Inês Barbosa Moreira

Bachelor of Science in Biology – Molecular Biology and Genetics

Impact of truncated *O*-glycans in cancer associated CD44 variants

Dissertation submitted in partial fulfillment of
the requirements for the degree of

Master of Science in
Biochemistry for Health

Adviser: Diana Campos, PhD, Junior Researcher, IPATIMUP/I3S

Co-adviser: Celso A. Reis, PhD, Principal Investigator, IPATIMUP/I3S

September, 2019

Inês Barbosa Moreira

Bachelor of Science in Biology – Molecular Biology and Genetics

Impact of truncated *O*-glycans in cancer associated CD44 variants

Dissertation submitted in partial fulfillment of
the requirements for the degree of

Master of Science in
Biochemistry for Health

Adviser: Diana Campos, PhD, Junior Researcher, IPATIMUP/I3S

Co-adviser: Celso A. Reis, PhD, Principal Investigator, IPATIMUP/I3S

Júri:

Presidente: Prof. Doutora Maria Teresa Nunes Mangas Catarino
Arguente: Prof. Doutora Maria Angelina de Sá Palma
Vogal: Doutora Diana Alexandra Vieira Campos

Departamento de Química da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

October, 2019

Impact of truncated *O*-glycans in cancer associated CD44 variants

All our dreams can come true, if we have the courage to pursue them –

Walt Disney

ACKNOWLEDGEMENTS

Throughout the writing of this dissertation, I have received a great deal of support. First, I would like to acknowledge Dr. Celso A. Reis for having received me in his laboratory and for giving me the opportunity to work in I3S/IPATIMUP, one of the best scientific research institutes in Portugal.

Above all, I would like to sincerely thank Diana Campos, my advisor, my mentor, my Jedi Master, for her patience and kindness in teaching me how to be a good scientist. Thank you for always seeing the potential I sometimes don't know I have. You are the source of inspiration to make scientific research my future career.

I would also like to acknowledge my colleagues from my Master at FCT-UNL and from the Glycobiology in Cancer team, particularly Filipe Pinto, for all the kind help with experiments. Also, I could not forget to thank the GlycoNinjas, for their amazing friendship. You were the source of fun during the insane year of my master's thesis. Thank you for choosing me as your friend, it's a real honor.

A special word to Porto, the wonderful city whose lights and fog forever stole my heart.

In addition, I would like to thank my beautiful parents and family. *Obrigada por me deixarem voar e seguir os meus sonhos, mantendo sempre os braços abertos para me darem colo e amor*! Also, to the best of friends in the history of best friends, Teresa e Sofia, for always supporting me, as only two dear clementines know how to do. Finally, to Manel, the love of my life. There are no words to describe how lucky I am to be with the smartest, bravest person I have ever met and how grateful I am that you finally found me. Thank you for loving and protecting me

ABSTRACT

O-glycosylation is a post-translational modification misregulated in cancer and often correlated with poor prognosis and low overall survival. Particularly, truncated *O*-glycans, Tn and STn, are known as pancarcinoma markers. Their presence in surface proteins is well established in gastrointestinal (GI) cancer, which is currently one of the leading causes of death worldwide. The hyaluronic acid receptor, CD44, has critical roles in carcinogenesis, including GI cancer. This heavily *O*-glycosylated protein contains a vast repertoire of alternatively spliced variants whose detection/presence is tested in this thesis. In colorectal and gastric cancer (GC) tissue cohorts, CD44v9 has a cancer-specific detection, showing co-expression with STn. Using GC cell line models, the presence of CD44 variants is also characterized, through different immunodetection methods. CD44 variants detection is being influenced by the expression of truncated *O*-glycans. In a *COSMC*$^{-/-}$ genetic background, with homogenous expression of Tn and STn structures, there is an increased protein detection of CD44v9. These particular CD44v9 glycoforms are also more sensitive to trypsin and display higher cleavage rates. This finding has special relevance in cancer

development, providing promising directions to identify a serum biomarker. Furthermore, it enables future work regarding the influence of CD44v9 glycoprofile on its malignant functions, adding higher level of specificity as a marker for GC diagnosis and therapeutic target.

Keywords: Tn; STn; gastrointestinal cancer; CD44v9; cancer biomarker.

RESUMO

A *O*-glicosilação é uma modificação pós-traducional que está desregulada no cancro, sendo frequentemente correlacionada com mau prognóstico e com uma redução global na sobrevida. Em particular, Tn e STn, *O*-glicanos truncados, são notórios marcadores globais de cancro. A sua presença em proteínas de superfície está bem descrita no cancro gastrointestinal (GI), que é uma das principais causas de morte no mundo, atualmente. O recetor do ácido hialurónico, CD44, desempenha um papel crítico na carcinogénese, incluindo no cancro GI. Esta proteína extensamente glicosilada contém um repertório amplo de variantes, resultante de fenómenos de *splicing* alternativo, cuja deteção/presença demonstra ser influenciada pela expressão aberrante de *O*-glicanos truncados, nesta tese. Em concreto, a presença de CD44v9 é específica de cancro em séries de tecidos de cancro gástrico e colorretal, mostrando coexpressão com STn. Adicionalmente, a presença de variantes de CD44 é caracterizada em modelos celulares de cancro gástrico. Num modelo genético de

COSMC $^{-/-}$, que apresenta uma expressão homogénea de glicanos Tn e STn, observa-se uma maior deteção de CD44v9. No mesmo modelo é demonstrado que o CD44v9 é mais sensível ao corte proteolítico da tripsina e, numa análise ao secretoma, tem maiores taxas de clivagem. Esta observação tem especial relevância num contexto de desenvolvimento de cancro, fornecendo orientações promissoras para identificar um novo biomarcador serológico. Assim, esta tese é o ponto de partida para estudos futuros sobre a influência do glicoperfil do CD44v9 nas suas funções malignas, sendo que esta associação proteína/glicanos truncados pode melhorar a sua especificidade como biomarcador de diagnóstico de cancro gástrico e alvo terapêutico.

Palavras-chave: Tn; STn; cancro gastrointestinal; CD44v9; biomarcador de cancro.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Abreviations

CD44	Cluster of differentiation 44
CD44ecd	Extracellular domain of CD44
CD44icd	Intracellular domain of CD44
CD44s	Standard CD44
CD44v	Variant CD44
CGDs	Congenital disorders of glycosylation
CRC	Colorectal cancer
CSC	Cancer stem cell
EMT	Epithelial to mesenchymal transition
ER	Endoplasmic reticulum
Fuc	Fucose
Gal	Galactose
GalNAc	*N*-acetylgalactosamine
GC	Gastric cancer
GI	Gastrointestinal
Glc	Glucose
GlcNAc	*N*-acetylglucosamine
HA	Hyaluronic acid
HNSCC	Head and neck squamous cell carcinoma
IF	Immunofluorescence
IHC	Immunohistochemistry
MAb	Monoclonal antibody
Man	Mannose
NAT	Normal adjacent tissue
NeuAc	*N*-acetylneuraminic acid
PD-L1	Programmed death-ligand 1
ROS	Reactive oxygen species
SC	Simple Cell
Ser	Serine
STn	Sialyl Tn
T	Thomsen-Friedenreich
TCGA	The cancer genome atlas
Thr	Threonine
TMA	Tissue microarray
Tn	T nouvelle antigen
WB	Western blot
WT	Wild type
xCT	Glutamate-cystine transporter

CHAPTER 1

GENERAL INTRODUCTION

1.1 CANCER

Cancer and Malignant Transformation

Cancer is a disease that involves dynamic changes in the genome, that has been the subject of a global focus of attention regarding scientific and clinical research. Despite recent major advances in the unraveling of cancer and the development of novel treatments, many patients die due to the disease [1]. The leading cause of cancer death is a failure of drug treatments, as cancer cells become resistant to those drugs, which consequently limits the efficiency of chemotherapeutic agents [2]. Moreover, this therapeutic failure results in tumor recurrence and metastasis, that corresponds to the spread of cancer cells from the primary tumor to surrounding or distant organs, which is the cause of approximately 90% of cancer deaths [3, 4].

To address the initial mechanisms underlying the emergence and development of malignant cells, Bishop and Weinberg identified in 1996 two classes of cancer genes that were altered in human and animal cancer cells. Those two classes of oncogenes were responsible for the production of cancer phenotypes in experimental models: some mutations produced oncogenes with dominant gain of function and some others originated tumor suppressor genes with recessive loss of function [5]. In humans, tumorigenesis is a multistep process that reflects somatic genetic alterations, driving the progressive transformation of normal cells into highly malignant byproducts [6].

There are more than 100 different types of cancer and within specific organs there are different cancer subtypes [7]. Cancer cells have defects in regulatory circuits, that usually control normal cell proliferation and homeostasis. In this way, the complex process that provokes malignant transformation is associated with a series of questions, including which and how regulatory circuits are affected.

It is important to acknowledge tumors as complex tissues, since cancer cells not only have suffered critical mutational alterations, but also have recruited and subverted normal cells to serve as active collaborators in their neoplastic agenda. In fact, the essential changes in cell physiology that dictate malignant growth involve a set of diverse mechanisms. Six acquired alterations were first described to cause the vast catalog of cancer cell phenotypes that exist, including auto-generation of growth signals, resistance to growth suppressors, unlimited replicative capacity (immortality), evasion of programmed cell death (apoptosis), induction of angiogenesis, and capability of tissue invasion and metastization [8]. Later on, in 2011, conceptual progress added two emerging hallmarks that are involved in the pathogenesis of virtually all cancers: capability to modify and

reprogram cellular energetic metabolism and to evade immunological destruction [9]. Additionally, there are two consequential characteristics of neoplasia underlying both core and emerging hallmarks: genome instability, responsible for the genetic diversity that drives their acquisition, and inflammation, which nurtures multiple hallmark capabilities [9].

The interaction between the malignant cells and all the supportive partners and features of the tumor microenvironment is absolutely critic to understand cancer pathogenesis and to develop novel and more effective therapies. Therefore, acknowledge that there is extensive applicability of these concepts will progressively affect the development of new means to treat human cancer.

Gastrointestinal Cancer

As previously mentioned, cancer is the global notation of a vast and heterogeneous set of diseases characterized by the uncontrolled growth of abnormal cells within a tissue. It is still the greatest challenge of modern medicine, responsible for 9.6 million deaths only in 2018 [10].

Gastrointestinal (GI) cancer is one of the most significant causes of cancer death. In Portugal, colorectal cancer (CRC) and gastric cancer (GC) are the third and sixth more incident types, respectively. In terms of mortality, the scenario aggravates itself: CRC places in second and GC in fifth (Figure 1) [11]. The diagnosis of both gastric and colorectal cancers entails invasive procedures, such as endoscopy and colonoscopy, respectively.

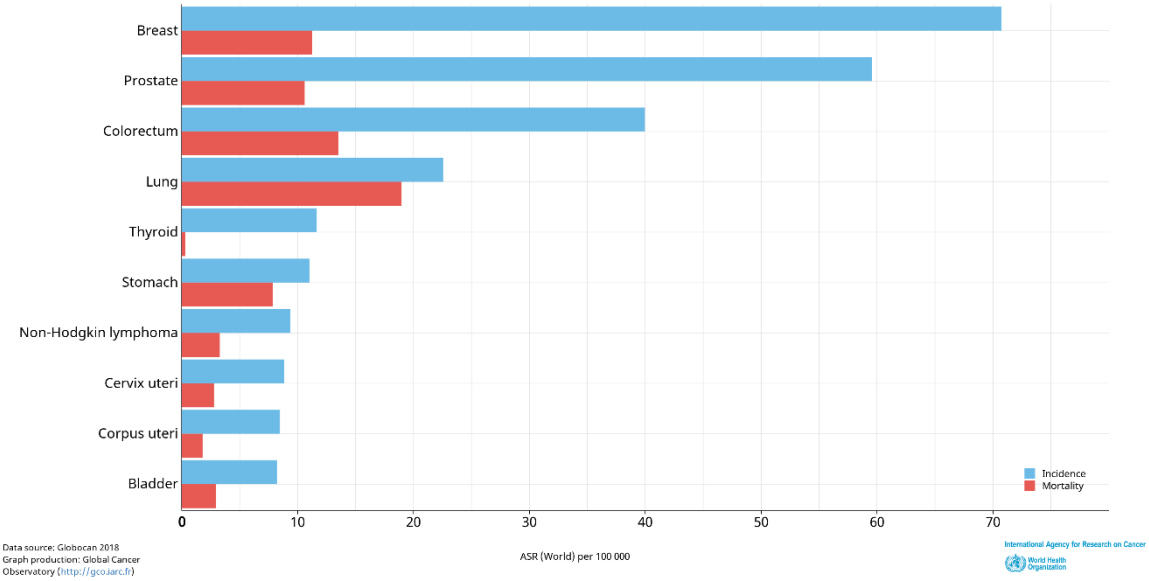


Figure 1. Estimated age-standardized incidence and mortality rates (World) in 2018, Portugal, both sexes, all ages [10, 13].

Clinical Aspects of Gastric Cancer

Gastric cancer is one of the major threats to public health worldwide, with >780,000 deaths every year [12], and with over 1 million new cases in 2018 [13]. It has overall poor prognosis, with a 5-year survival rate of 20%, which reflects the lack of knowledge of its etiological factors and pathogenesis [14]. Moreover, early-stage stomach cancer rarely causes specific symptoms. This is one of the main reasons GC is so hard to be early detected. Most of the symptoms include nausea, poor appetite and weight loss, which are more likely to be caused by things other than cancer, such as a stomach virus or an ulcer [15]. Therefore, GC is associated with a consequent late diagnosis and lack of effective treatments [14]. In terms of risk factors, the main ones include infection with *Helicobacter pylori (H. pylori)*, family history of stomach cancer or obesity [16].

As said before, delays in diagnosis still seem to be common, virtually affecting the stage of GC at diagnosis as well as patient's outcome. In most countries, where a screening program is not feasible, the diagnosis of GC and the

selection criteria for endoscopy depends on alarm symptoms reported by the patient, that are not sufficiently sensitive to detect malignancies [17].

Gastric adenocarcinomas represent more than 95% of gastric malignancies. Since 1965, the histological classification of GC has been largely based on Lauren's criteria, in which intestinal type and diffuse type adenocarcinoma are the two major histologic subtypes, plus indeterminate type as uncommon variant [18]. The relative frequencies are approximately 54% for intestinal type, 32% for diffuse type and 15% for the indeterminate type [19]. In the intestinal phenotype there is preservation of the glandular structure and cell polarity, whereas in the diffuse type these are lost [20]. The intestinal type is the most common form of GC and is associated with several risk factors, one of them being a cascade of neoplastic events triggered by *H. pylori* [16].

The diffuse type malignancies have no association with pre-neoplastic events and are thought to arise spontaneously. Tumors have poorly differentiated cells, with an infiltrative profile, resulting from low cell-cell adhesion related to E-cadherin loss of function [21, 22].

In terms of metastasis, intestinal type GC preferentially metastasizes haematogenously to the liver, whereas the diffuse type migrates to the initial segment of the small intestine and establishes metastasis in peritoneal surfaces [18, 21].

In 2014, The Cancer Genome Atlas (TCGA) Research Network published an effort to further dissect the differences between GC subtypes, aiming to find the key players of the distinct types of GC, consequently improving the repertoire of available treatments [23]. By characterizing its molecular and genomic basis, they reached four classes of GC:

i. Epstein-Barr virus positive tumors, which show DNA hypermetilation and increased PD-L1 expression;
ii. Microsatellite unstable tumors, resulting from germline mutations in DNA mismatch repair genes;
iii. Chromosomally unstable tumors, with increased aneuploidy;
iv. Genomically stable tumors, usually showing E-cadherin alterations.

Regarding treatment options, several pathways have been identified for each subtype described in the gastric cancer TGCA project, relatively different from each other and potentially targetable. Over the last decade, a number of novel agents of therapy have been examined in clinical trials, with largely disappointing results. This could be explained by the lack of molecularly selected trial populations or the weak predictive biomarkers that exist, within the context of a highly heterogeneous disease [24]. Despite that, the main treatment for GC patients continues to be surgery, either total or subtotal gastrectomy. Depending on the stage of the disease, chemotherapy or chemoradiation may be given before surgery, to try to shrink the tumor and make it easier to remove, and also as a treatment after surgery. In the metastatic setting, surgery should only be considered for palliation of symptoms [25]. Regardless of the heterogeneous features of this disease, all GC patients should be discussed at a multidisciplinary team meeting, concerning treatment options. Moreover, clinical trials and newer treatments may be an option and should always be considered.

Clinical Aspects of Colorectal Cancer

CRC is responsible for >860,000 deaths per year and, only in 2018, more than 1.8 million new cases have been diagnosed worldwide [26]. Unlike other types of cancer, it has not been identified a single risk factor accounting for the

majority of cases. In epidemiological studies, family history of CRC, inflammatory bowel disease, obesity, among others, have showed association with cancer incidence [27].

Colorectal cancer, sometimes called colon cancer, combines both colon and rectal cancers, which begins in the rectum. CRC usually begins as noncancerous polyps formed on the inside of the colon, that may be small and produce few, if any symptoms. For this reason, it is recommended regular screening tests to help prevent CRC, by identifying and removing polyps before they progress to cancer. If CRC do develop, symptoms like rectal bleeding, persistent abdominal discomfort or unexplained weight loss may appear [28]. In terms of treatment, there are many options available that help to control the disease, including surgery and radio, chemo and immunotherapy [29, 30].

CRCs are classified in terms of local invasion depth, lymph node involvement and presence of distant metastases. Also, they are discriminated accordingly between stages from I to IV. The presence of microsatellite instability or deficiencies in DNA mismatch repair genes can also be used to classify tumors. These genetic features result in higher immunoreactivity and immunotherapy has been tested for this cancer type [31].

CRC is associated with a cascade of malignant events which transform the normal epithelia into a cancer tissue. The inactivation of the adenomatous polyposis gene (*APC*), usually through somatic mutation, is the first step of the sequence of carcinogenesis, resulting in the formation of early adenomas. The second event is the acquisition of a mutation in the Kirsten-ras gene (*KRAS*), which causes the formation of a mature adenoma. Finally, mutations in the *Tp53* gene lead to loss of p53-mediated pathways of apoptosis and there is progression to carcinoma (Figure 2) [32].

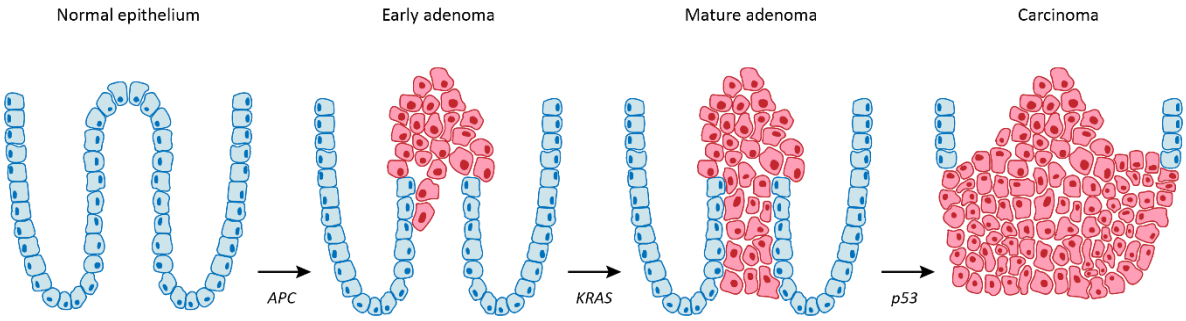


Figure 2. Schematic representation of colorectal carcinogenesis cascade. Adapted from [32].

1.2 KEY CANCER BIOMARKERS AND TREATMENT TARGETS

Currently Used Markers and New Potential Markers

Nowadays, antibodies are being used in the medical treatment of diseases, such as cancer. They are a remarkable class of proteins that recognize molecular shapes with unmatched specificity and range. Many of the current biomarkers recognize glycan structures or glycoproteins, a feature that will be reviewed in section 1.3. Some examples of markers for gastric and colorectal cancer treatment and clinical tests include CEA (Carcinoembryonic antigen), CA-125 (MUC 16), CA19-9 (sialyl-Lewis a), CA 72-4 (STn), Tn-antigen and IGF-1R (Insulin-like Growth Factor I Receptor) [33 - 36].

If there is a high level of a tumor marker, it usually suggests that cancer may be present in the body. However, by itself, a high tumor marker level is not enough to make a diagnosis, being one of the reasons why international guidelines have not yet accepted tumor markers in the process of diagnosis of gastric and colorectal cancers [37].

Today's technological age allows the usage of antibodies for virtually any target. Therefore, it is crucial to do an informed decision when choosing cancer specific therapeutic targets. One of the most promising approaches is to identify the cells with tumor-initiating capacities. Cancer cells within a tumor have a variety of fates that can range from death, proliferation, dormancy and/or exhaustion of clonal growth capacity [38]. However, the cellular origin of most solid tumors remains largely unknown. It has been speculated that distinct subtypes of cancer arise from different cells of origin at the time of tumor initiation.

In a normal context, adult stem cells have an important role in physiological processes in many different somatic tissues [39]. There are three different proprieties shared by all stem cells: self-renewal (i.e., at cell division, one or both daughter cells retain the same biological properties as the parental cell); the capacity to differentiate into multiple lineages; and the potential to proliferate extensively [40]. In fact, tumor cells have high proliferative capacity, phenotypic plasticity and aberrant differentiation [38] and the similarities with stem cells have given rise to the hypothesis that tumors should arise from undifferentiated stem or progenitor cells. Therefore, malignant cells harboring stem cells' features have been termed "cancer stem cells" (CSCs).

The CSC model proposes that tumors, like normal tissues, are organized in a cellular hierarchy. In this organized cellular system, CSCs are the only cells with unlimited proliferation potential and, therefore, capable of driving tumor growth and progression, cancer cell invasion and metastasis, and resistance to various forms of therapies, including radio, chemo and immunotherapy [40, 41].

Therapy failure and relapse should not be solely attributed to the acquisition of novel mutations, but also to the surviving cells' ability to regenerate the malignant tissue. In fact, some studies demonstrated that CSCs become enriched following therapy, suggesting a selection process [38]. CSCs display higher resistance to cytotoxic chemotherapy, when compared to non-CSCs, related to increased levels of drug transporters, enhanced DNA-damage repair mechanisms and the ability to escape the cytotoxic chemotherapy by maintaining a quiescent state. Combined with the capability of CSCs to evade the immune system [41] and to activate epithelial to mesenchymal transition (EMT) programs [20], it is clear that they are responsible for cancer aggressiveness, drug resistance and tumor relapse [38].

Gastric and colorectal CSCs are the basis for the onset of each correspondent type of cancer. They may be derived from bone marrow mesenchymal stem cells, or gastric/colorectal stem cells, respectively.

Therapies that specifically target CSCs hold great promise for improving survival and quality of life for cancer patients. In various cancers, these cells are mainly identified using cell-surface markers expressed on corresponding normal stem cells. Many specific molecular markers have been used for identification and isolation of gastric and colorectal CSCs, such as CD44, CD133, Lgr5 and ALDH1 [42, 43]. These cells are highly tumorigenic, chemoresistant and radioresistant, being critical in the metastasis and recurrence of GC and CRC and disease-free survival [44]. However, not all the receptors named above gather consensus on their usage as cancer biomarkers. For instance, controversial results were reported where the $CD133^-$ cells where considered to be more aggressive [45, 46]. Also, $Lgr5^+$ cells have been detected in both the population of normal intestinal stem cells and gastric cancer cells, in human, which is important to consider when the directed treatment is based exclusively on this biomarker [47]. Finally, observations were reported in which loss of ALDH expression was correlated with advanced stage of CRC [48].

Among the several known biomarkers, CD44 has unique features that make it one of the most promising tumor markers, with potential diagnostic and therapeutic applications. Therefore, those features will be carefully review in the next sections, being CD44 the focus of the work developed in this thesis.

Future therapies for gastric and colorectal cancers include the development of new targeted treatment therapies against membrane receptors, some of them already mentioned, to ultimately achieve the purpose of curing cancer patients. Moreover, molecular screening could help to fully characterize

cancers, identify the best prognostic panel of biomarkers and recognize patients suitable for specific targeted treatments. Nevertheless, the objective of future research is to detect biomarkers that can provide cost-effective and non-invasive cancer diagnosis.

CD44 as a CSCs Marker

Interestingly, the first demonstration of CSCs in a solid tumor (breast cancer) showed that they were $CD44^+/CD24^-$ [49] and the research for CSC-specific markers gained momentum [38]. As already mentioned, markers for CSCs are associated with cancer-progeny, but CD44 has been shown to be present in CSCs of several types of malignant tissues, such as head and neck [50], gastric [51, 52] and colon cancer [53, 54]. CD44 (Cluster of Differentiation 44) was first identified as a lymphocyte homing receptor and is expressed in several cell types, being the major receptor that binds hyaluronic acid (HA) [55]. The binding of CD44 to HA influences many functions of the protein, often related to cell-extracellular matrix crosstalk. It is a complex surface glycoprotein displaying high molecular weight heterogeneity, due to alternative splicing of a single *CD44* gene transcript and its differential glycosylation.

CD44 Alternative Splicing

Alternative splicing occurs in the great majority of human genes and is an intricate process where exons are removed (spliced out) in the pre-mRNA molecule, being often misregulated during cancer progression [56].

The genetic annotation of the *CD44* gene has been a controversial subject. It is consensually accepted that the full-length *CD44* gene contains 19 exons, of

which exons 1 to 5 and 15 to 19 are spliced together, encoding the two constant regions of the protein. Exons 6 to 14 can be alternatively spliced and assembled with the two constant regions to generate CD44 variants (CD44v), v2 to v10 (Figure 3). When only the constant mRNA regions are transcribed, the ubiquitously expressed standard isoform of CD44 (CD44s) is generated. In some published work and revisions, it is claimed that CD44 has a 20th exon, exon 5a, relative to the v1 isoform. However, in human, exon 5a encodes a stop codon and is not expressed [56].

CD44s has a protein domain (first 3 exons) that binds to hyaluronic acid (HA), a stem region (exons 4, 5, 15, and 16), a transmembrane (exon 17) and a cytoplasmatic tail (exons 18 and 19) (Figure 4A). The variant exons encode protein domains in the proximal plasma membrane extracellular region but the regulation of the process that generates the different forms is not completely understood [57]. As CD44s is ubiquitously expressed, its usefulness as a CSC marker may be limited. Special attention has been given to CD44v isoforms as CSC surface markers, for isolating and enriching CSCs in different types of cancer.

