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

Stefan Mereiter[1,2,†], Alvaro M. Martins[1,2], Catarina Gomes[1,2], Meritxell Balmaña[1,2], Joana A. Macedo[1,2‡] , Karol Polom[3,4], Franco Roviello[4], Ana Magalhães[1,2] and Celso A. Reis[1,2,5,6]

1 I3S – Instituto de Investigac_~ao e Inovac_~ao em Sa_ude, Universidade do Porto, Portugal
2 IPATIMUP – Institute of Molecular Pathology and Immunology, University of Porto, Portugal
3 Department of Surgical Oncology, Medical University of Gdansk, Poland
4 General Surgery and Surgical Oncology Department, University of Siena, Italy
5 Faculty of Medicine, University of Porto, Portugal
6 Instituto de Ci^encias Biom_edicas Abel Salazar, University of Porto, Portugal
†IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, VBC – Vienna BioCenter Campus, Vienna, Austria
‡3B's Research Group, I3Bs – Research Institute on Biomaterials, Biodegradables and Biomimetics, University of Minho, Braga, Portugal
*Corresponding Author: celsor@ipatimup.pt

ABSTRACT

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

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

Abbreviations: CS, chondroitin sulfate; HA, hyaluronan; HS, heparan sulfate; IP, immunoprecipitation; M3, mock transfected MKN45 cells and control of ST3; M6, mock transfected MKN45 cells and control of ST6; PLA, proximity ligation assay; RTK, receptor tyrosine kinase; SC, MKN45 SimpleCells; ST3,

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

ST3GAL4 transfected MKN45 cells; ST6, ST6GALNAC1 transfected MKN45 cells; STn, sialyl Thomsen-nouvelle antigen; T, Thomsen-Friedenreich antigen; Tn, Thomsen-nouvelle antigen; WB, western blot; WT, wild type MKN45 cells.

INTRODUCTION

CD44 is an exceptionally versatile and diverse glycoprotein and consequently often described as a whole protein family encoded by a single gene [1, 2]. The molecular weight of CD44 largely varies among human cells, and ranges from 80 kDa to more than 250 kDa [3, 4]. This heterogeneity has been mainly attributed to three mechanisms, namely (1) alternative splicing at the RNA level [2, 5, 6], (2) extensive glycosylation of the protein backbone [7-10], and (3) the possible proteolytic cleavage through metalloproteases at the cell surface [4, 11, 12].

The process of alternative splicing leads to the expression of variable exon products in the extracellular, membrane proximal region. Due to the 9 variable exons (v), numbered from v2 to v10, and through a complex underlying regulatory machinery that enables countless combinations, hundreds of functionally distinct splice-variants of CD44 exist [2, 13]. A single cell can express several different CD44 isoforms simultaneously and many intra- and extracellular cues have been shown to dictate this splicing signature [5, 14-16], making CD44 a sentinel for cellular and environmental alterations.

Functionally, CD44 is a key player of cellular adhesion and signal transduction by multiple means: (1) it is the main receptor of hyaluronan (HA), one of the chief components of the extracellular matrix [17, 18]; (2) it may carry heparan sulfate (HS) and chondroitin sulfate (CS) chains which are docking sites for various growth factors and proteases [8, 19, 20]; (3) it is a co-receptor to cell surface receptor tyrosine kinases (RTKs), such as MET, RON and HER2 [16, 21-23]; and (4) it can be a potent ligand for selectins in the extravasation process [10]. However, these biological functions can vary significantly among different CD44 isoforms. For instance, the efficiency of CD44 to bind HA is controlled by the alternatively spliced variable (v) chains and its glycosylation status [7, 24]. Further, the single heparan sulfate chain of CD44 is linked to the variant chain v3 and the peptide sequence that acts as co-receptor of MET and RON is encoded in the v6 variable exon [21-23]. In gastric cancer, CD44 is commonly upregulated and its variant form CD44v6 is frequently de novo expressed [15, 25, 26]. CD44 has been implicated in cancer stem cell properties, and cancer cells that overexpress CD44 display an increased metastatic potential and are associated with poor survival of the patients [15, 27]. Particularly the expression of the v6 variant form of CD44 has been associated in gastric cancer with distant metastasis, poor prognosis of patients but also with susceptibility to chemotherapy [15, 28, 29] (REF). Functionally, CD44v6 has been shown to play a crucial role in the mediation of (RTK) activation, such as MET and RON, and thereby to contribute to oncogenic features [21, 23]. Thus, CD44 in general and CD44v6 in particular has a central role in regulating both cellular adhesion and receptor activation has been shown to contribute to various cancer hallmarks [16].

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

CD44 and its variable forms are extensively glycosylated and may carry N-glycans, O-glycans and glycosyaminoglycans [8, 9, 30]. In fact, a major part of the molecular mass of CD44 is derived from these numerous glycan modifications [8, 9, 31]. Alterations in the glycosylation machinery is a key event in carcinogenesis and cancer progression [32]. In gastric cancer, the abnormal biosynthesis of O-glycans is a frequent event leading to the expression of short, truncated glycan-epitopes such as T, Tn or sialyl Tn (STn) [33-37]. Recently, we demonstrated that CD44v6 carries the cancer-associated glycan epitope STn in gastric cancer [38] and that gastric cancer patients present elevated serum levels of CD44v6 with STn when compared to healthy individuals [39]. Yet, despite the indications that glycosylation may have a strong modulatory effect on CD44, potentially driving malignancy, the functional impact of O-glycan truncation on CD44 has not been addressed in cancer.

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

MATERIALS AND METHODS

Cell lines

The gastric carcinoma cell line MKN45 was obtained from the Japanese Cancer Research Bank (Tsukuba, Japan) and was stably transfected with the full length human ST3GAL4 gene, the full length human ST6GALNAC1 gene or the corresponding empty vector pcDNA3.1 (Mock) as previously described [40, 41]. The MKN45 SimpleCells (SC) were obtained by targeting the C1GALTC1 (COSMC) gene by zinc-finger nuclease as previously described [38, 42]. All cells were grown in monolayer in uncoated cell culture flasks. Cells were maintained at 37°C in an atmosphere of 5% CO_2, in RPMI 1640 GlutaMAX, HEPES medium supplemented with 10% FBS. The media of cells transfected with expression vectors were supplemented by 0.5 mg/ml G418 (all from Invitrogen, Waltham, MA). 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.

Primary antibodies

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

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Immunoprecipitation and western blotting

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 (BioRad, Hercules, CA). For immunoprecipitation, protein G fast flow sepharose beads (GE healthcare, Little Chalfont, UK) were pre-incubated with the CD44 antibody and 500 µg of protein lysate were applied. Western blotting was performed as previously described [45], using the Mini-PROTEAN® Tetra Cell System (BioRad, Hercules, CA) and polyvinylidene difluoride membranes (GE Healthcare, Chicago, IL). Densitometry was performed with Image Lab (BioRad, Hercules, CA), values were normalized to tubulin and represented as relative values compared to the control.

