pulmonis UAB CTIP, mycoplasma penetrans HF-2 and mycoplasma synoviae 53.

Primary antibodies

In western blot, immunofluorescence (IF), flow cytometry and proximity ligation assay (PLA) experiments the following primary antibodies have been used: CD44 (156/3C11; Cell Signaling Technology, Danvers, MA, USA), CD44v6 (MA54; Invitrogen), RON (C-20; Santa Cruz Biotechnology, Dallas, TX, USA), pRON (Y1238/Y1239; R&D Systems, McKinley Place, MN, USA), Syndecan 1 (B-A38; Abcam, Cambridge, UK), Heparan Sulfate (10E4; USBiological, Swampscott, MA, USA), Chondroitin Sulfate (A-7; Santa Cruz Biotechnology), STn (clone TKH2 for western blot [43], clone B72.3 for flow cytometry and IF [44]).

Immunoprecipitation and western blotting

Cells were washed twice with PBS and directly collected in lysis buffer 17 (R&D Systems) additionally supplemented with 1 mM sodium orthovanadate, 1 mM phenylmethanesulfonylfluoride and protease inhibitor cocktail (Roche, Basel, Switzerland). Protein concentrations of lysates were determined by DC protein assay (BioRad, Hercules, CA, USA). For immunoprecipitation, protein G fast flow sepharose beads (GE Healthcare, Little Chalfont, UK) were preincubated with the CD44 antibody and 500 µg of protein lysate were applied. Western blotting was performed as previously described [45], using the Mini-PROTEAN® Tetra Cell System (BioRad) and polyvinylidene difluoride membranes (GE Healthcare, Chicago, IL, USA). Densitometry was performed with Image Lab (BioRad), values were normalized to tubulin and represented as relative values compared to the control.

Flow cytometry

Cells were detached using Gibco® versene solution (ThermoFisher, Waltham, MA, USA) and stained with primary antibodies for 1 h at 4 °C. Secondary antibodies or HA-fluorescein (Merck Millipore, Burlington, MA, USA) were incubated for 30 or 15 min respectively. Cells were strained, labelled with propidium iodide and measured using BD FACSCanto™ II (BD Biosciences, San Jose, CA, USA). Data were analysed using FlowJo (BD Biosciences).

Transcriptomics

The transcriptomic analysis was performed as previously described [45]. Briefly, total RNA extracts from cell lysates were isolated with TRI Reagent (Sigma-Aldrich, St. Louis, MO, USA). The mRNAs of over 20 000 primed targets were sequenced using Ion AmpliSeq Transcriptome Human Gene Expression Kit (Life Technologies, Carlsbad, CA, USA). The Ion Chef system was used for templating and the loaded chips were sequenced using the Ion Proton System (both from Life Technologies). Sequencing data were automatically transferred to the dedicated Ion Torrent server to generate sequencing reads. Reads quality and trimming was performed using Torrent Server v4.2 before read alignment with TMAP 4.2 (Life Technologies). The TS plugin CoverageAnalysis v4.2 was used to generate reads count. The sequencing was performed in duplicates and sequence reads were normalized to the total read count.

qRT-PCR

Total RNA extracts from cell lysates were isolated with TRI Reagent (Sigma-Aldrich) and converted into cDNA using the SuperScript® IV Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. Following

primers were used: CD44 for 5'-CCAATGCCTTTGATGG ACC-3', rev 5'-TCTGTCTGTGCTGTCCGGTGAT-3'; CD 44v3 for 5'-CGTCTTCAAATACCATCCAGCA-3', rev 5'-ATCTTCATCATCAATGCCTGA 3' and CD44v6 for 5'-G GCAACTCCTAGTAGTACAACG-3', rev 5'-GTCTTCTC TGGGTGTTTTGGC-3'. Expression of 18S (for 5'-CGCC GCTAGAGGTGAAATTC-3'; rev 5'-CATTCTTGGCAA TGCTTTCG-3') was used for the normalization.

Relative expression values and standard deviation (SD) have been calculated using the $\Delta\Delta C_T$ approach, as previously described [46,47].

Enzymatic glycosaminoglycan digestion

The glycosaminoglycans were removed from total cell lysates by applying Chondroitinase ABC (120 mU·mL⁻¹), Heparinase I (2.5 mU·mL⁻¹) and Heparinase III (2.5 mU·mL⁻¹) (all from Sigma Aldrich) in TBS per 550 µg protein lysate. The reaction mix was supplemented by 50 µM calcium acetate and incubated overnight at 37 °C in a final volume of 250 µL.

Secretome analysis

For the secretome analysis, cells were grown in medium without FBS supplementation. Conditioned media were collected after 48 h and were enriched in protein content using 10 kDa Amicon Purification System (Merck Millipore).

In silico analysis of O-glycosylation sites

The O-glycosylation sites were predicted by NETOGLYC 4.0 (<http://www.cbs.dtu.dk/services/NetOGlyc/>; [30]). Only O-glycosylation sites with a score higher than 0.5 were considered. Experimentally confirmed O-glycosylation sites were extracted from Glycodomainviewer (<http://glycodomain.glycomics.ku.dk/>; [30]).

Immunofluorescence and in situ proximity ligation assay

Cells were grown in µ-Chamber 12 well glass slides (IBIDI, Martinsried, Germany) and fixed with 4% paraformaldehyde for 15 min. Slides were blocked with 1/5 goat serum and 10% bovine serum albumin in PBS and stained overnight with the corresponding primary antibody. For pRON IF staining, fluorophore conjugated secondary antibody and DAPI were used for labelling. The *in situ* PLA of RON and CD44v6 in cell lines was performed in separate wells using Duolink® PLA Technology (Sigma Aldrich) according to the manufacturer's instructions.

For the combined PLA/IF, all primary antibodies were previously conjugated. The antibody B72.3 was biotinylated

using Pierce™ Antibody Biotinylation Kit for IP (Thermo-Fisher). Antibodies for CD44v6 and RON were conjugated with positive or negative strand oligonucleotides for PLA by Duolink® In Situ Probemaker (Sigma Aldrich). To evaluate the signal in gastric carcinoma, formalin-fixed paraffin embedded gastric carcinoma tissue slides were provided by the department of surgical oncology of the University of Siena (Italy). All procedures were performed after patients' written informed consent and approved by the local ethical committee. Three gastric carcinomas were selected that fulfilled following criteria: (a) The carcinoma had both, regions that were STn positive and regions that were STn negative. The STn positivity has previously been evaluated in this cohort [35], (b) Adjacent to the carcinoma there were normal gastric mucosa and intestinal metaplasia that would act as endogenous biological negative and positive controls respectively. First slides were dewaxed and rehydrated, and then blocked according to Duolink® PLA Technology kit. The slides were incubated overnight with the conjugated primary antibodies at 4 °C. Ligation and amplification of the PLA signal was performed according to the Duolink® PLA Technology kit. Finally, slides were incubated with Streptavidin-FITC and DAPI, and mounted using vectashield mounting medium (Vector Laboratories, Burlingame, CA, USA). All PLA, IF and combined PLA/IF samples were examined under a Zeiss Imager.Z1 Axio fluorescence microscope (Zeiss, Welwyn Garden City, UK). Images were acquired using a ZEISS AXIO CAM MRM and the AXIOVISION REL. 4.8 software (both from Zeiss). For visualization purposes, PLA images were enhanced using IMAGEJ [48] as follows: The red channel (PLA signal) was duplicated with one image being subjected to gaussian blur and then the difference between the two images was calculated (image calculator function). Contrast and brightness was adjusted and PLA spot size was amplified twice by the 'maximum' filter. The PLA signal was quantified using Duolink® ImageTool (Sigma Aldrich) and IMAGEJ [48].

Results

Truncation of O-glycans affects the molecular weight of CD44

To investigate the effect of O-glycan truncation on CD44, we applied genetically engineered cell line models derived from the gastric carcinoma cell line MKN45. The parental MKN45 cell line was selected as parental cell line to develop these models, since it expresses high levels of various CD44 splice variants and requires activity of the RTKs RON and MET, to which CD44 is a coreceptor [15,49]. We induced the truncation of O-glycans in MKN45 through three different mechanisms (Fig. 1A): (a) through overexpression of the α 2,3-sialyltransferase ST3GAL4 that led to

the early termination of O-glycan synthesis (henceforth referred to as ST3) [40,45,50]; (b) through overexpression of the α 2,6-sialyltransferase ST6GALNAC1 that led to a strong increase in the short, aberrant sialyl Tn (STn) glycan (henceforth referred to as ST6) [41]; and (c) through the knockout of the chaperon COSMC, which led to the abrogation of the initial O-glycan elongation step (henceforth referred to as SC for SimpleCell) [38].

