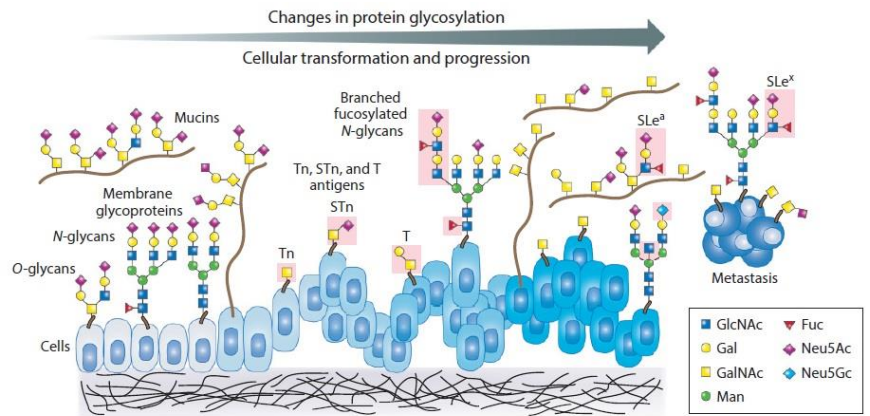




Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Ana Cíntia Silva Barreira

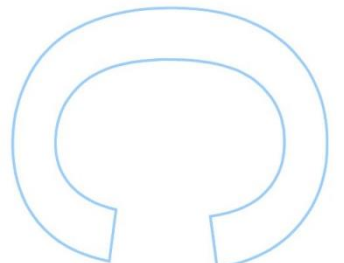
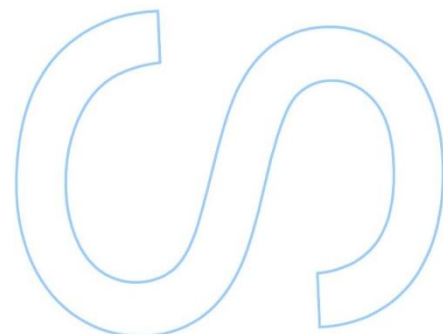
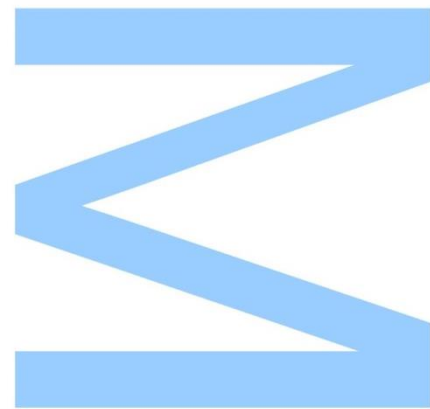
Mestrado Bioquímica
Glycobiology in cancer, IPATIMUP
2015

Orientador

Catarina de Sena Bastos Gomes, Post-Doc, Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP)

Coorientador

Celso Albuquerque Reis, Professor auxiliar convidado, Instituto de Ciências Biomédicas Abel Salazar (ICBAS)



Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,

Porto, ____ / ____ / ____

N

S

O

Role of Carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Laboratório de Glicobiologia no Cancro

IPATIMUP – Instituto de Patologia e Imunologia Molecular da Universidade do Porto

ICBAS – Instituto Ciências Biomédicas Abel Salazar da Universidade do Porto

FCUP – Faculdade de Ciências da Universidade do Porto

Orientadora: Doutora Catarina de Sena Bastos Gomes

Afiliação: IPATIMUP – Instituto de Patologia e Imunologia Molecular
da Universidade do Porto; Faculdade de Medicina da Universidade do Porto

Co-orientador: Professor Doutor Celso Albuquerque Reis

Afiliação: IPATIMUP – Instituto de Patologia e Imunologia Molecular
da Universidade do Porto; Faculdade de Medicina da Universidade do Porto; Instituto
de Ciências Biomédicas Abel Salazar

Dissertação de candidatura ao grau de Mestre em Bioquímica submetida à Faculdade
de Ciências da Universidade do Porto.



Porto, 2015

“You never know what’s around the corner.
It could be everything or it could be nothing.
You keep putting one foot in front of the other,
and then one day you look back and
you’ve climbed a mountain”

Tom Hiddleston

PERSONAL THANKS

Ao Professor Doutor Celso Reis pela oportunidade de realizar a minha dissertação de mestrado no grupo *Glycobiology in Cancer* e pela disponibilidade para esclarecimento de dúvidas de carácter científico.

À Doutora Catarina Gomes por toda a atenção antes e depois de entrar no laboratório. Por toda a disponibilidade mesmo quando ausente, por toda a vontade de transmitir os seus conhecimentos científicos. Pela compreensão de algumas falhas e confiança no trabalho por mim desenvolvido. Acima de tudo, porque todo este trabalho não seria possível sem a sua orientação.

Aos elementos do grupo *Glycobiology in Cancer* por tornarem esta estadia mais enriquecedora. Por toda a disponibilidade e apoio dentro e fora do laboratório. Por serem uma pequena família de braços abertos para receber novos elementos com vontade de crescer cientificamente. Um agradecimento especial aos “miúdos” do grupo, por toda a cumplicidade e por conseguirem sempre partilhar sorrisos nos momentos mais difíceis. Não podia desejar-lhes um futuro mais brilhante... *To my international friends a huge ‘gràcies i bona sort’ and ‘danke und viel Glück’.*

Ao Professor Doutor Sobrinho Simões pelas excelentes condições de acolhimento. A toda a equipa do IPATIMUP pela simpatia.