Flow cytometry

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

Transcriptomics

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

qRT-PCR

Total RNA extracts from cell lysates were isolated with TRI Reagent (Sigma-Aldrich, St. Louis, MO) and converted into cDNA using the SuperScript® IV Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol. Following primers were used: CD44 for5′-CCAATGCCTTTGATGGACC-3′, rev5′-TCTGTCTGTGCTGTCGGTGAT-3′; CD44v3 for5′-CGTCTTCAAATACCATCCAGCA-3′, rev5′ ATCTTCATCATCAATGCCTGA 3′ and CD44v6 for5′-GGCAACTCCTAGTAGTACAACG-3′, rev5′-GTCTTCTCTGGGTGTTTGGC-3′. Expression of 18S (for5'-CGCCGCTAGAGGTGAAATTC-3'; rev5'-CATTCTTGGCAAATGCTTTCG-3') was used for the normalization.

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

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Enzymatic glycosaminoglycan digestion

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

Secretome analysis

For the secretome analysis, cells were grown in medium without FBS supplementation. Conditioned media were collected after 48h and were enriched in protein content using 10 kDa Amicon Purification System (Merck Millipore, Burlington, MA).In silico analysis of O-glycosylation sites

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

Immunofluorescence and in situ proximity ligation assay

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

For the combined PLA/IF, all primary antibodies were previously conjugated. The antibody B72.3 was biotinylated using Pierce™ Antibody Biotinylation Kit for IP (ThermoFisher, Waltham, MA). Antibodies for CD44v6 and RON were conjugated with positive or negative strand oligonucleotides for PLA by Duolink® In Situ Probemaker (Sigma Aldrich, St.Louis, MO). To evaluate the signal in gastric carcinoma, formalin-fixed paraffin embedded gastric carcinoma tissue slides were provided by the department of surgical oncology of the University of Siena (Italy). All procedures were performed after patients' written informed consent and approved by the local ethical committee. Three gastric carcinomas were selected that fulfilled following criteria: 1. The carcinoma had both, regions that were STn positive and regions that were STn negative. The STn positivity has previously been evaluated in this cohort [35] 2. Adjacent to the carcinoma there were normal gastric mucosa and intestinal metaplasia that would act as endogenous biological negative and positive controls, respectively. First slides were dewaxed and rehydrated, and then blocked according to Duolink® PLA Technology kit. The slides were incubated overnight with the conjugated primary antibodies at 4°C. Ligation and amplification of the PLA signal was performed according to the Duolink® PLA Technology kit. Finally, slides were incubated with Streptavidin-FITC and DAPI, and mounted using vectashield mounting medium (Vector Laboratories, Burlingame, CA). All PLA, IF and combined PLA/IF samples were examined under a Zeiss Imager.Z1 Axio fluorescence microscope (Zeiss, Welwyn Garden City, UK). Images were acquired using a Zeiss Axio cam MRm and the AxioVision Rel.

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

4.8 software. For visualization purposes, PLA images were enhanced using ImageJ [48] as follows: The red channel (PLA signal) was duplicated with one image being subjected to gaussian blur and then the difference between the two images was calculated (image calculator function). Contrast and brightness was adjusted and PLA spot size was amplified twice by the "maximum" filter. The PLA signal was quantified using Duolink® ImageTool (Sigma Aldrich, St.Louis, MO) and ImageJ [48].

RESULTS

Truncation of O-glycans affects the molecular weight of CD44

To investigate the effect of O-glycan truncation on CD44, we applied genetically engineered cell line models derived from the gastric carcinoma cell line MKN45. The parental MKN45 cell line was selected as parental cell line to develop these models, since it expresses high levels of various CD44 splice variants and it is addicted to the activation of the RTKs RON and MET, to whom CD44 is a co-receptor [15, 49]. We induced the truncation of O-glycans in MKN45 through three different mechanisms (Fig 1a): (1) through the overexpression of the α2,3-sialyltransferase ST3GAL4 leading to the early termination of O-glycans (henceforth referred to as ST3) [40, 45, 50]; (2) through the overexpression of the α2,6-sialyltransferase ST6GALNAC1 leading to a strong increase in the short, aberrant sialyl Tn (STn) glycan (henceforth referred to as ST6) [41]; (3) through the knockout of the chaperon COSMC, which leads to the abrogation of the initial O-glycan elongation step (henceforth referred to as SC for SimpleCell) [38].

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

O-glycan truncation does not alter total CD44 levels or alternatively spliced variants

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Alternative splicing of CD44 can be induced through inter and extracellular stimuli and constitutes a major mechanism of CD44 modulation, strongly affecting CD44's molecular weight [16]. We analysed whether the induced aberrant glycosylation of our cell models triggered changes in expression and splicing of CD44 or other genes relevant in this process. For this purpose, we performed transcriptomic analysis of the cell models and revealed that CD44 expression levels were not significantly altered (Fig 2a). In addition, the relative expression levels of CD44v3 and of CD44v6 to total CD44, assessed by qRT-PCR, revealed that no significant alterations in alternative splicing of these isoforms were induced (Fig 2b). These results are supported by the fact that the expression of genes described to be involved in the alternative splicing of CD44 [13] were not notably altered (Fib 2c).

In addition, the subcellular localization of CD44 and CD44v6 seemed to be unchanged by the truncation of O-glycans (Fig 2d, f). The total protein levels of CD44 and CD44v6 measured by flow cytometry were comparable between the glyco-engineered cell lines and their respective controls (Fig 2e, g). Altogether, these results demonstrate that in the applied models the alterations of O-glycans did not significantly affect the CD44 mRNA and protein expression.

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

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

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

together, these findings show that CD44 is a major carrier of truncated O-glycans in gastric cancer cells, which underpins the observed mass shift in the glyco-engineered cell models.

Truncation of O-glycans increases the capacity of CD44 to bind hyaluronan and affects the shedding of CD44

Following the identification of CD44 as carrier of truncated O-glycans we assessed the capacity of our models to bind to HA. CD44 is known to be the major binding protein of HA in human cells [16]. Other receptors of HA are not appreciably expressed by our cell models (Fig 4a). The HA binding assay revealed that the truncation of O-glycans led to a striking increase of HA binding in our cell models (Fig 4b). Although the induced truncation of O-glycans was achieved through different molecular alterations, the increased binding was consistent among all models. This underlines the significant effect of O-glycan truncation on the HA binding capacity of CD44. In addition, it has been reported before that CD44 can be shed into the extracellular milieu by the cleavage of the juxtamembrane domain [11, 12, 52] and that changes in glycosylation can impact this process [53]. To assess whether the truncation of O-glycans affect the secretion of CD44 we have analysed the secretome of the glyco-engineered cell models (Fig 4c). According to expectation a mass shift of approximately 10 to 20 kDa was observed in all models, which corresponds to a loss of cytosolic, transmembrane and a small part of the extracellular domain. Interestingly, CD44 bearing shorter O-glycans appeared to be more abundantly shed than that of the control cell lines.