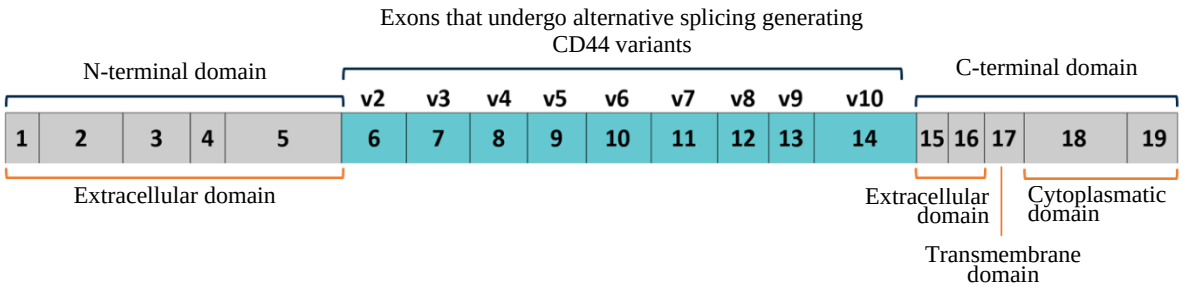


Figure 3. *CD44* Open Reading Frame and exon/protein domains correspondence. Adapted from [56].

CD44 Surface Cleavage

Shedding of cell-surface adhesion receptors is one of the modulator processes of cell–cell and cell–matrix adhesion, as well as in signaling pathways activation, which can transduce information of the crosstalk between the cell and its microenvironment. CD44 is a substrate of several proteases and can suffer sequential cleavages, resulting in the release of its extracellular domain (CD44ecd) and its intracellular domain (CD44icd). When the protein is intact, CD44ecd interacts with the extracellular matrix and the CD44icd associates with the actin cytoskeleton [55]. Upon extracellular cleavage, CD44icd can localize to the nucleus and regulate transcription of target genes that promote tumor development, including *CD44* itself. This establishes a positive feedback loop that can regulate *CD44* expression (Figure 4B) [58].

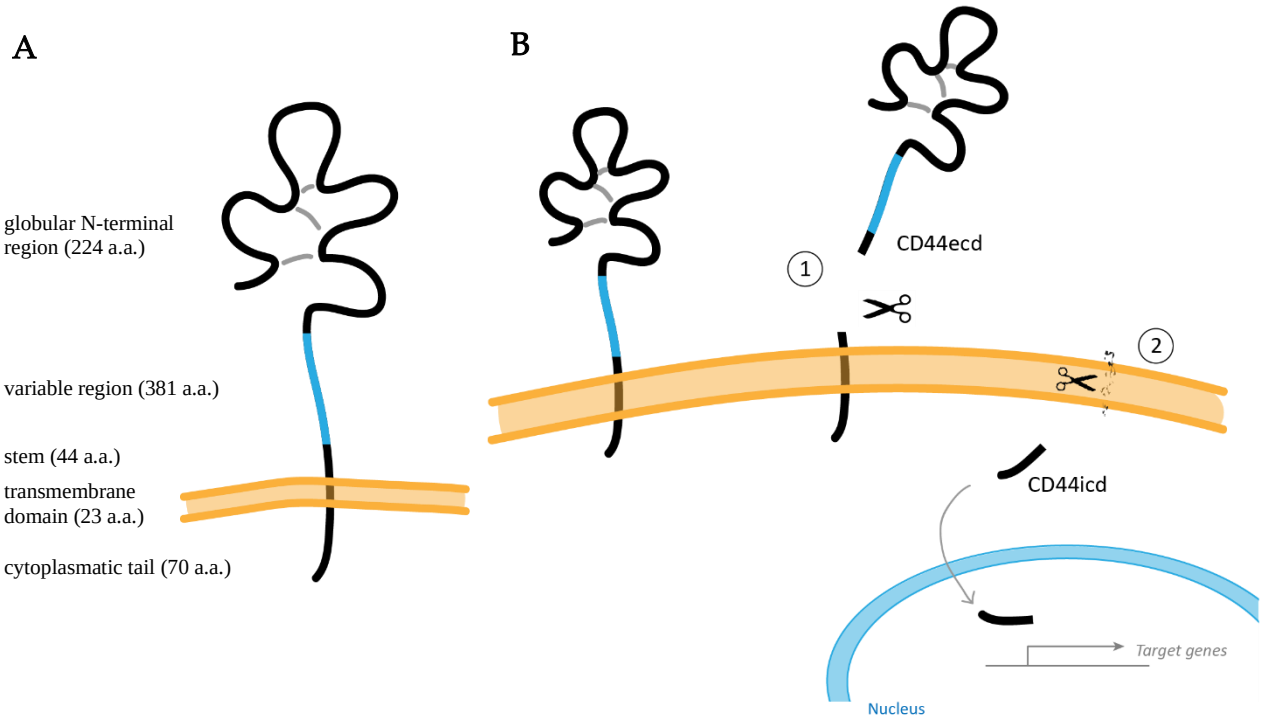


Figure 4. Schematic illustration of a full-length CD44 protein. **(A)** CD44 domains, relative positions and sizes are represented. The position of the three disulfide bonds is shown in grey. Adapted from [56]. **(B)** CD44 extracellular cleavage, that results in ① release of extracellular domains and, after a transmembrane cleavage, in ② release of intracytoplasmatic domain and its translocation to the nucleus.

CD44 shedding is observed in a variety of human tumors *in vivo*, including colon, and a strong positive correlation between the presence of CD44v and cleaved CD44 was observed [59]. As a consequence of cleavage, serum (cleaved) CD44 could be used as a biomarker of tumor burden, as it was suggested for gastric and colorectal cancer in the 90's [60].

Therapeutic applications based on CD44

As described in this thesis, CD44 has several features that make it a promising therapeutic target for cancer treatment. Its clinical applications are diverse and encompass different types of cancer. Antibodies targeting different epitopes of CD44 have been generated and used in pre-clinical studies. Roche has developed a novel functional monoclonal antibody (MAb) (RO5429083) which targets a glycosylated, conformation dependent epitope of CD44 in the constant region of the protein [61]. Based on pre-clinical studies it was approved and it is being currently used in clinical trials in several types of cancer, including GC and CRC [62].

The MAb U36 was shown to target the squamous cell-specific CD44v6 [63]. This clone was labeled with indium-111 and intravenously injected to nude mice bearing head and neck squamous cell carcinoma (HNSCC) xenographs, expressing CD44v6, and the MAb U36 had significantly higher uptake in tumors [64]. However, not all CD44-based therapeutic applications showed successful results. An example is the CD44v6-specific MAb, Bivatuzumab, conjugated with mertansine, a drug designed to treat patients with incurable HNSCC. The main limiting toxicity of this CD44-targeted drug delivery compound in Phase I clinical studies was severe skin reactions (desquamation), including one fatal case of toxic epidermal necrolysis due to massive apoptosis of skin keratinocytes.

These complications resulted from an abundant 'off-target' expression of the CD44v6 antigen in normal squamous cells, with the main toxicity of bivatuzumab mertansine being directed against the skin, and caused an early termination of clinical development of the MAbs [65].

Relevance of CD44 in Gastrointestinal Cancer

CD44 several forms (CD44s and the multiple CD44v) are one of the most promising targets for therapy, being of great functional importance in mechanisms of cancer development.

CD44 variants have been reported in gastrointestinal cancer and their expression was found to correlate with poor prognosis in patients. Back in 1993, Heider *et al.* described the expression of CD44v in sporadic gastric cancer tissues [66] and Mayer *et al.* demonstrated the positive correlation between CD44v9 expression at time of diagnosis and distant metastization in primary gastric carcinomas [67]. In seminal work, Ishimoto *et al.* demonstrated that the presence of CD44v conferred resistance to ROS in a panel of GI cancer cell lines. The same group had published, one year before, work where they found that CD44v (v6–10, v7–10, and v8–10) were preferentially expressed in gastric tumors of Gan mice (a genetic model for gastric tumorigenesis) [68]. Following those findings, they presented evidence that the CD44v isoforms interact with a glutamate-cystine transporter, xCT, to control the intracellular glutathione levels, which are related to protection against ROS. Indeed, forced expression of CD44v8–10, but not that of CD44s, promoted ROS resistance, enabling proliferation of gastric CSCs [69]. Lau *et al.* isolated CD44+ CSCs from patient tumor samples and observed that CD44v8-10 was the major CD44 variant found in those cells. Furthermore, they confirmed the previous findings relating CD44v8-10 with ROS resistance [70].

CD44v9 has been used in several cancer types, commonly as a CSC marker, with its presence often associated with poor prognosis and survival. In pancreatic cancer and adult T cell leukemia *in vitro* assays CD44v9 showed co-expression with the multidrug resistance protein 1, which conferred resistance to therapy [71, 72]. In duodenal adenocarcinoma tissue samples, CD44v9 positive cells showed low mitotic activity and increased resistance to apoptosis [73]. In ovarian cancer, the presence of cleaved CD44v8-10 in ascites fluid was associated with metastization [74]. CD44v9 is highly expressed in GC and consistently it is observed that its presence is related with elevated tumor sizes, poor prognosis and survival [75, 76]. Moreover, it has been identified as a predictor of tumor recurrence in several GC cohorts [75] and silencing of CD44v9 expression is a novel pathway for treating GC [77]. Nevertheless, its role in colorectal cancer has also deserved focus. Like in GC cell models, CD44v9+ cells generate aggressive tumors in tumorigenic assays using CRC cell lines [54]. The expression of CD44v9 was observed in circulating human CRC cells, that have CSC potential, and it was shown to be a potent marker in predicting recurrence, prognosis, and treatment efficacy [78]. The mechanistic importance of CD44v9 reveals a clear and promising target in biomarker discovery and therapeutic strategies in GI cancer.

1.3 GLYCOSYLATION

Overview

Glycosylation is the enzymatic process responsible for the attachment of glycans (carbohydrates) to proteins, lipids or other saccharides [33]. This major post-translational modification, that can act as a key regulatory mechanism controlling both physiological and pathological processes, occurs in the Endoplasmic Reticulum (ER)/Golgi compartment of essentially all cells. It is mediated by the coordinated action of a portfolio of different glycosyltransferase and glycosidase enzymes that, in a series of steps, form carbohydrate structures. Most of these steps depend on substrate availability, enzyme activity, levels of gene transcription and enzyme location within organelles [79]. The glycome, that comprehends the entire complement of glycan structures in an organism, differs from the genome and the proteome. In fact, in terms of expression, the glycome is produced in a non-templated manner and is intricately controlled at multiple levels in the ER and Golgi apparatus. The complexity is further increased with metabolism-dependent transport of monosaccharides, feed-back loops that control glycogene-expression and even the physical conditions within the ER and Golgi lumen features [80]. Thus, Glycobiology is the science that studies the intricate processes that gives rise to structure, biosynthesis and biology of glycans and glycoconjugates.

Glycan vast diversity arises from differences in monosaccharide composition, in linkage between monosaccharides, in branching structures, among others. In fact, glycoconjugates are primarily defined according to the nature of their non-glycosidic part, as well as the linkage between them [81, 82].

Within the main glycoconjugates are included the glycoproteins, where one or more glycans are covalently attached to a polypeptide backbone predominantly via nitrogen (*N*) and oxygen (*O*) linkages, in which case they are known as *N*-glycans or *O*-glycans, respectively [81, 83]. Other major classes include proteoglycans and glycosphingolipids (Figure 5) [33].

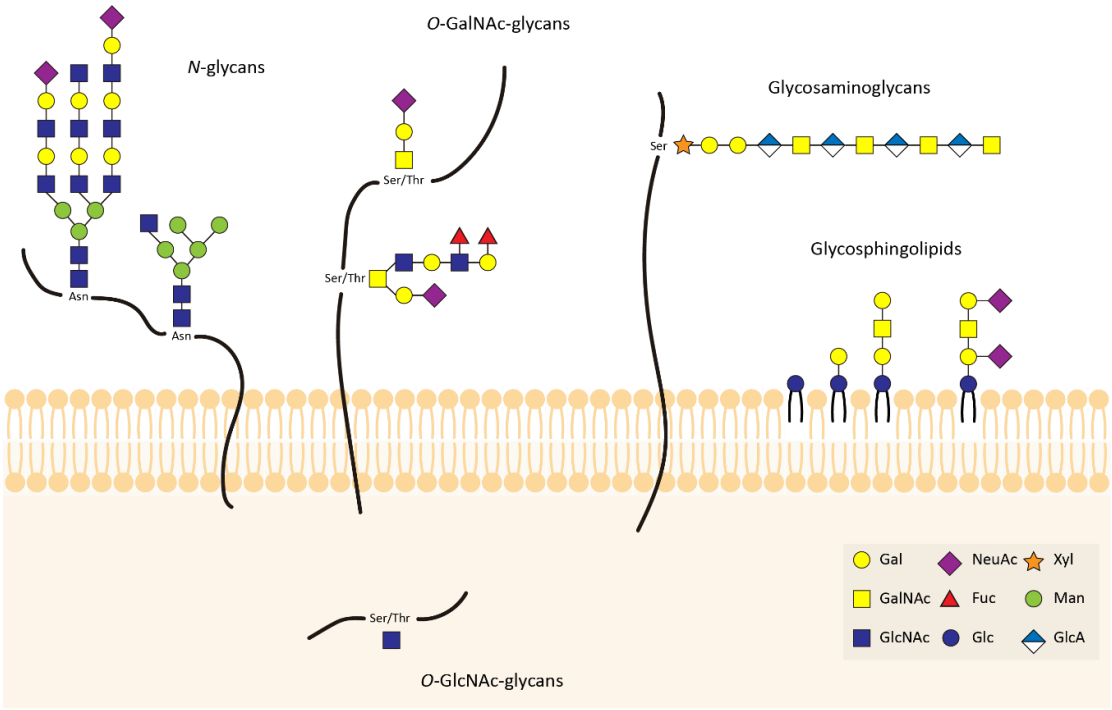


Figure 5. Schematic representation of the common classes of glycoconjugates in mammalian cells. Adapted from [33].

The most common constituents of glycans are hexoses, that are composed of six carbon atoms [84]. Glucose (Glc) is a convenient reference compound when comparing these structures, as many of them can be derived from glucose by a single epimerization (a change in the stereochemical configuration of a single carbon atom in a sugar), substitution or addition/loss. For example, mannose (Man) derives from an epimerization at the 2 position, while epimerization at the position 4 yields galactose (Gal). On the other hand, a substitution of the 2-hydroxyl group of glucose or galactose for an acetylated amino group gives rise

to *N*-acetylglucosamine (GlcNAc) or *N*-acetylgalactosamine (GalNAc), respectively, while the loss of the 6-hydroxyl group and a change of configuration creates fucose (Fuc). Finally, it is important to mention the existence of sialic acids, a large family of sugars. *N*-acetylneuraminic acid (NeuAc), the most common form, is a nine-carbon sugar acid that is usually found in a six-member ring configuration. The several and unique substituents extended from the ring will give rise to other forms of sialic acid [81].

The sugar components of glycoconjugates modulate a variety of functions in physiological and pathophysiological states, in addition to forming important structural features [80]. On the outmost surfaces of cellular and secreted macromolecules, where most glycans are found, they regulate receptor turnover, affinity and signaling [85]. These highly dynamic protein-bound glycans are also abundant in the nucleus and cytoplasm of cells, exerting regulatory effects [86]. Of the many molecular mechanisms that drive the formation of glycoconjugates, the next section will focus on the *O*-glycosylation, which is the most relevant for the scope of the work presented in this thesis.

O-type Glycosylation

O-glycosylation consists in the addition of a sugar molecule to the oxygen atom of serine (Ser) or threonine (Thr) residues in a protein. There can be several types of *O*-glycosylation, depending on the monosaccharides that are first attached to the protein amino acids and the type of extension that occurs from there. In humans, the monosaccharides that most commonly initiate *O*-glycosylation are GalNAc and GlcNAc [87]. The initial findings about this type of glycosylation were on mucins, heavily *O*-glycosylated proteins found in mucous secretions or in the membrane of epithelial cells [88]. These glycoproteins

have peptide repeats rich in *O*-GalNAc acceptors, which enables the conjugation of glycan structures, through *O*-GalNAcylation [80]. This type of *O*-glycosylation is one of the most diverse forms of posttranslational modification involving up to 50-100 distinct genes. The first step alone is performed by up to 20 polypeptide GalNAC-transferases [83] and, from there, the extension process can give rise to several different structures (Figure 6). The single GalNAc attached to a Ser or Thr is called the Tn antigen. When the Tn antigen is elongated with a ß1,3-Gal, by the C1GalT1 transferase, the T antigen structure is formed, which can be further elongated, resulting in a core 1 *O*-glycan. C1Gal-T1 β3galactosyltransferase is unique in requiring a private chaperone COSMC for folding and catalytic function. Core 2 *O*-glycans contain a branching GlcNAc in the T structure, that can also be extended. Core 3 and 4 structures have one or two GlcNAc-initiated branches, respectively, starting from the Tn antigen [87]. Although cores 1 to 4 correspond to the four most common *O*-GalNAc glycan core structures found in mucins, there are an additional four, designated cores 5 through 8 [89].

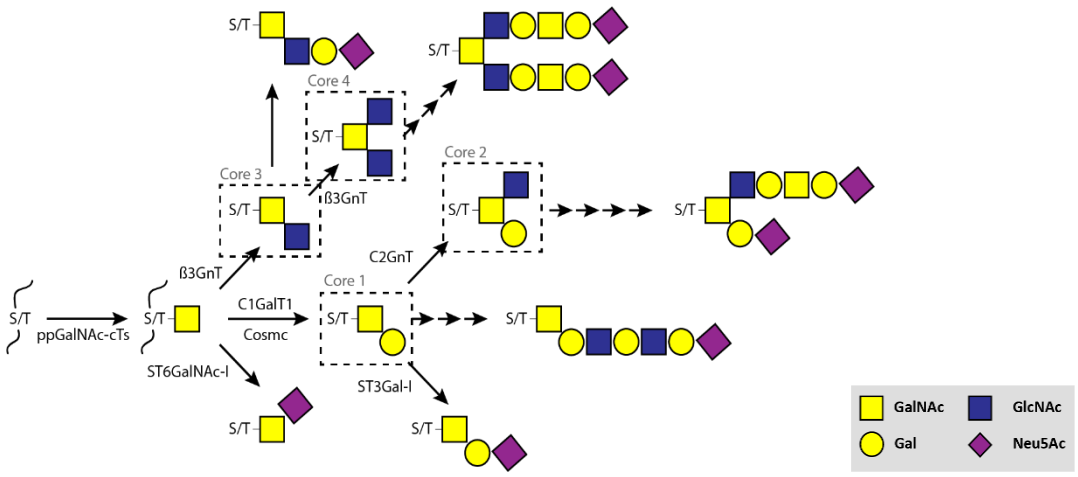


Figure 6. Schematic representation of the elongation process of *O*-GalNAcylation. Adapted from [90]

The elongation of the many initial core structures is carried out through the action of different glycosyltransferases that extend the existing structure with Gal, GlcNAc, sialic acid or even fucose, as the protein moves through the *cis*-, *medial*- and *trans*-Golgi compartments [87]. These GalNAc-linked glycan structures are abundant on many secreted and extracellular proteins and have been shown to participate in several key biological processes of organ development and function [88]. The extensive glycosylation protects both the glycoproteins and cellular surfaces from external stress, as it forms an important interface between epithelial cells and the external mucosal surfaces of the body. For instance, in secreted mucins on epithelial cells, *O*-glycans are responsible for hydration of the layer [91]. It is well established the role of mucin *O*-glycans in selectin-mediated cell adhesion in immunity [92]. Also, knockout mice for the C1GalT1 transferase died by embryonic day 14, due to impaired vasculature formation [93].

O-GlcNAc-linked glycans are typically found on intracellular compartments, including the cell nucleus, mitochondria and cytoplasm. Unlike the mucin-type *O*-glycans, the addition of GlcNAc does not occur in the Golgi apparatus and is typically not extended. Instead, it is regulated through the action of the *O*-linked GlcNActransferase (OGT) and the *O*-linked GlcNAcase (OGA), that perform a rapid cycle of addition and removal of GlcNAc from protein substrates [94]. This process is known to compete with protein phosphorylation at Ser and Thr residues, impacting several cellular regulatory functions, including cellular metabolism [95].

Briefly, another type of *O*-glycosylation consists in monosaccharides attached via *O*-mannose. This mannosyltransfer is catalyzed by a conserved

family of protein *O*-mannosyltransferases (PMTs) and takes place in the endoplasmic reticulum (ER) [96].

Glycosylation in disease

At this point, it is clear that glycosylation is critical for both physiological and pathological cellular functions [97]. Indeed, there are many examples of diseases associated with aberrant glycosylation or failure of carbohydrate recognition systems [98]. Being genetic in origin or acquired through somatic mutations, they can affect virtually all cells in the human body, having a systemic impact [99].

Congenital disorders of glycosylation (CDGs) are inherited conditions that are quite rare, being often embryonic lethal [100, 101]. However, they have been recognized due to changes in glycosylation of serum glycoproteins, such as the iron-binding blood plasma transferrin, that controls the level of free iron in biological fluids [102]. CDGs can be categorized into two different types, both derived from mutations in enzymes for *N*-glycans synthesis [103]. In the first one, normal *N*-glycans are attached to serum glycoproteins but in a small percentage of all the potential *N*-glycosylation sites. The second group consists of syndromes in which all or most of the usual glycosylation sites are utilized, but the attached glycans are reduced in size and complexity [104, 105]. Although these syndromes provide limited insights into the exact functions of glycans, they confirm that complex glycans have crucial roles in development [106].

Another disorder that is genetic in origin is the von Willebrand factor deficiency. It is a relatively common bleeding disorder that results from a two- to five-fold reduction of this factor in serum levels. The von Willebrand factor is essential for initiate the blood-clotting cascade but, in case of disease, the small

amount of the factor that is present, although having a normal activity, leads to excessive bleeding following wounding. Glycosylation plays a critical role regarding this deficiency. Galgt2 *N*-acetylgalactosaminyltransferase, an enzyme normally found in epithelial cells, can suffer a mislocalization to endothelial cells, due to a change in the coding region of the GalNAc transferase gene. This alteration occurs at a single genetic *locus* and the dominant mutant gene is inherited in simple Mendelian fashion [107]. Therefore, von Willebrand factor, that is made in endothelial cells, where it normally does not encounter that glycosyltransferase, will receive terminal GalNAc residues. This abnormally glycosylated factor will be then cleared from circulation. The mislocalization of GalNAc transferases is probable to cause aberrant glycosylation of not only the von Willebrand factor but several different proteins in the endothelial cells [108].

Further examples involve chemical reactions of sugars, immunological responses to them and diseases in which glycosylation changes are correlated with disease phenotype, but where a causal connection is not yet clear. In the normal forms of glycosylation, the addition of glycans to glycoproteins requires the action of specific glycosyltransferases. However, a direct chemical reaction of monosaccharides with proteins can occur. These so-called glycation reactions are generally inefficient, and their products have a very distinct nature when compared to the chemical features of *N*- and *O*-glycans [109]. Therefore, they are only observed under special circumstances, such as when glucose levels become very high in diabetic patients or in proteins that have long half-lives.

The impact of glycans extends into immunity and inflammation. As one would expect, there are glycoconjugates found on the surface of innate and adaptive immune cells, that can sense environmental signals. On the other hand, the immune receptors themselves recognize pathogen-associated glycans

expressed on the surface of microorganisms, including bacterial lipopolysaccharides, peptidoglycans, fungal mannans and teichoic acids [110]. Glycans are known to have different roles that are essential for the differentiation of B and T cells. In fact, the immunoglobulins that are secreted by B cells, and vital components of humoral immunity, are glycoproteins with specific glycosylation patterns [111, 112]. Defects on these patterns or alterations in glycan composition can contribute to immune dysregulation, leading to a range of autoimmune, infectious or chronic inflammatory diseases, such as systemic lupus erythematosus [113], HIV infections [114] or rheumatoid arthritis [115].

Another rare blood disorder of *O*-glycosylation, that is acquired and permanent, is the so-called Tn syndrome [116]. It is characterized by the presence of Tn immature antigens on glycoconjugates of all hematopoietic cell lineages and can affect both males and females at any age. In this disorder, the Tn antigen is over-represented due to a clonal somatic mutation in *COSMC*, which encodes a molecular chaperone essential for C1GalT1 function. This reduces the amount of active C1GalT1, leading to a decrease in the galactosylation of terminal GalNAc, thus accumulating Tn antigens [117, 118]. These truncated glycan structures are recognized by IgM antibodies, causing different degrees of anemia, leukopenia and thrombocytopenia in affected patients [116].

Finally, one of the subjects that deserves special attention in terms of glycobiology is Cancer. There are several well-known alterations in glycoconjugates associated with the malignant process, that will be discussed in detail further in the next section.

Alterations in the glycosylation can have remarkably subtle and indirect effects. Although these changes might be global, phenotype associated with such alterations are often most evident for one or a few proteins. To uncover the

defects and disorders involving glycosylation will bring us closer to understand its basic mechanisms, allowing the association between specific glycoconjugates and disease phenotypes [86]. Besides, knowing the importance of glycans in human cellular processes opens new possibilities in terms of diagnosis and treatment of several diseases.

1.4 GLYCOSYLATION IN CANCER

Tumor-associated Alterations

Cancer cells often present a range of glycosylation alterations that lead to oncofetal phenotypes, as they resemble glycoprofiles in early development [33]. Malignant transformation is associated with aberrant glycosylation virtually in all tissues, being the neo-synthesis and the formation of immature truncated *O*-glycan structures the most common phenotypes [33, 86].

Neo-synthesis results from the induction of transcription of specific genes involved in the expression of carbohydrate determinants. This leads to *de novo* expression of abnormal glycoprofiles, containing sialyl Lewis X structures for instance, which can promote the metastatic process [33, 84].

Complex and elongated *O*-glycans contribute to protein stability and function, so the absence of these structures has obvious functional implications. In fact, truncated *O*-glycans have been under the spotlight regarding cancer research. Therefore, their biological role in malignant tissues will be dissected below.

Other glycoprofile alterations associated with cancer cells are abnormal core fucosylation and increased *N*-glycan branching [33].

Mechanisms for *O*-Glycan Truncation

The surface expression of early intermediates of the *O*-glycosylation pathway, the so-called immature or truncated *O*-glycans, is one of the most common phenotypes of malignant cells. The expression of truncated *O*-glycans is strongly correlated with poor prognosis and low survival and are present in

most gastric [33, 119, 120] and colorectal cancers [121, 122]. In fact, the Tn, STn and T glycan structures have been considered pancarcinoma antigens [123]. Tn and STn promote cancer progression or protection from the surveillance of the immune system, hence being valuable therapeutic targets in clinical treatment [33].

A number of different mechanisms may underlie this characteristic change in glycosylation, including altered expression of glycosyltransferases involved in branching and capping [124, 125], general reorganization of glycosyltransferase topology [126, 127], as well as fluctuations in cellular pH [128, 129].