Aos amigos de sempre... À família que encontrei na faculdade, por todo o apoio, por existir sempre um tempo para estarmos juntos e compensar tudo o resto. Por saber que estavam sempre comigo mesmo quando não era possível e saber que a saudade era recíproca. Aos jantares e aos “sócios” da Casa do Povo por proporcionarem momentos inesquecíveis. Obrigada por não existirem intervalos entre nós.

Ao Diogo por toda a paciência e compreensão. Por tornar os meus dias mais completos. Obrigada pelo que és e por aquilo que sou quando estou contigo.

Aos meus avós por todo o carinho e por me ajudarem de todas as formas possíveis, sem que sejam necessárias palavras. Ao meu pai, por saber que posso contar com ele em qualquer momento.

À minha mãe por ser inexplicavelmente única e por sorte, minha.

ABSTRACT/RESUMO

Nowadays glycosylation is considered the most common and complex post-translation modification giving rise to a huge variety of glycan structures attached to glycoconjugates. Glycans are usually involved in multiple cellular and molecular processes playing a role in several physiological and pathological events, such as cancer.

Aberrant glycosylation has been considered as a hallmark of cancer and one of the most common alterations observed is the expression of sialylated structures, as a result from abnormal expression of sialyltransferases. SLe^x antigen is commonly expressed in cancer and its expression has been associated with cancer progression and aggressiveness, such as in aggressive type of gastric cancer.

Previously, our group demonstrated the expression of sLe^x in a gastric carcinoma cell line transfected with a gene encoded for the enzyme ST3Gal IV. In this cell line, carcinoembryonic antigen (CEA) was identified as the main protein carrier of sLe^x. Thus, the general aim of this study was to understand the role of CEA glycosylation, specifically the presence of sLe^x structures in gastric carcinoma.

In this study, we reported the expression of sLe^x in CEA molecules in ST3Gal IV expression' cells when compared to Mock cells. The characterization of glycans present in CEA molecules demonstrated the presence of complex type N-glycans and possibly O-glycans containing sLe^x structures.

In addition, we identified Perlecan and CD44v6 as putative carriers of sLe^x in MST3Gal IV cells. Furthermore, we confirmed the increase activation of c-Met and Src in MST3Gal IV cells when compared to Mock cells. This increased c-Met activation was also observed in Mock cells cultured with condition medium suggesting a possible role of secreted sLe^x protein carriers in the activation of this receptor. Our preliminary results using a mice model demonstrated a tendency of SLe^x expressing cells to promote the formation of bigger tumors.

Glycobiology could be a shortcut into the maze observed in cancer. The identification of aberrant glycosylation in cancer can bring new insights on the carcinogenesis process and contribute to the development of new serological biomarker assays for gastric cancer diagnosis and new therapeutically targets.

Keywords: gastric cancer, glycosylation, ST3Gal IV, sLe^x, invasion, CEA, CD44, Perlecan.

Actualmente, a glicosilação é considerada a modificação pós-tradução mais comum e complexa, que resulta numa diversa variedade de estruturas de glicanos associadas a glicoconjugados. Os glicanos estão normalmente envolvidos em diversos processos celulares e moleculares desempenhando um papel importante em processos fisiológicos e patológicos assim como em cancro.

A glicosilação aberrante tem sido considerada um “hallmark of cancer” e uma das alterações mais comuns em cancro é a expressão de estruturas sialiladas, como resultado da expressão aberrante de sialiltransferases. A expressão de sLe^x é muito comum em cancro estando associada a progressão e agressividade tumoral, como por exemplo em casos mais agressivos de carcinoma gástrico.

Previamente, o nosso grupo demonstrou a expressão de sLe^x numa linha celular de carcinoma gástrico transfetada com o gene que codifica a enzima ST3Gal IV. Na mesma linha celular, CEA foi identificado como a principal proteína transportadora de sLe^x. Assim, o principal objetivo deste trabalho baseou-se no estudo do papel da glicosilação de CEA em cancro gástrico, especificamente a presença de estruturas sLe^x. Neste estudo, verificamos a expressão de sLe^x no CEA e identificamos novas proteínas como sendo transportadoras de sLe^x nas células MST3Gal IV quando comparando com as células Mock. A caracterização de glicanos presentes no CEA revelou a presença de N-glicanos do tipo complexo assim como a possível presença de sLe^x em O-glicanos.

Para além disso, identificamos Perlecan e CD44v6 como possíveis transportadores de sLe^x nas células MST3Gal IV. A ativação de c-Met e Src foi confirmada nas células MST3Gal IV quando comparado com as células Mock, e o aumento de ativação de c-Met foi também observado nas células Mock cultivadas em meio condicionado. Este resultado sugere um possível papel de proteínas secretadas que expressam sLe^x na ativação de c-Met. Os resultados preliminares do modelo animal demonstraram uma tendência para a formação de tumores maiores em ratinhos injetados com as células MST3Gal IV.

A Glicobiologia poderá representar um atalho dentro do “labirinto” observado no cancro. A identificação de estruturas glicosiladas aberrantes em cancro pode trazer novos conhecimentos do processo de carcinogénese contribuir para o desenvolvimento de novos testes serológicos de diagnóstico de cancro gástrico assim como novos alvos terapêuticos.

Palavras-chave: cancro gástrico, glicosilação, ST3Gal IV, sLe^x, invasão, CEA, CD44, Perlecan.