We compared the SC clone with wild-type MKN45 (WT), and the two overexpression clones ST3 and ST6 with the mock-transfected clones, M3 and M6 respectively. Indeed, the early termination of O-glycan synthesis showed to significantly impact the molecular weight of CD44 (Fig. 1B). Whereas WT, M3 and M6 express predominantly CD44 with molecular weight above 150 kDa (heavy CD44), the glyco-engineered cell line models expressed a CD44 protein with less than 150 kDa (light CD44). The light CD44 was approximately 50–70 kDa smaller than the heavy CD44 and presented a more defined band indicating more homogenous CD44 glycoforms. In order to investigate whether the observed shift could be induced by reduced elongation of O-glycans, we predicted *in silico* the number of O-glycosylation sites of CD44 (Fig. 1C). The full-length CD44 is predicted to carry 146 O-glycan sites in the stem region being 41 of these sites already experimentally proven (Fig. 1C,D) [30]. The M3 and ST3 models have previously been characterized by O-glycomic analysis and thus allowed us to estimate the induced alterations [45,50]. In this regard, the major O-glycans of M3 with the composition HexNAc₅Hex₅Neu5Ac₂ and HexNAc₄Hex₄Neu5Ac₂, with the average masses of 2427.2 and 2061.9 Da, respectively, were altered in the ST3 model to HexNAc₂Hex₂Neu5Ac₂, with an average mass of 1331.2 Da. This leads to an average mass loss of 913.4 Da per O-glycan, which, considering the high amount of O-glycan sites, could explain a total mass shift of 50–70 kDa.

O-glycan truncation does not alter alternative splicing, or total expression levels of CD44

Alternative splicing of CD44 can be induced through intracellular and extracellular stimuli and constitutes a major mechanism of CD44 modulation, strongly affecting CD44 molecular weight [16]. We analysed whether the induced aberrant glycosylation of our cell models triggered changes in expression and splicing of CD44 or other genes relevant in this process. For this purpose, we performed transcriptomic analysis of the cell models and revealed that CD44 expression levels

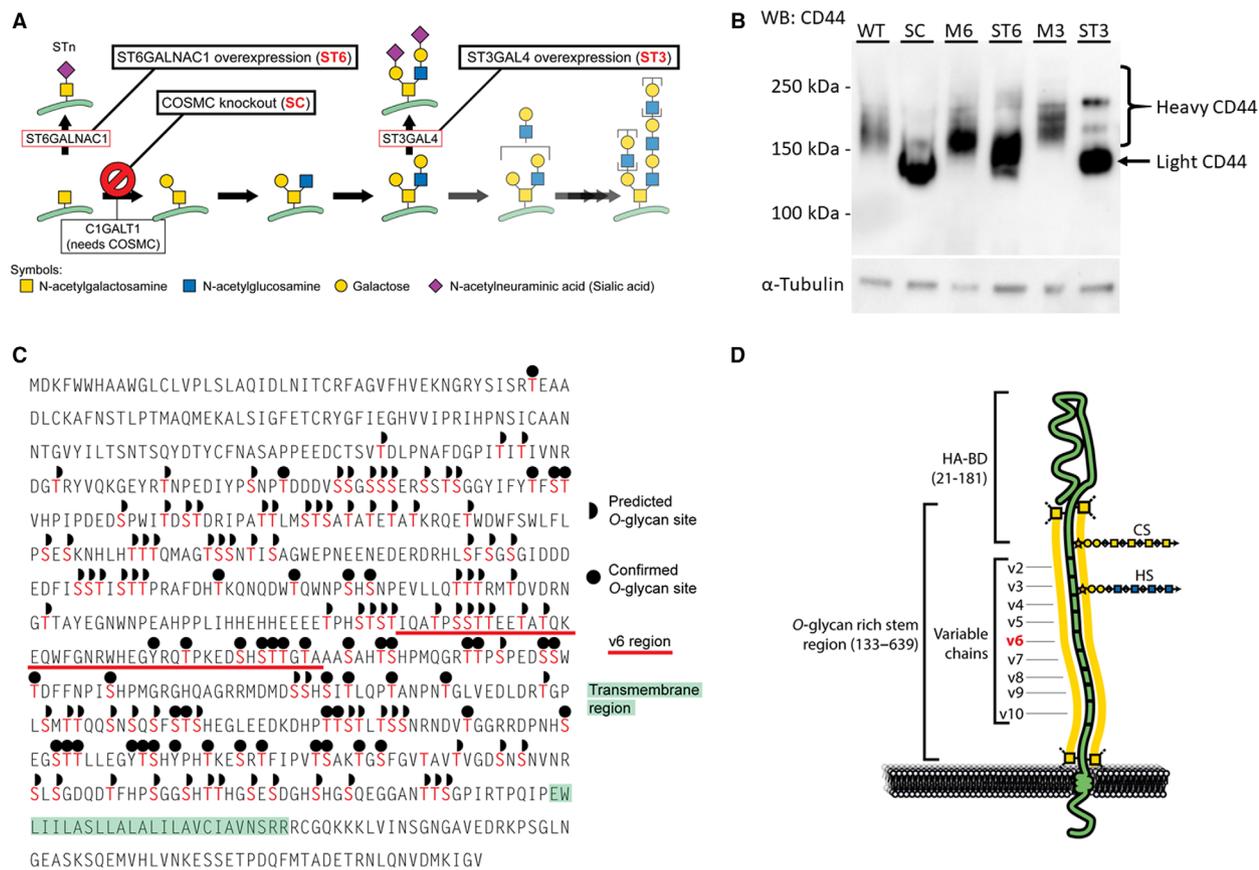


Fig. 1. The effect of O-glycan truncation on CD44. (A) Schematic representation of the mucin-type O-glycan biosynthesis pathway. The mechanisms by which O-glycans have been truncated in the applied MKN45 gastric carcinoma cell line models (SC, ST6 and ST3) are indicated. (B) Western blot analysis of CD44 in cell lysates of glyco-engineered MKN45 models (SC, ST6 and ST3) and their control counterparts (WT, M6 and M3). The glyco-engineered cell lines express primarily CD44 of < 150 kDa (light CD44) as opposed to the above 150 kDa (heavy CD44) of the control cell lines. (C) O-glycan site annotation of the full-length CD44 amino acid sequence. Half circles depict *in silico* predicted O-glycan sites and full circles depict experimentally confirmed O-glycan sites. The variable exon v6 is underscored in red and the transmembrane domain is highlighted in green. (D) Schematic illustration of a full-length CD44 glycoprotein. The hyaluronan binding domain (HA-BD), variable chains (v2-10) as well as the O-glycan-rich stem region, CS and HS are annotated. The information according to publications and UniProt database [8,30,64,65].

were not significantly altered (Fig. 2A). In addition, the relative expression levels of CD44v3 and of CD44v6 to total CD44, as assessed by qRT-PCR, revealed that no significant alterations in alternative splicing of these isoforms were induced (Fig. 2B). These results are supported by the fact that the expression of genes so far reported to be involved in the alternative splicing of CD44 [13] were not notably altered (Fig. 2C).

In addition, the subcellular localization of CD44 and CD44v6 seemed to remain unchanged despite the truncation of O-glycans (Fig. 2D,F). The total protein levels of CD44 and CD44v6, as measured by flow cytometry were comparable between the glyco-engineered cell lines and their respective controls (Fig. 2E,

G). Altogether, these results demonstrate that in the applied models the alterations of O-glycans did not significantly affect the CD44 mRNA and protein expression.