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

CD44v6 is a co-receptor of the RON receptor tyrosine kinase and mediates its activation [21]. The oncogenic hyperactivation of RON has been described as a common event in gastric cancer and it is associated with tumour progression, angiogenesis and chemoresistance [54]. In order to test if the altered glycosylation of CD44 influences the v6 association with RON, we performed a series of in situ PLAs. Remarkably, the colocalization of CD44v6 and RON was significantly more frequent in cells with shorter O-glycans than in the control cell lines (Fig 5). The significantly elevated colocalization events were concomitant with the increased activation of the RON receptor in these cells, as depicted by pRON immunolabelling (Fig 5a-c). This indicates an involvement of the truncation of O-glycans in the activation of RON, through the facilitated interaction of CD44v6 and RON. In order to validate this association in patients' clinical samples, we developed a combinatory approach with RON/CD44v6 PLA and FITC-labelling of STn in tumour tissue sections. This innovative approach enabled the in situ evaluation of the colocalization of RON, CD44v6 and the truncated O-glycan epitope STn (Fig 6). An association between tumour areas that were STn positive and increased amounts of RON/CD44v6 colocalization events was observed in all the gastric carcinoma cases evaluated (Fig 6). Interestingly, extensive colocalization of STn and RON/CD44v6 was also evident in the areas of intestinal metaplasia (a pre-neoplastic lesion of the gastric mucosa [55]), albeit to a lower degree than observed in the carcinoma (Fig 6b). No significant histological differences between STn positive and negative areas were evident (Supplementary Fig 2

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

DISCUSSION

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

In this manuscript, we describe for the first time the impact of O-glycan truncation on molecular features of CD44 and highlight new functional implications of these glycans. We have performed a comparative analysis of three differently glyco-engineered MKN45 models: ST3GAL4 overexpressing (ST3), ST6GALNAC1 overexpressing (ST6) and COSMC knock-out (SC from SimpleCell), and compared them to their corresponding control cell lines (mock and WT). MKN45 WT and mock expresses primarily core 2 O-glycans with 2 or more LacNAc repeats [45, 50]. This glycosylation profile is reflective of some gastric carcinomas [57]. Although different genes were targeted in our glyco-engineered models, the early termination of O-glycans was a common feature of the three cell line models.

ST3GAL4 is a sialyltransferase that affects both N- and O-glycans [45, 50]. The upregulation of ST3GAL4 in MKN45 leads to an increase of α2,3 sialylation and a decrease of α2,6 sialylation on N-glycans and significantly decreases the amount of bisected structures [45, 50]. On O-glycans, ST3GAL4 leads to the earlier termination due to the increase of one specific structure: the di-sialylated core 2 structure [45, 50]. Both the ST6GALNAC1 overexpression and the COSMC deletion induce alteration on O-glycans and lead to an upregulation of STn by two differential mechanisms [38, 41]. ST6GALNAC1 is a sialyltransferase that adds sialic acid onto Tn to form STn and thereby, leads to the early termination of the O-glycan chains [41]. COSMC on the other hand, is a dedicated chaperone of the core 1 galactosyltransferase (C1GALT1) [58]. Without its chaperone, C1GALT1 is dysfunctional and fails to elongate the initial N-acetylgalactosamine (which constitutes the Tn antigen) to core 1. In SC, the lack of COSMC causes therefore the accumulation of Tn and STn antigens [38, 59]. The fact that STn expression was initiated in MKN45 both, after the overexpression of ST6GALNAC1 but also after the knockout of COSMC (with endogenous levels of ST6GALNAC1) highlights that the STn expression arises due to the relative activity of competing glycosyltransferases. In addition, other mechanisms besides the relative expression of glycosyltransferases have been described to induce the truncation of O-glycans in cancer [32, 60].

The four mechanisms that could underpin the molecular weight changes of CD44, which span over 50 kDa, are (1) alternative splicing, (2) proteolytic cleavage, (3) altered O-glycosylation and (4) changes in the modification with glycosaminoglycan chains. MKN45 expresses mostly CD44v8-10 splicing isoforms, known as CD44E, named after its common expression in epithelial cells, but also

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

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

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

We also ruled out that alterations in glycosaminoglycans accounted for the light version of CD44. Our results point unequivocally towards the truncation of O-glycosylation as the underlying mechanism for the expression of the light CD44 isoform. This is supported by the mass shift observed in the likewise highly O-glycosylated proteoglycan Syndecan 1 (Fig 3B). Of note, we made the observation that western blot band intensities were significantly increased when lighter glycoforms of CD44 or Syndecan 1 were expressed. This might stem from increased transfer efficiencies of smaller proteins and from the more focused nature of the band. The CD44 amounts detected by western blot analysis were not fully overlapping with those of the FACS analysis as well as with the transcription levels. The latter two methods are quantitative and revealed no significant changes. These results support that protein glycosylation alterations could interfere with protein quantifications assessed by western blot.

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

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

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

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

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

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

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

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

REFERENCES

1 Screaton, G. R., Bell, M. V., Jackson, D. G., Cornelis, F. B., Gerth, U. and Bell, J. I. (1992) Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. Proc Natl Acad Sci USA. 89, 12160-12164

2 Marhaba, R. and Zoller, M. (2004) CD44 in cancer progression: adhesion, migration and growth regulation. J Mol Histol. 35, 211-231

3 Bourguignon, L. Y., Gunja-Smith, Z., Iida, N., Zhu, H. B., Young, L. J., Muller, W. J. and Cardiff, R. D. (1998) CD44v(3,8-10) is involved in cytoskeleton-mediated tumor cell migration and matrix metalloproteinase (MMP-9) association in metastatic breast cancer cells. Journal of cellular physiology. 176, 206-215

4 Murakami, D., Okamoto, I., Nagano, O., Kawano, Y., Tomita, T., Iwatsubo, T., De Strooper, B., Yumoto, E. and Saya, H. (2003) Presenilin-dependent gamma-secretase activity mediates the intramembranous cleavage of CD44. Oncogene. 22, 1511-1516

5 Branco da Cunha, C., Klumpers, D. D., Koshy, S. T., Weaver, J. C., Chaudhuri, O., Seruca, R., Carneiro, F., Granja, P. L. and Mooney, D. J. (2016) CD44 alternative splicing in gastric cancer cells is regulated by culture dimensionality and matrix stiffness. Biomaterials. 98, 152-162

6 Azevedo, R., Gaiteiro, C., Peixoto, A., Relvas-Santos, M., Lima, L., Santos, L. L. and Ferreira, J. A. (2018) CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. Clinical proteomics. 15, 22

7 Bennett, K. L., Modrell, B., Greenfield, B., Bartolazzi, A., Stamenkovic, I., Peach, R., Jackson, D. G., Spring, F. and Aruffo, A. (1995) Regulation of CD44 binding to hyaluronan by glycosylation of variably spliced exons. The Journal of cell biology. 131, 1623-1633

8 Jackson, D. G., Bell, J. I., Dickinson, R., Timans, J., Shields, J. and Whittle, N. (1995) Proteoglycan forms of the lymphocyte homing receptor CD44 are alternatively spliced variants containing the v3 exon. The Journal of cell biology. 128, 673-685

9 English, N. M., Lesley, J. F. and Hyman, R. (1998) Site-specific de-N-glycosylation of CD44 can activate hyaluronan binding, and CD44 activation states show distinct threshold densities for hyaluronan binding. Cancer research. 58, 3736-3742