TABLE OF CONTENTS

List of figures and tables	15
List of abbreviations	19
Introduction.....	23
1. Glycobiology and its importance in cells	24
2. Glycoconjugates.....	26
2.1 Mucin Type O-glycosylation biosynthesis	26
2.2 N-glycosylation.....	28
2.3 The importance of sialylated structures at cell's surface.....	29
2.4 Aberrant glycosylation and its consequences.....	31
3. Cancer and glycosylation: the bittersweet relation.. ..	32
3.1 A general view of cancer.	32
3.1.1 Gastric cancer.....	32
3.2 Glycosylation in cancer	33
3.2.1 Glycoconjugates in cancer	37
4. Tumor biomarker.. ..	39
4.1 Carcinoembryonic antigen (CEA).. ..	39
4.1.1 hnRNPM-CEA, a key molecule for the metastatic process?.....	40
Aim.	43
Material and Methods	45
Results	55
1. Characterization of CEA glycosylation <i>in vitro</i> and in clinical samples.. ..	56
1.1 Characterization of CEA glycosylation in glycoengineered gastric carcinoma cells.. ..	56
1.2 Characterization of CEA glycosylation in clinical samples	61
2. Identification of putative new protein targets which can be involved in MST3Gal IV invasive phenotype.....	62
2.1 Role of secreted sLe ^x expressing proteins on c-Met and Src activation	62
2.2 Identification and glycan characterization of differential expressed proteins in MST3Gal IV cells	63
3. Evaluation of tumorigenic potential of sLe ^x expressing cells	72
Discussion	75
Future Perspectives	83
Bibliography.....	87
Supplementary data.....	101

LIST OF FIGURES AND TABLES

Figures

Fig. 1 – Cell surface glycoconjugates.....	24
Fig. 2 – Mucin type O-glycosylation	27
Fig. 3 – Complexity of N-Glycans.....	29
Fig. 4 – Sialylated structures.....	30
Fig. 5 – Changes on protein glycosylation observed during cancer progression.	35
Fig. 6 – Metastatic process of cancer cells.....	36
Fig. 7 – Interaction of tumor cells and Kupfer cells in the liver.....	41
Fig. 8 –CD44 specific primers.....	52
Fig. 9 – SLe ^x expression in gastric carcinoma cell line stably transfected with ST3Gal IV.....	56
Fig. 10 – SLe ^x expression upon PNGase F and Endo H enzymatic digestion.	57
Fig. 11 – Glycoprotein analysis in Mock and MST3Gal IV cells.....	57
Fig. 12 – Expression of CEA glycoprotein in both Mock and MST3Gal IV cells in different conditions.....	58
Fig. 13 – Western blot analysis of CEA and sLe ^x in CEA immunoprecipitates.....	59
Fig. 14 – Glycan characterization of CEA immunoprecipitated molecules using PNGase F and Endo H enzymatic digestion.....	60
Fig. 15 – CEA and sLe ^x expression in gastric carcinoma tumor samples and colon biopsy.....	61
Fig. 16 – Western blot analysis of total and phosphorylated c-Met and Src expression levels in gastric carcinoma cells.....	62
Fig. 17 – Role of sLe ^x secreted proteins in the activation of c-Met and Src proteins....	63
Fig. 18 – Coomassie Blue staining of secretome samples from MST3Mock and MST3GalIV	64
Fig. 19 – Expression of Perlecan in different condition in Mock and MST3Gal IV cells... ..	64
Fig. 20 – mRNA expression of Perlecan from Mock and MST3GalIV using domain specific primers.....	65
Fig. 21 – Effect of enzymatic digestion, using PNGase F and Endo H in Perlecan isoforms	66
Fig. 22 – Western blot analysis of Perlecan immunoprecipitation in total cell lysates and secretome from Mock and MST3Gal IV cells	67
Fig. 23 – Expression of CD44 and CD44v6 containing isoforms in Mock and MST3Gal IV cells... ..	68
Fig. 24 – PCR analysis of CD44 transcripts in Mock and MST3Gal IV cells	69

Fig. 25 – CD44v6 western blot analysis after treatment with PNGase F and Endo H in Mock and MST3GalIV total cell lysates and secretome	70
Fig. 26 – Evaluation of sLe ^x expression in CD44v6 from Mock and MST3Gal IV total cell lysates and secretome	71
Fig. 27 – Tumorigenic potential of Mock and MST3Gal IV cells in athymic nude mice	72
Fig. 28 – Tumor growth kinetics of Mock and MST3GalIV injected mice	73
Fig. 29 – HnRNPM4 expression in Mock and MST3Gal IV cell lysates	73
Fig. 30 – Immunofluorescence analysis to evaluate the expression sLe ^x and hnRNPM4 in Mock and MST3Gal IV gastric carcinoma cells	74

Tables

Table 1 – Composition of 7,5% tris-glycine SDS-polyacrylamide gel	47
Table 2 – List of primary antibodies used with the respective dilutions.	48
Table 3 – PNGase F and Endo H enzymatic reaction mixtures for CEA immunoprecipitated samples.	49
Table 4 – PNGase F and Endo H enzymatic reaction mixtures for cell lysates and secretome samples	50
Table 5 – Primers for PCR amplification of Perlecan cDNA	52
Table 6 – Primers for PCR amplification of CD44 cDNA..	52
Table 7 – Proteins identified by mass spectrometry regarding sLe ^x antigen detection by western blot..	58

Supplementary data

Supplementary figure 1 - CEA is a sLe ^x carrier in gastric carcinoma tissues.....	103
Supplementary table 1 – Expression of CEA and sLe ^x in gastric carcinoma tissues and expression of CEA/sLe ^x by PLA.....	104