CD44 is a major carrier of truncated O-glycans in gastric cancer cells

Despite the different glycosylation alterations induced in the three glyco-engineered cell models, we observed a similar shift in the electrophoretic mobility of CD44. We therefore hypothesized that the underlying mechanisms that account for the loss of molecular weight are equivalent among these models. Having excluded alternative splicing as a possible explanation, glycosylation

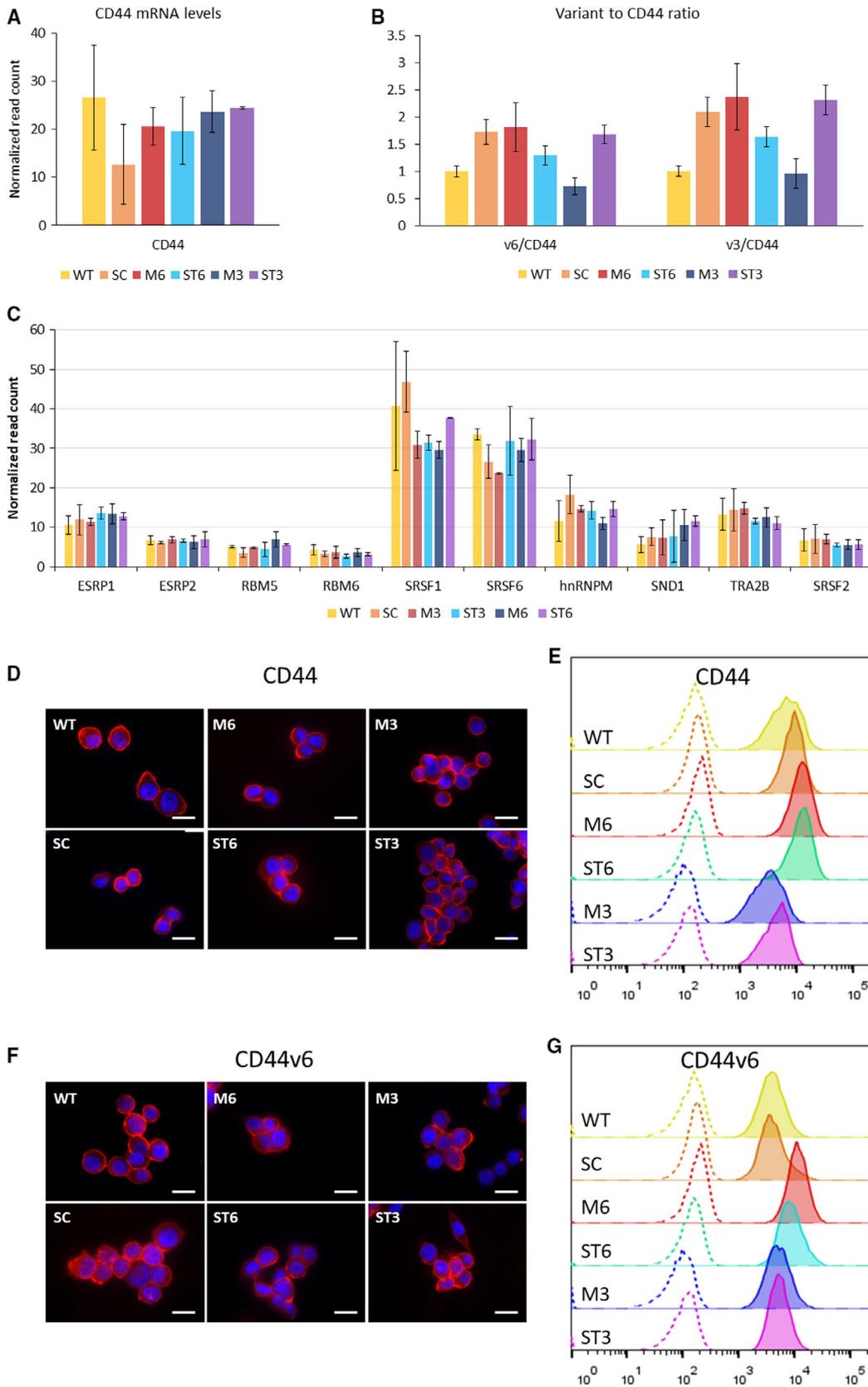


Fig. 2. CD44 expression analyses. (A) CD44 mRNA levels of glyco-engineered cell lines (SC, ST6 and ST3) and their control counterparts (WT, M6 and M3) are shown. The expression data were extracted from whole-transcriptome analysis, performed in duplicates and represented as average of the normalized reads \pm SD. (B) Analysis of the mRNA expression of variants CD44v3 and CD44v6 normalized to total CD44 expression by qRT-PCR. Analysis was performed in two biological replicates with three technical replicates each and is shown as average \pm SD. Comparison of each glyco-engineered cell line (SC, ST6 and ST3) and its respective control counterparts (WT, M6 and M3) showed no statistical significant alterations (Student's *t*-test, $P > 0.05$). (C) Expression levels of the splicing factors *ESRP1*, *ESRP2*, *RBM5*, *RBM6*, *SRSF1*, *SRSF6*, *hnRNPM*, *SND1*, *TRA2B*, *SRSF2*, which are known to be involved in CD44 alternative splicing. The expression data were extracted from whole-transcriptome analysis, performed in duplicates and represented as average of the normalized reads \pm SD. (D) IF staining of CD44 of the six cell line models. CD44 is shown in red and DAPI in blue. The scale bar is 20 μ m. (E) Flow cytometry analysis of CD44 expression in MKN45 glyco-engineered cell lines (SC, ST6 and ST3) as compared to their control cell lines (WT, M6 and M3 respectively). The negative controls are shown in dotted lines. (F) IF staining of CD44v6 of the six cell line models. CD44v6 is shown in red and DAPI in blue. The scale bar is 20 μ m. (G) Flow cytometry analysis of CD44v6 in MKN45 glyco-engineered cell lines (SC, ST6 and ST3) as compared to their control cell lines (WT, M6 and M3 respectively). The negative controls are shown in dotted lines.

remained as the modification that could account for such molecular weight differences. Alterations in occupancy and composition of the glycan chains can significantly influence glycoprotein/proteoglycan molecular weight. To test whether this drastic shift was caused by differential addition of HS and CS glycosaminoglycans, we performed complete digestion of HS, CS or a combination of both on the cell lysates. The efficiency of these digestions was shown by the loss of HS-specific and CS-specific antibody binding in dot-blot assays (Fig. S1). The digestion of HS and CS had no significant effect on CD44 and therefore excludes alterations in glycosaminoglycan occupancy or length as the cause for the difference between light and heavy CD44 (Fig. 3A). As an additional control of HS digestion, Syndecan 1 which is a highly O-glycosylated protein [30] known to carry HS [51] was also evaluated. The western blot detection of Syndecan 1 at under 100 kDa demonstrates the efficient release of HS (Fig. 3B). Interestingly, Syndecan 1 showed a sharper band with lower molecular weight in the glyco-engineered cell models compared to their controls (Fig. 3B). This difference in molecular weight is analogous to what was observed for CD44. This observation further underlines the involvement of O-glycan truncation in the striking change in CD44's molecular weight in these cell models. We then used the SC and ST6 models to assess the extent to which CD44 is decorated with truncated O-glycans. Both models result in the overexpression of STn [38,41], a truncated O-glycan epitope that can be detected by specific monoclonal antibodies [43,44]. We confirmed a marked overexpression of STn by flow cytometry and showed by western blot that ST6 and SC express STn primarily on specific proteins (Fig. 3C,D). The STn staining of immunoprecipitated CD44 revealed that, among the proteins observed to be modified with STn in SC and

ST6, the light CD44 is the most abundant carrier of STn (Fig. 3D). Taking together, these findings show that CD44 is a major carrier of truncated O-glycans in gastric cancer cells, which underpins the observed mass shift in the glyco-engineered cell models.

Truncation of O-glycans increases hyaluronan binding and affects CD44 shedding

Following the identification of CD44 as carrier of truncated O-glycans we assessed the capacity of CD44 carrying truncated O-glycans to bind to HA. CD44 is known to be the major binding protein of HA in human cells [16]. Other receptors of HA are not appreciably expressed in our cell models (Fig. 4A). The HA binding assay revealed that the truncation of O-glycans led to a striking increase in HA binding in our cell models (Fig. 4B). Although the induced truncation of O-glycans was achieved through different molecular alterations, the increased binding was consistent among all models. This underlines the significant effect of O-glycan truncation on the HA binding capacity of CD44.

In addition, it has been previously reported that CD44 can be shed into the extracellular milieu by the cleavage of the juxtamembrane domain [11,12,52] and that changes in glycosylation may impact this process [53]. To assess whether the truncation of O-glycans affect the secretion of CD44 we have analysed the secretome of the glyco-engineered cell models (Fig. 4C). According to expectation a mass shift of approximately 10–20 kDa was observed in all models, which corresponds to a loss of cytosolic, transmembrane and a small part of the extracellular domain. Interestingly, CD44 bearing shorter O-glycans appeared to be more abundantly shed than that of the control cell lines.