10 Dimitroff, C. J., Lee, J. Y., Fuhlbrigge, R. C. and Sackstein, R. (2000) A distinct glycoform of CD44 is an L-selectin ligand on human hematopoietic cells. Proc Natl Acad Sci U S A. 97, 13841-13846

11 Chetty, C., Vanamala, S. K., Gondi, C. S., Dinh, D. H., Gujrati, M. and Rao, J. S. (2012) MMP-9 induces CD44 cleavage and CD44 mediated cell migration in glioblastoma xenograft cells. Cellular signalling. 24, 549-559

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

12 Anderegg, U., Eichenberg, T., Parthaune, T., Haiduk, C., Saalbach, A., Milkova, L., Ludwig, A., Grosche, J., Averbeck, M., Gebhardt, C., Voelcker, V., Sleeman, J. P. and Simon, J. C. (2009) ADAM10 is the constitutive functional sheddase of CD44 in human melanoma cells. The Journal of investigative dermatology. 129, 1471-1482

13 Prochazka, L., Tesarik, R. and Turanek, J. (2014) Regulation of alternative splicing of CD44 in cancer. Cellular signalling. 26, 2234-2239

14 Banky, B., Raso-Barnett, L., Barbai, T., Timar, J., Becsagh, P. and Raso, E. (2012) Characteristics of CD44 alternative splice pattern in the course of human colorectal adenocarcinoma progression. Molecular cancer. 11, 83

15 da Cunha, C. B., Oliveira, C., Wen, X., Gomes, B., Sousa, S., Suriano, G., Grellier, M., Huntsman, D. G., Carneiro, F., Granja, P. L. and Seruca, R. (2010) De novo expression of CD44 variants in sporadic and hereditary gastric cancer. Lab Invest. 90, 1604-1614

16 Orian-Rousseau, V. and Sleeman, J. (2014) CD44 is a multidomain signaling platform that integrates extracellular matrix cues with growth factor and cytokine signals. Advances in cancer research. 123, 231-254

17 Misra, S., Hascall, V. C., Markwald, R. R. and Ghatak, S. (2015) Interactions between Hyaluronan and Its Receptors (CD44, RHAMM) Regulate the Activities of Inflammation and Cancer. Frontiers in immunology. 6, 201

18 Richter, U., Wicklein, D., Geleff, S. and Schumacher, U. (2012) The interaction between CD44 on tumour cells and hyaluronan under physiologic flow conditions: implications for metastasis formation. Histochemistry and cell biology. 137, 687-695

19 Jones, M., Tussey, L., Athanasou, N. and Jackson, D. G. (2000) Heparan sulfate proteoglycan isoforms of the CD44 hyaluronan receptor induced in human inflammatory macrophages can function as paracrine regulators of fibroblast growth factor action. The Journal of biological chemistry. 275, 7964-7974

20 Yu, W. H., Woessner, J. F., Jr., McNeish, J. D. and Stamenkovic, I. (2002) CD44 anchors the assembly of matrilysin/MMP-7 with heparin-binding epidermal growth factor precursor and ErbB4 and regulates female reproductive organ remodeling. Genes & development. 16, 307-323

21 Matzke, A., Herrlich, P., Ponta, H. and Orian-Rousseau, V. (2005) A five-amino-acid peptide blocks Met- and Ron-dependent cell migration. Cancer research. 65, 6105-6110

22 Sherman, L. S., Rizvi, T. A., Karyala, S. and Ratner, N. (2000) CD44 enhances neuregulin signaling by Schwann cells. The Journal of cell biology. 150, 1071-1084

23 Matzke-Ogi, A., Jannasch, K., Shatirishvili, M., Fuchs, B., Chiblak, S., Morton, J., Tawk, B., Lindner, T., Sansom, O., Alves, F., Warth, A., Schwager, C., Mier, W., Kleeff, J., Ponta, H., Abdollahi, A. and Orian-Rousseau, V. (2016) Inhibition of Tumor Growth and Metastasis in Pancreatic Cancer Models by Interference With CD44v6 Signaling. Gastroenterology. 150, 513-525 e510

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

24 Stamenkovic, I., Aruffo, A., Amiot, M. and Seed, B. (1991) The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. The EMBO journal. 10, 343-348

25 Cancer Genome Atlas Research, N. (2014) Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 513, 202-209

26 Mayer, B., Jauch, K. W., Gunthert, U., Figdor, C. G., Schildberg, F. W., Funke, I. and Johnson, J. P. (1993) De-novo expression of CD44 and survival in gastric cancer. Lancet. 342, 1019-1022

27 Takaishi, S., Okumura, T., Tu, S., Wang, S. S., Shibata, W., Vigneshwaran, R., Gordon, S. A., Shimada, Y. and Wang, T. C. (2009) Identification of gastric cancer stem cells using the cell surface marker CD44. Stem Cells. 27, 1006-1020

28 Yamaguchi, A., Goi, T., Yu, J., Hirono, Y., Ishida, M., Iida, A., Kimura, T., Takeuchi, K., Katayama, K. and Hirose, K. (2002) Expression of CD44v6 in advanced gastric cancer and its relationship to hematogenous metastasis and long-term prognosis. Journal of surgical oncology. 79, 230-235

29 Pereira, C., Ferreira, D., Lemos, C., Martins, D., Mendes, N., Almeida, D., Granja, P., Carneiro, F., Almeida, R., Almeida, G. M. and Oliveira, C. (2018) CD44v6 expression is a novel predictive marker of therapy response and poor prognosis in gastric cancer patients. bioRxiv, 468934

30 Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B. and Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. The EMBO journal. 32, 1478-1488

31 Dasgupta, A., Takahashi, K., Cutler, M. and Tanabe, K. K. (1996) O-linked glycosylation modifies CD44 adhesion to hyaluronate in colon carcinoma cells. Biochemical and biophysical research communications. 227, 110-117

32 Pinho, S. S. and Reis, C. A. (2015) Glycosylation in cancer: mechanisms and clinical implications. Nature reviews. Cancer. 15, 540-555

33 David, L., Nesland, J. M., Clausen, H., Carneiro, F. and Sobrinho-Simoes, M. (1992) Simple mucin-type carbohydrate antigens (Tn, sialosyl-Tn and T) in gastric mucosa, carcinomas and metastases. APMIS. Supplementum. 27, 162-172

34 Mereiter, S., Balmana, M., Gomes, J., Magalhaes, A. and Reis, C. A. (2016) Glycomic Approaches for the Discovery of Targets in Gastrointestinal Cancer. Frontiers in oncology. 6, 55

35 Mereiter, S., Polom, K., Williams, C., Polonia, A., Guergova-Kuras, M., Karlsson, N. G., Roviello, F., Magalhaes, A. and Reis, C. A. (2018) The Thomsen-Friedenreich Antigen: A Highly Sensitive and Specific Predictor of Microsatellite Instability in Gastric Cancer. Journal of clinical medicine. 7