LIST OF ABBREVIATIONS

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Asn – Asparagine

C2GnT – core 2 β 1-6 N-acetylglucosaminyltransferase

c-Myc – myelocytomatosis oncogene

CA19-9 – carbohydrate antigen 19-9 or sLe^a

CA125 – carbohydrate antigen 125 or Muc16

CAM model – chicken embryo chorioallantoic membrane model

CD44s – CD44 standard form

CD44v – CD44 variant isoform

cDNA – complementary DNA

CEA – carcinoembryonic antigen

CEACAM – cell adhesion molecules

CEAR – CEA receptor

CD66e – cluster of differentiation or CEA

Dol-P – dolichol phosphate

Dol-P-P-GlcNAc – dolichol pyrophosphate N-acetylglucosamine

DNA – deoxyribonucleic acid

DTT - dithiothreitol

ECM – extracellular matrix

EDTA – ethylenediamine tetraacetic acid

EGF – epidermal growth factor

EMT – epithelial-mesenchymal transition

Endo H – endoglycosidase H

ER – endoplasmic reticulum

FBS – fetal bovine serum

Fuc – fucose

Fuc-T – fucosyltransferase

G418 – gentamicin

Gal – galactose

GalNAc – N-acetylgalactosamine

GalNAcT – N-acetylgalactosamine

Gal-T – galactosyltransferase

GlcNAcT or GnT – N-acetylglucosamine transferase or N-acetylglucosaminyltransferases

Glc – Glucose

HGFR – hepatocyte growth factor receptor

hnRNPM – heterogeneous nuclear ribonucleoprotein M

HSPG – heparan sulfate proteoglycan

IP – immunoprecipitation

Man – mannose

MET – mesenchymal-epithelial transition

MHC – major histocompatibility complex

Mock – MKN45 cells transfected with an empty vector

mRNA – Messenger RNA

MST3Gal IV – MKN45 cells transfected with ST3Gal IV gene

Neu5Ac - N-acetylneuraminic acid

NaCl – sodium chloride

p-Met – phosphorylated Met (activated form of c-Met)

P/S – penicillin-streptomycin

p-Src – phosphorylated Src (activated form of Src)

PBS – phosphate buffer saline

PCR – polymerase chain reaction

PLA – Proximity ligation assay

PMSF – phenylmethylsulfonyl fluoride

PNGase F – peptide-N-glycosidase F

ppGalNAc-Ts – UDPGalNAc-polypeptide N-acetylgalactosaminyl-transferases

RNA – Ribonucleic acid

Sia / SA – Sialic Acid

SDS – dodecyl sulfate

SDS-PAGE – Polyacrylamide gel electrophoresis

Ser – serine

sLe^x – sialyl lewis x antigen

Src – proto-oncogene tyrosine-protein kinase Src

ST3Gal – Galactose α 2,3-sialyltransferase

ST6Gal – Galactose α 2,6 sialyltransferase

ST6GalNAc – N-acetylgalactosamine α 2,6 sialyltransferase

ST8Sia – Sialic acid α 2,8 sialyltransferase

ST – sialyl T antigen

STn – sialyl Tn antigen

Thr – threonine

Tris-HCl – Tris-hydrochloride

UDP-GalNAc – UDP sugar donor N-acetylgalactosamine

INTRODUCTION

1. Glycobiology and its importance in cells

According with the last years, it is clear the relevant attention that scientific research has given to the dogma's biology: conversion of DNA into RNA and then into proteins. This centered investigation resulted in a devaluation of other cellular processes which are also very important to cell's function, such as glycosylation¹.

Glycobiology has a long history that began with investigations of the basic constituents of cells and the nature of the polysaccharide carbohydrate component². Later it was possible to understand that every free-living cell and every cell type within multicellular organisms is covered with a dense and complex layer of glycans (fig.1). These glycans are commonly attached to proteins, lipids (and nucleic acids) through an enzymatic reaction called glycosylation. It is the nature of glycosylation, for instance the linkage between the monosaccharides units and the linkage to their non-glycosyl part, (protein or lipid) which increases the biological diversity and complexity of glycans^{3,4}.

Nowadays research in this area is increasing, with the emerging of the term "glycomics" as the complete set of glycans and glycoconjugates which are made by a cell or organism under specific conditions⁵.

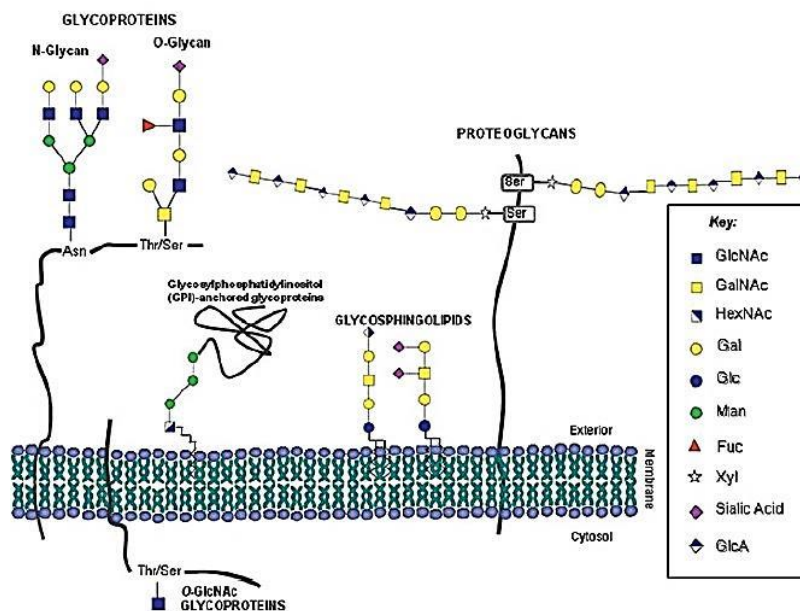


Fig.1 – Cell surface glycoconjugates⁶. Glycans can be found in many different forms attached to proteins and lipids being called glycoproteins and glycolipids, respectively. Proteins can be glycosylated in a particular amino acid resulting in O- and N-glycans (chapter 2) and can also be found as membrane linked to a phosphatidylinositol called glycosylphosphatidylinositol (GPI)-anchored proteins and as proteoglycans. Glycolipids are mostly represented by glycosphingolipids which are ceramide-linked glycans.