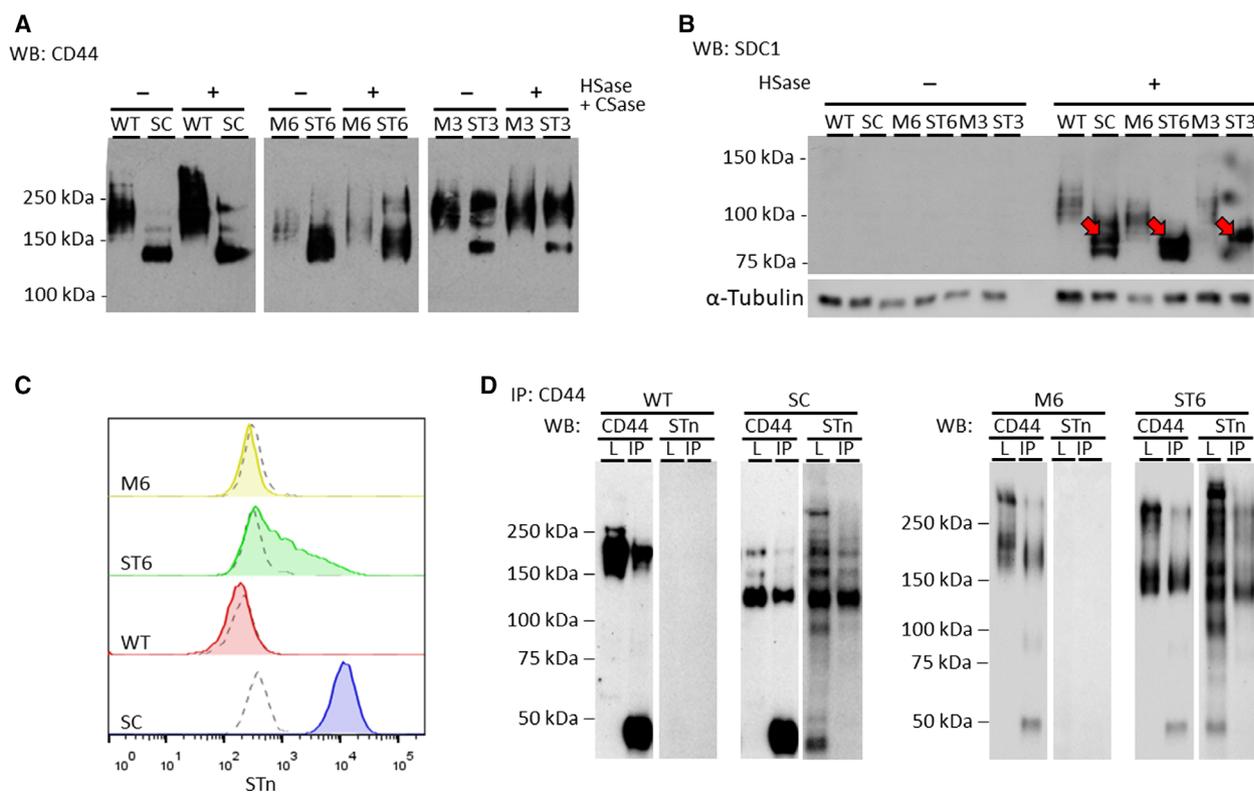


Fig. 3. Glycosaminoglycan digestion and sialyl Tn analysis of CD44. (A) Western blot analysis of CD44 before and after incubation with heparinase I and III (HSase) and chondroitinase ABC (CSase). The glyco-engineered cell lines SC, ST6 and ST3 are shown adjacent to their respective control cell lines WT, M6 and M3. (B) Western blot analysis of syndecan 1 (SDC1) before and after incubation with HSase and CSase. SDC1 was detected with this antibody only after the release of glycosaminoglycan chains and reveals a similar mass shift (indicated by red arrows) as previously observed for CD44. (C) Flow cytometry analysis of STn expression in MKN45 glyco-engineered cell lines, SC and ST6, as compared to their control cell lines, WT and M6. The negative controls are shown in dotted lines. (D) Immunoprecipitation of CD44 from glyco-engineered cell lines, SC and ST6, and their respective control cell lines WT and M6. The total lysate (L) of the cell lines that has been used as input for the immunoprecipitation is shown in each blots' left lane and the immunoprecipitated CD44 (IP) is loaded on each blots right lane. Each sample was blotted for CD44 and STn.

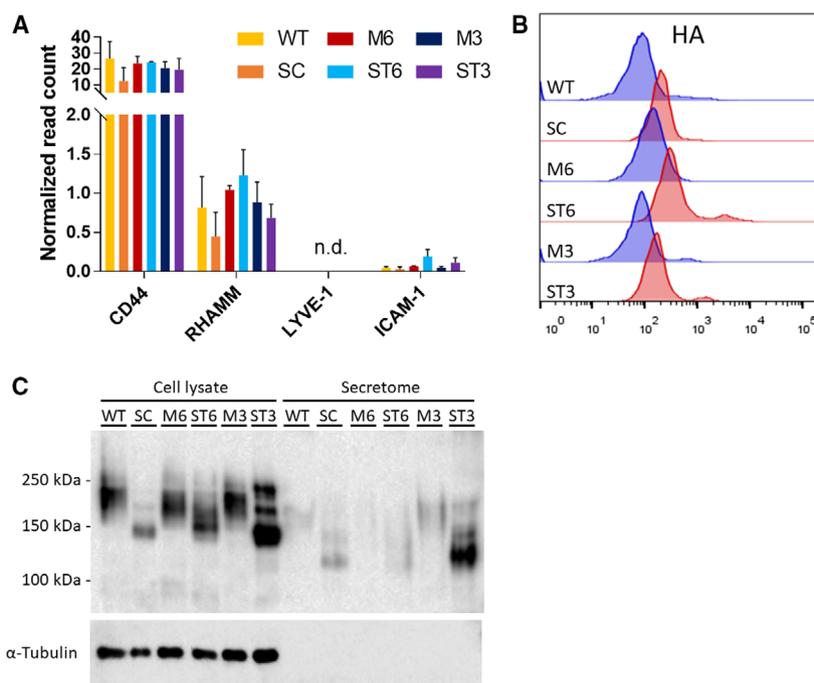
Truncation of O-glycans increases the CD44 and RON colocalization and results in increased RON activation

CD44v6 is a coreceptor of the RON receptor tyrosine kinase and mediates its activation [21]. The oncogenic hyperactivation of RON has been described as a common event in gastric cancer and it is associated with tumour progression, angiogenesis and chemoresistance [54]. In order to test if the altered glycosylation of CD44 influences the v6 association with RON, we performed a series of *in situ* PLAs. Remarkably, the colocalization of CD44v6 and RON was significantly more frequent in cells with shorter O-glycans than in the control cell lines (Fig. 5). The significantly elevated colocalization events were concomitant with the increased activation of the RON receptor in these cells, as depicted by pRON immunolabelling (Fig. 5A–C).

This indicates an involvement of the truncation of O-glycans in the activation of RON, through the facilitated interaction of CD44v6 and RON.

In order to validate this association in clinical samples, we developed a combinatory approach with RON/CD44v6 PLA and FITC-labelling of STn in tumour tissue sections from patients with gastric cancer. This approach enabled the *in situ* evaluation of the colocalization of RON, CD44v6 and the truncated O-glycan epitope STn (Fig. 6). An association between tumour areas that were STn positive and increased amounts of RON/CD44v6 colocalization events was observed in all the gastric carcinoma cases evaluated (Fig. 6). Interestingly, extensive colocalization of STn and RON/CD44v6 was also evident in the areas of intestinal metaplasia (a preneoplastic lesion of the gastric mucosa [55]), albeit to a lower degree than

Fig. 4. CD44 binding to HA altered through O-glycan truncation. (A) Expression levels of other HA binding proteins (CD44, RHAMM, LYVE-1 and ICAM-1) are shown in the six cell line models. The expression data were extracted from whole-transcriptome analysis, performed in duplicates and represented as average of the normalized reads \pm SD. (B) Flow cytometry analysis of HA-fluorescein binding to MKN45 glyco-engineered cell lines with truncated O-glycans (SC, ST6 and ST3) and their respective control cell lines (WT, M6 and M3). (C) Comparative western blot analysis of CD44 and tubulin in total cell lysate and in conditioned medium (secretome). MKN45 glyco-engineered cell lines with truncated O-glycans (SC, ST6 and ST3) and their respective control cell lines (WT, M6 and M3) are shown.



observed in the carcinoma (Fig. 6B). No significant histological differences between STn positive and negative areas were evident (Fig. S2).