36 Persson, N., Stuhr-Hansen, N., Risinger, C., Mereiter, S., Polonia, A., Polom, K., Kovacs, A., Roviello, F., Reis, C. A., Welinder, C., Danielsson, L., Jansson, B. and Blixt, O. (2017) Epitope mapping of a new anti-Tn antibody detecting gastric cancer cells. Glycobiology. 27, 635-645

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

37 Marcos, N. T., Bennett, E. P., Gomes, J., Magalhaes, A., Gomes, C., David, L., Dar, I., Jeanneau, C., DeFrees, S., Krustrup, D., Vogel, L. K., Kure, E. H., Burchell, J., Taylor-Papadimitriou, J., Clausen, H., Mandel, U. and Reis, C. A. (2011) ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. Frontiers in bioscience. 3, 1443-1455

38 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

39 Oliveira, F. M. S., Mereiter, S., Lonn, P., Siart, B., Shen, Q., Heldin, J., Raykova, D., Karlsson, N. G., Polom, K., Roviello, F., Reis, C. A. and Kamali-Moghaddam, M. (2018) Detection of post-translational modifications using solid-phase proximity ligation assay. New biotechnology. 45, 51-59

40 Carvalho, A. S., Harduin-Lepers, A., Magalhaes, 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

41 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

42 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 O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. Nat Methods. 8, 977-982

43 Kjeldsen, T., Clausen, H., Hirohashi, S., Ogawa, T., Iijima, H. and Hakomori, S. (1988) Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2---6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope. Cancer research. 48, 2214-2220

44 Thor, A., Ohuchi, N., Szpak, C. A., Johnston, W. W. and Schlom, J. (1986) Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. Cancer research. 46, 3118-3124

45 Mereiter, S., Magalhaes, A., Adamczyk, B., Jin, C., Almeida, A., Drici, L., Ibanez-Vea, M., Gomes, C., Ferreira, J. A., Afonso, L. P., Santos, L. L., Larsen, M. R., Kolarich, D., Karlsson, N. G. and Reis, C. A. (2016) Glycomic analysis of gastric carcinoma cells discloses glycans as modulators of RON receptor tyrosine kinase activation in cancer. Biochimica et biophysica acta. 1860, 1795-1808

46 Livak, K. J. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25, 402-408

47 Ling, D., Pike, C. J. and Salvaterra, P. M. (2012) Deconvolution of the confounding variations for reverse transcription quantitative real-time polymerase chain reaction by separate analysis of biological replicate data. Analytical biochemistry. 427, 21-25

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

48 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods. 9, 676-682

49 Benvenuti, S., Lazzari, L., Arnesano, A., Li Chiavi, G., Gentile, A. and Comoglio, P. M. (2011) Ron kinase transphosphorylation sustains MET oncogene addiction. Cancer research. 71, 1945-1955

50 Mereiter, S., Magalhaes, A., Adamczyk, B., Jin, C., Almeida, A., Drici, L., Ibanez-Vea, M., Larsen, M. R., Kolarich, D., Karlsson, N. G. and Reis, C. A. (2016) Glycomic and sialoproteomic data of gastric carcinoma cells overexpressing ST3GAL4. Data in brief. 7, 814-833

51 Couchman, J. R., Gopal, S., Lim, H. C., Norgaard, S. and Multhaupt, H. A. (2015) Fell-Muir Lecture: Syndecans: from peripheral coreceptors to mainstream regulators of cell behaviour. International journal of experimental pathology. 96, 1-10

52 Nagano, O. and Saya, H. (2004) Mechanism and biological significance of CD44 cleavage. Cancer science. 95, 930-935

53 Gasbarri, A., Del Prete, F., Girnita, L., 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

54 Catenacci, D. V., Cervantes, G., Yala, S., Nelson, E. A., El-Hashani, E., Kanteti, R., El Dinali, M., Hasina, R., Bragelmann, J., Seiwert, T., Sanicola, M., Henderson, L., Grushko, T. A., Olopade, O., Karrison, T., Bang, Y. J., Kim, W. H., Tretiakova, M., Vokes, E., Frank, D. A., Kindler, H. L., Huet, H. and Salgia, R. (2011) RON (MST1R) is a novel prognostic marker and therapeutic target for gastroesophageal adenocarcinoma. Cancer biology & therapy. 12, 9-46

55 Jencks, D. S., Adam, J. D., Borum, M. L., Koh, J. M., Stephen, S. and Doman, D. B. (2018) Overview of Current Concepts in Gastric Intestinal Metaplasia and Gastric Cancer. Gastroenterology & hepatology. 14, 92-101

56 Thapa, R. and Wilson, G. D. (2016) The Importance of CD44 as a Stem Cell Biomarker and Therapeutic Target in Cancer. Stem cells international. 2016, 2087204

57 Jin, C., Kenny, D. T., Skoog, E. C., Padra, M., Adamczyk, B., Vitizeva, V., Thorell, A., Venkatakrishnan, V., Linden, S. K. and Karlsson, N. G. (2017) Structural Diversity of Human Gastric Mucin Glycans. Mol Cell Proteomics. 16, 743-758

58 Wang, Y., Ju, T., Ding, X., Xia, B., Wang, W., Xia, L., He, M. and Cummings, R. D. (2010) Cosmc is an essential chaperone for correct protein O-glycosylation. Proc Natl Acad Sci U S A. 107, 9228-9233

59 Ju, T., Lanneau, G. S., Gautam, T., Wang, Y., Xia, B., Stowell, S. R., Willard, M. T., Wang, W., Xia, J. Y., Zuna, R. E., Laszik, Z., Benbrook, D. M., Hanigan, M. H. and Cummings, R. D. (2008) Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. Cancer research. 68, 1636-1646

60 Gomes, J., Mereiter, S., Magalhaes, A. and Reis, C. A. (2017) Early GalNAc O-Glycosylation: Pushing the Tumor Boundaries. Cancer cell. 32, 544-545

61 Shi, M., Dennis, K., Peschon, J. J., Chandrasekaran, R. and Mikecz, K. (2001) Antibody-induced shedding of CD44 from adherent cells is linked to the assembly of the cytoskeleton. Journal of immunology. 167, 123-131

62 Bartolazzi, A., Nocks, A., Aruffo, A., Spring, F. and Stamenkovic, I. (1996) Glycosylation of CD44 is implicated in CD44-mediated cell adhesion to hyaluronan. The Journal of cell biology. 132, 1199-1208

63 Yao, H. P., Zhou, Y. Q., Zhang, R. and Wang, M. H. (2013) MSP-RON signalling in cancer: pathogenesis and therapeutic potential. Nature reviews. Cancer. 13, 466-481

64 Hurt-Camejo, E., Rosengren, B., Sartipy, P., Elfsberg, K., Camejo, G. and Svensson, L. (1999) CD44, a cell surface chondroitin sulfate proteoglycan, mediates binding of interferon-gamma and some of its biological effects on human vascular smooth muscle cells. The Journal of biological chemistry. 274, 18957-18964