One of the most explored mechanisms is the reduction of core 1 elongation, due to a blockade in the production of the T structure. Work on Tn positive cancer cell lines and patients with Tn syndrome revealed loss-of-function mutations in the *COSMC* gene and loss of T-synthase [121]. Promoter hypermethylation of *COSMC* was also identified in Tn-positive human pancreatic cancers [130]. Also, defective COSMC function results in aggregation and proteasomal degradation of the T-synthase, blocking glycan formation in the Tn antigen stage [118].

If not further modified by the T-synthase, the Tn antigen faces only the action of the ST6GalNAc glycosyltransferases. The expression of the sialylated form of Tn, STn, which serves as a biosynthetic stop, is likely controlled by the sialyltransferase ST6GalNAc-I [127]. Although this enzyme is also expressed in normal cells, overexpression appears to be needed to override the elongation pathway [125]. In fact, overexpression of ST6GalNAc-I is seen in intestinal metaplasia of GC and colon carcinomas, which colocalizes with STn positive cells [131].

Another mechanism for *O*-glycan truncation is the mislocalization and/or expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases

(ppGalNAcTs), the enzymes responsible for the UDP-GalNAc addition to a Ser or Thr in the peptide backbone. For instance, in murine colon carcinoma cells, loss of ppGalNAcT 3 resulted in *O*-glycan truncation and promoted metastasis to the liver [132].

Glycoproteins expressing the abnormal *O*-glycan Tn antigen may have altered structure and stability and be stably expressed on the membrane. Therefore, the expression of Tn and STn antigens in glycoproteins, instead of normal elongated *O*-glycans, can lead to altered expression and function of the protein carrier.

Models to Study Glycosylation in Cancer

Glycans play critical functional roles in host–pathogen interaction, progression of malignancies and embryonic development. Models to study the glycosylation contribution in biological processes are necessary for research.

Mouse knockout models and transgenics have been used to study the impact of specific carbohydrate deficiencies in several processes [133]. Cell lines derived from tumors have also been employed for the research of glycosylation's role. Jurkat cells, derived from human leukemia, have a mutation in the *COSMC* gene and were used for the initial studies of *O*-glycosylation truncation [134]. Interestingly, it was observed that in CRC bulk cell cultures Tn positive cells were present as a subpopulation, harbouring mutations in *COSMC* [121].

The use of programmable nucleases revolutionized all fields of scientific research, and the study of glycans was no exception. Zinc finger nucleases (ZFNs) were used to generate knockout cell lines for *COSMC* [135], the so-called SimpleCell (SC) model, including in MKN45 gastric cancer cells (Figure 7A, B) [119]. Recently, a library of gRNAs targeting human genes of the glycosylation

machinery using CRISPR/Cas9 was published [136].

Other genetic tools, such as RNA-interference and plasmid-induced gene expression, were also shown to be able to generate cell lines to investigate glycosylation modifications [131, 137]. Examples of these include the knockdown of *COSMC*, which enriches *O*-glycan structures to the Tn and STn antigens [131]. Also, *ST6GalNAc-I* overexpressing plasmids were stably transfected in MKN45 GC cells, so the Tn glycan is preferentially used as a target by the ST6GalNAc-I transferase, favoring the formation of STn and competing with the *O*-glycan elongation process (Figure 7A, C) [135].

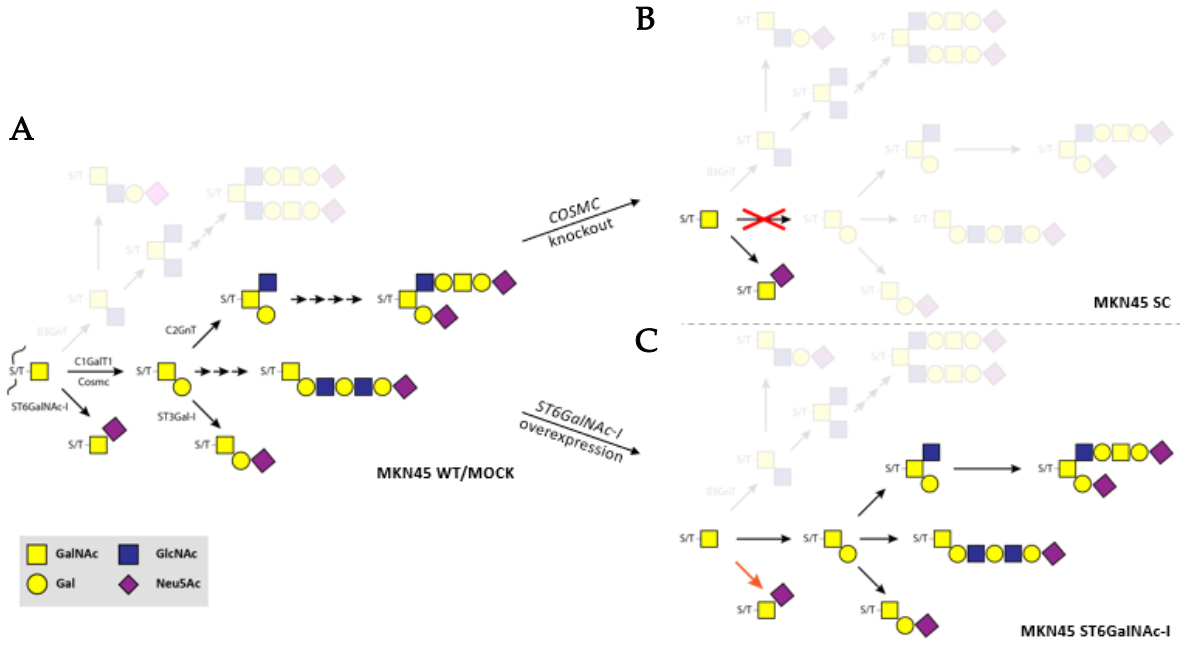


Figure 7. Schematic representation of the mucin-type *O*-glycan biosynthesis pathway **(A)** in the MKN45 gastric carcinoma cell line control models, MKN45 wild type (WT) and MKN45 MOCK transfected cells. The mechanisms by which *O*-glycans have been truncated in the applied MKN45 cell line models gave rise to **(B)** the MKN45 SimpleCell model and **(C)** the MKN45 ST6GalNAC-I model. MKN45 cells naturally have the absence of core 3 and 4 *O*-glycans expression.

These models are essential to study global phenotypes of cancer, including the glycosylation role in specific proteins, as it will be discussed in the next section.

CD44 Glycoforms in Cancer

As previously mentioned, CD44 is a global marker for CSCs and a potential cancer biomarker for diagnosis and treatment. Its structural heterogeneity derives from alternative splicing events but is further amplified by extensive and often isoform-specific glycosylation. CD44 variable isoforms may carry *N*-glycans (9 predicted and validated sites), *O*-glycans (146 predicted and 56 validated sites) and glycosaminoglycans (2 predicted and validated sites) (Figure 8) [55].

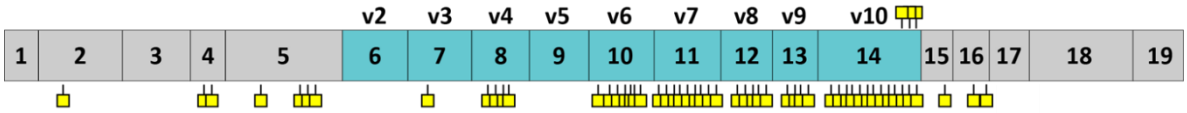


Figure 8. Validated *O*-glycan sites annotation of the full-length *CD44* Open Reading Frame of CD44. The information is according to the GlycoDomain viewer database.

There is missing knowledge on the influence of the glycoprofile of CD44 on its function. Gasbarri *et al.* modulated the *N*- and *O*-glycosylation pathways and observed, in melanoma cell lines, that spontaneous CD44 shedding was dependent on the presence of its *O*-glycosites on the membrane-proximal domain [138]. Mereiter *et al.* showed that the affinity of CD44 to HA was higher when cells expressed immature *O*-glycan structures, which promoted cell motility and invasion [139].

Using the SC model, a glycoproteomics strategy to explore potential biomarker *O*-glycoproteins was published, including the secretome of two GC cell lines, AGS and MKN45. CD44 was identified in the gastric cancer SCs as a potential secreted biomarker *O*-glycoprotein with immature *O*-glycans. This was validated in cancer patients' sera and the staining patterns of CD44 and STn

overlapped in gastric tumor sections. Interestingly, the single *O*-glycosite identified by mass spectrometry in CD44 was in the v9 exon [119].

Glycosylation has gained more and more attention in several research fields, including cancer development and progression, of which, with technological advances, has begun to have its role further dissected.

CHAPTER 2

OBJECTIVES

Colorectal and gastric carcinomas are the third and fifth most common cause of cancer death worldwide and remain among the leading cancers in incidence and mortality. This reflects the late diagnosis of these types of cancer, due to appearance of clinical symptoms only when the cancer has progressed. To overcome this problem and improve patient care and overall prognosis, discovery of new and better characterized biomarkers is pivotal.

The lack of success in validating many of the new promising targets such as CD44v, often arise from impaired methods of its detection/targeting. Structural and functional heterogeneity, generated by alternative splicing and glycosylation, makes CD44 one of the most variable surface molecules. This provides the explanation for its various cellular functions, but it also may explain the failure in validate CD44 as an established cancer target. This works aims to

show how this structural heterogeneity cannot be neglected on detection methods of this target.

General Aim

To characterize CD44 variants expression in gastric and colorectal cancers and to evaluate the potential of CD44v9 cancer-associated glycoforms detection methods for future clinical application in gastrointestinal cancer.

Specific Aims

1. *Characterize CD44 variants and truncated O-glycan antigens in gastric and colorectal cancer tissues*

From a biomarker perspective, it is necessary to investigate selected targets in tissue samples as well as their cancer specific glycoforms. Therefore, in chapter 4 section 4.1 it is performed an immunohistochemical characterization of CD44v9 in colorectal tumor tissue microarrays (TMAs) and gastric carcinoma tissues. Co-expression and co-localization of CD44v9 and the truncated *O*-glycan antigens Tn and STn is evaluated in a gastric cancer series.

2. *Evaluate CD44 isoforms expression in gastric and colorectal cancer cell line models to assess whether it correlates with a glycoprofile*

During the process of novel biomarkers discovery, *in vitro* characterization is crucial to validate immunohistochemical analysis of patients' samples. Thus, in chapter 4 section 4.2 the influence of *O*-glycosylation in molecular features of CD44 and its variant isoforms is assessed, using the SimpleCell and the ST6GalNAc-I overexpressing models, representing different strategies of *O*-

glycosylation elongation disruption. Furthermore, it is evaluated whether truncated *O*-glycans influence the transcriptional levels of CD44 genes.

3. *Detect CD44 extracellular cleavage and its relationship with O-glycan truncation and establish its biological biomarker potential.*

O-glycoproteins constitute many in use serum biomarkers and the detection of specific changed *O*-glycans, mainly truncated ones, may provide increased specificity of these structures. CD44 is cleaved *in vivo* and CD44v9 was detected in patient's sera. In chapter 4 section 4.3 it is evaluated CD44s and CD44v9 isoforms extracellular cleavage according to the complexity degree of *O*-glycans they express, in different gastric cancer cell lines. Moreover, the secretome analysis of those cell lines is performed, targeting the CD44/CD44 isoforms that were secreted/shed from the tumor cells and detectable in cell culture media.

CHAPTER 3

MATHERIALS AND METHODS

3.1 CELL LINES

The gastric carcinoma cell lines MKN45 and AGS were obtained from the Japanese Collection of Research Bioresources and ATCC, respectively. The MKN45 and AGS SimpleCells (MKN45 SC and AGS SC) were obtained by targeting the *C1GALT1* (*COSMC*) gene using zinc-finger nuclease precise gene editing as previously described [119, 135]. Besides, MKN45 cell line was stably transfected with the full length human *ST6GALNAC1* gene (MKN45 ST6GalNAc I) or the corresponding empty vector pcDNA3.1 (MKN45 Mock) as previously described [140, 127]. The colorectal carcinoma cell lines RKO and LS-174T were obtained from ATCC.

All cells were grown in monolayer in uncoated cell culture flasks. Cells were maintained at 37 ˚C in an atmosphere of 5% CO_2, in different media: AGS WT/SC, MKN45 WT/SC and RKO in RPMI 1640 GlutaMAX™, HEPES medium

(Alfagene); MKN45 Mock (ST6) and MKN45 ST6GalNAC I also in RPMI 1640 GlutaMAX™, HEPES medium, supplemented with 0.5 mg/ml of G418 (Invitrogen, Waltham, MA) to select transduced cells; LS-174T in Dulbecco's Modified Eagle Medium (DMEM, high glycose/pyruvate; Alfagene). All media were supplemented with 10% heat-inactivated FBS (Fetal Bovine Serum; Labclinics). Cell culture medium was replaced every two to three 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.

3.2 PRIMARY ANTIBODIES

In western blot, dot blot, immunofluorescence, flow cytometry and immunohistochemistry (IHC) experiments the following primary MAbs have been used: CD44, CD44s PanS, CD44v9, STn, Tn and Tubulin (Table 1).

Table 1. List of antibodies.

Antibody	Clone	Isotype	Host	Specificity/ Sensitivity	Immunogen	Source	Ref.
CD44	156-3C11	IgG2a	Mouse	Detects endogenous levels of total CD44 protein	Human CD44 transfected cells	Cell Signaling	[141]
CD44s Pan Specific	691534	IgG2a	Mouse	Detects human CD44s	Human CD44s transfected cells	R&D Systems	[142]
CD44v9	RV3	IgG2a	Rat	Detects human CD44 variants isoform 9 (v9)	Human CD44 variant isoform 8-10 transfected cells	Abnova	N.A.
Tn	1E3	IgG1	Mouse	Detects human Tn-antigen	N.A.	Non-commercial	U.D.

STn	B72.3	IgG1	Mouse	Detects human STn-antigen	Membrane-enriched fraction of human breast carcinoma cells	Non-commercial	[143]
Tubulin	DM1A	IgG1	Mouse	Detects the ~60kDa tubulin protein found in all eukaryotic cells	Purified human brain α-Tubulin	Merck	[144]

N.A. = not available; U.D. = Unpublished Data (Clausen and Hakomori)

3.3 IMMUNOFLUORESCENCE

Immunofluorescence (IF) is a common microscopy-based technique that uses specific antibodies to label fluorescent dyes to specific biomolecules targets within a cell.

Cells were grown in coverslips, in 12 well plates, and fixed in 4% paraformaldehyde for 10 minutes as previously described and frozen at -20 ˚C until immunofluorescence performance [145]. Coverslips were permeabilized with PBS 0.2 % Triton X-100 for 10 minutes at RT, washed three times with PBS 1x and blocked with goat serum (DAKO, Denmark) diluted 1:5 in PBS 10% BSA for 30 minutes at RT. Cells were incubated overnight at 4 ˚C with the corresponding primary MAbs: Tn and STn (undiluted); CD44s PanS (10 µg/ml) and CD44v9 (3 µg/ml), both diluted in PBS 5% BSA. After being washed 3 times, coverslips were incubated 1h at RT and in the dark with secondary antibodies: FITC-conjugated rabbit anti-mouse immunoglobulin (DAKO) diluted 1:100 in PBS 5% BSA for Tn, STn and CD44s PanS; donkey anti-rat Alexa Fluor® 488 immunoglobulin (Invitrogen, Eugene, OR) diluted 1:500 in PBS 5% BSA, for CD44v9. Cells were washed 3 times with PBS 1x, incubated with DAPI (4′,6-diamidino-2-phenylindole) diluted 1:100 in PBS for 10 minutes at RT, washed 2 times and the coverslips were mounted with Vectashield mounting

medium for fluorescence (Vector, Burlingame, CA). Negative controls were performed by substitution of the primary antibody with PBS 5% BSA. Cells were visualized and images were acquired using Zeiss Optical Microscope

3.4 WESTERN BLOTTING

The western blot (WB), or protein immunoblot, is an analytical technique that detects specific proteins in a biological sample or extract.

Cells were washed twice with PBS and directly collected in lysis buffer 17 (R&D Systems, McKinley Place, MN) 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 (Bio-Rad, Hercules, CA), based on the Bradford dye-binding method [146] and a total amount of 25 µg of protein extract were used in each experiment. Samples were mixed with 6.25 µL Loading Buffer (Bio-Rad), 2.25 µL of DTT and denaturated in dry bath for 10 minutes at 90 °C, before being resolved in a 4-15% polyacrylamide precast gel (Mini-PROTEAN®TGX™) at 100V for 1 hour, in 10x TrisGlycine/SDS buffer (Bio-Rad). Proteins were transferred to polyvinylidene difluoride (PVDF) membranes (GE Healthcare, Chicago, IL) at 330 mA for 1 hour, in 10x Tris/Glycine buffer (Bio-Rad), using a Mini-PROTEAN® Tetra Cell System (Bio-Rad, Hercules, CA). The membranes were blocked with TBS with 0.1% TWEEN (TBS-T) containing 5% skim milk, followed by an overnight incubation with primary antibodies (CD44, CD44sPanS or CD44v9, at 1:1000, 1:2500 or 1:1000, respectively) in TBS-T. Membrane were incubated with peroxidase AffiniPure goat anti-mouse IgG (H+L; Jackson ImmunoResearch) in case of CD44 total and CD44s PanS primary antibodies, and with goat anti-rat IgG-HRP (Santa Cruz Biotechnology, CA,

USA) in case of CD44v9, at 1:5000 or 1:2000, respectively, in TBS-T. The signal was detected with ECL™ Western Blotting detection reagents (Amersham™, GE Healthcare) and high performance chemiluminescence films (Amersham Hyperfilm™ ECL, GE Healthcare), exposed in a Hypercassette™ (Amersham, Biosciences) and revealed and fixed with developer and fixer solutions, respectively. Densitometry was performed with GS-800 Calibrated Densitometer (Bio-Rad), values were normalized to tubulin and represented as relative values compared to the control, using the Image Lab software (Biorad). Results shown in Figure 14 are from a n = 2 independent experiments (for CD44 total antibody) or from a n = 4 independent experiments (for CD44s PanS and CD44v9 antibodies).

3.5 DOT BLOTTING

A dot blot represents a simplification of the western blot technique in which the proteins to be detected are not first separated by electrophoresis.

Protein lysates (from MKN45 WT/SC, MKN45 Mock/ST6GalNAc I and AGS WT/SC) were diluted in water for a total volume of 11 µl. PVDF membranes were prepared drawing grids by pencil to indicate the region that is going to be blotted. Samples were spotted onto the membrane at the center of each grid. To minimize the area that the solution penetrates, the procedure was made slowly, in two times (5.5 µl for each). The membrane was left to dry for 30 minutes. The blocking, incubation in primary and secondary antibodies, and signal detection procedures were performed as previously explained (see Western blot, materials and methods). Results shown in Figure 16 are from one experiment.

3.6 PROTEIN ENZYMATIC DEGLYCOSYLATION

Enzymatic deglycosylation consist in the removal of carbohydrates from glycoproteins to simplify the analysis of the peptide portion of the glycoprotein, for instance.

Three different deglycosylation processes were performed, to remove different glycans from 25 µg of total cell protein lysates (RKO, MKN45 WT and MKN45 SC).

N-linked glycans were removed by denaturing the protein lysates with 10x glycoprotein denaturing buffer (New England Biolabs) at 100 ˚C for 10 minutes, chilling one ice, and by applying a deglycosylation mix composed of 10 % G7 Glyco buffer (New England Biolabs), 10% NP-40 and 5% PNGase F (New England Biolabs), for a total volume of 30 µl. *N*-linked and *O*-linked glycans were removed by denaturing the protein lysates with 10x glycoprotein denaturing buffer at 100 ˚C for 10 minutes, chilling one ice, and by applying a deglycosylation mix composed of 10 % Glyco buffer 2 (New England Biolabs), 10% NP-40 and 5% Protein Deglycosylation Mix II (New England Biolabs), for a total volume of 30 µl. The enzymes included in the deglycosylation enzyme cocktail are *O*-glycosidase, PNGaseF, α2-3,6,8,9 Neuraminidase A, β1-4 Galactosidase S and β-M-Acetylhexosaminidase. Sialic acids were removed by directly applying Neuraminidase (0.2 U/ml) (Sigma-Aldrich, St. Louis, MO) diluted in PBS 10x, for a total volume of 26,5 µl.

All samples were incubated overnight at 37 ˚C in a total volume of 30 µl and analyzed in a Western blot for CD44v9 characterization. Densitometry was performed with GS-800 Calibrated Densitometer (Bio-Rad), values were normalized to tubulin and represented as relative values compared to the control

(MKN45 WT without deglycosylation treatment), using the Image Lab software (Biorad). Results shown in Figure 15 are from one experiment.

3.7 SECRETOME ANALYSIS

The term secretome denotes all the proteins secreted by a cell to the extracellular space [147].

For the secretome analysis, cells were grown in medium without FBS supplementation. Conditioned media were collected after 48h and were enriched in protein content using 10 kDa Amicon Purification System (Merck Millipore, Burlington, MA). Afterwards, samples were quantified and analyzed in a Western Blot for CD44 total, CD44s PanS and CD44v9 characterization. Densitometry was performed with GS-800 Calibrated Densitometer (Bio-Rad), values were normalized to the staining of the bands with Ponceau S (Sigma-Aldrich) upon protein transfer to the PVDF membrane and represented as relative values compared to the control (MKN45 WT), using the Image Lab software (Biorad). Results shown in Figure 21 are from one experiment.

3.8 FLOW CYTOMETRY

Fluorescent-activated cell sorting (FACS) is a fluorescent-based method that allows the quantification of specific markers in a population of cells.

Cells were detached using Gibco® Versene solution (ThermoFisher, Waltham, MA) or trypsin solution (trypsin 0.05% EDTA, Alfagene) washed with FACS buffer (PBS 1x with 1% FBS) and centrifuged at 300g for 5 minutes at 4 ˚C, followed by the staining of dead cells with Fixable Viability Dye APC-Cy7 (FVD 1:2000, ThermoFisher, Waltham, MA) for 30 minutes on ice. Afterwards, cells

were submitted to one of the following protocols: fixation only or fixation and permeabilization protocols. In the first protocol, cells were stained with primary antibodies diluted in FACS buffer (CD44v9 1:100; CD44s PanS 1:200) for 30 minutes on ice. Cells were washed and were incubated with secondary antibodies (α-mouse and α-rat Alexa Fluor® 488 1:200, Invitrogen, Eugene, OR) for CD44s PanS or CD44v9, respectively, for 30 minutes on ice. Finally, the cells were fixed with 2% PFA for 30 minutes at room temperature (RT), centrifuged and ressuspended in FACS buffer for analysis by flow cytometry. For the fixation and permeabilization protocol, cells were first fixed, permeabilized with 0.5% Saponine (Sigma-Aldrich) diluted in PBS for 10 minutes at RT, and stained with primary and secondary antibodies diluted in 0.5% Saponine/PBS. Cells were ressuspended in FACS buffer for analysis. After the staining cells were analyzed in a FACS Canto II using the FACSDiva software (BD Biosciences, San Jose, CA). Cells were interrogated for APC-Cy7 using a 633 nm laser and a 780/60 band-pass filter and for Alexa Fluor® 488 using a 488 nm laser and a 530/30 band-pass filter. Data were analyzed using FlowJo software version vX.10.0. (Tree Star Inc, USA). For the MFI differences, t-tests were done. All the data were analyzed using the GraphPad Prism 6 software (GraphPad Software). Results were considered statistically significant with $p<0.05$. Results shown in Figures 17 and 20 are from a n = 2 independent experiments (for MKN45 WT and SC) or n = 1 (MKN45 MOCK and ST6GalNAc I).

3.9 RNA EXTRACTION, cDNA SYNTHESIS AND PCR REACTION

Complementary DNA (cDNA) is DNA synthesized from a single-stranded RNA template, like messenger RNA (mRNA), in a reaction catalyzed by the enzyme reverse transcriptase [148].

Polymerase chain reaction (PCR) is a common method used to exponentially amplify a specific nucleic acid sequence, generating thousands to millions of more copies of that particular DNA segment [149].

Cell lines (MKN45 WT/SC, MKN45 Mock/ST6GalNAc I, RKO and LS-174-T) were grown in 6 well plates and washed with ice cold PBS 1x. 500 µl of TRI Reagent (guanidine thioacyanate and phenol; Sigma-Aldrich) were added in order to lyse the cells and dissolve DNA, RNA and proteins. After being completely detached, cells were transferred to *eppendorfs*, 100 µl of chloroform was added and rested for 15 minutes on ice, to promote the separation of the phases. Cellular extracts were centrifuged for 15 minutes at 12000 g at 4 ˚C, the aqueous phase was transferred to a clean eppendorf, 250 µl of isopropanol was added and the samples were set to rest 10 minutes on ice (or stored at -80 ˚C until the protocol is resumed). RNA extracts were centrifuged 10 minutes at 12000 g / 4 ˚C, the supernatant discarded, and the pellet ressuspended in 800 µl of ice-cold ethanol 100%. The centrifugation processes were repeated, and the pellet ressuspended in 500 µl of ice-cold ethanol 75%, in order to rehydrate the sample. After one last centrifugation, the supernatant was removed, and the pellet was left to dry for 30 minutes. Afterwards, the pellet was ressuspended in ultrapure H_2O (nuclease free water, Ambion®) and quantified using a Nanodrop 1000 UV-Vis Spectrophotometer system (Thermo Scientific) and stored at -80 ˚C.

Total RNA from cellular samples was transcribed into complementary DNA (cDNA) using 2000 ng of RNA. RNA was diluted in nuclease free H_2O, and 1 µl of random primers (50 µM, Thermo Scientific) and 1 µl of deoxynucleotides (dNTPs, NZY Tech) (10 mM) were added, in a total of 11 µl. In the T100™ Thermo Cycler (Bio-Rad), the samples were heated at 65 ˚C for 5 minutes, followed by 1 minute on ice. Afterwards, 4 µl of 5x SSIV buffer, 1 µl of DTT reagent (100 mM),

1 µl of RNase OUT recombinant RNase Inhibitor (40 U/µl) and 1 µl of SuperScript® IV Reverse Transcriptase (200 U/µl; Invitrogen) were added to a reaction tube, briefly centrifuged to mix the components 7 µl of this reaction mix were added to each solution containing the annealed RNA. The samples were subjected to PCR on the T100™ Thermo Cycler (23 ˚C x 10 min, 52 ˚C x 20 min, 80 ˚C x 10 min, 4 ˚C x ∞; Bio-Rad).