Discussion

CD44 is a multifunctional transmembrane protein that is abundantly decorated by a variety of glycan moieties [8,9,30]. In gastric cancer, high expression levels of CD44 are found in about half of all primary tumours and are associated with the presence of distant metastases, tumour recurrence and increased mortality [25,26]. Moreover, an increasing body of evidence demonstrates that CD44 is a gastric cancer stem cell marker, since CD44 expressing cancer cells are competent to perform self-renewal and to produce differentiated progeny [27]. CD44 functions as major HA binding protein, supporting the adhesion and migration on the extracellular matrix [17]. CD44 has also been shown to act as coreceptor for several RTKs and to elevate cancer cell resistance to therapy [56]. In particular, the variant form CD44v6 was directly linked to RON activation [21].

In this manuscript, we describe for the first time the impact of O-glycan truncation on molecular features of CD44 and highlight new functional implications of these glycans. We have performed a comparative analysis of three differently glyco-engineered MKN45 models: ST3GAL4 overexpressing (ST3), ST6GALNAC1 overexpressing (ST6) and COSMC knock-out

(SC from SimpleCell), and compared them to their corresponding control cell lines (mock and WT). MKN45 WT and mock expresses primarily core 2 O-glycans with 2 or more LacNAc repeats [45,50]. This glycosylation profile is reflective of some gastric carcinomas [57]. Although different genes were targeted in our glyco-engineered models, the early termination of O-glycans was a common feature of the three cell line models.

ST3GAL4 is a sialyltransferase that affects both N- and O-glycans [45,50]. The upregulation of ST3GAL4 in MKN45 leads to an increase in α 2,3 sialylation and a decrease in α 2,6 sialylation on N-glycans and significantly decreases the amount of bisected structures [45,50]. On O-glycans, ST3GAL4 leads to the earlier termination due to the increase in one specific structure: the di-sialylated core 2 structure [45,50]. Both the ST6GALNAC1 overexpression and the COSMC deletion induce alteration on O-glycans and lead to an upregulation of STn by two differential mechanisms [38,41]. ST6GALNAC1 is a sialyltransferase that adds sialic acid onto Tn to form STn and thereby, leads to the early termination of the O-glycan chains [41]. COSMC on the other hand, is a dedicated chaperone of the core 1 galactosyltransferase (C1GALT1) [58]. Without its chaperone, C1GALT1 is dysfunctional and fails to elongate the initial *N*-acetylgalactosamine (which constitutes the Tn antigen) to core 1. In SC, the lack of COSMC causes therefore the accumulation of Tn and STn antigens [38,59]. The fact that STn

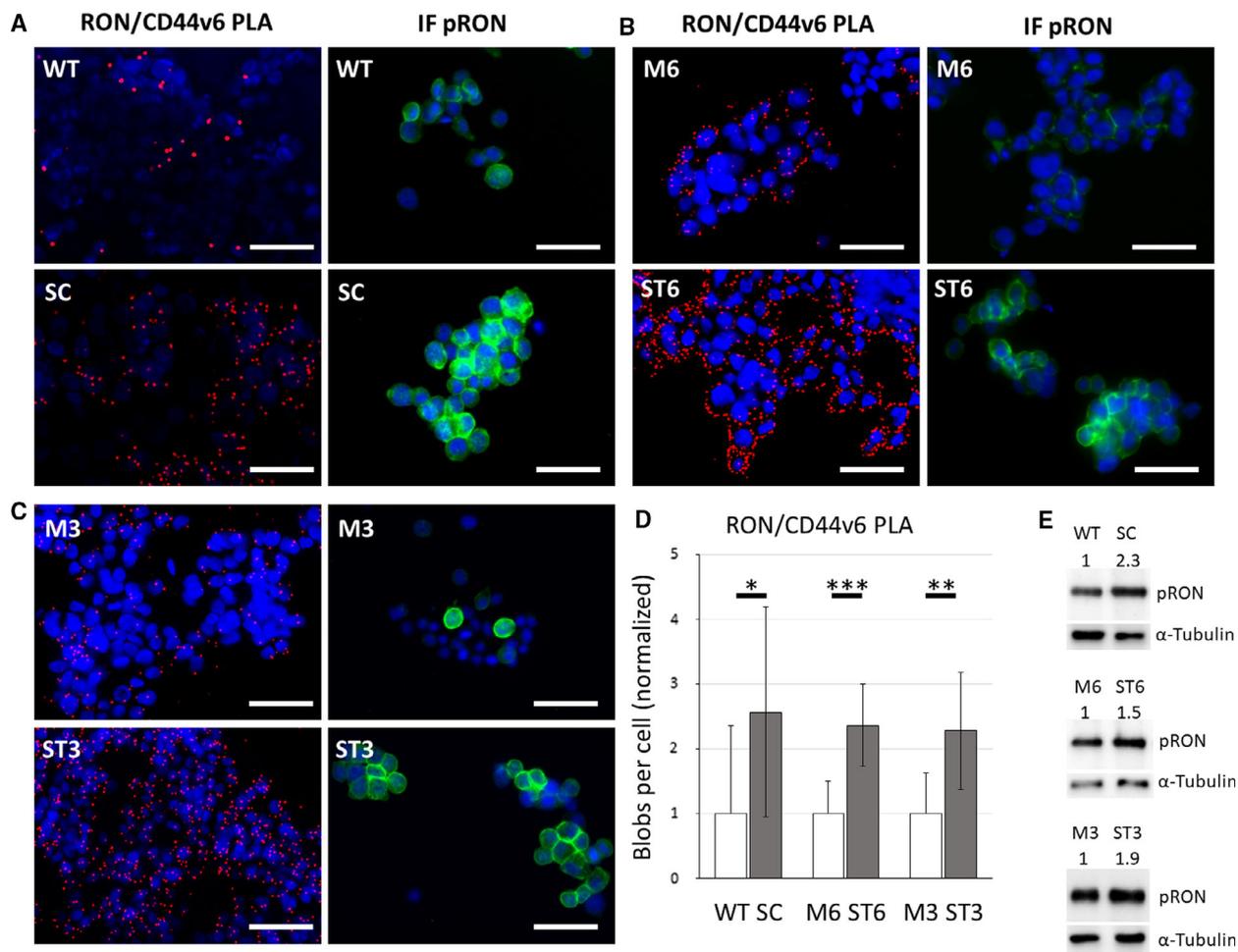


Fig. 5. Colocalization analysis between CD44v6 and RON as well as RON activation. (A–C) PLA for the receptor tyrosine kinase RON and its coreceptor CD44v6 (left panels). Red PLA dots (blobs) indicate colocalization event of RON and CD44v6. In addition, IF staining of pRON for each respective cell line is shown in green (right panels). DAPI is shown in blue. The scale bar indicates 50 μ m. (D) Quantification of RON/CD44v6 PLA blobs per cell. Values have been normalized to the respective control cell lines (WT, M6 and M3) and are represented as average \pm SD. Statistical significance was determined by Student's *t*-test (*P*-value * < 0.05; ** < 0.01; *** < 0.001). (E) Western blot analysis of pRON and tubulin in cell lysates of glyco-engineered MKN45 models (SC, ST6 and ST3) and their control counterparts (WT, M6 and M3). The normalized densitometric quantification of pRON amounts are shown in numbers above the bands.

expression was initiated in MKN45 both, after the overexpression of ST6GALNAC1 but also after the knockout of COSMC (with endogenous levels of ST6GALNAC1) highlights that the STn expression arises due to the relative activity of competing glycosyltransferases. In addition, other mechanisms besides the relative expression of glycosyltransferases have been described to induce the truncation of O-glycans in cancer [32,60].

The four mechanisms that could underpin the molecular weight changes of CD44, which span over 50 kDa, are (a) alternative splicing, (b) proteolytic cleavage, (c) altered O-glycosylation and (d) changes in the modification with glycosaminoglycan chains.