65 UniProt Consortium, T. (2018) UniProt: the universal protein knowledgebase. Nucleic acids research. 46, 2699

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Figure 1

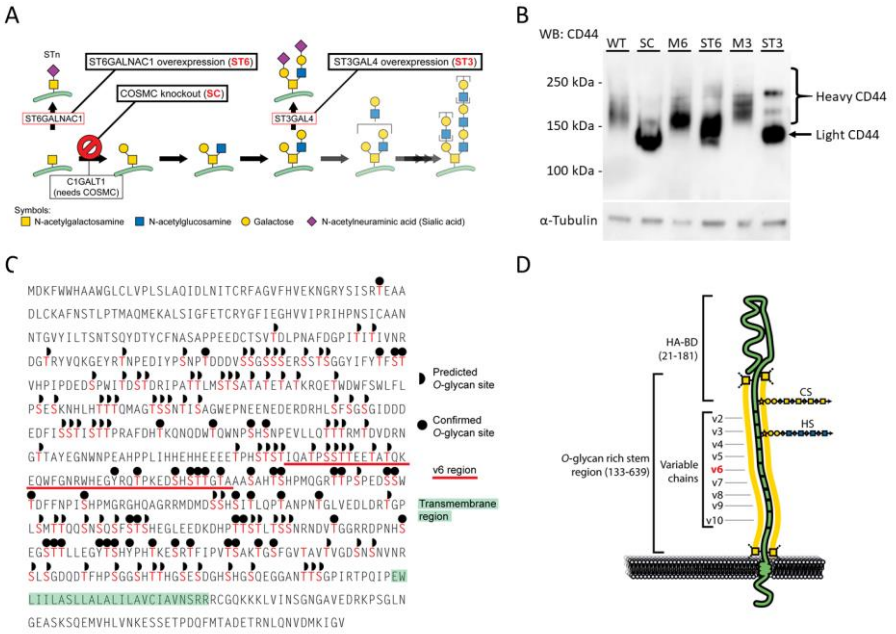


Fig. 1. The effect of O-glycan truncation on CD44.

(A) Schematic representation of the mucin-type O-glycan biosynthesis pathway. The mechanisms by which O-glycans have been truncated in the applied MKN45 gastric carcinoma cell line models (SC, ST6 and ST3) are indicated.

(B) Western blot analysis of CD44 in cell lysates of glyco-engineered MKN45 models (SC, ST6 and ST3) and their control counterparts (WT, M6 and M3). The glyco-engineered cell lines express primarily CD44 of less than 150 kDa (light CD44) as opposed to the above 150 kDa (heavy CD44) of the control cell lines.

(C) O-glycan site annotation of the full-length CD44 amino acid sequence. Half circles depict in silico predicted O-glycan sites and full circles depict experimentally confirmed O-glycan sites. The variable exon v6 is underscored in red and the transmembrane domain is highlighted in green.

(D) Schematic illustration of a full-length CD44 glycoprotein. The hyaluronan binding domain (HA-BD), variable chains (v2-10) as well as the O-glycan rich stem region, chondroitin sulfate (CS) and heparan sulfate (HS) are annotated. The information according to publications and UniProt database [8, 30, 64, 65].

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Figure 2

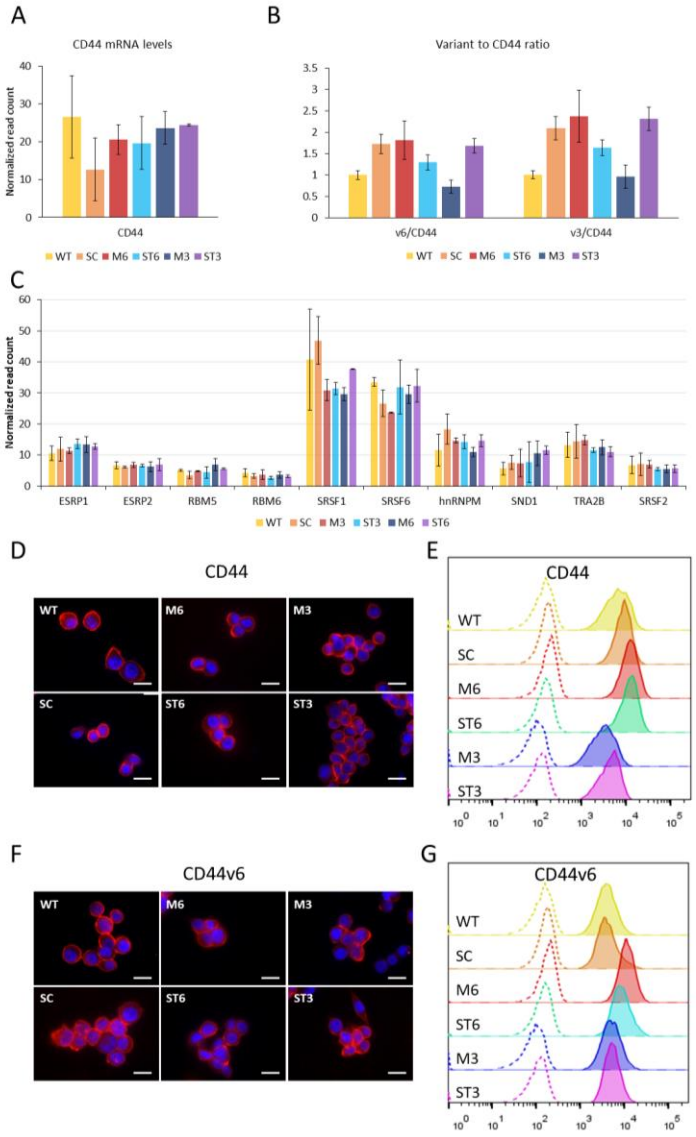


Fig. 2 CD_{44} expression analyses.
(A) CD_{44} mRNA levels of glyco-engineered cell lines (SC, ST6 and ST_3) and their control counterparts (WT, M6 and M_3) are shown. The expression data was extracted from whole transcriptome analysis, performed in duplicates and represented as average of the normalized reads ± SD.
(B) Analysis of the mRNA expression of variants $CD_{44}v_3$ and $CD_{44}v_6$ normalized to total CD_{44} expression by qRT-PCR. Analysis was performed in 2 biological replicates with 3 technical replicates each and is shown as average ± SD. Comparison of each glyco-engineered cell line (SC, ST6 and ST_3) and its respective control counterparts (WT, M6 and M_3) showed no statistical significant alterations (Student's t-test, $p > 0.05$).
(C) Expression levels of the splicing factors $ESRP_1$, $ESRP_2$, RBM_5, RBM6, $SRSF_1$, SRSF6, hnRNPM, SND_1, TRA_2B, $SRSF_2$, which are known to be involved in CD_{44} alternative splicing. The expression data was extracted from whole transcriptome analysis, performed in duplicates and represented as average of the normalized reads ± SD.
(D) IF staining of CD_{44} of the six cell line models. CD_{44} is shown in red and DAPI in blue. The scale bar is 20 µm.
(E) Flow cytometry analysis of CD_{44} expression in MKN_{45} glyco-engineered cell lines (SC, ST6 and ST_3) as compared to their control cell lines (WT, M6 and M_3, respectively). The negative controls are shown in dotted lines.
(F) IF staining of $CD_{44}v_6$ of the six cell line models. $CD_{44}v_6$ is shown in red and DAPI in blue. The scale bar is 20 µm.
(G) Flow cytometry analysis of $CD_{44}v_6$ in MKN_{45} glyco-engineered cell lines (SC, ST6 and ST_3) as compared to their control cell lines (WT, M6 and M_3, respectively). The negative controls are shown in dotted lines.