It is not clear this universality of cell-surface and secreted glycosylation, but it may suggest that evolution has selected glycans as being the most flexible and diverse molecules located at the interface between cells and the extracellular milieu⁷.

Protein glycosylation is the most common and complex posttranslational modification involved in many physiological events⁸, and over 50% of proteins are glycosylated⁹. It is important to understand that this modification is an enzymatic process that occurs in the endoplasmic reticulum (ER), Golgi and cytoplasm in opposite to a nonenzymatic chemical reaction of glycation (or “glucosylation”) which refers to the non-enzymatic covalent linkage of saccharides such as glucose to proteins¹⁰⁻¹³. Protein glycosylation is therefore controlled by various factors like cellular concentration and localization of nucleotide sugars and expression, localization and competition of the involving enzymes¹⁴.

Glycans reveal to be very important to cells due to their involvement in multiple cellular and molecular processes like protein folding and trafficking, cell-cell and cell-matrix interactions (for example through interaction with lectins¹⁵), cellular differentiations and cell adhesion⁶, immune response¹⁶⁻¹⁸, receptor activation, signal transductions, endocytosis¹³ and also their involvement in embryonic development and cellular activation in vertebrates which are typically accompanied by changes in cellular glycosylation profiles¹⁹. For instance, in the case of immune response, specific glycoforms have been described to be involved in the folding, quality control and assembly of peptide-loaded major histocompatibility complex (MHC) antigens and T-cell receptor complex²⁰, and also involved in host-pathogen interaction²¹. In addition, glycosylation of secreted proteins may also provide solubility, hydrophilicity, and negative charge, thus reducing unwanted nonspecific intermolecular interactions in extracellular spaces and also protecting against proteolysis⁷.

Thus, not only is the glycome a more complex process than the genome, transcriptome or proteome, but it is also more dynamic, showing considerable variances with cell's type and status (health vs disease)¹⁴.

2. Glycoconjugates

In this phase, it is important to reemphasize that glycan chain structures are not encoded in the genome as protein sequences. In the case of glycans, they are secondary gene products and it is known that only a percent of genes in the human genome are involved in the glycosylation process⁷.

Glycosylation involves the synthesis of glycans which are assemblies of oligosaccharides or polysaccharides and exist as free form or in covalent complexes with lipids or proteins²². These complexes named glycoconjugates (fig. 1) can be classified into glycoproteins, glycosphingolipids, proteoglycans and glycosylphosphatidylinositol-linked proteins, which are then organized into several types according with the glycosylation linkage site. In the family of glycoproteins there are two types of glycans: O-linked glycans and N-linked glycans⁷.

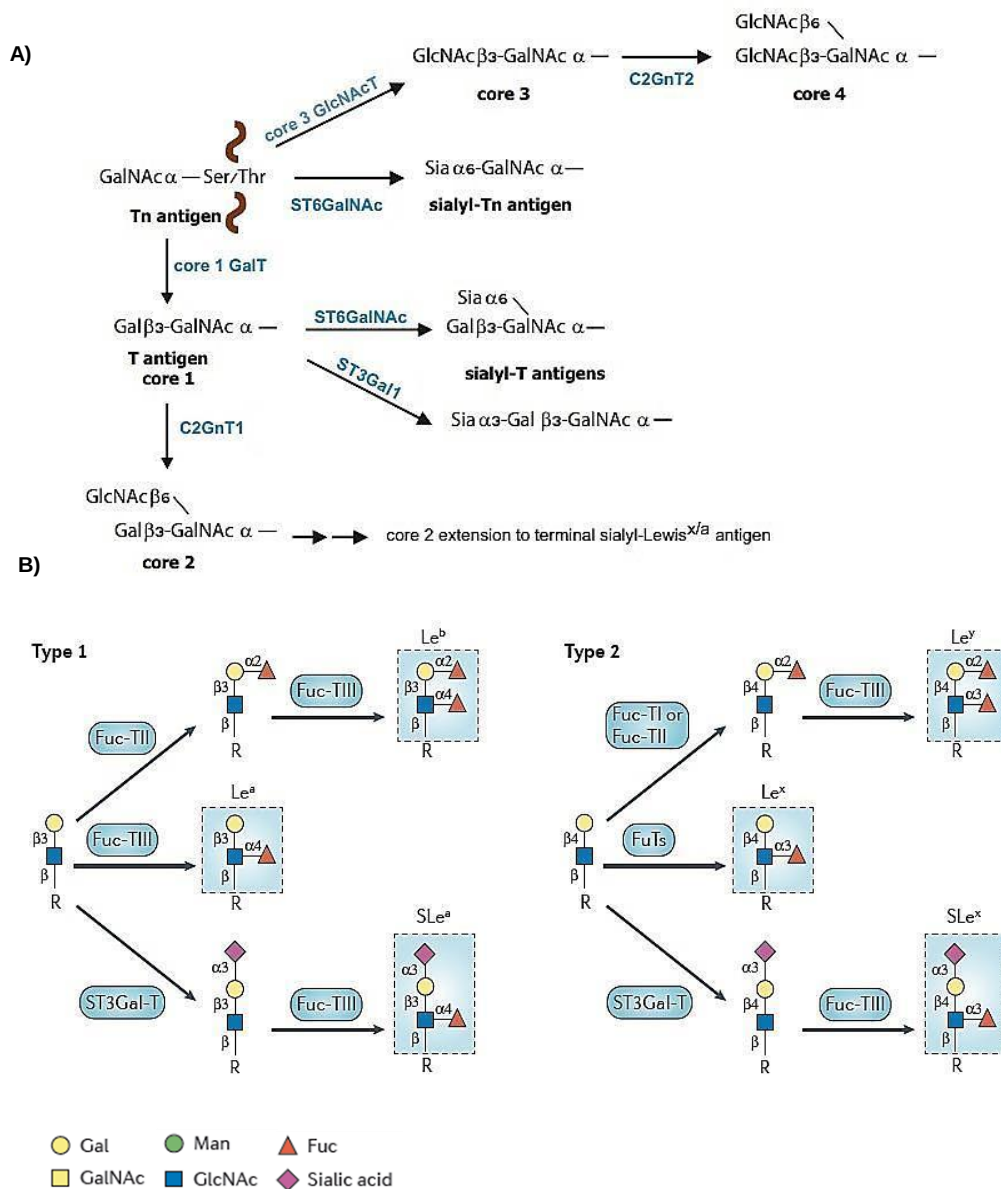
2.1 Mucin Type O-glycosylation biosynthesis