MKN45 expresses mostly CD44v8-10 splicing isoforms, known as CD44E, named after its common expression in epithelial cells, but also CD44s (without variant chains) and CD44v6 [15]. In order to validate alternative splicing as cause for the molecular weight shift we performed qRT-PCR analysis for total CD44 and V3, V6 variants. Since no major alterations on the expression levels were found within the different cell models, we excluded this hypothesis. In addition, expression analysis of already described CD44 splicing factors did not show any statistical difference within our models, supporting this conclusion (Fig. 2).

CD44 can be subjected to sequential proteolytic cleavages resulting in the release of the extracellular

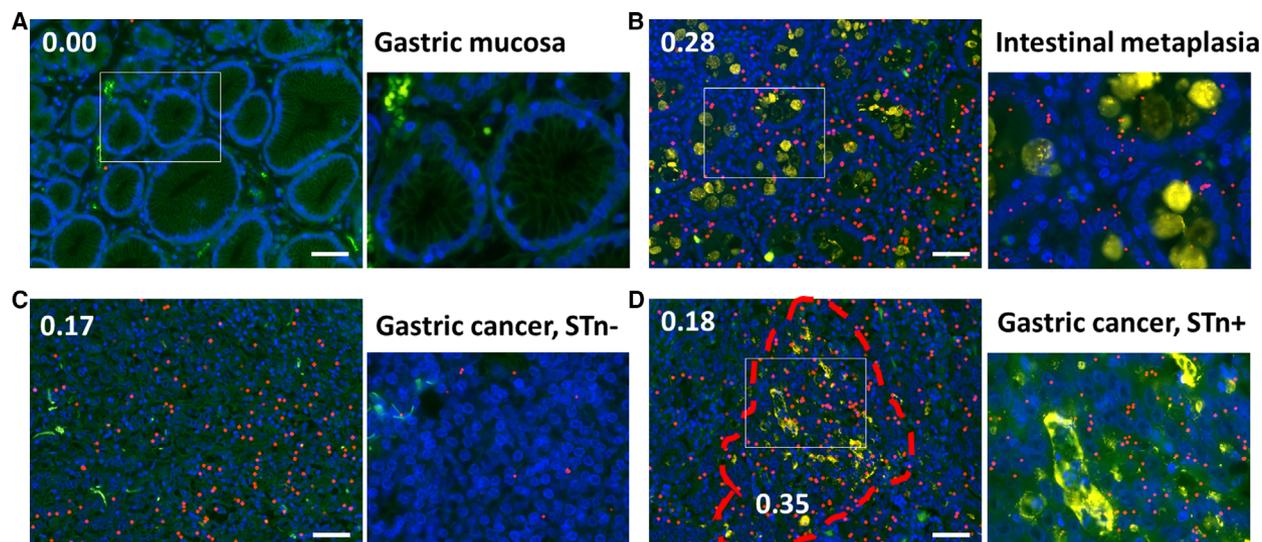


Fig. 6. PLA analysis of CD44v6/RON combined IF staining of STn in human gastric tissue sections. (A–D) Combined PLA/IF analysis in human gastric tissue sections. The PLA was performed with anti-RON and anti-CD44v6 antibody detecting their colocalization (red blobs). The number of PLA blobs per cell were quantified and are shown as numerical values within the image. The IF detection of STn is shown in yellow. Nuclear staining with DAPI is shown in blue and the green autofluorescence has been used to visualize tissue architecture. The scale bar indicates 50 μm . (A) Representation of the gastric mucosa adjacent to gastric carcinoma. RON/CD44v6 colocalization events were absent and no STn was expressed by gastric epithelial cells. (B) Representation of intestinal metaplasia adjacent to gastric carcinoma. The colocalization between RON and CD44v6 was detected and has 0.28 blobs per cell. Goblet cells of intestinal metaplasia are known to express high amounts of STn as shown. (C) Gastric carcinoma without STn expression. Limited amount of colocalization between RON and CD44v6 has been detected and has 0.17 blobs per cell. (D) Gastric carcinoma with localized expression of STn. A higher number (2-fold differences) of CD44v6/RON colocalization events, 0.35 blobs per cell, were detected in the tumour area with STn compared to the directly adjacent STn negative tumour areas, with 0.18 blobs per cell. The area that has been considered STn positive is framed in dashed lines.

domain (ECD). Several proteases can be involved in the shedding of the ECD, such as MT1–MMP, MMP9, ADAM-10 and ADAM-17 [11,12,52]. For some proteases, the cleavage sites have been confined to the stem region, which releases the ectodomain of CD44. The susceptibility of CD44 to be shed has previously been reported to be affected by glycosylation [53]. Here we show that the truncation of O-glycans can increase the cleavage susceptibility of CD44. Interestingly, it has been shown that the proteolytic cleavage can be triggered by the binding action of CD44 [61]. In that sense, the modulation of the CD44 interaction with HA is expected to further increase the shedding. However, we can exclude proteolytic shedding as the cause of the observed mass change in the glyco-engineered cell models.

We also ruled out that alterations in glycosaminoglycans accounted for the light version of CD44. Our results point unequivocally towards the truncation of O-glycosylation as the underlying mechanism for the expression of the light CD44 isoform. This is supported by the mass shift observed in the likewise highly O-glycosylated proteoglycan Syndecan 1

(Fig. 3B). Of note, we made the observation that western blot band intensities were significantly increased when lighter glycoforms of CD44 or Syndecan 1 were expressed. This might stem from increased transfer efficiencies of smaller proteins and from the more focused nature of the band. The CD44 amounts detected by western blot analysis were not fully overlapping with those of the FACS analysis as well as with the transcription levels. The latter two methods are quantitative and revealed no significant changes. These results support that protein glycosylation alterations could interfere with protein quantifications assessed by western blot.

Experiments in which immunoprecipitated CD44 was incubated with exoglycosidases failed to completely normalize the electrophoretic behaviour of CD44 among the different cell models (data not shown). This observation is likely due to the high glycosidase resistance of densely glycosylated proteins. Presently, no single glycosidase can release all O-glycans and high density of extended O-glycans entails considerable resistance to sequentially acting exoglycosidases, rendering gentle de-O-glycosylation a

challenging task. Harsh chemical approaches can result in the efficient release of O-glycans, but lead to the chemical modification and partial degradation of the peptide backbone.

Alterations of the glycosylation machinery leading to the truncation of O-glycans are a common feature in cancer [32]. Here, we demonstrate that O-glycan truncation may contribute to malignant phenotypes through the significant modulation of CD44 functional activity. In this regard, we showed that cell models with shorter O-glycans display an univocal increase in HA binding capacity. This result is in line with previous observations that changes in glycosylation of CD44 alters its capacity to act as HA receptor [31,62]. Here, we demonstrate for the first time this effect in the gastric cancer context, supporting that it might represent a critical mechanism in promoting a migratory proinvasive cancer cell phenotype.

Moreover, we discovered a striking increase in the colocalization events between CD44v6 and RON when O-glycans were truncated. Concomitant with the enhanced colocalization, we observed an elevated activation of RON in these cell lines. CD44v6 is known to mediate the activation of RON, a receptor tyrosine kinase important for the invasive growth in various cancers [63]. We therefore hypothesize that the truncation of O-glycans presents a novel oncogenic mechanism of RON activation through the mediation of CD44 and RON interactions. The increased colocalization of CD44v6 and RON in tumour areas of gastric cancer patients that express short truncated O-glycans supports an important mechanism in cancer that ought to be further investigated. CD44 is also involved in the activation of other RTKs, such as MET, EGFR, VEGFR and HER2 [16,22,23]. The targeting of these RTKs through monoclonal antibodies or small molecules is currently applied or under clinical investigation as treatment for cancer patients. Our data support that the expression of truncated O-glycan epitopes might constitute a new additional biomarker for treatment stratification of cancer patients.

Acknowledgements

We thank Diana Campos, Catharina Steentoft and Henrik Clausen for supporting the development of the SimpleCell (SC) cell line model. The authors acknowledge the support of José Luis Costa and Mafalda Rocha from the i3S Genomics Platform (GenCore), as well as the support by the Advanced Light Microscopy Platform and the Translational Cytometry Platform of i3S.