To perform the PCR, cDNA is amplified using gene-specific primers to generate many copies of that specific DNA sequence. For the PCR reaction mix preparation, 2 µl of 10x reaction buffer for supreme NZYTaq polymerase (NZY Tech), 0.6 µl of $MgCl_2$ (50 mM, NZY Tech), 0.4 µl of dNTPs (10 mM), 0.5 µl of Forward Primer (10 µM) (Table 2), 0.5 µl of Reverse Primer (10 µM), 0.2 µl of supreme NZYTaq II DNA polymerase (500 U; nzytech) and 14,8 µl of nuclease-free H_2O were added to 1 µl of cDNA from each cell line, in a total of 20 µl. PCR reaction was performed in a T100™ Thermal Cycler (Bio-Rad) following the program: 94 ˚C for 5 minutes during 1 cycle (initialization); 94 ˚C for 30 seconds (denaturation), 55 ˚C for 60 seconds (annealing), 72 ˚C for 75 seconds (elongation), during 34 cycles; 72 ˚C for 10 minutes during 1 cycle (final elongation) and 4 ˚C for an indefinite time (final hold). Samples were run in a 1.5% agarose gel and the DNA bands were visualized in a Molecular Imager® (Gel Doc™ XR+ Imaging System, Bio-Rad). Relative quantification was determined using the Image Lab software (Biorad). All primers were ordered from Integrated DNA Technologies. Results shown in Figures 18 and 19 are from one experiment.

Table 2. Probes used in PCR for the amplification of sequences of interest. Adapted from [150].

Primer designation	Primer sequence 5′ to 3′	Expected amplicon size (bp)
Exon 3 *forward*	TGGGTTCATAGAAGGGCATG	***Exon 3 - Exon 16*** ***CD44s[1]:*** 565 ***CD44v9[2]***: 655 ***CD44v10[2]***: 769 ***CD44v8-9[3]***: 757 ***CD44v8-10[3]***: 961 ***CD44v6+v8-10[3]***: 1090 ***CD44v6-10[3]***: 1222 ***CD44v3-10[3]***: 1579
v2 *forward*	GCTACAGCAACTGAGACAGC	***v2 - Exon 16*** **v2[2]**: 243 **v2+v7-10[3]**: 771 **v2-10[3]**: 1257
v3 *forward*	ACGTCTTCAAATACCATCTC	***v3 - Exon 16*** **v3[2]**: 255 **v3+v8-10[3]**: 651 **v3-10[3]**: 1143
v4 *forward*	CCAGGACTGGACCCAGTGGAA	***v4 - Exon 16*** **v4[2]**: 205 **v4+v7-10[3]**: 733 **v4-10[3]**: 979
v5 *forward*	GTAGACAGAAATGGCACCAC	***v5 - Exon 16*** **v5[2]**: 246 **v5-7+v10[3]**: 711 **v5-6+v8-10[3]**: 771 **v5+v7-10[3]**: 774 **v5-10[3]**: 903
v6 *forward*	CAGGCAACTCCTAGTAGTAC	***v6 - Exon 16*** **v6[2]**: 258 **v6-8[3]**: 492 **v6+v7+v10[3]**: 594 **v6-9[3]**: 582 **v6+v8-10[3]**: 654 **v6-10[3]**: 786
v7 *forward*	CCAGCCATCCAATGCAAGGA	***v7 - Exon 16*** **v7[2]**: 248 **v7+v8[3]**: 350 **v7+v9[3]**: 338 **v7+v10[3]**: 452 **v7-10[3]**: 644

v8 *forward*	CGCTTCAGCCTACTGCAAATCC	***v8 - Exon 16*** **v8**[2]: 209 **v8-9**[3]: 299 **v8+v10**[3]: 413 **v8-10**[3]: 503
v9 *forward*	GCTTCTCTACATCACATGAAGG	***v9 - Exon 16*** **v9**[2]: 203 **v9+v10**[3]: 407
v10 *forward*	AGGAATGATGTCACAGGTGG	***v10 - Exon 16*** **v10**[2]: 335
Exon 16 *reverse*	ATTTGGGGTGTCCTTATAGG	

[1] No inclusion of variant exon

[2] Inclusion of single variant exon

[3] Inclusion of multiple variant exons

3.10 IMMUNOHISTOCHEMISTRY

Immunohistochemistry (IHC) is a common microscopy-based technique that uses specific antibodies that labeled an antigen target in cells of a tissue section.

Data from 18 patients treated in the University Hospital São João and IPO-Porto, between 1990 and 1994 with GC diagnosis and submitted to surgical treatment was collected prospectively and their use approved by the local Ethical committees. The data collected from clinical examinations included: age, gender, gastric tumor classification, sinus histology, follicular hyperplasia and Ming classification. Expression of CD44v9, Tn and STn was evaluated in 12 cases of human gastric carcinomas (10 intestinal, 5 diffuse sub-type and 3 atypical) [18] and CD44v9 immunohistochemical expression was correlated with the available clinicopathological data. Additionally, the histological patterns of CD44v9

immunohistochemical expression were evaluated in 6 cases of normal adjacent tissue of gastric cancer patients.

Data from 312 patients treated in the Braga Hospital, between 2005 and 2010 with CRC diagnosis and submitted to surgical treatment was collected prospectively. The data collected from clinical and preoperative diagnostic examinations included: age, gender, clinical presentation, oncologic history, tumor localization, histological type, macroscopic appearance and preoperative staging. The tumor localization was recorded and classified as colon and rectum (between anal verge and 15 cm at rigid rectoscopy). Histopathological reports included: tumor differentiation, lymphatic and blood vessel invasion, tumor staging and recurrence. The level of positive lymph nodes was not described in all specimens. The histological type of CRC was determined by two experienced pathologists. A series of formal-fixed, paraffin-embedded tissues from these patients was analyzed by immunohistochemistry for CD44v9 expression. Slides from all 312 specimens were reviewed and mapped and tissue microarrays (TMA) were built using a manual tissue arrayer (MTA-1 Beecher Instrument, Silver Spring, MD, USA). Representative areas of the CRC lesions were selected and cores of 1.0 mm in diameter were twice sampled and arranged at 0.3 mm from each other in the recipient paraffin block. A database was built for every block produced, including the coordinates of each core and case of origin. Comparison of expression in tumor versus normal cells was also possible since, in most cases, the same paraffin section contained both neoplastic and normal colonic epithelium. CD44v9 immunohistochemical expression was correlated with the available clinicopathological data.

Immunostaining was performed by the modified avidin-biotin-peroxidase complex (ABC) method [151]. Paraffin sections were dewaxed in

xylene, rehydrated with a gradual decreasing concentration of ethanol (100%-70%) and washed with water. For CD44v9 staining only, heat-induced antigen retrieval was carried out in sodium citrate buffer (10 mM citric acid, pH 6.0) for 40 minutes in a steam cooker (Ufesa), left to cool for 10 minutes at RT and washed with PBS 1x. Sections were then treated with 10 % hydrogen peroxide (H_2O_2) in methanol for 10 minutes to block the endogenous peroxidase activity. After washing the slides in PBS 1x, sections were incubated with rabbit serum 1:5 diluted in PBS 10% bovine serum albumin (BSA) in a humid chamber for 30 minutes at RT. The tissues were incubated with primary antibodies to CD44v9 (1:500 diluted in PBS 5% BSA) or Tn and STn (undiluted) at 4 ˚C in a humid chamber, overnight. After three washing steps with PBS 1x, sections for immunohistochemistry were incubated with biotin-labelled secondary rabbit anti-rat (CD44v9) or rabbit anti-mouse (Tn and STn) (DAKO) diluted 1:200 in PBS 5% BSA, for 30 minutes at RT and with ABC kit (Vectastain®, Vector Laboratories) for 30 minutes at RT. After three washing steps with PBS 1x, sections were stained for 2-3 minutes with 0.05% 3,3 diaminobenzidine tetrahydrochloride (DAB) containing 0.01% H_2O_2. Finally, sections were counterstained with Mayers' haematoxylin solution, washed with water, dehydrated with a gradual increasing concentration of ethanol (70% - 100%) and xylol, and mounted. Negative controls were performed by substitution of the primary antibody with PBS 5% BSA. Slides were examined and images acquired using a Zeiss Optical Microscope.

CHAPTER 4

RESULTS

4.1 GASTRIC AND COLORECTAL CANCER TISSUES IMMUNOHISTOCHEMICAL CHARACTERIZATION

Gastric cancer tissues

Gastric cancer tissue sections from a total of 12 patients were used to perform a preliminary immunohistochemical (IHC) analysis of the *in situ* expression of CD44v9. The cohort characteristics are described in Table 3, according to CD44v9 expression. All the histological parameter classification was done by a certified pathologist. Histological analysis was performed to discriminate samples in their tumor classification, both for the Lauren parameters and the newly identified ones [18, 23]; sinus histology; presence of follicular hyperplasia; and Ming classification, that classifies tumors according to

their growth pattern. There were no significant differences in the CD44v9 expression among these parameters, which is expected given the small number of cases of this series (Table 3). With the expansion of the series it is important to further test some of the potential trends.

Table 3. Assessment of associations between CD44v9 expression and clinical data in primary GC. P values were determined for statistical significance using the Fisher's exact test.

		CD44v9 IHC		
		Tumor cells		
	n	Negative/weak (%)	Moderate/strong (%)	*p value*
Gender				
Male	7	3 (42.9)	4 (57.1)	0.575
Female	5	1 (20.0)	4 (80.0)	
Age (Years)				
≤ 45	1	0 (0.0)	1 (100.0)	1.000
> 45	11	4 (36.4)	7 (63.6)	
Lauren Classification				
Diffuse	3	1 (33.3)	2 (66.7)	0.851
Intestinal	7	2 (28.6)	5 (71.4)	
Atypical	2	1 (50.0)	1 (50.0)	
New Classification				
Glandular	6	3 (50.0)	3 (50.0)	0.545
Mix	6	1 (16.7)	5 (83.3)	
Sinus Histology				
minimum/absent	9	4 (44.4)	5 (55.6)	0.490
abundant/moderate	3	0 (0.0)	3 (100.0)	
Follicular hyperplasia				
minimum/absent	1	1 (100.0)	0 (0.0)	0.333
abundant/moderate	11	3 (27.3)	8 (72.7)	
Ming Classification				
expanding growth pattern	5	1 (20.0)	4 (80.0)	0.575
infiltrating growth pattern	7	3 (42.9)	4 (57.1)	

One of the main goals of the work presented in this thesis was to assess the relationship of CD44v9 expression to a glycoprofile of truncated *O*-glycans. Therefore, the co-expression profile of the isoform and the Tn and STn structures was assessed. Co-expression of CD44v9 and both *O*-glycan structures did not show a statistically significant association. However, co-expression with STn showed a *p*-value of 0.091, possibly indicating a trend (Table 4).

Table 4. Assessment of associations between co-expression of CD44v9 and the *O*-glycan Tn and STn antigens in gastric tumor cases.

glycan antigen	Tn				STn			
	n	Negative/ weak (%)	Moderate/ strong (%)	*p* value	n	Negative/ weak (%)	Moderate/ strong (%)	*p* value
CD44v9								
negative	4	1 (25.0)	3 (75.0)	0.333	4	2 (50.0)	2 (50.0)	0.091
positive	8	0 (0.0)	8 (100.0)		8	0 (0.0)	8 (100.0)	

To further detail the co-existence of CD44v9 and a glycoprofile featuring truncated *O*-glycans, the co-localization patterns of the isoform and the Tn and STn structures expression was analyzed (Figure 9). Sequential slides were used to perform the three staining protocols and the degree of co-localization was classified independently by trained researchers. The results showed several co-localization profiles (Figure 9 A). In fact, a moderate/high colocalization between CD44v9 and Tn was observed in 75% (6 out of 8) of the cases (Figure 9 B). Concerning STn, co-localization was detected in 57% (4 out of 7) (Figure 9 C).

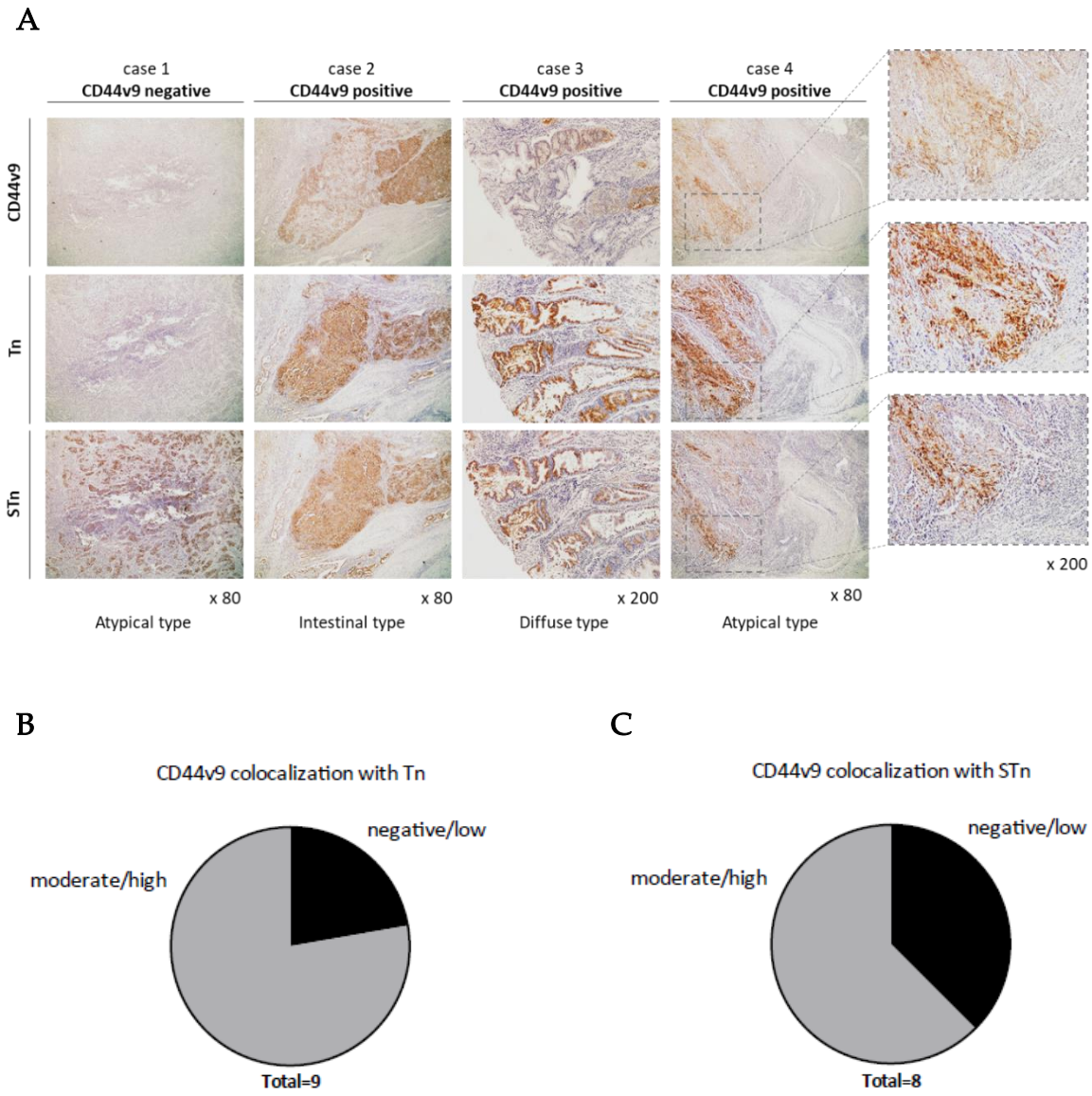


Figure 9. Co-localization of CD44v9 and the truncated *O*-glycan Tn and STn antigens in gastric cancer samples. (A) Representative immunohistochemical co-localized expression of CD44v9 and the Tn and STn antigens in the same cases of GC (x 80 and x 200 magnification). (B) Statistical analysis of the co-localization of CD44v9 and the Tn antigen. (C) Statistical analysis of the co-localization of CD44v9 and the STn antigen.

Interestingly, on normal adjacent tissue (NAT) of some sections, NAT samples showed CD44v9 expression in a specific pattern, limited to a frontline of linearly connected cells. The histological profile points to CD44v9 expression in a tumor basal membrane-like or CD44+ stem cell manner (Figure 10).

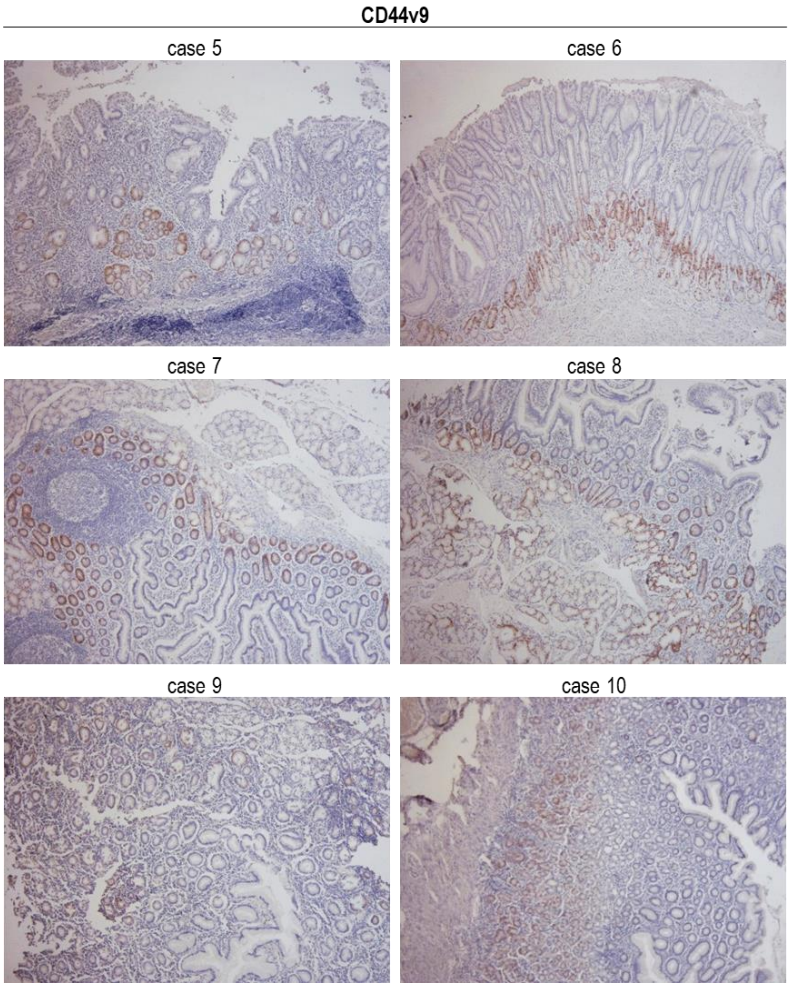


Figure 10. Immunohistochemical expression of CD44v9 in gastric cancer normal adjacent tissue samples (original magnification x 50).

Colorectal cancer tissues

The malignant function of CD44v9 was also observed in CRC-focused studies [54, 78]. Therefore, samples from 312 cases of CRC from the Braga Hospital were organized in tumor TMAs, and the expression of CD44v9 was evaluated. The cohort characteristics are described in Table 5, such as gender, age, tumor presentation and others. The majority of tumors were localized in the colon and most subjects were negative for familial history of CRC (Table 5).

Table 5. Assessment of associations between CD44v9 expression and clinical data in primary CRC. P values were determined for statistical significance using the Fisher's exact test.

			CD44v9 IHC		
		n	Tumor cells		
			Negative/weak (%)	Moderate/strong (%)	*p value*
Gender					
	Male	194	135 (69.6)	59 (30.4)	0.320
	Female	118	75 (63.6)	43 (36.4)	
Age (Years)					
	≤ 45	15	11 (73.3)	4 (26.7)	0.781
	> 45	297	199 (67.0)	98 (33.0)	
Presentation					
	Asymptomatic	59	39 (66.1)	20 (33.9)	0.878
	Symptomatic	253	171 (67.6)	82 (32.4)	
Localization					
	Colon	236	158 (66.9)	78 (33.1)	0.889
	Rectum	76	52 (68.4)	24 (31.6)	
Macroscopic appearance					
	Polyploid	157	102 (65.0)	55 (35.0)	0.310
	Ulcerative	67	45 (67.2)	22 (32.8)	
	Infiltrative	35	27 (77.1)	8 (22.9)	
	Exophytic	21	17 (81.0)	4 (19.0)	
Familiar history of CRC					
	Negative	256	171 (66.8)	85 (33.2)	0.825
	Positive	25	16 (64.0)	9 (36.0)	
CEA (ng/mL)					
	≤ 10	214	139 (65.0)	75 (35.0)	0.437
	> 10	48	34 (70.8)	14 (29.2)	

The samples were classified into CD44v9 negative/weak and moderate/strong expression. Moreover, the results from the IHC protocol to detect CD44v9 were matched with histological and clinical parameters, and

associations were analyzed (Figure 11 and Table 6). There was a statistically significant association (*p*-value = 0.023) between the tumor size and CD44v9 expression, with higher percentage of CD44v9 cases in larger tumors. Histological type, differentiation degree, lymphatic invasion and cancer recurrence did not show any association with CD44v9 presence. Vessel invasion showed a close to significant association (*p*-value = 0.132), where CD44v9 positive malignancies present lower levels of vascular structures within the tumor. Although not statistically significant, there seems to be a trend with tumor staging, with higher percentage of CD44v9 positive cases in higher cancer stages.

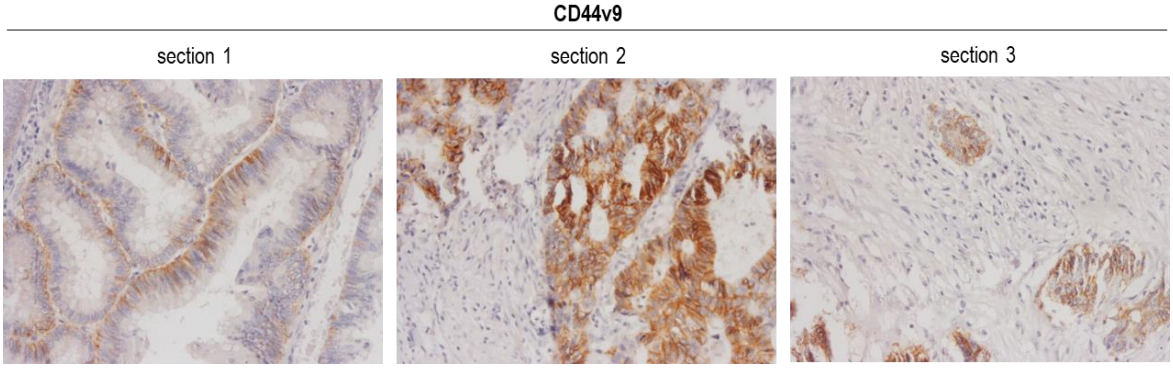


Figure 11. Representative immunohistochemical expression of CD44v9 in colorectal cancer TMA samples (original magnification x 200).

Table 6. Assessment of associations between CD44v9 expression and pathological data in primary CRC. * indicates p-value <0.050

	CD44v9 IHC			
	n	Tumor cells		*P value*
		Negative/weak (%)	Moderate/strong (%)	
Tumor size (cm)				
≤ 4.5	179	129 (72.1)	50 (27.9)	0.023*
> 4.5	115	68 (59.1)	47 (40.9)	

Histological Type				
Adenocarcinoma	281	188 (66.9)	93 (33.1)	0.477
A. Mucinous	28	19 (67.9)	9 (32.1)	
Signet ring and mucinous	3	3 (100.0)	0 (0.0)	
Differentiation				
Well/moderate	272	186 (68.4)	86 (31.6)	0.422
Poorly/Undifferentiated	33	19 (57.6)	14 (42.4)	
Stage				
0	2	2 (100.0)	0 (0.0)	0.437
1	48	36 (75.0)	12 (25.0)	
2	118	79 (66.9)	39 (33.1)	
3	99	65 (65.7)	34 (34.3)	
4	41	24 (58.5)	17 (41.5)	
Lymphatic invasion				
Absent	133	87 (65.4)	46 (34.6)	0.617
Present	159	109 (68.6)	50 (31.4)	
Vessel invasion				
Absent	173	111 (64.2)	62 (35.8)	0.132
Present	126	92 (73.0)	34 (27.0)	
Recurrence				
No	266	179 (67.3)	87 (32.7)	1.000
Yes	46	31 (67.4)	15 (32.6)	

4.2 CD44v DIFFERENTIAL CHARACTERIZATION IN GASTROINTESTINAL CANCER CELL LINES EXPRESSING TRUNCATED *O*-GLYCANS

To explore potential alterations of CD44 protein regulation in cells with *O*-glycosylation truncation, previously established cell line models were used. The MKN45 (diffuse type gastric carcinoma) and the AGS (intestinal type gastric

carcinoma) isogenic SimpleCell models were developed using ZFNs to disrupt *COSMC* [119, 135]. In addition, MKN45 were stably transfected with a *ST6GalNAc-I* overexpressing plasmid (MKN45 ST6GalNAc I cells) and a mock control (MKN45 MOCK cells), giving rise to an alternate *O*-glycan truncation model [131]. Cells were cultured in coverslips and the expression of truncated *O*-glycans was evaluated in these models by immunofluorescence. The 1E3 monoclonal antibody, targeting the Tn structure, and the B72.3 targeting the STn (Figure 12) were used. As expected, IFs showed positive signal for Tn and STn in the SC models, with more intensity than the control counterparts, where the signal was absent or low for both structures. The RKO cell line (colon carcinoma) was also evaluated for the presence of Tn and STn and found to have absent or low expression of both.

The same set of cells was submitted to IF protocols for two CD44 isoforms: CD44s and CD44v9. The monoclonal antibody 2C5 clone was used to detect CD44s and the RV3 for CD44v9. According to the manufacturer, the 2C5 clone recognizes an epitope in the invariant N-terminal region of CD44, present in all CD44 isoforms, with higher affinity to CD44s [142]. The RV3 clone recognizes the peptide sequence of the v9 exon, that may be present in several isoforms (CD44v3-10 or CD44v8-10, for instance). Immunofluorescence analysis revealed all the MKN45 cell models, WT, SC, MOCK and ST6GalNAc I, had expression of CD44s and CD44v9. In contrast, none of the isoforms tested was detected in AGS models. The most intense signal for CD44s was seen in RKO cells, that in turn show no CD44v9 expression (Figure 13).