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Figure 3

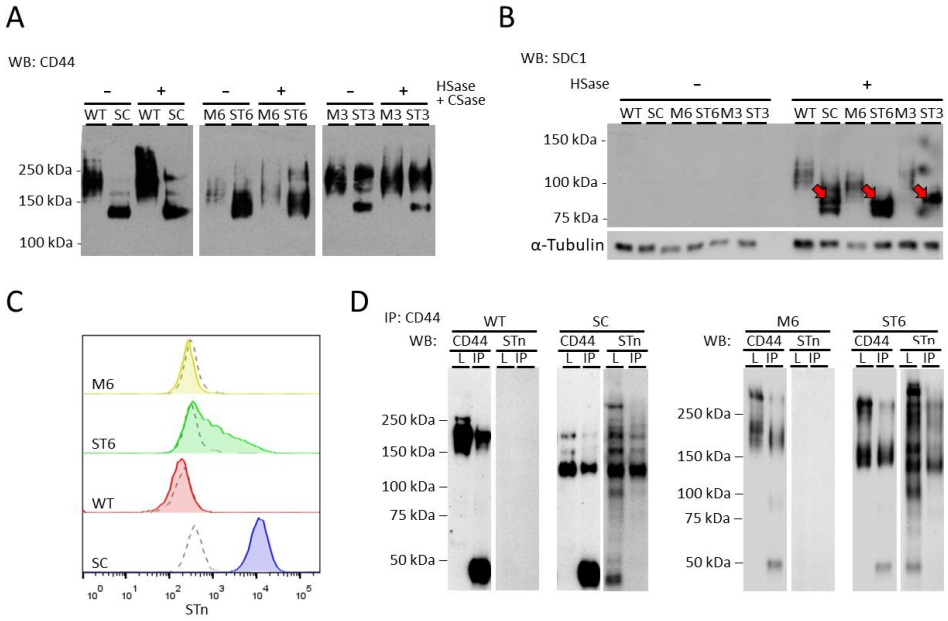


Fig. 3 Glycosaminoglycan digestion and sialyl Tn analysis of CD44.
(A) Western blot analysis of CD44 before and after incubation with heparinase I and III (HSase) and chondroitinase ABC (CSase). The glyco-engineered cell lines SC, ST6 and ST3 are shown adjacent to their respective control cell lines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

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Figure 4

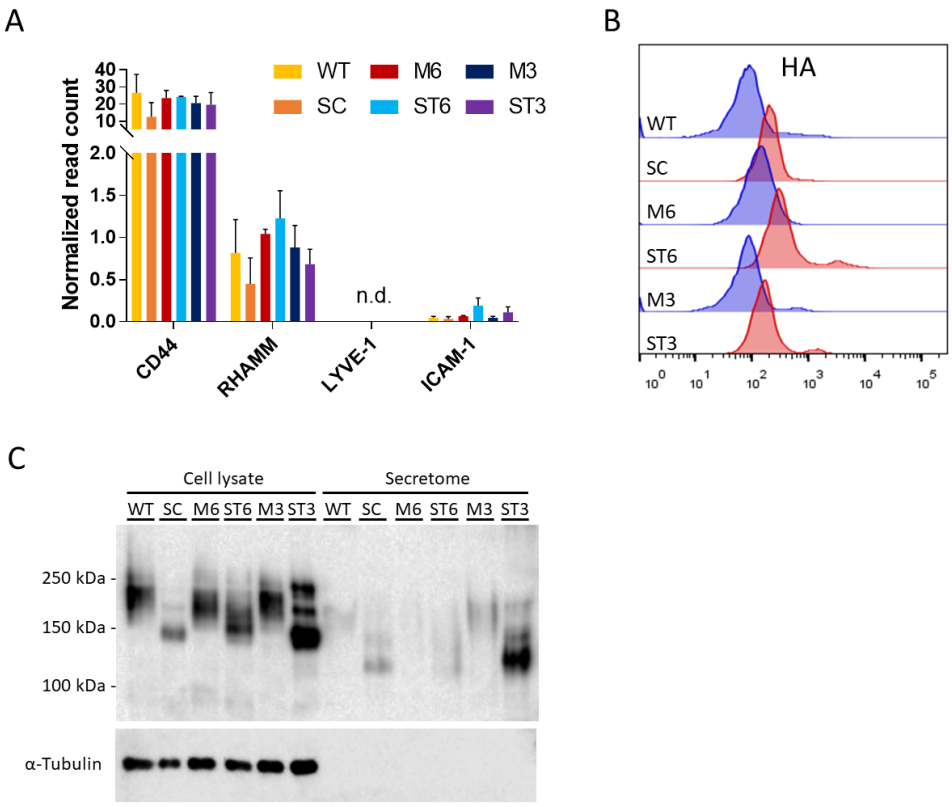


Fig. 4 CD_{44} binding to HA altered through O-glycan truncation.
(A) Expression levels of other HA binding proteins (CD_{44}, RHAMM, LYVE-1 and ICAM-1) are shown in the six cell line models. The expression data was extracted from whole transcriptome analysis, performed in duplicates and represented as average of the normalized reads ± SD.
(B) Flow cytometry analysis of HA-fluorescein binding to MKN_{45} glyco-engineered cell lines with truncated O-glycans (SC, ST6 and ST_3) and their respective control cell lines (WT, M6 and M_3).
(C) Comparative western blot analysis of CD_{44} and tubulin in total cell lysate and in conditioned medium (secretome). MKN_{45} glyco-engineered cell lines with truncated O-glycans (SC, ST6 and ST_3) and their respective control cell lines (WT, M6 and M_3) are shown.