Funding

This work was funded by FEDER funds through the Operational Programme for Competitiveness Factors-COMPETE (POCI-01-0145-FEDER-016585; POCI-01-0145-FEDER-007274; POCI-01-0145-FEDER-028489) and National Funds through the Foundation for Science and Technology (FCT), under the projects: PTDC/BBB-EBI/0567/2014 (to CAR), PTDC/MED-ONC/28489/2017 (to AM), UID/BIM/04293/2013 and CEECIND/02760/2017 (to SM); and the project NORTE-01-0145-FEDER-000029, supported by Norte Portugal Regional Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). We acknowledge the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 748880 (to MB). The authors acknowledge the support by Gastric Glyco Explorer Initial Training Network (European Union Seventh Framework Programme GastricGlycoExplorer project, grant number 316929).

Author contributions

SM, AM and CAR conceived and supervised the study; SM and AM designed experiments; SM, AMM, CG, MB and JAM performed experiments; KP and FR provided clinical specimens and contributed to scientific discussions; SM wrote the manuscript; All authors made manuscript revisions.

References

- 1 Sreaton GR, Bell MV, Jackson DG, Cornelis FB, Gerth U and Bell JI (1992) Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc Natl Acad Sci USA* **89**, 12160–12164.
- 2 Marhaba R and Zoller M (2004) CD44 in cancer progression: adhesion, migration and growth regulation. *J Mol Histol* **35**, 211–231.
- 3 Bourguignon LY, Gunja-Smith Z, Iida N, Zhu HB, Young LJ, Muller WJ and Cardiff RD (1998) CD44v (3,8-10) is involved in cytoskeleton-mediated tumor cell migration and matrix metalloproteinase (MMP-9) association in metastatic breast cancer cells. *J Cell Physiol* **176**, 206–215.
- 4 Murakami D, Okamoto I, Nagano O, Kawano Y, Tomita T, Iwatsubo T, De Strooper B, Yumoto E and Saya H (2003) Presenilin-dependent gamma-secretase activity mediates the intramembranous cleavage of CD44. *Oncogene* **22**, 1511–1516.

- 5 Branco da Cunha C, Klumpers DD, Koshy ST, Weaver JC, Chaudhuri O, Seruca R, Carneiro F, Granja PL and Mooney DJ (2016) CD44 alternative splicing in gastric cancer cells is regulated by culture dimensionality and matrix stiffness. *Biomaterials* **98**, 152–162.
- 6 Azevedo R, Gaiteiro C, Peixoto A, Relvas-Santos M, Lima L, Santos LL and Ferreira JA (2018) CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. *Clin Proteomics* **15**, 22.
- 7 Bennett KL, Modrell B, Greenfield B, Bartolazzi A, Stamenkovic I, Peach R, Jackson DG, Spring F and Aruffo A (1995) Regulation of CD44 binding to hyaluronan by glycosylation of variably spliced exons. *J Cell Biol* **131**, 1623–1633.
- 8 Jackson DG, Bell JI, Dickinson R, Timans J, Shields J and Whittle N (1995) Proteoglycan forms of the lymphocyte homing receptor CD44 are alternatively spliced variants containing the v3 exon. *J Cell Biol* **128**, 673–685.
- 9 English NM, Lesley JF and Hyman R (1998) Site-specific de-N-glycosylation of CD44 can activate hyaluronan binding, and CD44 activation states show distinct threshold densities for hyaluronan binding. *Can Res* **58**, 3736–3742.
- 10 Dimitroff CJ, Lee JY, Fuhlbrigge RC and Sackstein R (2000) A distinct glycoform of CD44 is an L-selectin ligand on human hematopoietic cells. *Proc Natl Acad Sci USA* **97**, 13841–13846.
- 11 Chetty C, Vanamala SK, Gondi CS, Dinh DH, Gujrati M and Rao JS (2012) MMP-9 induces CD44 cleavage and CD44 mediated cell migration in glioblastoma xenograft cells. *Cell Signal* **24**, 549–559.
- 12 Andereg U, Eichenberg T, Parthaune T, Haiduk C, Saalbach A, Milkova L, Ludwig A, Grosche J, Averbeck M, Gebhardt C *et al.* (2009) ADAM10 is the constitutive functional sheddase of CD44 in human melanoma cells. *J Invest Dermatol* **129**, 1471–1482.
- 13 Prochazka L, Tesarik R and Turanek J (2014) Regulation of alternative splicing of CD44 in cancer. *Cell Signal* **26**, 2234–2239.
- 14 Banky B, Raso-Barnett L, Barbai T, Timar J, Becsagh P and Raso E (2012) Characteristics of CD44 alternative splice pattern in the course of human colorectal adenocarcinoma progression. *Mol Cancer* **11**, 83.
- 15 da Cunha CB, Oliveira C, Wen X, Gomes B, Sousa S, Suriano G, Grellier M, Huntsman DG, Carneiro F, Granja PL *et al.* (2010) De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest* **90**, 1604–1614.
- 16 Orian-Rousseau V and Sleeman J (2014) CD44 is a multidomain signaling platform that integrates extracellular matrix cues with growth factor and cytokine signals. *Adv Cancer Res* **123**, 231–254.
- 17 Misra S, Hascall VC, Markwald RR and Ghatak S (2015) Interactions between hyaluronan and its receptors (CD44, RHAMM) regulate the activities of inflammation and cancer. *Front Immunol* **6**, 201.
- 18 Richter U, Wicklein D, Geleff S and Schumacher U (2012) The interaction between CD44 on tumour cells and hyaluronan under physiologic flow conditions: implications for metastasis formation. *Histochem Cell Biol* **137**, 687–695.
- 19 Jones M, Tussey L, Athanasou N and Jackson DG (2000) Heparan sulfate proteoglycan isoforms of the CD44 hyaluronan receptor induced in human inflammatory macrophages can function as paracrine regulators of fibroblast growth factor action. *J Biol Chem* **275**, 7964–7974.
- 20 Yu WH, Woessner JF Jr, McNeish JD and Stamenkovic I (2002) CD44 anchors the assembly of matrilysin/MMP-7 with heparin-binding epidermal growth factor precursor and ErbB4 and regulates female reproductive organ remodeling. *Genes Dev* **16**, 307–323.
- 21 Matzke A, Herrlich P, Ponta H and Orian-Rousseau V (2005) A five-amino-acid peptide blocks Met- and Rndependent cell migration. *Can Res* **65**, 6105–6110.
- 22 Sherman LS, Rizvi TA, Karyala S and Ratner N (2000) CD44 enhances neuregulin signaling by Schwann cells. *J Cell Biol* **150**, 1071–1084.
- 23 Matzke-Ogi A, Jannasch K, Shatirishvili M, Fuchs B, Chiblak S, Morton J, Tawk B, Lindner T, Sansom O, Alves F *et al.* (2016) Inhibition of tumor growth and metastasis in pancreatic cancer models by interference with CD44v6 signaling. *Gastroenterology* **150**, 513–525. e510.
- 24 Stamenkovic I, Aruffo A, Amiot M and Seed B (1991) The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. *EMBO J* **10**, 343–348.
- 25 Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209.
- 26 Mayer B, Jauch KW, Gunthert U, Figdor CG, Schildberg FW, Funke I and Johnson JP (1993) De-novo expression of CD44 and survival in gastric cancer. *Lancet* **342**, 1019–1022.
- 27 Takaishi S, Okumura T, Tu S, Wang SS, Shibata W, Vigneshwaran R, Gordon SA, Shimada Y and Wang TC (2009) Identification of gastric cancer stem cells using the cell surface marker CD44. *Stem Cells* **27**, 1006–1020.
- 28 Yamaguchi A, Goi T, Yu J, Hirono Y, Ishida M, Iida A, Kimura T, Takeuchi K, Katayama K and Hirose K (2002) Expression of CD44v6 in advanced gastric cancer and its relationship to hematogenous metastasis and long-term prognosis. *J Surg Oncol* **79**, 230–235.