O-linked glycosylation is a post-translational modification involved in protein maturation with the attachment of sugars to serine (Ser) and threonine (Thr), and with a less importance, to hydroxyproline and hydroxylysine¹¹ in a step-wise manner. The Ser- or Thr-linked oligosaccharides are present on many glycoproteins and predominate on secreted or membrane bound mucins (large glycoproteins), the most common type of O-glycosylation observed^{11,22} (fig. 2A).

Mucin type O-glycosylation takes place in the Golgi, beginning with the transfer of N-acetylgalactosamine (GalNAc) from a sugar donor (UDP-GalNAc) to the hydroxyl group of Ser/Thr residues which results in a glycosidic bond mediated and controlled by a large family of UDPGalNAc-polypeptide N-acetylgalactosaminyl-transferases (ppGalNAc-Ts) leading to the formation of the Tn antigen²³. After this first step, it can follow three different pathways: 1) the formation of the core 1 structure also known as the T antigen, through the addition of galactose (Gal) to GalNAc mediated by core 1 synthase (core1 GalT); 2) the formation of the core 3 structure as an elongation of the core 1 structure, with the addition of N-acetylglucosamine (GlcNAc) mediated by a glucosaminyltransferase (Gn-T); or 3) the formation of sialyl-Tn antigen which is formed due to the addition of a sialic acid through the activity of a sialyltransferase. The addition of a sialic acid to a carbohydrate structure stops any further processing of the oligosaccharide chain, which is also the case of the synthesis of sialylated core 1 structures sialyl-T and disialyl-T antigens²⁴⁻²⁷ (fig. 2A).

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

The biosynthesis of the core 2 and core 4 structures is a result of the elongation of core 1 and 3, respectively, mediated by glucosaminyltransferases. These four core structures (core 1-4) are the most common in mammals being core 2 produced in many epithelial and haematopoietic cells, which its extension by two different enzymes (β 3- or β 4-transferases, Gal-Ts and Gn-Ts) results into two types of chains (type 1 and 2 chains). Type 1 and type 2 chains are characterized by the formation of variable repetitive disaccharide structures of GlcNAc and Gal residues in β 1,3 and β 1,4 linkage respectively. The activity of different enzymes such as fucosyltransferases and sialyltransferases with distinct glycosylation linkage site of action in the type 1 and 2 chains results into sialyl Lewis antigens: sialyl Lewis A (sLe^a) and sialyl Lewis X (sLe^x)^{22,25,26,28}.



Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Fig. 2 – Mucin type O-glycosylation. (A) Schematic representation of the biosynthetic pathways of most common mucin-type core structures²⁹. (B) Schematic representation of sLe^a (in the left) and sLe^x synthesis (in the right)⁴. sLe^a and sLe^x are a result of sialylation and fucosylation of type 1 and 2 chains, respectively. In the case of sLe^x the addition of α 2,3 SA by different sialyltransferases such as ST3Gal-IV followed by the addition of α 1,3 Fuc by α 1,3-fucosyltransferase. C2GnT, core 2 β 1-6 N-acetylglucosaminyltransferase; GalT, galactosyltransferase; GlcNAcT, N-acetylglucosaminyltransferase; ST6GalNAc, α 2,6-sialyltransferase which act on N-acetylgalactosamine (GalNAc) and results in the synthesis of sialic acids α 2,6, such as STn; ST3Gal-T, α 2,3-sialyltransferase which act on galactose and results in the synthesis of sialic acids α 2,3, for instance sLe^x and sLe^a; Fuc-TIII, fucosyltransferase.

2.2 N-glycosylation

The N-glycosylation is also a post-translational modification which results in the attachment of sugars to asparagine (Asn) residues through a N-glycosidic bond. In this case, Asn must be located in a consensus sequence Asn-X-Ser/Thr, where X can be any amino acid with the exception of proline³⁰. This process begins at the cytosolic surface of the endoplasmic reticulum (ER) membrane, with the assembling of the precursor oligosaccharide that begin with the addition of GlcNAc from a sugar donor (UDP-GlcNAc) to the lipid-like precursor dolichol phosphate (Dol-P) which generates dolichol pyrophosphate N-acetylglucosamine (Dol-P-P-GlcNAc). This initial precursor continues with the addition of monosaccharides GlcNAc and mannose (Man) catalyzed by various glycosyltransferases. This reaction results in Dol-P-P-GlcNAc₂Man₅ which structure flips to the luminal side of the ER and receives more four residues of mannose and three residues of glucose (Glc) resulting into Dol-P-P-GlcNAc₂Man₉Glc₃. Each of these sugar additions are catalyzed by specific glycosyltransferases located on both sides of the ER membrane and after the full assembly, the glycan is transferred in block to a nascent peptide chain in the consensus region, within the ER^{8,17,31,32}.