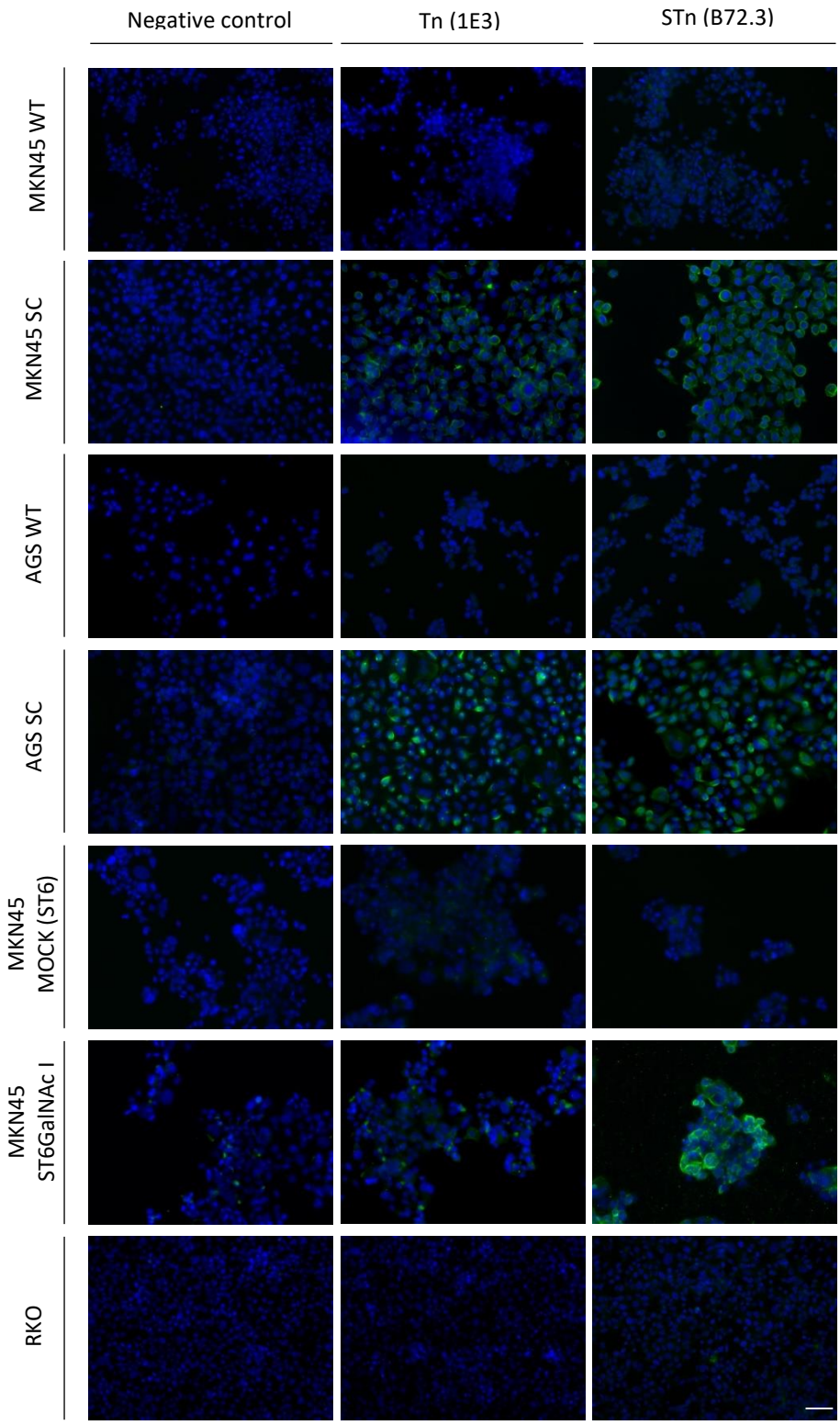


Figure 12. Immunofluorescence analysis of Tn and STn antigens expression in GC (MKN45 and AGS) and CRC (RKO) cancer cell lines. Nuclei are shown in blue and *O*-glycan structures in green. The scale bar corresponds to 50 μm.

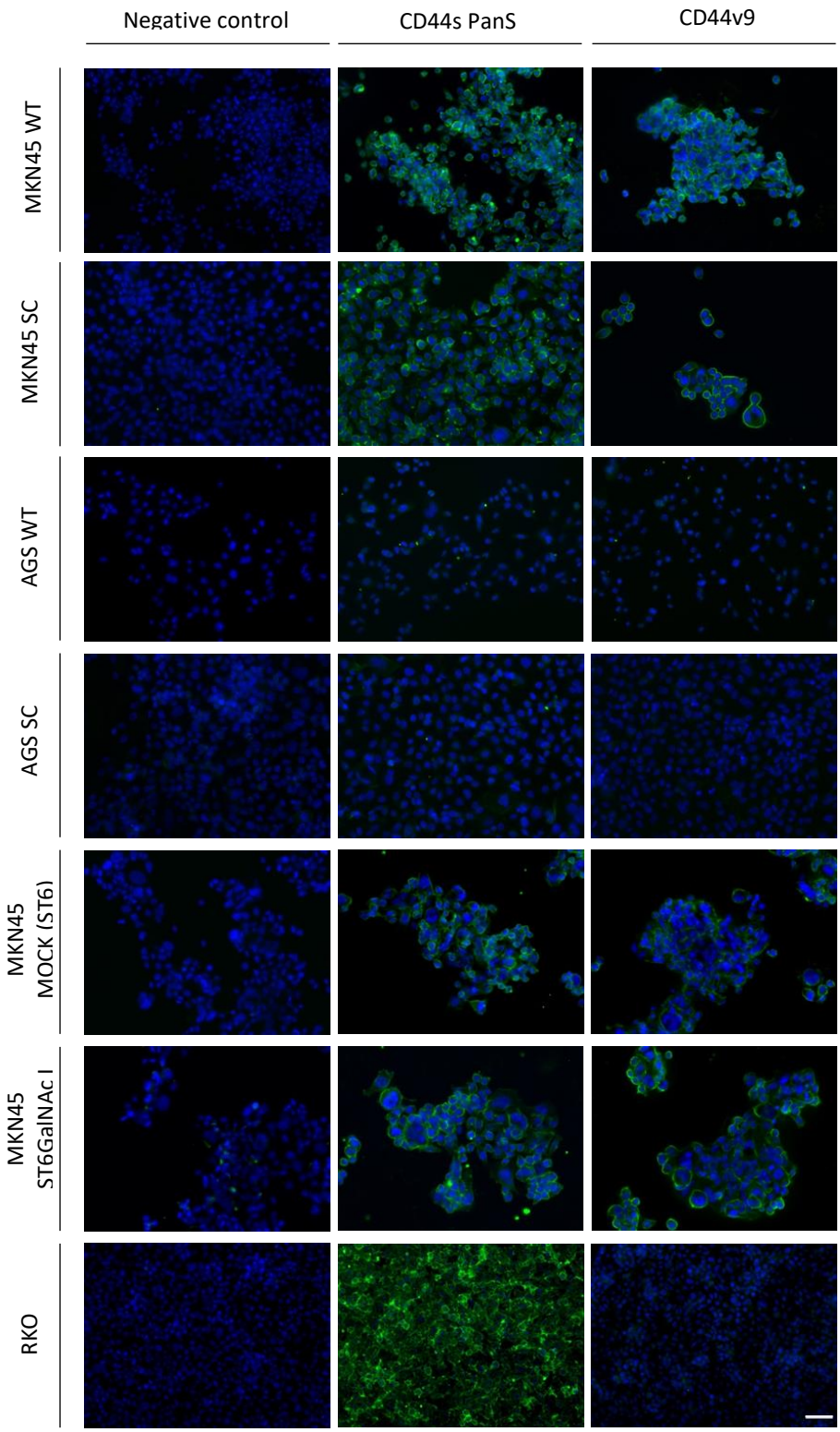


Figure 13. Immunofluorescence analysis of CD44s PanS and CD44v9 expression in GC (MKN45 and AGS) and CRC (RKO) cancer cell lines. Nuclei are shown in blue and CD44 isoforms in green. The scale bar corresponds to 50 μm.

With the CD44s and CD44v9 expression first assessed by IF, protein extracts were made and Western blot analysis was done to detect protein levels of total CD44, CD44s and CD44v9 (Figure 14). The monoclonal antibody that was used to recognize total CD44 was the clone that detects all CD44 protein isoforms. The expression profiles revealed that MKN45 WT cells express high levels of total CD44 and CD44s, whereas the presence of CD44v9 was weakly detected. The SC model expressed total CD44 at levels below 20% of the WT cells (Figure 14 A), and no CD44s (Figure 14 B), but expressed an average of 3-fold higher protein amounts of CD44v9 (Figure 14 C). Interestingly, there were no significant differences between the MOCK and ST6GalNAc I cells expression profiles of tested CD44 variants, although there is a tendency for increased CD44v9 expression in the ST6GalNAc I models (Figure 14 C). The WB analysis enables the interpretation of the molecular weights of the protein isoforms detected. Both the WT and MOCK total CD44 proteins are detected between 150 and 250 kDa, whereas in the SC there is a discrete band bellow 150 kDa, and an enrichment of total CD44 in the 150 kDa range in the ST6GalNAc I model (Figure 14 A). The same profile is detected for CD44s, except for the SC model, that shows no CD44s protein expression, as mentioned. Focusing in the CD44v9 blot, again the WT and MOCK show similar expression profiles. In the SC, CD44v9 is predominantly detected at the same molecular weight as total CD44 and in the ST6GalNAc I there is an enrichment of protein in the same molecular weight, but larger proteins are also detected (Figure 14 C).

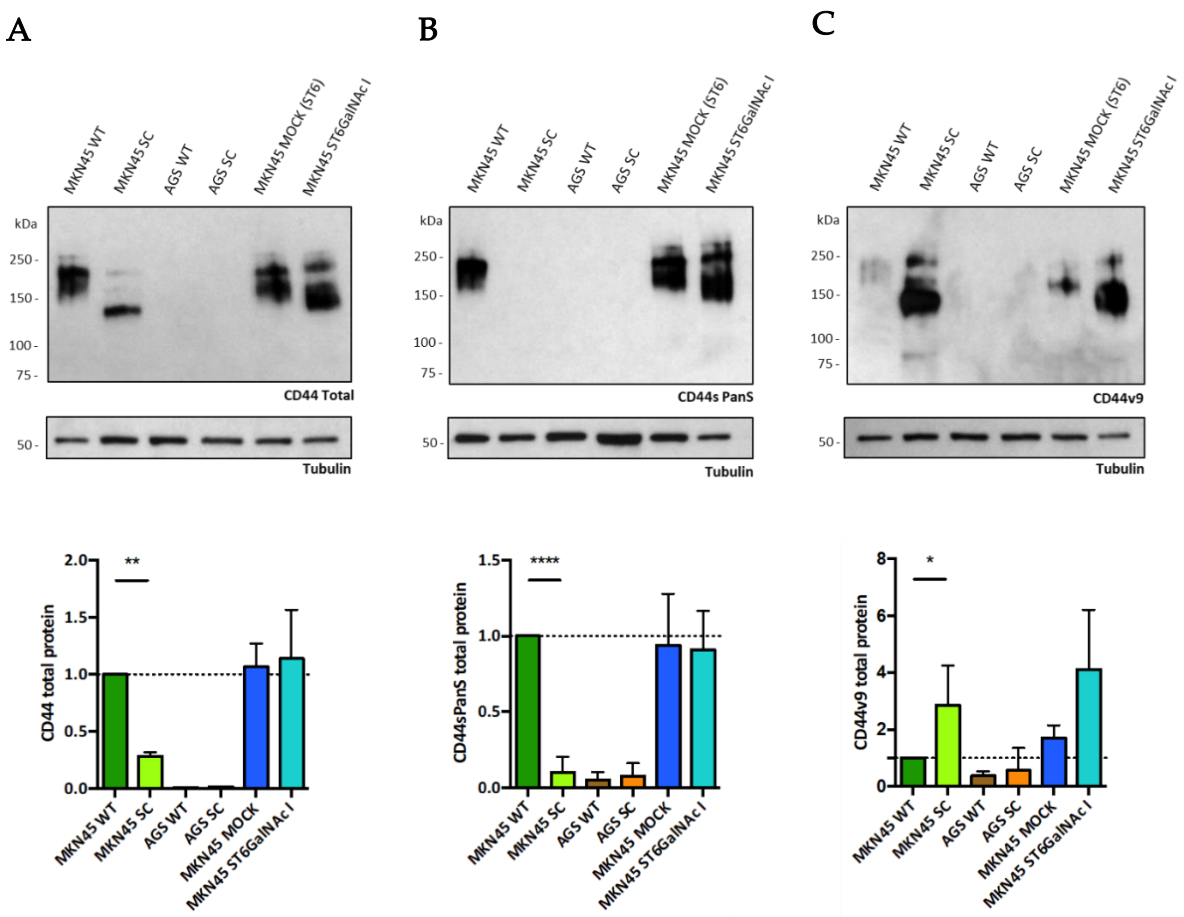


Figure 14. Western blot analysis of total CD44 and isoforms in protein extracts from gastric cancer cell line models of *O*-glycan truncation (MKN45 SC, AGS SC, MKN45 ST6GalNAc I) and their control counterparts (MKN45 WT, AGS WT and MKN45 MOCK). **(A)** Total CD44 protein detection and relative quantification. **(B)** CD44s PanS protein detection and relative quantification. **(C)** CD44v9 protein detection and relative quantification. For quantifications, the intensity of all protein bands was normalized to tubulin and MKN45 WT cell line. * indicates p-value <0.0500, ** <0.00500, **** <0.00005.

The intriguing results regarding the different immunoreactivity of the MAbs used as evaluated in IF and WB analysis, led to further investigate their reactivity in these models. Most commercial MAbs, including MAbs to CD44 variants, are raised by non-glycosylated recombinant proteins and, hence, there's the possibility that glycosylation could result in both altered CD44v protein conformation and masking of specific proteins on the peptide backbone.

Since WT cells showed CD44s and CD44v9 expression in the IF assays, but significantly reduced protein presence in the WB, when compared to the SC model, deglycosylation assays were employed on the protein extracts, to evaluate whether the elongated glycans in the WT proteins could interfere with CD44 isoforms detection. Protein extracts were incubated with deglycosylation enzyme mixes, which included neuraminidase, PNGase F and a commercial deglycosylation mix (Figure 15).

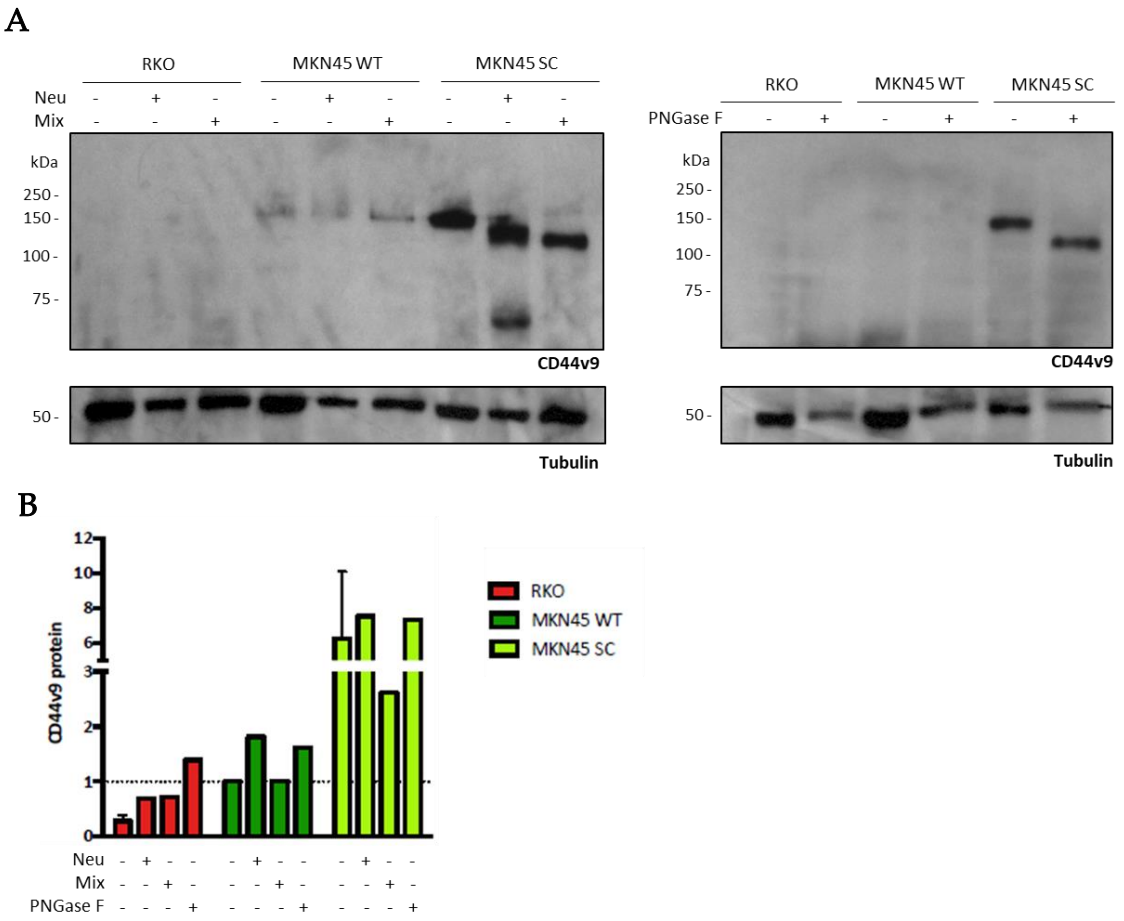


Figure 15. Impact of deglycosylation treatment in CD44v9 detection in protein extracts from colorectal cancer (RKO) and gastric cancer cell line models (MKN45 WT and SC). **(A)** Western blot analysis of CD44v9 detection after treatment with neuraminidase (Neu), a deglycosylation mix (Mix) and PNGase F. **(B)** Relative quantification of western blot results. Intensity of all protein bands was normalized to tubulin and MKN45 WT cell line without deglycosylation treatments. Intensity of the band correspondent to MKN45 SC without treatment consists of the mean between the 2 experiments.

Neuraminidase removes terminal sialic acid molecules in both *N*- and *O*-glycan structures. The PNGase F enzyme cleaves all *N*-glycans, which are known to exist in CD44: 6 sites in the standard protein, 1 in the v5 exon and 2 in the v10 exon. Finally, the commercial deglycosylation mix contains an enzyme cocktail, described in section 3.6, which cleaves the majority of glycan structures. Protein extracts from RKO cells were used as a negative control for the CD44v9 deglycosylation assays, since this cell line shows CD44s but no CD44v9 expression. In the MKN45 WT extracts the CD44v9 detection was not improved by any of the treatments, nor the shift in the molecular weight of the CD44v9 band. The removal of glycan structures should generate detection at a lower molecular weight. For the SC extracts it was observed a shift in molecular weight for all mixes, which indicated that the CD44v9 protein, in these cells, maintains *N*-glycosylation structures, and terminal sialic acids (presumably in the STn glycan and *N*-glycans).

As it was previously mentioned, the CD44 receptor is heavily *O*-glycosylated thanks to its mucin-like repeats in the variable region. In MKN45 WT background these *O*-glycan structures are complex and elongated. In CD44v the degree of glycosylation is increased as extra glycosylation sites are provided by additional inserted exon products. All together, in MKN45 WT, CD44 is expected to have a high molecular weight, that can interfere in technical aspects of the WB method, such as, the gel electrophoresis (entrance in the gel, for instance) and transfer.

To exclude this hypothesis, dot blotting analysis was performed. Dot blots are simplified protocols of WBs with the exception that the proteins to be detected were not separated according to its molecular weight in an electrophoresis step, nor transferred to a membrane from a polyacrylamide gel.

Instead, protein extracts are loaded directly to the membrane and detection is performed. For this, MKN45 WT and SC extracts were used, as well as AGS WT and SC (as a negative control) and MKN45 MOCK and ST6GalNAc I. The detection of the STn glycan was used as an internal experimental control. The results showed STn expression for both SC models and for ST6GalNAc I cells. The detection of total CD44, CD44s and CD44v9 followed the pattern of the WBs already showed (Figure 16).

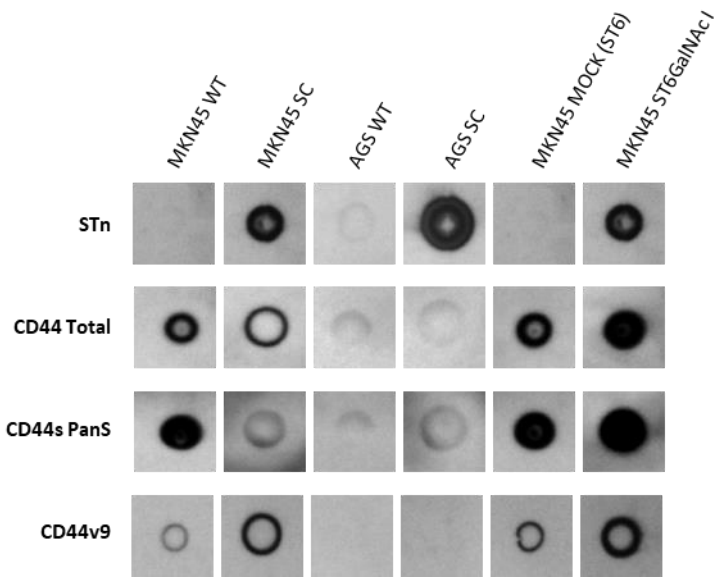


Figure 16. Dot blot analysis of STn antigen, total CD44, CD44s PanS and CD44v9 isoform in protein extracts from gastric cancer cell line models of *O*-glycan truncation (MKN45 SC, AGS SC, MKN45 ST6GalNAc I) and their control counterparts (MKN45 WT, AGS WT and MKN45 MOCK).

Since the IF gave conflicting results with WB and Dot Blot analysis regarding CD44v9 protein expression and CD44s presence in the *O*-glycan truncation models, flow cytometry analysis was done for both membrane and total CD44s and CD44v9. Cells were stained and fixed for the analysis of the membrane protein and fixed, permeabilized and stained for the analysis of the membrane and intracellular fractions. Permeabilization was done with Saponin, which enables the formation of pores in the membrane, so detection antibodies

can enter the cytoplasm (Figure 17 A). Since the permeabilization with this method is transient, the same Saponin concentration was maintained during the permeabilization and staining steps.

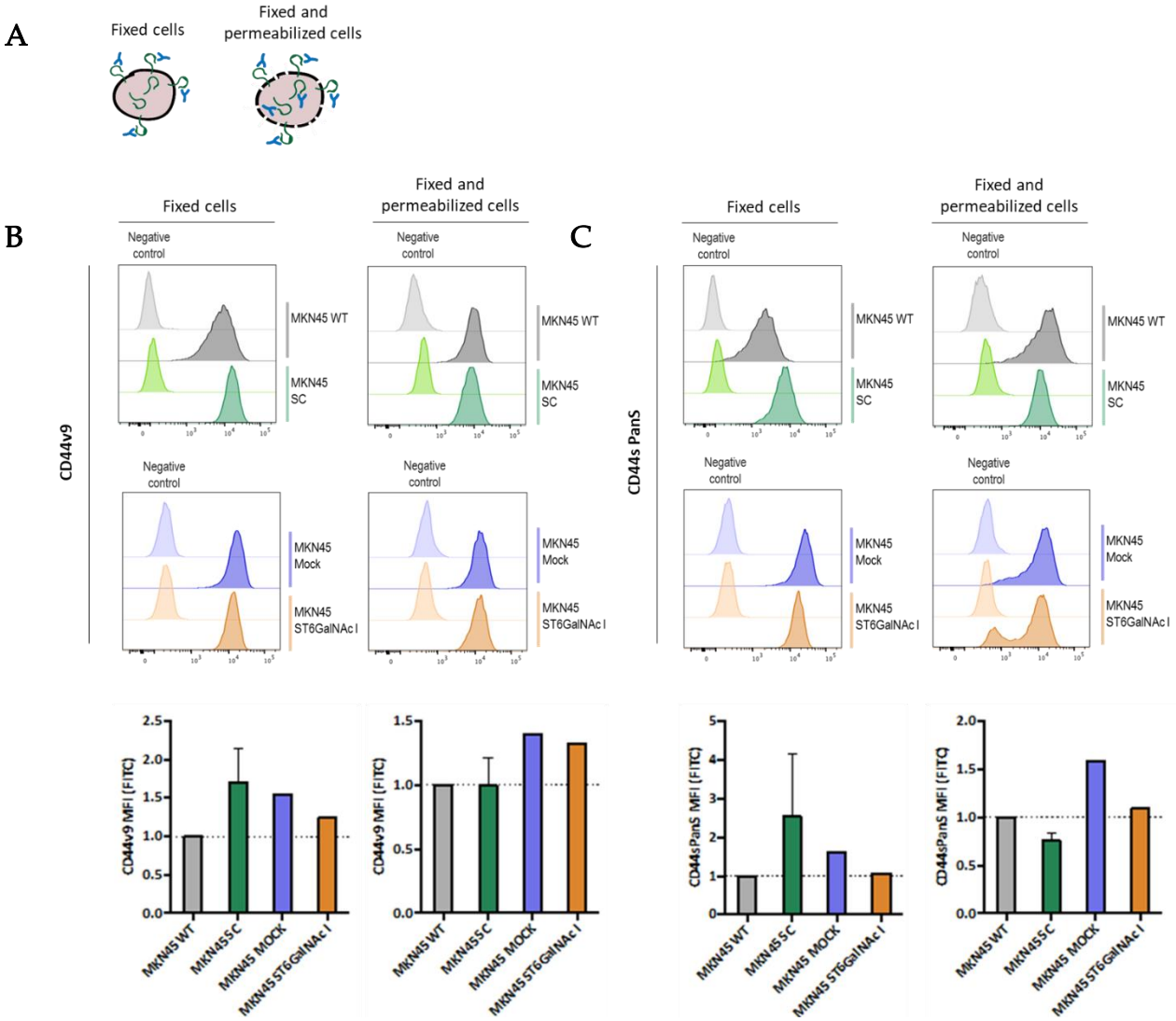


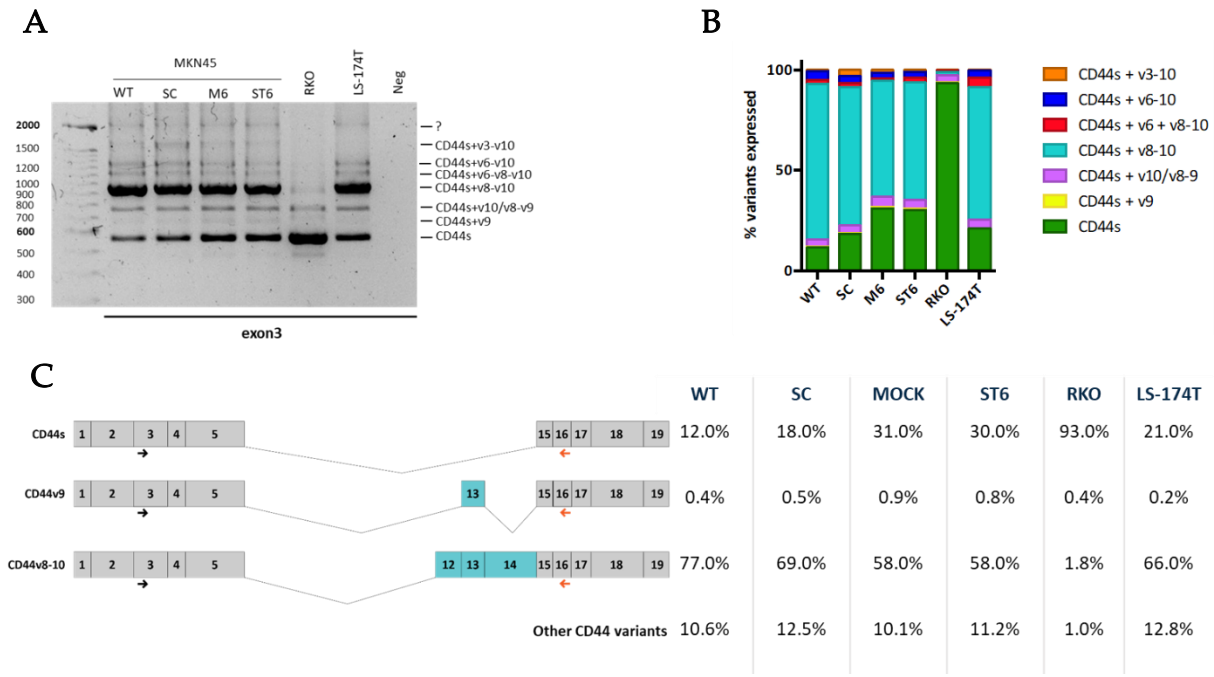
Figure 17. Flow cytometry analysis of CD44v9 and CD44s PanS isoforms expression in gastric cancer cell line models of *O*-glycan truncation (MKN45 SC and MKN45 ST6GalNAc I) and their control counterparts (MKN45 WT and MKN45 MOCK). **(A)** Schematic representation of staining protocols: surface fraction (left) and both surface and intracellular fractions (right). **(B)** Histograms of CD44v9 fluorescence signal of used cell lines and respective negative controls (top) and relative MFI quantification (bottom). **(C)** Histograms of CD44s PanS fluorescence signal of used cell lines and respective negative controls (top) and relative MFI quantification (bottom). All MFI values were subtracted the negative control intensity and normalized to the MKN45 WT cell line.