- 29 Pereira C, Ferreira D, Lemos C, Martins D, Mendes N, Almeida D, Granja P, Carneiro F, Almeida R, Almeida GM *et al.* (2018) CD44v6 expression is a novel predictive marker of therapy response and poor prognosis in gastric cancer patients. *bioRxiv* 468934 [PREPRINT].
- 30 Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L *et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**, 1478–1488.
- 31 Dasgupta A, Takahashi K, Cutler M and Tanabe KK (1996) O-linked glycosylation modifies CD44 adhesion to hyaluronate in colon carcinoma cells. *Biochem Biophys Res Comm* **227**, 110–117.
- 32 Pinho SS and Reis CA (2015) Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540–555.
- 33 David L, Nesland JM, Clausen H, Carneiro F and Sobrinho-Simoes M (1992) Simple mucin-type carbohydrate antigens (Tn, sialosyl-Tn and T) in gastric mucosa, carcinomas and metastases. *APMIS Suppl* **27**, 162–172.
- 34 Mereiter S, Balmana M, Gomes J, Magalhaes A and Reis CA (2016) Glycomic approaches for the discovery of targets in gastrointestinal cancer. *Front Oncol* **6**, 55.
- 35 Mereiter S, Polom K, Williams C, Polonia A, Guergova-Kuras M, Karlsson NG, Roviello F, Magalhaes A and Reis CA (2018) The Thomsen-Friedenreich antigen: a highly sensitive and specific predictor of microsatellite instability in gastric cancer. *J Clin Med* **7**. <https://doi.org/10.3390/jcm7090256>
- 36 Persson N, Stuhr-Hansen N, Risinger C, Mereiter S, Polonia A, Polom K, Kovacs A, Roviello F, Reis CA, Welinder C *et al.* (2017) Epitope mapping of a new anti-Tn antibody detecting gastric cancer cells. *Glycobiology* **27**, 635–645.
- 37 Marcos NT, Bennett EP, Gomes J, Magalhaes A, Gomes C, David L, Dar I, Jeanneau C, DeFrees S, Krustup D *et al.* (2011) ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front Biosci* **3**, 1443–1455.
- 38 Campos D, Freitas D, Gomes J, Magalhaes A, Steentoft C, Gomes C, Vester-Christensen MB, Ferreira JA, Afonso LP, Santos LL *et al.* (2015) Probing the O-glycoproteome of gastric cancer cell lines for biomarker discovery. *Mol Cell Proteomics* **14**, 1616–1629.
- 39 Oliveira FMS, Mereiter S, Lonni P, Siart B, Shen Q, Heldin J, Raykova D, Karlsson NG, Polom K, Roviello F *et al.* (2018) Detection of post-translational modifications using solid-phase proximity ligation assay. *New Biotechnol* **45**, 51–59.
- 40 Carvalho AS, Harduin-Lepers A, Magalhaes A, Machado E, Mendes N, Costa LT, Matthiesen R, Almeida R, Costa J and Reis CA (2010) Differential expression of alpha-2,3-sialyltransferases and alpha-1,3/4-fucosyltransferases regulates the levels of sialyl Lewis a and sialyl Lewis x in gastrointestinal carcinoma cells. *Int J Biochem Cell Biol* **42**, 80–89.
- 41 Marcos NT, Pinho S, Grandela C, Cruz A, Samyn-Petit B, Harduin-Lepers A, Almeida R, Silva F, Morais V, Costa J *et al.* (2004) Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated sialyl-Tn antigen. *Can Res* **64**, 7050–7057.
- 42 Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, Bennett EP, Mandel U, Wandall H, Levery SB and Clausen H (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* **8**, 977–982.
- 43 Kjeldsen T, Clausen H, Hirohashi S, Ogawa T, Iijima H and Hakomori S (1988) Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2—6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope. *Can Res* **48**, 2214–2220.
- 44 Thor A, Ohuchi N, Szpak CA, Johnston WW and Schlom J (1986) Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. *Can Res* **46**, 3118–3124.
- 45 Mereiter S, Magalhaes A, Adamczyk B, Jin C, Almeida A, Drici L, Ibanez-Vea M, Gomes C, Ferreira JA, Afonso LP *et al.* (2016) Glycomic analysis of gastric carcinoma cells discloses glycans as modulators of RON receptor tyrosine kinase activation in cancer. *Biochem Biophys Acta* **1860**, 1795–1808.
- 46 Livak KJ and Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402–408.
- 47 Ling D, Pike CJ and Salvaterra PM (2012) Deconvolution of the confounding variations for reverse transcription quantitative real-time polymerase chain reaction by separate analysis of biological replicate data. *Anal Biochem* **427**, 21–25.
- 48 Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682.
- 49 Benvenuti S, Lazzari L, Arnesano A, Li Chiavi G, Gentile A and Comoglio PM (2011) Ron kinase transphosphorylation sustains MET oncogene addiction. *Can Res* **71**, 1945–1955.
- 50 Mereiter S, Magalhaes A, Adamczyk B, Jin C, Almeida A, Drici L, Ibanez-Vea M, Larsen MR, Kolarich D, Karlsson NG *et al.* (2016) Glycomic and sialoproteomic data of gastric carcinoma cells overexpressing ST3GAL4. *Data Brief* **7**, 814–833.

- 51 Couchman JR, Gopal S, Lim HC, Norgaard S and Multhaupt HA (2015) Fell-Muir Lecture: Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. *Int J Exp Pathol* **96**, 1–10.
- 52 Nagano O and Saya H (2004) Mechanism and biological significance of CD44 cleavage. *Cancer Sci* **95**, 930–935.
- 53 Gasbarri A, Del Prete F, Girnita L, Martegani MP, Natali PG and Bartolazzi A (2003) CD44s adhesive function spontaneous and PMA-inducible CD44 cleavage are regulated at post-translational level in cells of melanocytic lineage. *Melanoma Res* **13**, 325–337.
- 54 Catenacci DV, Cervantes G, Yala S, Nelson EA, El-Hashani E, Kanteti R, El Dinali M, Hasina R, Bragelmann J, Seiwert T *et al.* (2011) RON (MST1R) is a novel prognostic marker and therapeutic target for gastroesophageal adenocarcinoma. *Cancer Biol Ther* **12**, 9–46.
- 55 Jencks DS, Adam JD, Borum ML, Koh JM, Stephen S and Doman DB (2018) Overview of current concepts in gastric intestinal metaplasia and gastric cancer. *Gastroenterol Hepatol* **14**, 92–101.
- 56 Thapa R and Wilson GD (2016) The importance of CD44 as a stem cell biomarker and therapeutic target in cancer. *Stem Cells Int* **2016**, 2087204.
- 57 Jin C, Kenny DT, Skoog EC, Padra M, Adamczyk B, Vitizeva V, Thorell A, Venkatakrisnan V, Linden SK and Karlsson NG (2017) Structural diversity of human gastric mucin glycans. *Mol Cell Proteomics* **16**, 743–758.
- 58 Wang Y, Ju T, Ding X, Xia B, Wang W, Xia L, He M and Cummings RD (2010) Cosmc is an essential chaperone for correct protein O-glycosylation. *Proc Natl Acad Sci USA* **107**, 9228–9233.
- 59 Ju T, Lanneau GS, Gautam T, Wang Y, Xia B, Stowell SR, Willard MT, Wang W, Xia JY, Zuna RE *et al.* (2008) Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Can Res* **68**, 1636–1646.
- 60 Gomes J, Mereiter S, Magalhaes A and Reis CA (2017) Early GalNAc O-glycosylation: pushing the tumor boundaries. *Cancer Cell* **32**, 544–545.
- 61 Shi M, Dennis K, Peschon JJ, Chandrasekaran R and Mikecz K (2001) Antibody-induced shedding of CD44 from adherent cells is linked to the assembly of the cytoskeleton. *J Immunol* **167**, 123–131.
- 62 Bartolazzi A, Nocks A, Aruffo A, Spring F and Stamenkovic I (1996) Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. *J Cell Biol* **132**, 1199–1208.
- 63 Yao HP, Zhou YQ, Zhang R and Wang MH (2013) MSP-RON signalling in cancer: pathogenesis and therapeutic potential. *Nat Rev Cancer* **13**, 466–481.
- 64 Hurt-Camejo E, Rosengren B, Sartipy P, Elfsberg K, Camejo G and Svensson L (1999) CD44, a cell surface chondroitin sulfate proteoglycan, mediates binding of interferon-gamma and some of its biological effects on human vascular smooth muscle cells. *J Biol Chem* **274**, 18957–18964.
- 65 UniProt Consortium T (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **46**, 2699.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Confirmation of the efficiency of heparinase and chondroitinase digest.

Fig. S2. Haematoxylin and eosin (H&E) staining in sialyl Tn (STn) negative or positive gastric cancer region as well as in intestinal metaplasia.