N-glycan processing is carried out in the ER and then in the Golgi, which involves other modification including: removal and addition of sugar residues catalyzed by glycosyltransferases; action of a chaperone protein to assist protein folding; and another series of sugar additions to the core glycan structure. This addition and removal of sugar residues in the precursor oligosaccharide gives rise to three main types of glycans: high mannose, hybrid and complex N-glycans (fig. 3A). Mature N-linked glycans consists of a core structure containing two GlcNAc and three Man residues, which two of Man can be elongated with antenna formed by galactose, GalNAc, fucose and sialic acid. As O-glycans, the addition of saccharides to the branch structures of N-glycans can also result in the synthesis of epitopes as sLe^a and sLe^x^{8,11,19,33} (fig. 3B).

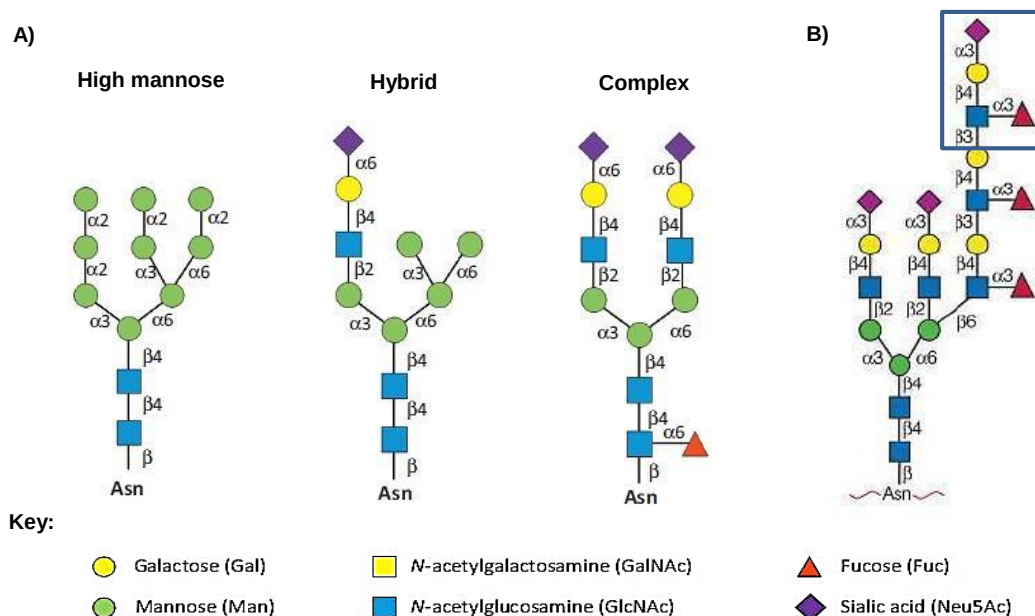


Fig. 3 – Complexity of N-Glycans. (A) N-glycosylation consists in three main types of N-glycans³⁴: 1) high mannose type containing only branched mannose residues attached to the pentasaccharide core structure; 2) Hybrid type containing mannose residues in one of the branches and a complex branch initiated with GlcNAc residue; 3) Complex oligosaccharides containing two branches which begin with the addition of two GlcNAc residues catalyzed by N-acetylglucosaminyltransferases (GnTs). Complex and hybrid oligosaccharides are decorated with the addition of sialic acid, fucose, galactose, N-acetylgalactosamine and sulfate; (B) The addition of diverse types of saccharides as sialic acid and fucose to N-glycans can also result in the synthesis of epitopes such as sLe^x, represented in the blue square¹⁹).

2.3 The importance of sialylated structures at cell's surface

Sialylated structures are characterized by containing the saccharide sialic acid (Sia) which comprises a family of more than fifty carbohydrates derived from neuraminic acid and result in the sharing of a similar nine carbon backbone. Sialic acids are enzymatically attached to O- or N-linked glycans on proteins or glycolipids inside the cell, which the most common variant is N-acetylneuraminic acid (Neu5Ac)³⁵.

The level of surface sialylation depends on the action of various types of enzymes such as sialyltransferases that catalyze the addition of sialic acids to proteins and lipids and also sialidases or neuraminidase enzymes which are responsible for the cleavage of Sia during glycoprotein degradation¹⁵. Regarding the addition of Sia to glycoproteins, there are four major families of sialyltransferases which are classified according to the carbohydrate linkage where they transfer the Sia residues: ST6Gal (Galactose α 2,6 sialyltransferase), ST3Gal (Galactose α 2,3 sialyltransferase), ST6GalNAc (N-acetylgalactosamine α 2,6 sialyltransferase) ST8Sia (Sialic acid α 2,8 sialyltransferase)^{36,37} (fig. 4).

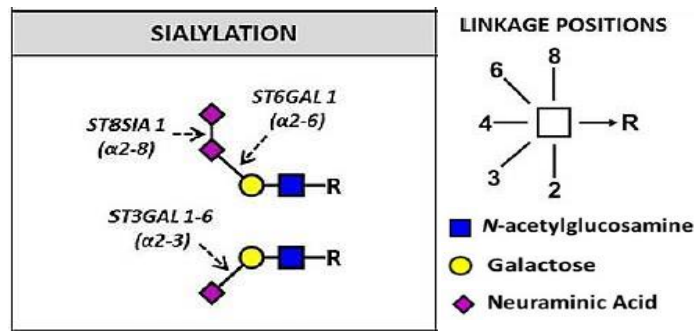


Fig. 4 – Sialylated structures³⁷. It is possible to see the linkage positions where the addition of a sialic acid occurs by sialyltransferases. Concerning the described enzymes, only ST3Gal III and IV are both involved in O- and N-glycans sialylation.