The MNK45 WT cells showed more heterogenous and reduced levels of membrane CD44v9 expression when compared to the SC model, given by the

mean fluorescence intensity (MFI) of the antibody (Figure 17 B, top left histogram and bottom left graph). There were no differences between the MOCK and ST6GalNAc CD44v9 membrane expression profiles (Figure 17 B, bottom left histogram and bottom left graph). When the detection of intracellular CD44v9 was enabled, there was no significant differences between any of the models, indicating differential distributions of the protein in the SC model (Figure 17 B, right histograms and bottom right graph). Regarding CD44s, the SC model tends to an increased membrane CD44s, compared to the WT cells, and there are no differences between MOCK and St6GalNAc I (Figure 17 C, left histograms and bottom left graph). The total protein staining revealed that there were no differences between the models, regarding CD44s protein presence (Figure 17 C, right histograms and bottom right graph).

The final analysis made was on the relative *CD44* mRNA isoforms expressed in the cell lines used. Primers were designed so specific variants would be amplified on the cDNA from total RNA extracts. The PCR reactions had an exon/variant specific forward primer and a common reverse primer targeting exon 16 (present in all isoforms) (Table 1). In total, 10 different forward primers were used. The MKN45 WT, SC, MOCK and ST6GalNAc RNA was used, as well as RNA from RKO and LS-174T cells. The LS-174T cell line is derived from CRC and has a genetically dual population in bulk: *COSMC*$^{+/+}$ or *COSMC*$^{-/-}$ [152].

The PCR products for the several isoforms were separated according to the molecular weight in an agarose gel electrophoresis and the band sizes were matched with *in silico* analysis of the mRNA sizes, after alternative splicing (Table 1). The PCR for exon 3 was designed to amplify all the isoforms in the RNA extracts, since it had forward and reverse primers targeting the two standard regions of the *CD44* transcript (Figure 18).



	WT	SC	MOCK	ST6	RKO	LS-174T
CD44s	12.0%	18.0%	31.0%	30.0%	93.0%	21.0%
CD44v9	0.4%	0.5%	0.9%	0.8%	0.4%	0.2%
CD44v8-10	77.0%	69.0%	58.0%	58.0%	1.8%	66.0%
Other CD44 variants	10.6%	12.5%	10.1%	11.2%	1.0%	12.8%

Figure 18. *CD44* mRNA levels of GC cell lines, MKN45 WT (WT), MKN45 SC (SC), MKN45 MOCK (M6) and MKN45 ST6GalNAc I (ST6), and CRC cell lines, RKO and LS-174T. **(A)** Analysis of the total set of *CD44* isoforms expressed in the different cell lines. **(B)** Percentage of detected variants expression, relative to the total PCR products of each cell line. **(C)** Schematic representation of three main *CD44* isoforms (left) and respective percentage of relative expression for the different cell lines (right).

The *CD44v8-10* was the predominant isoform in the MKN45 cell line models, with relative presences ranging from 58 to 77% of the total transcript (Figure 18 B and C). The *CD44s* was the 2nd most expressed isoform in the MKN45 cells, with larger presence for the MOCK and ST6GalNAc I models (Figure 18 B and C). LS-174T followed a similar expression pattern than the MKN45 cells, whereas the RKO cells expressed 93% of the *CD44s* isoform, which was according to the strong IF signal already observed in Figure 13 (Figure 18 B and C).

To increase sensitivity of the assay, other primers were used to further dissect isoform expression (Figure 19). The SC model stood out from the other MKN45 cell lines in the increased expression of the *CD44v3-10* and *CD44v5-10* transcripts (Figure 19 A and B). The LS-174T cell line revealed itself similar to MKN45 in the other variants' expression (Figure 19). The differences on RKO variants' expression continued to be detected. These cells expressed higher levels of the single exon variants *CD44v3*, *CD44v5* and *CD44v6* (Figure 19 A, B and C).

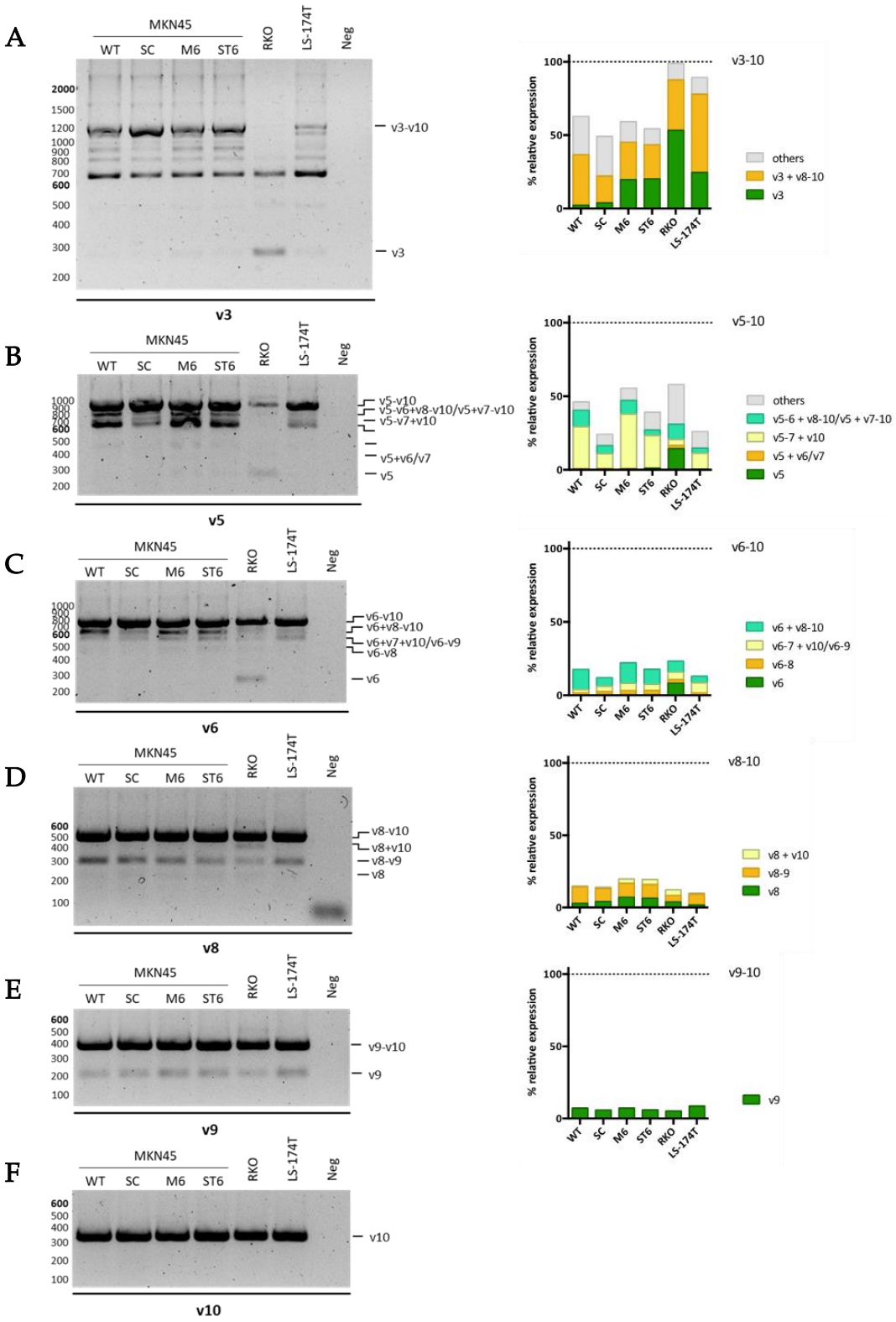


Figure 19. *CD44* mRNA levels of GC cell lines, MKN45 WT (WT), MKN45 SC (SC), MKN45 MOCK (M6) and MKN45 ST6GalNAc I (ST6), and CRC cell lines, RKO and LS-174T, using different variant specific forward primers: **(A)** v3; **(B)** v5; **(C)** v6; **(D)** v8; **(E)** v9; **(F)** v10. Percentage of detected variants expression, relative to the total PCR products of each cell line is represented on the right.

4.3 CD44v EXTRACELLULAR CLEAVAGE AND ITS RELATIONSHIP WITH *O*-GLYCAN TRUNCATION

CD44 proteolytic cleavage at its extracellular region is known to have biological implications in the regulation of the CD44 dependent cell-matrix interaction and signaling pathways [58]. As it was mentioned, *O*-glycan truncation was shown to increase total CD44 surface cleavage [139], but there was missing knowledge about specific isoforms.

The cell lines used in the work of this thesis are adherent cells and their detachment from culture flasks is required for their maintenance and to place cells in a single-cell suspension for flow cytometry analysis. The detachment protocol is based on the disruption of cell-cell and cell-flask adhesion linkages, either by chelator agents (that remove ions that mediate interactions) or proteolytic cleavage of the adhesion proteins. For the flow cytometry analysis showed in Figure 17, Versene, a commercial chelator agent was used, to preserve surface protein integrity. To test the sensitivity of CD44 variants against extracellular proteases, the CD44v9 and CD44s presence was assessed by flow cytometry comparing the detachment of cells using either Versene or Trypsin. In the cells treated with Trypsin, the levels of surface CD44v9 detection was reduced to 50% in the MKN45 WT, whereas in the SC was decreased to 5%. For the MOCK and ST6GalNAc I cells the results were similar, showing reduction to 20-30% in the Trypsin condition. The reduction in the Trypsin-treated cells had fewer extend when both surface and intracellular fraction were analyzed (Figure 20 A). When the CD44s was probed in the Trypsin-treated cells there was a reduction of 2-fold in fluorescent signal for the MKN45 WT cells. Interestingly, CD44s was not detected with these conditions in the SC model. There was a small

reduction in the MFI signal for the MOCK and ST6GalNAc I cells (Figure 20 B). When the cells were permeabilized prior to staining, there was no alteration for the WT cells and CD44s was detected in the SC model. The MOCK and ST6GalNAc I cells showed reduction of total (surface and intracellular) CD44s following Trypsin treatment (Figure 20 B).

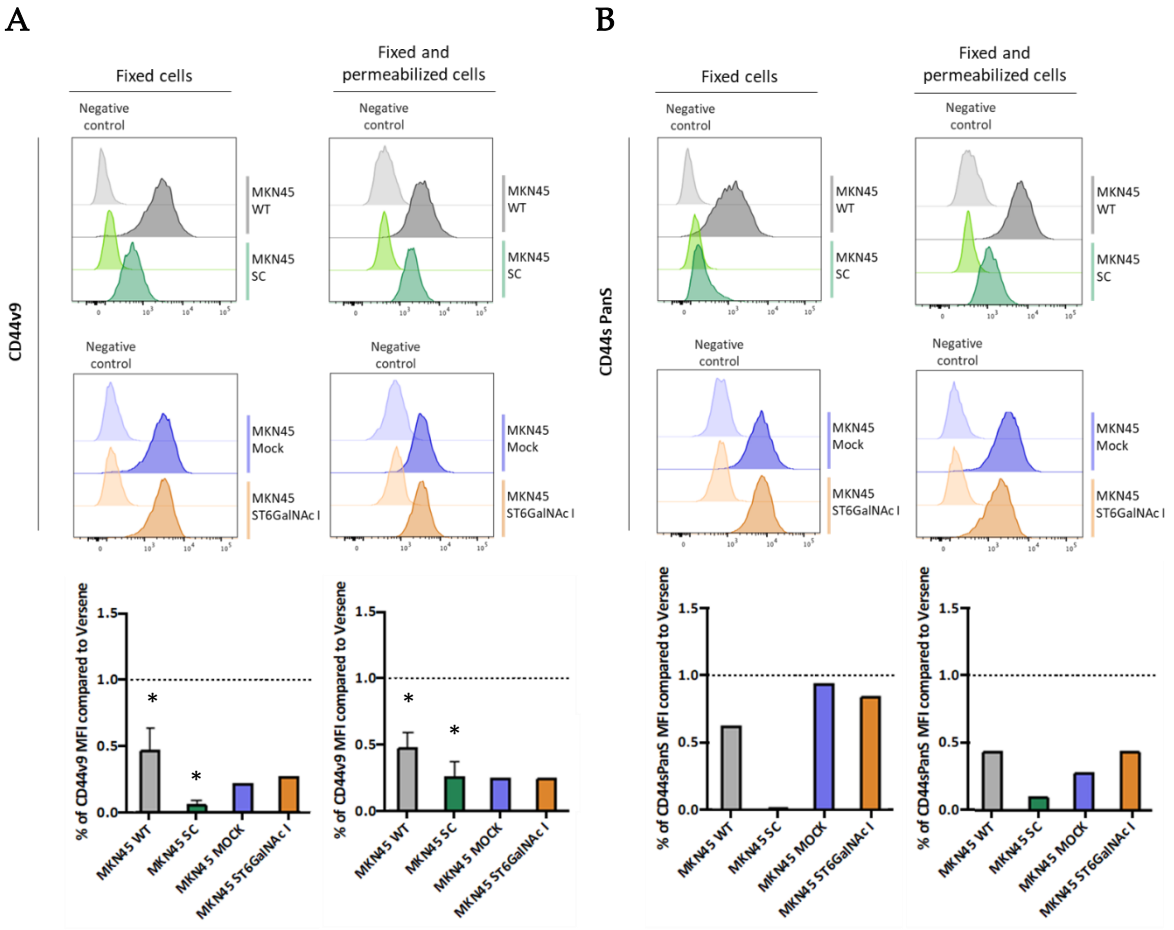


Figure 20. Flow cytometry analysis of CD44v9 and CD44s PanS isoforms expression in gastric cancer cell line models of *O*-glycan truncation (MKN45 SC and MKN45 ST6GalNAc I) and their control counterparts (MKN45 WT and MKN45 MOCK), after detachment using Trypsin. **(A)** Histograms of CD44v9 fluorescence signal of used cell lines and respective negative controls (top) and relative MFI quantification compared to Versene (bottom). **(B)** Histograms of CD44s PanS fluorescence signal of used cell lines and respective negative controls (top) and relative MFI quantification compared to Versene (bottom). All MFI values were subtracted the negative control intensity and normalized to the MKN45 WT cell line treated with Versene. * indicates p-value <0.050

The Trypsin-resistance results indicated that CD44 variants sensitivity to proteases could be altered in the genetic background responsible for the

generation of glycoprofiles enriched with immature structures. The presence of cleaved total CD44, CD44v9 and CD44s were assessed by WB of the protein extracts of the cells and the culture media (Figure 21 A). This protocol involved the culture of cells in FBS-free media for 48 hours and the removal of proteins bellow 10 kDa. As it was mentioned, the WB analysis enables the detection of the molecular weights of the proteins, relevant for cleavage experiments, since the cleaved protein is always smaller than the intact one. The band intensity was normalized to the total Ponceau staining. The relative protein presence, compared to the extract of MKN45 WT cells, was analyzed. The ST6GalNAc I showed increased cleavage of total CD44 and CD44s. For the CD44v9 cleavage, the SC and the ST6GalNAc cells showed a 4-fold increase in CD44v9 media detection (Figure 21 B).

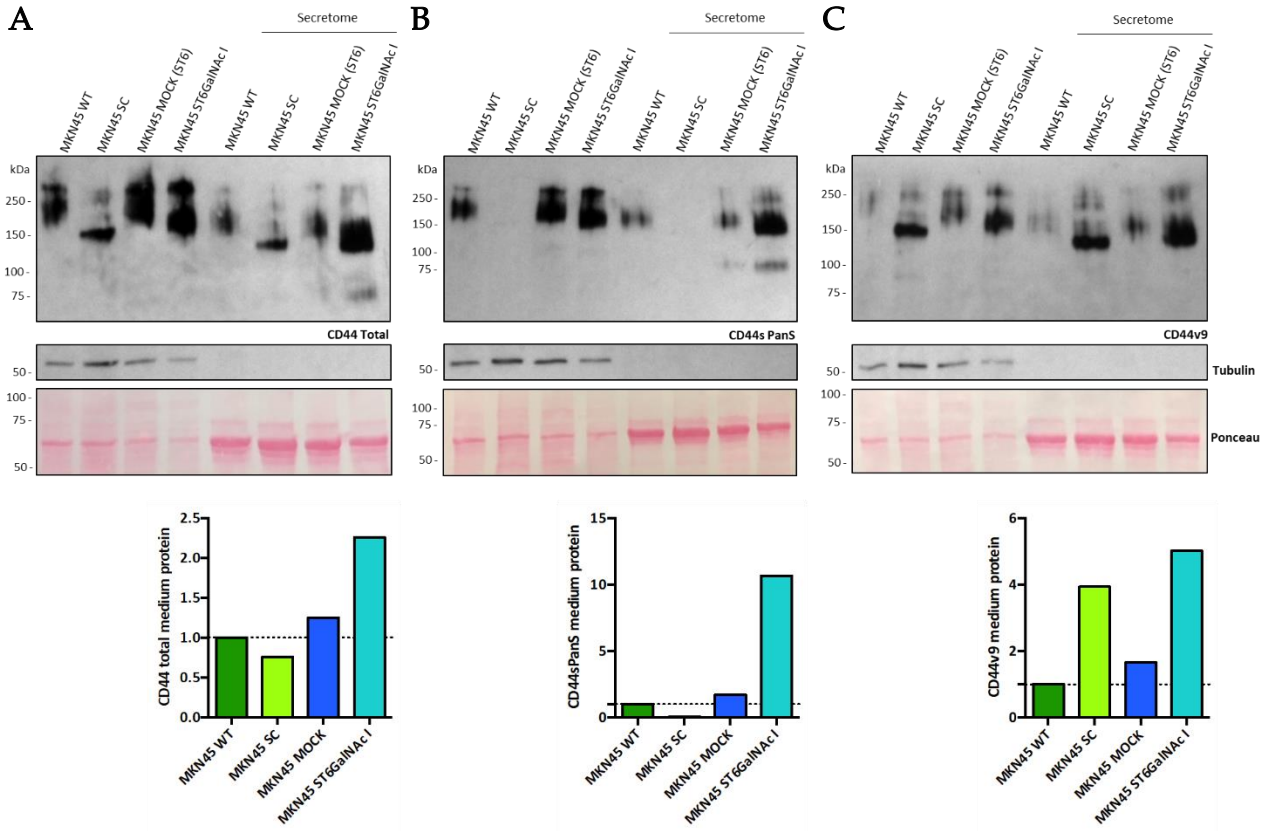


Figure 21. Comparative western blot analysis of total CD44 and isoforms and tubulin in total protein extracts and in conditioned medium (secretome), from gastric cancer cell line models of *O*-glycan truncation (MKN45 SC, AGS SC, MKN45 ST6GalNAc I) and their control counterparts (MKN45 WT, AGS WT and MKN45 MOCK). **(A)** Total CD44 protein detection and relative quantification. **(B)** CD44s PanS protein detection and relative quantification. **(C)** CD44v9 protein detection and relative quantification. For quantifications, the intensity of all protein bands was normalized to the total Ponceau staining and MKN45 WT cell line.

CHAPTER 5

GENERAL DISCUSSION

CD44 is one of the most variable surface molecules, in terms of function and structure. Alternative splicing, often altered in cancer cells [56], generates a vast repertoire of *CD44* transcripts that translate into different protein variants, containing different profiles of post-translational modifications [55, 58]. CD44 presence in both GC and CRC has been shown in different cohorts and its variant CD44v9 has also deserved attention, being associated with tumor progression and aggressiveness [75, 78, 153]. Interestingly, in previous work of the Glycobiology in Cancer group, CD44 was identified in the SC model of a gastric cancer cell line as a potential secreted biomarker *O*-glycoprotein with immature *O*-glycans, which was validated in cancer patients' sera. In the same work, the single *O*-glycosite identified by mass spectrometry in GC patient's serum was in the v9 exon [119].

In this thesis, tissue samples and cell lines from GI cancers were used to characterize the global relationship between CD44v9 and CD44s expression and a truncated *O*-glycoprofile.

I. CD44v9 expression in GI cancer tissues

CD44v9 has been shown to have critical roles in GI carcinogenesis [69, 70, 75, 78, 153] and its presence was evaluated in a GC series of patients available in the group. For this, IHC analysis was first performed in tissue samples of the target malignancies, GC and CRC. The expression of CD44v9 and the Tn and STn structures were assessed in a GC series of patients and associations were evaluated. The staining pattern of the targets tested followed what had been described for CD44v9 [153] and truncated *O*-glycans [131]. Interestingly, there is a trend of CD44v9 and STn co-expression (p-value = 0.091) in the GC series. Although being a small series, all cases with CD44v9 expression showed STn presence (Table 3). Moreover, co-localization of Tn and STn with CD44v9 was observed for the majority of the cases, suggesting that the CD44v9 proteins present in the tissue could be glycosylated with those structures. Proximity ligation assays (PLAs) targeting both truncated *O*-glycans and CD44v9 should reveal the veracity of the previous hypothesis. PLA is a technique were the distance between two antibody targets is measured, having the sensitivity do determine whether CD44v9 is sugar-coated with those glycans structures. In previous work, one of the CD44 variants, CD44v6, was also shown to be aberrantly glycosylated with the STn glycan in GC tissue sections, by PLA [119]. Additionally, CD44 proteins containing the STn structure were detected in both GC tissue protein lysates and matching sera. In this case, the higher levels of serum CD44-STn were shown to be able to differentiate between patients and

healthy subjects. However, it did not occur for tissue CD44-STn [154]. Therefore, the focus on CD44 specific isoform variants combined with cancer-related glycan structures could pave the way for new biomarker molecules in GC. The results of this thesis present CD44v9 as a potential target. Interestingly, CD44v9 was detected in normal adjacent tissue (NAT) of cancer patients. The profile of staining resembles the front-line basal membrane of a tumor, previously seen in GC samples [153]. Moreover, it could also be the staining of proliferative cells in gastric villi, that express normal and elongated *O*-glycans. Generally, in the human gut epithelium, the CD44s, CD44v6 and CD44v4-10 isoforms are found, being the last variant expressed usually by intestinal stem cells [155]. The CD44v9 positive cells in the NAT could be taking part in the regeneration of the gastric epithelium. Therefore, their glycoprofile could be a discriminating factor, meaning that CD44v9 is not expected to carry STn in non-tumorous cells. To elucidate, an IHC analysis for STn, structure that functions as proxy for cancer cells, could be performed, clarifying the true identity of CD44v9 expressing cells in the NAT.

Colorectal TMAs were also used to address the existence and extend of CD44v9 expression. Analysis of the CD44v9 profile of detection in the samples was performed to investigate any clinical associations. There was a statistically significant association of CD44v9 expression and tumor size (Figure 11 and Table 5). This agrees with the notions that CD44v9 positive tumors are more aggressive. In addition, a positive trend is observed (p-value = 0.132) between CD44v9 expression and the absence of vessel invasion. What can be hypothesized is that the presence of CD44v9, which provides resistance to ROS [69], as it was discussed in the Introduction, inhibits angiogenesis. Elevated levels of intracellular ROS were shown to induce vessel formation through the vascular

endothelial growth factor (VEGF) axis [156]. To confirm this, the evaluation of VEGF presence could be inversely related with CD44v9. In fact, CD44v9 presence could be used as an additional parameter for patient differentiation in anti-VEGF targeted therapies in CRC [157]. Moreover, the absence of CD44v9 in stage 0 CRC indicates that it is a cancer specific marker. The noteworthy 2-fold increase in CD44v9 positive sample percentage within the early stage 1 and the late stage 4 disease also represents the CD44v9 relationship with cancer progression and aggressiveness. Also following this trend in the GC series is the expanding growth pattern showed in CD44v9 positive cases (4 out of 5), according to the Ming classification. Nevertheless, future work on the expression of Tn and STn in the TMAs will be crucial to further investigate the relationship between truncated *O*-glycans and CD44v9.

II. CD44 variants presence characterization in truncated *O*-glycan models

Given these results, it is clear that CD44v9 has a major role in gastric and colorectal cancer and that the combination with *O*-glycan truncation could present itself as a novel promising mechanism of malignancy. To dissect these mechanisms, alterations in the cancer-related CD44v9 [157] and the ubiquitously expressed CD44s [158] isoforms expression were evaluated in truncated *O*-glycan models that are described in section 1.4 of the Introduction chapter. Also, identification of CD44 isoforms involved in all these processes is possible mainly using CD44v specific monoclonal antibodies. The present work assesses the immunoreactivity of some of these MAbs. Immunodetection of CD44 could be impaired essentially by: (i) post-translational modifications that can mask exon-specific epitopes, (ii) conformational changes due to both alternative splicing and post-translational modifications; (iii) specific exon splicing; and (iv) protein

cleavage/shedding. All of these features are cell type specific and thus, likely to impact immunodetection, and they were all tested in this work.