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Figure 5

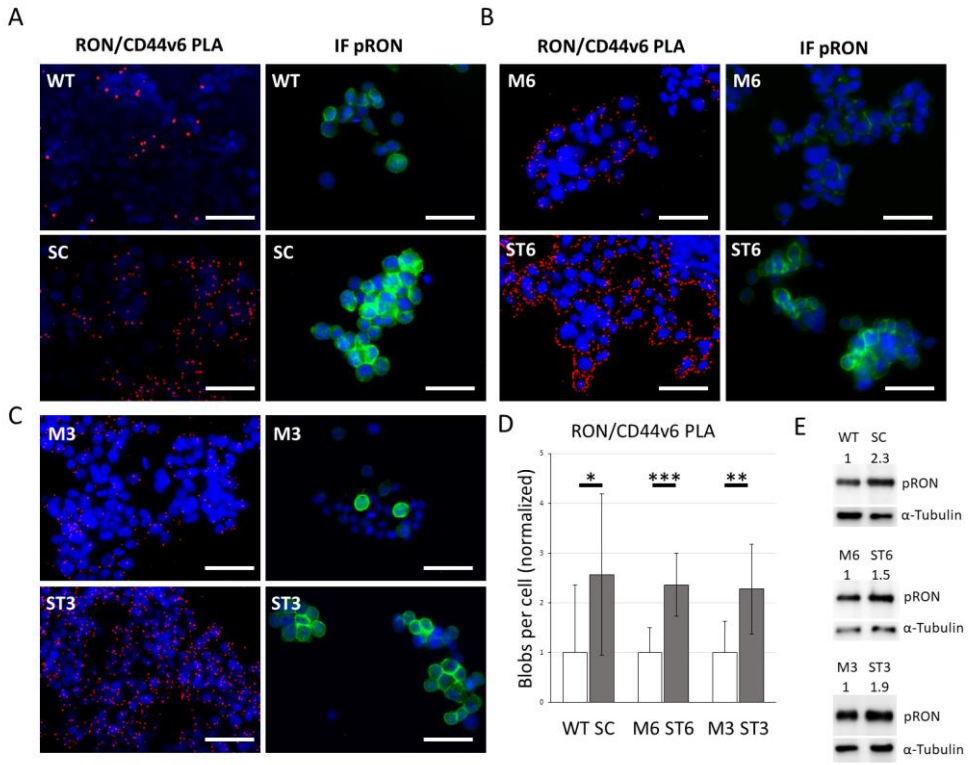


Fig. 5 Colocalization analysis between CD44v6 and RON as well as RON activation.
(A-C) PLA for the receptor tyrosine kinase RON and its co-receptor CD44v6 (left panels). Red PLA dots (blobs) indicate colocalization event of RON and CD44v6. In addition, IF staining of pRON for each respective cell line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

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Figure 6

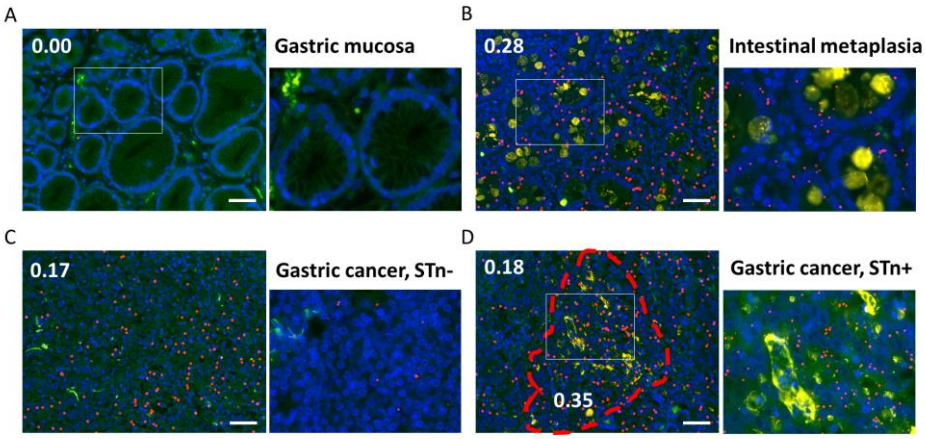
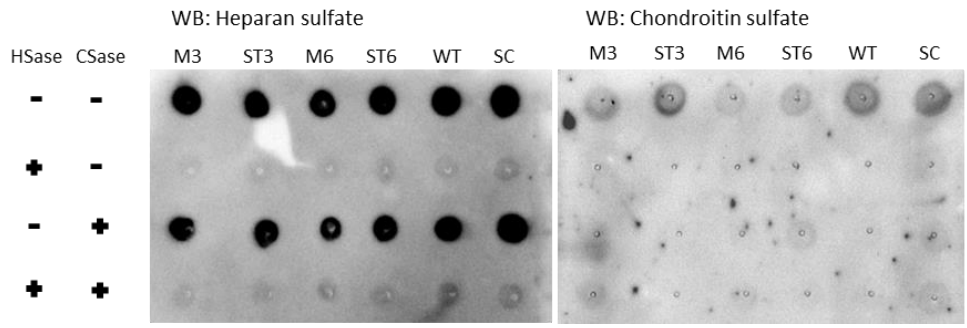


Fig. 6 PLA analysis of CD44v6/RON combined IF staining of STn in human gastric tissue sections.
(A-D) Combined PLA/IF analysis in human gastric tissue sections. The PLA was performed with anti-RON and anti-CD44v6 antibody detecting their colocalization (red blobs). The number of PLA blobs per cell were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

INSTITUTO
DE INVESTIGAÇÃO
E INOVAÇÃO
EM SAÚDE
UNIVERSIDADE
DO PORTO

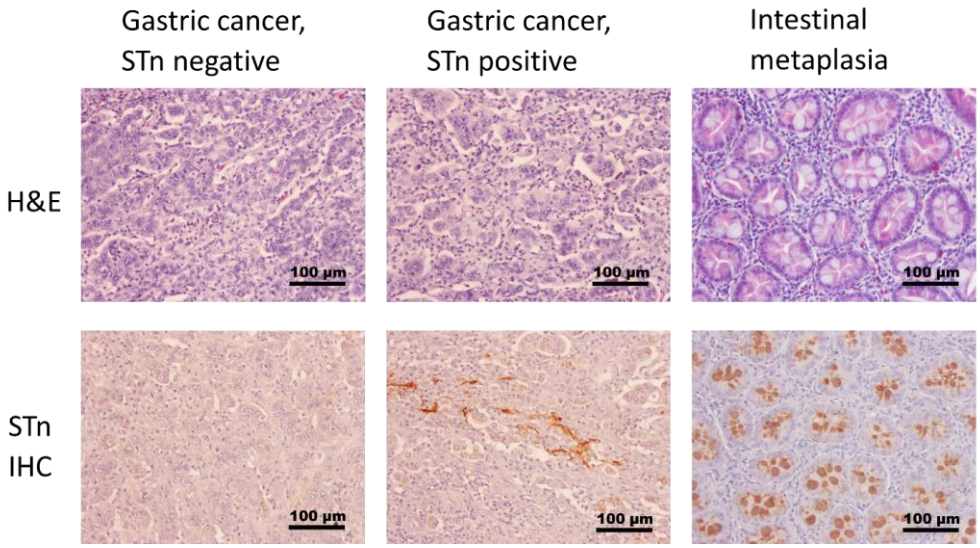
Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt

Supplementary Figure 1.



Confirmation of the efficiency of heparinase and chondroitinase digest.
Dot blot analysis of whole cell lysates before and after heparinase I and III (HSase) and/or chondroitinase ABC (CSase) digest. 1 µg of lysate was loaded per dot onto a PVDF membrane and incubated with anti-HS antibody (left) or anti-CS antibody (right).

Supplementary Figure 2.



Hematoxylin and eosin (H&E) staining in sialyl Tn (STn) negative or positive gastric cancer region as well as in intestinal metaplasia. No significant histological differences between STn negative and positive areas were evident. In addition, immunohistochemistry (IHC) staining of STn is shown in respective tissue regions.

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal
+351 220 408 800
info@i3s.up.pt
www.i3s.up.pt