The initial approach was to perform IF analysis of the two target CD44 variants presence in cultured cells. This analysis gave the preliminary profiles of CD44v9 and CD44s expression that guided following work. The AGS cell models, being negative for CD44 presence, were further used as a negative control.

Western blot analysis revealed distinct patterns of total CD44, CD44s and CD44v9 protein detection. It is known that glycans increase molecular weight. Therefore, it would be expected the same profile of molecular weight distribution to be shifted to lower sizes for the *O*-glycan truncation models, given the lack of complex elongated structures. This is observed between the MOCK and ST6GalNAc I models in the total CD44, CD44s and CD44v9 blots. The shift is further seen in the WBs comparing the WT and SC. However, a CD44 sharper band was observed in the total CD44 WB in the SC model. This band may correspond to a condensation of the CD44 pattern shown in the WT, due to the homogeneous profile of the SC cells, composed exclusively of truncated *O*-glycans. Additionally, CD44s detection is absent in the SC, which conflicts with the IF results. Both WB results for CD44 total and CD44s show that the SC and ST6GalNac I models, although being used for the same purpose of *O*-glycan truncation, have distinct glycoprofiles regarding CD44 isoforms detection.

CD44v9 detection both in WT and MOCK is significantly reduced when compared with the correspondent models with truncated *O*-glycans (Figure 14). Since Dot blot analysis validated all WB results, it excludes the possibility of technical problems concerning protein electrophoresis and transfer (Figure 16).

Therefore, the results could indicate that CD44 variants have differential protein regulation in the truncated *O*-glycan models.

Antibodies used in experimental procedures detect protein epitopes that can be influenced by post-translational structures, including glycans. IF and WB gave conflicting results regarding CD44 isoforms, that could be a product of glycan-dependent antibody recognition. Therefore, deglycosylation assays were done for CD44v9. In the WT, the deglycosylation treatment did not improve CD44v9 detection and the expected size shift towards lower molecular weight did not occur. This could be explained by the inefficient removal of *O*-glycan structures. In this experiment, GAGs were not removed because they are conserved among cell models and are not expected to influence antibody detection. Also, since the *N*-glycosylation pathway is theoretically unchanged in the SC model, the PNGase F treatment should not improve detection in WT protein extracts, which is observed. In the SC model, the *O*-glycans do not contribute to the significant increase in molecular weight of CD44, as they are truncated. However, the detected band is the same for PNGase F and the deglycosylation mix, which further validates that the mix is not properly working in its fullest, as it appears to only be removing *N*-glycans. This is not surprising given that the *O*-glycan structures removal requires a vast repertoire of glycosidases, that sequentially remove individual monosaccharides (Figure 15). In addition, as CD44 is heavily glycosylated, containing 146 predicted sites of *O*-glycosylation [139], it is reasonable to think that the commercial cocktail does not provide the efficiency required. On the other hand, *N*-glycans have a conserved structure which enables full removal using a single glycosidase. Besides, CD44 only has 9 *N*-glycosylation sites (6 in CD44s, 1 in CD44v5 and 2 in CD44v10), a much smaller number when compared to the *O*-glycosylation sites

[79, 80]. To further dissect the impact of *O*-glycosylation elongation in antibody recognition, *in vitro O*-glycosylation inhibition assays should be performed. Moreover, those assays would also give insights to CD44s PanS antibody detection.

The cellular localization of the CD44 variants studied in this thesis could give insights on the differences observed on the WB analysis. To achieve this, surface and intracellular staining were performed for CD44v9 and CD44s and evaluated by flow cytometry (Figure 17). Total CD44 is not included as previous published work showed no differences between the cell lines [139]. The SC model shows an increase in surface CD44v9, when compared to the WT, that is not observed in the total protein staining. This reveals a differential distribution of the protein between cell models, with the SC model having elevated CD44v9 levels at the surface. Also, this can be linked to already described malignant features of the SC model [120]. Interestingly, between the MOCK and ST6GalNAc I this is not observed, showing that different truncated *O*-glycans profiles have distinct influence in CD44v9 distribution. Regarding CD44s, the SC model shows the highest protein expression at the cell surface but the lowest total protein presence, which contradicts WB results.

To assess whether alternative splicing of the *CD44* gene transcript is being influenced by truncated *O*-glycans, PCRs where done. What is observed is that *O*-glycan truncation does not interfere with the global *CD44* profile of expression (Figure 18), given by the comparison of MKN45 WT and MOCK with SC and ST6GalNAc I, respectively. Two CRC cell lines were included for the variant relative expression evaluation, RKO and LS-174T. LS-174T showed similar variant presence compared to MKN45 models. This enables the use of this cell line for the validation of the CD44 behavior, described in this thesis, for *O*-glycan

truncation in colorectal cancer. Regarding RKO, the vast majority of *CD44* transcripts corresponds to the standard CD44s isoform, without variant exons, whis is in agreement with IF results.

O-glycan truncation impact on *CD44* alternative splicing is observed when PCR sensitivity is increased for some specific variants (Figure 19 A-C). These specific splicing events could be linked to receptor signaling pathways, which are known to be influenced by their glycoprofile, mediating receptor clustering and turnover rates [33, 79, 80]. The mainly expressed CD44 isoform containing the CD44v9 exon is the CD44v8-10. The relative quantification showed that its transcript is not differentially detected between MKN45 cell lines. This proves that the CD44v9 distinct behavior that was observed in the previous assays can only be explained by post-translational regulation, as opposed to alterations in gene transcription.

Analyzing all the results obtained in subchapter 4.2, the recognition of CD44v9 by monoclonal antibodies revealed to have glycan-variant specific dependencies [159]. On the other hand, the CD44s PanS antibody detection inconsistencies between the different techniques performed could be attributed to a dependency regarding protein conformation, associated to a specific glycosylation profile. In fact, SC engineering could, *per se*, contribute to the change in conformation of glycosylated proteins such as CD44. Together with the fact that the three-dimensional protein structure was denatured for WB analysis, this could explain the lack of CD44s presence in the SC model, but antibody recognition when it is used in assays without protein denaturation, like IF or flow cytometry. The analysis of the RNA results (Figure 18) for the MKN45 WT, MOCK and ST6GalNAc I revealed that the total CD44 protein repertoire is not exclusively composed by CD44s. However, given the analysis of the WB band

profiles of those cell lines, it reveals the exact same patterns using total CD44 or CD44s PanS antibodies. This opens the question of CD44s PanS antibody specificity. One approach to clarify the antibody's target would be to transfect CD44 negative cells, AGS cell line for instance, with either *CD44s* or *CD44v8-10* expression plasmids, and perform a WB analysis for total CD44, CD44s and CD44v8-10 proteins. If the staining observed by the CD44s PanS antibody would contain any signal in the CD44v8-10-transfected cells, the specificity of the antibody could be interrogated. Furthermore, to assess the glycan structures influence on recognition, the same experimental setup could be done in the AGS SC model.

The increased detection of CD44v9 in the SC model could either be explained by a more efficient antibody recognition of the receptor containing truncated *O*-glycans, or by an enrichment of the protein in these cells. The detection issue will be further dissected in future work, already mentioned, with the inhibition of the *O*-glycosylation pathway. In fact, the CD44v9-specific MAb clone RV3 used in this thesis has been applied in GC research. Kodama *et al.* studied the expression of three CD44v in GC tissue samples: CD44s, CD44v6 and CD44v9 (using the RV3 clone). The expression of all the CD44 variants at the tumor invasive front showed association with prognosis. The expression of CD44v9 in the whole tumor, equivalent to the profile observed in the GC series presented in this thesis, was the only that showed association with poor survival in the cohort studied [160]. Based on the results of section 4.1, the co-expression of CD44v9 and Tn or STn evaluation could increase even further the prognostic impact of this variant.

The other explanation implies a misregulation of CD44v9 dynamics in the SC model. Assuming no differences in protein translation between the WT and

SC cell lines, the CD44v9 containing isoforms could be more stable in MKN45 SC cells. Immature glycans were shown to decrease protein stability and promote degradation [33]. However, CD44v9 containing isoforms could be enriched in the SC model because they have the highest half-life. Supporting this, when intracellular CD44v9, the set of newly translated proteins, is the staining target (Figure 17), there are no differences between the models. This hypothesis could be demonstrated in assays where protein synthesis and degradation are inhibited, with the evaluation of CD44v9 presence in the different models.

III. CD44 variants extracellular cleavage in truncated *O*-glycans glycoprofiles

One of the main post-translational events of the CD44 protein is its extracellular cleavage [58]. Using a simple assay, the protease sensitivity of CD44v9 and CD44s was evaluated in the presence of truncated *O*-glycans. Trypsin treatment showed a decrease in surface protein detection by flow cytometry, for all models. However, for the SC model it was drastic. This indicated that *O*-glycans have an impact on trypsin proteolytical cleavage. Elongated and complex *O*-glycans seem to protect surface proteins from trypsin recognition and cleavage, when compared to truncated structures (Figure 20), as seen in other work [161].

Additionally, the secretome analysis showed that, whenever the *O*-glycosylation pathway was blocked, in the SC and ST6GalNAc I models, the secreted/shed CD44v9 detection was promoted, compared to WT or MOCK counterparts. Interestingly, only in the SC model, this was not observed for the total CD44, neither for CD44s, as it was for ST6GalNAc I cells, where all CD44 proteins tested had higher secreted fractions. Once again, SC and ST6GalNAc I cell models showed different behaviors. The last result reveals increased cleavage

rates of CD44ecd in immature *O*-glycan composed glycoprofiles, which leads to the release of the CD44icd from the membrane and consequential translocation to the nucleus. Nuclear CD44icd can associate with transcription complexes and activate the transcription of several genes [58]. Interestingly, CD44icd-depedent gene activation shares many of the targets of the hypoxia-inducible factor-1alpha (HIF1a) transcription factor [162]. Cell viability and growth in hypoxic conditions is a prominent feature of cancer cells [9] and has been related with the glycosylation profile in malignant tissues [163]. In fact, previous work showed in bladder cancer cells that hypoxia enhanced the expression of the STn structure, which in turn increased malignant characteristics [164]. The overexpression of truncated *O*-glycans, which leads to increased CD44icd-dependent gene activation (due to increased cleavage), could be one of the responsible mechanisms involved in the maintenance of truncated *O*-glycans expression in cancer cells, in a positive feedback loop. This hypothesis could have implications *in vivo*. In fact, results showed a co-localization of CD44v9 with Tn and STn in GC tissue samples. Knowing that truncated *O*-glycans facilitate CD44 extracellular cleavage, this event could be observed *in vivo*, enabling the use of cleaved CD44v9 with truncated *O*-glycans as a serum biomarker for tumor progression. This could also explain why this specific variant was found in sera of GC patients [165]. Therefore, the ELISA protocol for the detection of this specifically glycosylated CD44 isoform could be applied in clinical practice, with clear implications for GC patient's care and overall survival.

Glycobiology adds an additional layer of complexity to biological processes, as it is a non-templated post-translational modification. The potential alterations due to glycosylation changes in cancer constitute a major point of attention in research and clinal practice.

The use of specific CD44v for cancer treatment was already addressed in the Introduction chapter. One example of its failure is based on the expression of a cancer-related variant, CD44v6, in the human skin epithelium. CD44v9 was also shown to be expressed in the human skin epithelial cells [166], what could compromise its use in GC therapeutic or biomarker applications. However, when the glycoprofile of the cells is taken into account, CD44v9 with truncated *O*-glycans could be cancer-unique, as the human skin epithelium was shown to be STn and Tn negative [130]. Therefore, it would have been crucial to consider in the referred CD44v6 therapeutic target, that the glycoprofiles of skin and cancer cells are completely different, which could be used to differentiate them. This failed clinical trial exemplifies how important glycosylation is to consider when choosing biomarkers and therapeutic targets for clinical applications.

CONCLUSIONS

In this thesis, results have been generated on the impact of truncated *O*-glycans in cancer associated CD44 variants, suggesting that these glycan structures promote oncogenic features of CD44. The main conclusions of the work are:

- CD44v9 has a cancer-specific detection in gastric and colorectal cancer tissue cohorts, showing co-expression with STn in the GC samples, and a statistically significant association with tumor size in CRC TMAs;
- Truncated *O*-glycans influence molecular features of CD44 and its variant isoforms: antibody detection, molecular weight and cellular distribution of CD44v9 and CD44s;
- The presence of CD44v9 protein or its recognition is clearly enhanced by immature *O*-glycans but not *N*-glycans;
- The transcription of CD44v9 and CD44s does not seem to be influenced by truncated *O*-glycans;
- *O*-glycans modulate protease sensitivity of CD44v9 and CD44s.

These findings have special relevance in cancer development, providing promising directions to identify a serum biomarker and a target for treatment, and demonstrating that glycosylation cannot be neglected in detecting and targeting those markers.

FUTURE PROSPECTS

This thesis provides the baseline knowledge for future studies regarding the influence of CD44v9 glycoprofile on its malignant functions, as this protein/truncated glycans association may improve its specificity as a biomarker and therapeutic target. The work presented will be submitted to a peer-review journal, resulting in a published article. Moreover, the risen hypothesis deserves clarification. Therefore, a one-year project was designed to unravel CD44 glycoforms in GI cancer, with the following aims:

1. Detect a CD44v9 specific glycoform in the serum of GI cancer patients;
2. Characterize *in vitro* the CD44 protein dynamins in a context of *O*-glycosylation truncation;
3. Establish the contribution of CD44 pathways in the cancer-related aberrant immature *O*-glycans presence.

This project will fulfill, in a very specific way, several gaps that currently exist regarding the impact of glycosylation in important cancer biomarkers.

Finally, the long-term prospect is the implementation in the clinical practice of a biomarker based on the presence of CD44v9 glycoforms both in malignant tissue and the serum of GI cancer patients.

REFERENCES

[1] Makena, M., Ranjan, A., Thirumala, V. and Reddy, A. (2018). Cancer stem cells: Road to therapeutic resistance and strategies to overcome resistance. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 18, 30476-30479.

[2] Tredan, O., Galmarini, C. M., Patel, K. and Tannock, I. F. (2007). Drug resistance and the solid tumor microenvironment. *J Natl Cancer Inst*, 99, 1441-1454.

[3] Qian, C., Mei, Y. and Zhang, J. (2017). Cancer metastasis: issues and challenges, *Chinese Journal of Cancer*, 36 (1).

[4] Sporn, M. B. (1996). The war on cancer. *Lancet*, 347, 1377–1381.

[5] Bishop, J. M. and Weinberg, R. A. (1996). Molecular Oncology. New York: *Scientific American, Inc*.

[6] Kinzler, K. W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell*, 87, 159–170.

[7] Stratton, M., Campbell, P. and Futreal, P. (2009). The cancer genome. *Nature*, 458(7239), 719-724.

[8] Hanahan, D. and Weinberg, R. (2000). The Hallmarks of Cancer. *Cell*, 100(1), 57-70.

[9] Hanahan, D. and Weinberg, R. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5), 646-674.

[10] World Health Organization. (2019). *WHO Cancer Control Programme*. (online) Available at: https://www.who.int/cancer/en/ (Accessed 19 Jul. 2019)

[11] Gco.iarc.com. (2019). *Cancer today*. [online] Available at: https://gco.iarc.com/today (Accessed 9 Jul. 2019)].

[12] Stewart, B. and Wild, C.P. (eds.), International Agency for Research on Cancer, WHO. (2014) World Cancer Report 2014 [Online].

[13] Who.int. (2019). Cancer. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/cancer (Accessed 10 Jul. 2019).

[14] Sitarz, R., Skierucha, M., Mielko, J., Offerhaus, J., Maciejewski, R., and Polkowski, W. (2018). Gastric cancer: epidemiology, prevention, classification, and treatment. *Cancer Management and Research*, 10, 239–248.

[15] Maconi, G., Manes, G. and Porro, G. (2008). Role of symptoms in diagnosis and outcome of gastric cancer. *World Journal of Gastroenterology*, 14(8), 1149.

[16] Kelley, J. and Duggan, J. (2003). Gastric cancer epidemiology and risk factors. *Journal of Clinical Epidemiology*, 56(1), 1-9.

[17] Correa, P., Piazuelo, M. B. and Camargo, M. C. (2004). The future of gastric cancer prevention. *Gastric Cancer*, 7, 9–16.

[18] Lauren, P. (1965). The Two Histololgica Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand*, 64, 31-49.

[19] Polkowski, W., van Sandick, J. W., Offerhaus G. A., ten Kate, F. J., Mulder, J., Obertop, H., and Van Lanschot, J. B. (1999). Prognostic value of Lauren classification and c-erbB-2 oncogene overexpression in adenocarcinoma of the esophagus and gastroesophageal junction. *Annals of surgical oncology*, 6 (3), 290-297.

[20] Cai, Z., Cao, Y., Luo, Y., Hu, H. and Ling, H. (2018). Signalling mechanism(s) of epithelial-mesenchymal transition and cancer stem cells in tumour therapeutic resistance. *Clin Chim Acta*, 483, 156-163.

[21] Van Cutsem, E., Sagaert, X., Topal, B., Haustermans, K., and Prenen, H. (2016). Gastric cancer. *The Lancet*, 388(10060), 2654-2664.

[22] Carneiro, P., Fernandes, M. S., Figueiredo, J., Caldeira, J., Carvalho, J., Pinheiro, H., Leite, M, Melo, S., Oliveira, P., Simoões-Correia, J., Oliveira, M. J., Carneiro, F., Figueiredo, C., Paredes, J., Oliveira, C. and Seruca, R. (2012). E-cadherin dysfunction in gastric cancer - Cellular consequences, clinical applications and open questions. *FEBS Letters*, 586(18), 2981–2989.

[23] The Cancer Genome Atlas Research, N., et al. (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*, 513(7517), 202-209.

[24] Fontana, E. and Smyth, E. (2016). Novel targets in the treatment of advanced gastric cancer: a perspective review. *Therapeutic Advances in Medical Oncology*, 8(2), 113-125.

[25] Coburn, N., Cosby, R., Klein, L., Knight, G., Malthaner, R., Mamazza, J., Mercer, C. and Ringash, J. (2018). Staging and surgical approaches in gastric cancer: A systematic review. *Cancer Treatment Reviews*, 63, 104-115.

[26] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin, in press*. The online GLOBOCAN 2018 database is accessible at http://gco.iarc.fr/, as part of IARC's Global Cancer Observatory].

[27] Brenner, H., Kloor, M. and Pox, C. (2014). Colorectal cancer. *The Lancet*, 1–13.

[28] Winawer, S., Zauber, A., Ho, M., O'Brien, M., Gottlieb, L., Sternberg, S., Waye, J., Schapiro, M., Bond, J., Panish, J., Ackroyd, F., Shike, M., Kurtz, R., Hornsby-Lewis, L., Gerdes, H. and Stewart, E. (1993). Prevention of Colorectal Cancer by Colonoscopic Polypectomy. *New England Journal of Medicine*, 329(27), 1977-1981.

[29] Simpson, J. and Scholefield, J. (2008). Treatment of colorectal cancer: surgery, chemotherapy and radiotherapy. *Surgery (Oxford)*, 26(8), 329-333.

[30] Koido, S., Ohkusa, T., Homma, S., Namiki, Y., Takakura, K., Saito, K., Ito, Z., Kobayashi, H., Kajihara, M., Uchiyama, K., Arihiro, S., Arakawa, H., Okamoto, M., Gong, J. and Tajiri, H. (2013). Immunotherapy for colorectal cancer. *World journal of gastroenterology*, 19(46), 8531–8542.

[31] Ganesh, K., Stadler, Z. K., Cercek, A., Mendelsohn, R. B., Shia, J., Segal, N. H., and Diaz, L. A. (2019). Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nature Reviews Gastroenterology & Hepatology*, 1–15.

[32] Smith, G., Carey, F. A., Beattie, J., Wilkie, M. V., Lightfoot, T. J., Coxhead, J., Garner, R. C., Steele, R. and Wolf, C. R. (2002). Mutations in APC, Kirsten-ras, and p53—alternative genetic pathways to colorectal cancer. *Proceedings of the National Academy of Sciences*, 99(14), 9433–9438.

[33] Pinho, S. S. and Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, 15(9), 540–555.

[34] Mariampillai, A., Cruz, J., Suh, J., Sivapiragasam, A., Nevins, K. and Hindenburg, A. (2017). Cancer Antigen 72-4 for the Monitoring of Advanced Tumors of the Gastrointestinal Tract, Lung, Breast and Ovaries. *Anticancer Research*, 37(7).

[35] Filella, X., Fuster, J., Molina, R., Grau, J. J., Grande, L. and Ballesta, A. M. (1994) Tag-72, CA 19.9 and CEA as Tumor Markers in Gastric Cancer. *Acta Oncologica*, 33(7), 747-751.

[36] Vacante, M., Borzì, A. M., Basile, F., and Biondi, A. (2018). Biomarkers in colorectal cancer: Current clinical utility and future perspectives. *World journal of clinical cases*, 6(15), 869–881.

[37] Waddell, T., Verheij, M., Allum, W., Cunningham, D., Cervantes, A., and Arnold, D. (2013). Gastric cancer: ESMO–ESSO–ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 24(6), 57-63.

[38] Lytle, N. K., Barber, A. G., and Reya, T. (2018). Stem cell fate in cancer growth, progression and therapy resistance. *Nature Reviews Cancer*, 1–12.

[39] Jordan, C. T., Guzman, M. L., and Noble, M. (2006). Cancer Stem Cells. *The New England Journal of Medicine*, 355, 1253–1261.

[40] Pardal, R., Clarke, M. F. and Morrison, S. J. (2003). Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer*, 3(12), 895-902.

[41] Irvin, D. K., Jouanneau, E., Duvall, G., Zhang, X. X., Zhai, Y., Sarayba, D., Seksenyan, A., Panwar, A., Black, K. L. and Wheeler, C. J. (2010). T cells enhance stem-like properties and conditional malignancy in gliomas. *PLoS One*, 5(6), e10974.

[42] Dalerba, P., Dylla, S. J., Park, I. K., Liu, R., Wang, X., Cho, R. W., Hoey, T., Gurney, A., Huang, E. H., Simeone, D. M., Shelton, A. A., Parmiani, G., Castelli, C. and Clarke, M. F. (2007). Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A*, 104, 10158–10163.

[43] Wang, J. Y., Chang, C. C., Chiang, C. C., Chen, W. M. and Hung, S. C. (2012). Silibinin suppresses the maintenance of colorectal cancer stem-like cells by inhibiting PP2A/AKT/mTOR pathways. *J Cell Biochem*, 113, 1733–1743.

[44] Zhou, Y., Xia, L., Wang, H., Oyang, L., Su, M., Liu, Q., Lin, J., Tan, S., Tian, Y., Liao, Q. and Cao, D. (2017). Cancer stem cells in progression of colorectal cancer. *Oncotarget*, 9(70).

[45] O'Brien, C. A., Pollett, A., Gallinger, S., and Dick, J. E. (2006). A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature*, 445(7123).

[46] Shmelkov, S. V., Butler, J. M., Hooper, A. T., Hormigo, A., Kushner, J., Milde, T., St Clair, R., Baljevic, M., White, I., Jin, D. K., Chadburn, A., Murphy, A. J., Valenzuela, D. M., *et al.* (2008). CD133 expression is not restricted to stem cells, and both CD133+ and CD133- metastatic colon cancer cells initiate tumors. *J Clin Invest*, 118, 2111–2120.

[47] Zheng, Z. X., Sun, Y., Bu, Z. D., Zhang, L. H., Li, Z. Y., Wu, A. W., Ji, J. F., *et al.* (2013). Intestinal stem cell marker LGR5 expression during gastric carcinogenesis. *World journal of gastroenterology*, 19(46), 8714–8721.

[48] Hessman, C. J., Bubbers, E. J., Billingsley, K. G., Herzig, D. O. and Wong, M. H. (2012). Loss of expression of the cancer stem cell marker aldehyde dehydrogenase 1 correlates with advanced-stage colorectal cancer. *Am J Surg*, 203, 649–653.

[49] Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, 100(7), 3983–3988.

[50] Bourguignon, L. Y. W., Wong, G., Earle, C., and Chen, L. (2012). Hyaluronan-CD44v3 Interaction with Oct4-Sox2-Nanog Promotes miR-302 Expression Leading to Self-renewal, Clonal Formation, and Cisplatin Resistance in Cancer Stem Cells from Head and Neck Squamous Cell Carcinoma. *The Journal of Biological Chemistry*, 287(39), 32800–32824.

[51] Sun, M., Zhou, W., Zhang, Y. Y., Wang, D. L., and Wu, X. L. (2013). CD44+ gastric cancer cells with stemness properties are chemoradioresistant and highly invasive. *Oncology Letters*, 5(6), 1793–1798.

[52] Takaishi, S., Okumura, T., Tu, S., Wang, S. S. W., Shibata, W., Vigneshwaran, R., *et al.* (2009). Identification of Gastric Cancer Stem Cells Using the Cell Surface Marker CD44. *Stem Cells*, 27(5), 1006–1020.

[53] Ohata, H., Ishiguro, T., Aihara, Y., Sato, A., Sakai, H., Sekine, S., *et al.* (2012). Induction of the Stem-like Cell Regulator CD44 by Rho Kinase Inhibition Contributes to the Maintenance of Colon Cancer-Initiating Cells. Cancer Research, 72(19), 5101–5110.

[54] Kimura, Y., Goi, T., Nakazawa, T., Hirono, Y., Katayama, K., Urano, T., and Yamagughi, A. (2013). CD44variant exon 9 plays an important role in colon cancer initiating cells. *Oncotarget*, 4(5), 785–791.

[55] Ponta, H., Sherman, L., and Herrlich, P. A. (2003). CD44: From adhesion molecules to signalling regulators. *Nature Reviews Molecular Cell Biology*, 4(1), 33–45.

[56] Prochazka, L., Tesarik, R., and Turanek, J. (2014). Regulation of alternative splicing of CD44 in cancer. *Cellular Signalling*, 26(10), 1–6.

[57] Wu, K., Xu, H., Tian, Y., Yuan, X., Wu, H., Liu, Q., and Pestell, R. (2015). The role of CD44 in epithelial-mesenchymal transition and cancer development. *OncoTargets and Therapy*, 3783–10.

[58] Nagano, O., and Saya, H. (2004). Mechanism and biological significance of CD44 cleavage. *Cancer Sci*, 95(12), 930–935.

[59] Okamoto, I., Tsuiki, H., Kenyon, L. C., Godwin, A. K., Emlet, D. R., Holgado-Madruga, M., *et al*. (2010). Proteolytic Cleavage of the CD44 Adhesion Molecule in Multiple Human Tumors. *The American Journal of Pathology*, 160(2), 441–447.

[60] Yoo, C. H., and Noh, S. H. (2003). The Serum Assay of Soluble CD44 Standard, CD44 Variant 5, and CD44 Variant 6 in Patients with Gastric Cancer. *Cancer Research and Treatment*, 1, 3–8.

[61] National Cancer Institute. (2019). *NCI Drug Dictionary*. [online] Available at: https://www.cancer.gov/publications/dictionaries/cancer-drug/def/anti-cd44-monoclonal-antibody-ro5429083 [Accessed 1 Sep. 2019].

[62] Chen, C., Zhao, S., Karnad, A., and Freeman, J. W. (2018). The biology and role of CD44 in cancer progression: therapeutic implications. *Journal of hematology & oncology*, 11(1), 64.

[63] Van Hal, N. L. W., Van Dongen, G. A. M. S., Rood-Knippels, E. M. C., Van Der Valk, P., Snow, G. B., and Brakenhoff, R. H. (1996). Monoclonal antibody U36, a suitable candidate for clinical immunotherapy of

squamous-cell carcinoma, recognizes a CD44 isoform. *International Journal of Cancer*, 68(4), 520–527.

[64] Sandstrom, K., Nestor, M., Ekberg, T., Engstrom, M., Anniko, M. and Lundqvist, H. (2008). Targeting CD44v6 expressed in head and neck squamous cell carcinoma: preclinical characterization of an 111In-labeled monoclonal antibody. *Tumour Biol.*, 29(3), 137–144.

[65] Tijink, B., Buter, J., de Bree, R., Giaccone, G., Lang, M., Staab, A., Leemans, C. and van Dongen, G. (2006). A Phase I Dose Escalation Study with Anti-CD44v6 Bivatuzumab Mertansine in Patients with Incurable Squamous Cell Carcinoma of the Head and Neck or Esophagus. *Clinical Cancer Research*, 12(20), 6064-6072.

[66] Heider, K., Dammrich, J., Skroch-Angel, P., Muller-Hermelink, H., Vollmers, H. P., Herrlich, P. A., and Ponta, H. (1993). Differential Expression of CD44 Splice Variants in Intestinal- and Diffuse-Type Human Gastric Carcinomas and Normal Gastric Mucosa. *Cancer Research*, 53, 4197–4203.

[67] Mayer, B., Jauch, K. W., Gunthert, U., Fidgor, C. G., Schildberg, F. W., Funke, I., and Johnson, J. P. (1993). De-novo expression of CD44 and survival in gastric cancer. *The Lancet*, 1019–1022.

[68] Ishimoto, T., Oshima, H., Oshima, M., Kai, K., Torii, R., Masuko, T., *et al*. (2010). CD44+ slow-cycling tumor cell expansion is triggered by cooperative actions of Wnt and prostaglandin E 2in gastric tumorigenesis. *Cancer Science*, 101(3), 673–678.

[69] Ishimoto, T., Nagano, O., Yae, T., Tamada, M., Motohara, T., Oshima, H., *et al*. (2011). CD44 Variant Regulates Redox Status in Cancer Cells by Stabilizing the xCT Subunit of System xc- and Thereby Promotes Tumor Growth. *Cancer Cell*, 19(3), 387–400.

[70] Lau, W. M., Teng, E., Chong, H. S., Lopez, K. A. P., Tay, A. Y. L., Salto-Tellez, M., *et al*. (2014). CD44v8-10 Is a Cancer-Specific Marker for Gastric Cancer Stem Cells. *Cancer Research*, 74(9), 2630–2641.

[71] Kiuchi, S., Ikeshita, S., Miyatake, Y., and Kasahara, M. (2015). Pancreatic cancer cells express CD44 variant 9 and multidrug resistance protein 1 during mitosis. *Experimental and Molecular Pathology*, 98(1), 41–46.

[72] Miyatake, Y., Sheehy, N., Ikeshita, S., Hall, W. W., and Kasahara, M. (2015). Anchorage-dependent multicellular aggregate formation induces CD44 high cancer stem cell-like ATL cells in an NF-κB- and vimentin-dependent manner. *Cancer Letters*, 357(1), 355–363.

[73] Maruyama, Y., Uehara, T., Daikuhara, S., Kobayashi, Y., Nakajima, T., Matsumoto, A., *et al.* (2015). Clinicopathological characterisation of duodenal adenocarcinoma with high CD44 variant 9 expression. *Pathology*, 47(7), 647–652.

[74] Sosulski, A., Horn, H., Zhang, L., Coletti, C., Vathipadiekal, V., Castro, C. M., *et al.* (2016). CD44 Splice Variant v8-10 as a Marker of Serous Ovarian Cancer Prognosis. *PLoS ONE*, 11(6), e0156595–18.

[75] Yamakawa, Y., Kusuhara, M., Terashima, M., Kinugasa, Y., Sugino, T., Abe, M., *et al.* (2017). CD44 variant 9 expression as a predictor for gastric cancer recurrence: immunohistochemical and metabolomic analysis of surgically resected tissues. *Biomedical Research Tokyo*, 38(1), 41–52.

[76] Choi, E. S., Kim, H., Kim, H. P., Choi, Y., and Goh, S. H. (2017). CD44v8-10 as a potential theranostic biomarker for targeting disseminated cancer cells in advanced gastric cancer. *Scientific Reports*, 7(1), E359–10.

[77] Jang, B. I., Li, Y., Graham, D. Y. and Cen, P. (2011). The Role of CD44 in the Pathogenesis, Diagnosis, and Therapy of Gastric Cancer. *Gut Liver*, 5, 397-405.

[78] Katoh, S., Goi, T., Naruse, T., Ueda, Y., Kurebayashi, H., Nakazawa, T., *et al.* (2015). Cancer Stem Cell Marker in Circulating Tumor Cells: Expression of CD44 Variant Exon 9 Is Strongly Correlated to Treatment Refractoriness, Recurrence and Prognosis of Human Colorectal Cancer. *Anticancer Research*, 35, 239–244.

[79] Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., *et al.* (2015). In *Essentials of Glycobiology. 3rd edn.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

[80] Varki, A. (2017). Biological roles of glycans. *Glycobiology*, 27, 3–49.

[81] Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (2008). In *Essentials of Glycobiology. 2nd edn.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

[82] Cummings, R. D. (2009). The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.*, 5, 1087–1104.

[83] Clausen, H. and Bennett, E. P. (1996). A family of UDP-GalNAc: polypeptide N acetylgalactosaminyl-transferases control the initiation of mucin-type O linked glycosylation. *Glycobiology*, 6, 635–646.

[84] Roseman, S. (2001). Reflections on glycobiology. *Journal of Biological Chemistry*, 276, 41527-41542.

[85] Pereira, M., Alves, I., Vicente, M., Campar, A., Silva, M., Padrão, N., Pinto, V., Fernandes, Â., Dias, A. and Pinho, S. (2018). Glycans as Key Checkpoints of T Cell Activity and Function. *Frontiers in Immunology*, 9(2754).

[86] Reily, C., Stewart, T. J., Renfrow, M. B. and Novak, J. (2019). Glycosylation in health and disease. *Nature Reviews Nephrology*, 15, 346-366.

[87] Brockhausen, I. and Stanley, P. (2017). In *Essentials of Glycobiology (ed. Varki, A. et al.)*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press., 113–123.

[88] Bennett, E., Mandel, U., Clausen, H., Gerken, T., Fritz, T. and Tabak, L. (2011). Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, 22(6), 736–756.

[89] Brockhausen, I., Schachter, H. and Stanley, P. (2009). O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, et al., editors. *Essentials of Glycobiology. 2nd edition.* Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, Chapter 9.

[90] Chia, J., Goh, G. and Bard, F. (2016). Short O-GalNAc glycans: regulation and role in tumor development and clinical perspectives. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1860(8), 1623-1639.

[91] Royle, L., Matthews, E., Corfield, A., Berry, M., Rudd, P. M., Dwek, R. A., and Carrington, S. D. (2008). Glycan structures of ocular surface mucins in man, rabbit and dog display species differences. *Glycoconjugate Journal*, 25(8), 763–773.

[92] Rossi, F., Corbel, S., Merzaban, J., Carlow, D., Gossens, K., Duenas, J., So, L., Yi, L. and Ziltener, H. (2005). Recruitment of adult thymic progenitors is regulated by P-selectin and its ligand PSGL-1. *Nature Immunology*, 6(6), 626-634.

[93] Xia, L., Ju, T., Westmuckett, A., An, G., Ivanciu, L., McDaniel, J. M., Lupu, F., Cummings, R. D. and McEver, R. P. (2004). Defective angiogenesis and fatal embryonic hemorrhage in mice lacking core 1-derived O-glycans. *Journal Cell Biol.*, 164, 451–459.

[94] Zachara, N., Akimoto, Y. and Hart, G. W. (2017). In *Essentials of Glycobiology (ed. Varki, A. et al.)*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. 239–251.

[95] Van der Laarse, S. A. M., Leney, A. C. and Heck, A. J. R. (2018). Crosstalk between phosphorylation and O-GlcNAcylation: friend or foe. *FEBS J.*, 285, 3152–3167.

[96] Ma, J. and Hart, G. W. (2014). *O*-GlcNAc profiling: from proteins to proteomes. *Clin. Proteomics*, 11, 8.

[97] Ohtsubo, K. and Marth, J. D. (2006). Glycosylation in cellular mechanisms of health and disease. *Cell*, 126, 855–867.

[98] Dennis, J. W., Granovsky, M., and Warren, C. E. (1999). Protein glycosylation in development and disease. *BioEssays*, 21, 412-421.

[99] Hansen, L., Lind-Thomsen, A., Joshi, H., Pedersen, N., Have, C., Kong, Y., Wang, S., Sparso, T., Grarup, N., Vester-Christensen, M., Schjoldager, K., Freeze, H., Hansen, T., Pedersen, O., Henrissat, B., Mandel, U., Clausen, H., Wandall, H. and Bennett, E. (2014). A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology*, 25(2), 211-224.

[100] Monticelli, M., Ferro, T., Jaeken, J., Dos Reis Ferreira, V. and Videira, P. A. (2016). Immunological aspects of congenital disorders of glycosylation (CDG): a review. *J. Inherit. Metab. Dis.*, 39, 765–780.

[101] Ng, B. G. and Freeze, H. H. (2018). Perspectives on glycosylation and its congenital disorders. *Trends Genet*, 34, 466–476.

[102] Crichton, R. R. and Charloteaux-Wauters, M. (1987). Iron transport and storage. *European Journal of Biochemistry / FEBS*, 164 (3), 485–506.

[103] Al Teneiji, A., Bruun, T. U., Sidky, S., Cordeiro, D., Cohn, R. D., Mendoza-Londono, R., ... & Mercimek-Mahmutoglu, S. *et al.* (2017). Phenotypic and genotypic spectrum of congenital disorders of glycosylation type I and type II. *Mol. Genet. Metab.*, 120, 235–242.

[104] Péanne, R., de Lonlay, P., Foulquier, F., Kornak, U., Lefeber, D., Morava, E., Pérez, B., Seta, N., Thiel, C., Van Schaftingen, E., Matthijs, G. and

Jaeken, J. (2018). Congenital disorders of glycosylation (CDG): Quo vadis?. *European Journal of Medical Genetics*, 61(11), 643-663.

[105] Hennet, T. and Cabalzar, J. (2015). Congenital disorders of glycosylation: a concise chart of glycocalyx dysfunction. *Trends Biochem. Sci.*, 40, 377–384.

[106] Schachter, H. (2001). Congenital disorders involving defective N-glycosylation of proteins. *Cellular and Molecular Life Sciences*, 58, 1085-1104.

[107] Mohlke, K. L., Purkayastha, A. A., Westrick, R. J., Smith, P. L., Petryniak, B., Lowe, J. B., and Ginsburg, D. (1999). Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosylatransferase. *Cell*, 96, 111-120.

[108] Freeze, H. H. and Westphal, V. (2001). Balancing N-linked glycosylation to avoid disease. *Biochimie*, 83, 791-799.

[109] Barresi, R. and Campbell, K. P. (2003). Dystroglycan: from biosynthesis to pathogenesis of human disease. *Journal of Cell Science*, 119, 199-207.

[110] Bucala, R. and Cerami, A. (1992). Advanced glycosylation: chemistry, biology and implications for diabetes and aging. *Advances in Pharmacology*, 23, 1-34.

[111] Biermann, M., Griffante, G., Podolska, M., Boeltz, S., Stürmer, J., Muñoz, L., Bilyy, R. and Herrmann, M. (2016). Sweet but dangerous – the role of immunoglobulin G glycosylation in autoimmunity and inflammation. *Lupus*, 25(8), 934-942.

[112] Matsubara, N., Imamura, A., Yonemizu, T., Akatsu, C., Yang, H., Ueki, A., Watanabe, N., Abdu-Allah, H., Numoto, N., Takematsu, H., Kitazume, S., Tedder, T., Marth, J., Ito, N., Ando, H., Ishida, H., Kiso, M. and Tsubata, T. (2018). CD22-Binding Synthetic Sialosides Regulate B Lymphocyte Proliferation Through CD22 Ligand-Dependent and Independent Pathways, and Enhance Antibody Production in Mice. *Frontiers in Immunology*, 9.

[113] Tomana, M., Schrohenloher, R. E., Reveille, J. D., Arnett, F. C. and Koopman, W. J. (1992). Abnormal galactosylation of serum IgG in patients with systemic lupus erythematosus and members of families with high frequency of autoimmune diseases. *Rheumatol. Int.*, 12, 191–194.

[114] Moore, J., Wu, X., Kulhavy, R., Tomana, M., Novak, J., Moldoveanu, Z., Brown, R., Goepfert, P. and Mestecky, J. (2005). Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS*, 19(4), 381-389.

[115] Van Zeben, D., Rook, G., Hazes, J., Zwinderman, A., Zhang, Y., Ghelani, S., Rademacher, T. and Breedveld, F. (1994). Early agalactosylation of IgG is associated with a more progressive disease course in patients with rheumatoid arthritis: results of a follow-up study. *Br. J. Rheumatology*, 33(1), 36-43.

[116] Berger, E. G. (1999). Tn-syndrome. *Biochim. Biophys. Acta*, 1455, 255–268.

[117] Ju, T. and Cummings, R. D. (2005). Protein glycosylation: chaperone mutation in Tn syndrome. *Nature*, 437.

[118] Ju, T. and Cummings, R. D. (2002). A unique molecular chaperone Cosmc required for activity of the mammalian core 1 β3-galactosyltransferase. *Proc. Natl Acad. Sci. USA*, 99, 16613–16618.

[119] Campos, D., Freitas, D., Gomes, J., Magalhaes, A., Steentoft, C., Gomes, C., Vester- Christensen, M. B., Ferreira, J. A., Afonso, L. P., Santos, L. L., Pinto de Sousa, J., Mandel, U., Clausen, H., Vakhrushev, S. Y. and Reis, C. A. (2015). Probing the O-glycoproteome of gastric cancer cell lines for biomarker discovery. *Mol Cell Proteomics.*, 14, 1616-1629

[120] Freitas, D., Campos, D., Gomes, J., Pinto, F., Macedo, J. A., Matos, R., Mereiter, S., Pinto, M. T., Polónia, A., Gartner, F., Magalhães, A. and Reis, C. A. (2019). O-glycans truncation modulates gastric cancer cell signaling and transcription leading to a more aggressive phenotype. *EBioMedicine*, 40, 349-362.

[121] Sun, X., Ju, T., and Cummings, R. D. (2018). Differential expression of Cosmc, T- synthase and mucins in Tn-positive colorectal cancers. *BMC cancer*, 18(1), 827.

[122] Neves, M., Azevedo, R., Lima, L., Oliveira, M. I., Peixoto, A., Ferreira, D., *et al.* (2018). Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: a novel biomarker and an analytical tool for precision oncology applications. *New BIOTECHNOLOGY*, 1–38.

[123] Reis, C. A., Osorio, H., Silva, L., Gomes, C. and David, L. (2010). Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol*, 63, 322-329.

[124] Burchell, J., Poulsom, R., Hanby, A., Whitehouse, C., Cooper, L., Clausen, H., Taylor-Papadimitriou, J., *et al.* (1999). An α2,3-sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas. *Glycobiology*, 9(12), 1307-1311.

[125] Sewell, R., Bäckström, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Taylor-Papadimitriou, J., *et al.* (2006). The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *Journal of Biological Chemistry*, 281(6), 3586-3594.

[126] Gill, D. J., Tham, K. M., Chia, J., Wang, S. C., Steentoft, C., Clausen, H., Bard, F. A. *et al.* (2013). Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. *Proceedings of the National Academy of Sciences*, 110(34), E3152-E3161.

[127] Marcos, N. T., Pinho, S., Grandela, C., Cruz, A., Samyn-Petit, B., Harduin-Lepers, A., Almeida, R., Silva, F., Morais, V., Costa, J., Kihlberg, J., Clausen, H. and Reis, C. A. (2004) Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated sialyl-Tn antigen. *Cancer research*, 64, 7050-7057.

[128] Axelsson, M. A., Karlsson, N. G., Steel, D. M., Ouwendijk, J., Nilsson, T., and Hansson, G. C. (2001). 462 Neutralization of pH in the Golgi apparatus causes redistribution of glycosyltransferases and 463 changes in the O-glycosylation of mucins. *Glycobiology*, 11(633-644), 464.

[129] Hassinen, A., Pujol, F. M., Kokkonen, N., Pieters, C., Kihlström, M., Korhonen, K., and Kellokumpu, S. (2011). Functional organization of Golgi N-and O-glycosylation pathways involves pH-dependent complex formation that is impaired in cancer cells. *Journal of Biological Chemistry*, 286(44), 38329-38340.

[130] Radhakrishnan, P., Dabelsteen, S., Madsen, F. B., Francavilla, C., Kopp, K. L., Steentoft, C., *et al.* (2014). Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences*, 111(39), E4066–E4075.

[131] Marcos, N. T., Bennett, E. P., Gomes, J., Magalhães, A., Gomes, C., David, L., *et al.* (2011). ST6GalNAc-I controls expression of sialyl-Tn antigen in gastroinstestinal tissues. *Frontiers in Bioscience*, 1443–1455.

[132] Kato, K., Takeuchi, H., Kanoh, A., Miyahara, N., Nemoto-Sasaki, Y., Morimoto-Tomita, M., *et al.* (2010). Loss of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase 3 and reduced O-glycosylation in colon carcinoma cells selected for hepatic metastasis. *Glycoconjugate Journal*, 27(2), 267–276.

[133] Orr, S. L., Le, D., Long, J. M., Sobieszczuk, P., Ma, B., Tian, H., *et al.* (2013). A phenotype survey of 36 mutant mouse strains with gene-targeted defects in glycosyltransferases or glycan-binding proteins. *Glycobiology*, 23(3), 363–380.

[134] Piller, V., Piller, F. and Fukuda, M. (1990). Biosynthesis of truncated O-glycans in the T cell line Jurkat. Localization of O-glycan initiation. *J Biol Chem.*, 265(16), 9264–9271.

[135] Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B. and Clausen, H. (2011). Mining the Oglycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods.*, 8, 977-982.

[136] Narimatsu, Y., Joshi, H. J., Yang, Z., Gomes, C., Chen, Y.-H., Lorenzetti, F. C., *et al.* (2018). A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology*, 28(5), 295–305.

[137] Hofmann, B. T., Schlüter, L., Lange, P., Mercanoglu, B., Ewald, F., Fölster, A., *et al.* (2015). COSMC knockdown mediated aberrant O-glycosylation promotes oncogenic properties in pancreatic cancer. *Molecular Cancer*, 14(109), 1–15.

[138] Gasbarri, A., Del Prete, F., Martegani, M. P., Natali, P. G., 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 Research*, 13, 325–337.

[139] Mereiter, S., Martins, Á. M., Gomes, C., Balmaña, M., Macedo, J. A., Polom, K., Roviello, F., Magalhães, A. and Reis, C. A. (2019). O-glycan truncation enhances cancer-related functions of CD44 in gastric cancer. *FEBS Letters*, 58, 3736–3723.

[140] Carvalho, A. S., Harduin-Lepers, A., Magalhães, A., Machado, E., Mendes, N., Costa, L. T., Matthiesen, R., Almeida, R., Costa, J. and Reis, C. A. (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.

[141] Denning, S.M. *et al.* (1995) CD44 and CD45R Cluster report. In *Leucocyte Typing V. White cell differentiation antigens*. Eds Schlossman, S.F. *et al.* Oxford University Press. Volume 2, AS10, 1713 – 1719.

[142] Fox, S. B., Fawcett, J., Jackson, D. G., Collins, I., Gatter, K. C., Haris, A. L., *et al.* (1994). Normal Human Tissues, in Addition to Some Tumors, Express Different CD44 Isoforms. *Cancer Research*, 54, 4539–4546.

[143] Colcher, D., Hand, P. H., Nuti, M. and Schlom, J. A. (1981). Spectrum of monoclonal antibodies reactive with human mammary tumor cells. *Proc. Natl. Acad. Sci. USA*, 78, 3199–3203.

[144] Gagarin, D., Yang, Z., Butler, J., Wimmer, M., Du, B., Cahan, P., and McCaffrey, T. A. (2005). Genomic profiling of acquired resistance to apoptosis in cells derived from human atherosclerotic lesions: potential role of STATs, cyclinD1, BAD, and Bcl-XL. *Journal of molecular and cellular cardiology*, 39(3), 453-465.

[145] Marcos, N.T., Cruz, A., Silva, F., Almeida, R., David, L., Mandel, U., Clausen, H., Von Mensdorff-Pouilly, S., and Reis, C. A. (2003). Polypeptide GalNAc-transferases, ST6GalNAc-transferase I, and ST3Gal-Transferase I expression in gastric carcinoma cell lines. *J Histochem Cytochem*, 51, 761-771.

[146] Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*, 72, 248–254.

[147] Agrawal, G. K., Jwa, N. S., Lebrun, M. H., Job, D. and Rakwal, R. (2010). Plant secretome: unlocking secrets of the secreted proteins. *Proteomics*, 10(4), 799-827.

[148] Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 226(5252), 1209-1211.

[149] Bartlett, J. M. S. and Stirling, D. (2003). A short history of the Polymerase Chain Reaction. PCR protocols. *Methods in Molecular Biology*, 226(2), 3-6.

[150] Carrier, J. L., Javadi, P., Bourrier, E., Camus, C., Ségal-Bendirdjian, E., and Karniguian, A. (2012). cFos mediates cAMP-dependent generation of ROS and rescue of maturation program in retinoid-resistant acute promyelocytic leukemia cell line NB4-LR1. *PloS one*, 7(11), e50408.

[151] Hsu, S. M., Raine, L., and Fanger, H. X. (1981). Use of avidin-biotin-peroxidase complex (ABC) in immunoperoxidase techniques: a comparison between ABC and unlabeled antibody (PAP) procedures. *Journal of Histochemistry & Cytochemistry*, 29(4), 577-580.

[152] Ju, T., Lanneau, G. S., Gautam, T., Wang, Y., Xia, B., Stowell, S. R., *et al.* (2008). Human Tumor Antigens Tn and Sialyl Tn Arise from Mutations in Cosmc. *Cancer Research*, 68(6), 1636–1646.

[153] Hirata, K., Suzuki, H., Imaeda, H., Matsuzaki, J., Tsugawa, H., Nagano, O., *et al.* (2013). CD44 variant 9 expression in primary early gastric cancer as a predictive marker for recurrence. *British Journal of Cancer*, 109(2), 379–386.

[154] de Oliveira, F. M. S., Mereiter, S., Lönn, P., Siart, B., Shen, Q., Heldin, J., *et al.* (2017). Detection of Post-Translational Modifications Using Solid-Phase Proximity Ligation Assay. *New BIOTECHNOLOGY*, 1–33.

[155] Ma, L., Dong, L., and Chang, P. (2018). CD44v6 engages in colorectal cancer progression. *Nature Publishing Group*, 1–13.

[156] Xia, C., Meng, Q., Liu, L. Z., Rojanasakul, Y., Wang, X. R., and Jiang, B. H. (2007). Reactive Oxygen Species Regulate Angiogenesis and Tumor Growth through Vascular Endothelial Growth Factor. *Cancer Research*, 67(22), 10823–10830.

[157] Hopirtean, C., and Nagy, V. (2018). Optimizing the Use of Anti-VEGF targeted therapies in patients with metastatic colorectal cancer: Review of Literature. *Clujul Medical*, 91(1), 12–17.

[158] Wu, Y., Zhang, C., Yu, Z., Zheng, G., Wang, S., Liu, Y., *et al.* (2015). CD44 family proteins in gastric cancer: a meta-analysis and narrative review. *Int J Clin Exp Med*, 8(3), 3595–3606.

[159] Martegani, M. P., Del Prete, F., Gasbarri, A., Natali, P. G., and Bartolazzi, A. (1998). Structural Variability of CD44v Molecules and Reliability of

Immunodetection of CD44 Isoforms Using mAbs Specific for CD44 Variant Exon Products. *The American Journal of Pathology*, 154(1), 291–300.

[160] Kodama, H., Murata, S., Ishida, M., Yamamoto, H., Yamaguchi, T., Kaida, S., *et al.* (2016). Prognostic impact of CD44-positive cancer stem-like cells at the invasive front of gastric cancer. *British journal of cancer*, 116(2), 186–194.

[161] Tsuji, K., Ojima, M., Otabe, K., Horie, M., Koga, H., Sekiya, I., and Muneta, T. (2017). Effects of Different Cell-Detaching Methods on the Viability and Cell Surface Antigen Expression of Synovial Mesenchymal Stem Cells. *Cell Transplantation*, 26(6), 1089–1102.

[162] Miletti-González, K. E., Murphy, K., Kumaran, M. N., Ravindranath, A. K., Wernyj, R. P., Kaur, S., *et al*. (2012). Identification of Function for CD44 intracytoplasmic Domain (CD44-ICD). *Journal of Biological Chemistry*, 287(23), 18995-19007.

[163] Silva-Filho, A. F., Sena, W. L. B., Lima, L. R. A., Carvalho, L. V. N., Pereira, M. C., Santos, L. G. S., *et al*. (2017). Glycobiology modification in intratumoral Hypoxia: The Breathless Side of Glycans Interaction. *Cellular physiology and Biochemistry*, 41(5), 1801-1829.

[164] Peixoto, A., Fernandes, E., Gaiteiro, C., Lima, L., Azevedo, R., Soares, J., *et al.* (2016). Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. *Oncotarget*, 7(39), 63138-63157.

[165] Guo, Y. J., Liu, G., Wang, X., Jin, D., Wu, M., Ma, J., and Sy, M. (1994). Potential Use of Soluble CD44 in Serum as Indicator of Tumor Burden and Metastasis in Patients with Gastric or Colon Cancer. *Cancer Research*, 54, 422–426.

[166] Teye, K., Numata, S., Ishii, N., Krol, R. P., Tsuchisaka, A., Hamada, T., *et al*. (2016). Isolation of All CD44 Transcripts in Human Epidermis and Regulation of Their Expression by Various Agents. *PLoS ONE*, 11(8), e0160952–21.