Sialylated glycoconjugates play important roles in cells and participate in diverse physiological functions. Sialic acids can act as biological mask by shielding recognition sites and contributing to the recognition of cells as being 'self' or/and operating as biological recognition sites (ligands) for hormones, lectins, antibodies and adhesion molecules³⁵. Another important characteristic of Sia is the presence of a negatively charged at physiological pH, contributing for the potential to alter protein conformation. In addition, sialylation has been reported as being involved in the regulation and retention of various receptors at the cell surface¹⁵.

As mentioned before, the expression of sialylated structures is involved in many physiological processes. One mechanism that has been well described is the capacity of leukocytes to role within blood vessels and extravasate to the target tissues due to the expression of selectins ligands, sLe^x and sLe^a structures, at cell surface. ST3Gal IV has been described as one of the major sialyltransferases involved in the biosynthesis of selectin ligands required for leukocyte trafficking³⁸, such as the expression of sLe^x^{39,40}. This sialylated structure represents one of the most important $\alpha 2,3$ sialylated structure (fig. 3B). It appears at a specific stage of early embryonal development⁴¹⁻⁴³ and later is also expressed on leukocytes with an important role during inflammatory response. As a matter of fact, many studies found increased levels of ST3Gal IV expression associated with increased sLe^x epitopes in human models of inflammation in vitro^{44,45} and cancer^{46,47}. On the contrary, the expression of sLe^x is typically low in normal epithelial cells^{48,49}.

Taking all this facts in consideration, sialylated structures like sLe^x appear to be important regulators of molecular and cellular interactions³⁵.

2.4 Aberrant glycosylation and its consequences

At this point, it is possible to recognize the importance of secreted or cell's surface glycoconjugates and how their variable and dynamic nature allows the generation of the organism biological diversity and complexity. Although glycans are involved in many physiological functions required for normal development, there are also evidences of its involvement in disease. For instance sialic acids, the outermost monosaccharides units on the glycan chains, are often recognition sites for pathogens⁵⁰⁻⁵³. Thus, some abnormal functions are related with glycosylation alterations or defects in its biosynthesis, resulting for instance in a fail on the recognition of cells being 'self and triggering inflammatory diseases⁵⁴. In addition to this, one of the most studied examples is the association between alteration in glycosylation and cancer¹.

In cancer, aberrant glycosylation of glycoconjugates leads to the expression of cancer associated antigens which can be found in many molecules such as attaching and anchoring molecules, receptors, enzymes or secreted glycoconjugates. These alterations usually change the normal cellular interactions with their microenvironment and cell signaling pathways³⁴ involved in pathological events such as cell growth, differentiation, adhesion, motility and invasion⁵⁵. In conclusion, altered glycosylation of cell surface glycolipids, membrane associated or secreted glycoproteins are universal modifications in cancer.

3. Cancer and glycosylation: the bittersweet relation

3.1 A general view of cancer

Cancer is a leading cause of death worldwide and according with 2012 statistics there were an estimated 14.1 million new cases of cancer and 8.2 million cancer related deaths (http://globocan.iarc.fr/Pages/fact_sheets_population.aspx). It is predicted there will be 23.6 million new cancer cases in the following years⁵⁶.

Cancer is a generic term to identify a group of diseases that can affect any part of the body, resulting in a top five higher incidence cancers including lung, breast, colorectal, prostate and gastric cancer (<http://www.cancerresearchuk.org/cancer-info/cancerstats/world/incidence/>; http://globocan.iarc.fr/Pages/fact_sheets_population.aspx).

Although this disease arises from one single normal cell, tumors have been recognized as organs due to their heterogeneity and which approaches them to the complexity of normal healthy tissues. The transformation into a tumor cell is a multi-step process (oncogenesis) that includes genetic and epigenetic changes resulting in the appearance of characteristics which favors cell's survival, such as persistent growth signals, evasion of apoptosis, angiogenesis, invasion and metastasis⁵⁷. Regarding metastasis, this is a well-recognized process which allows the dissemination of tumor cells from their primary tumor site to a distant site^{57,58}. Again, this is possible due to changes in cell phenotype where epithelial cells exhibit reduced cell-cell adhesion and increased motility via EMT (epithelial-mesenchymal transition) which promotes the detachment of tumor cells from the primary site. Thereafter, tumor cells undergo the reverse transition MET (mesenchymal-epithelial transition) at the site of metastases and subsequently colonize forming secondary tumors^{58,59}.

3.1.1 Gastric cancer

According to recent statistics, gastric cancer is the fifth most common malignancy in the world and the third leading cause of cancer-related deaths responsible for 723,000 deaths worldwide⁵⁶. The incidence of gastric cancer differs worldwide where developing countries represent more than 70% of the cases, and half of the world cases occur in Eastern Asia. Thus, the highest mortality rates occur in Eastern Europe, Central and South America, besides Eastern Asia⁵⁶. In our case, Portugal has one of

the highest rates of incidence and mortality in Europe. These statistical evidences suggest gastric cancer as an important target of scientific research.

Gastric cancer has been associated with several factors such as age, sex⁶⁰, smoking⁶¹ and diets which include high salt intake and high content nitrates, such as vegetables^{62,63}. In addition, the major trigger of gastric carcinogenesis is the infection of *helicobacter pylori*⁶⁴. This gram negative bacterium infection has been also correlated with the large consumption of salted food, which enhances *pylori* colonization in the stomach⁶⁵. In fact, *pylori* is thought to be present in the stomach of half the global population⁶⁶.

The adhesion of the bacterium to gastric mucosa is mediated by bacterial proteins named adhesins, which specifically recognize and bind to host glycosylated epitopes such as Lewis and sialyl Lewis antigens⁶⁷. In fact, a recent study showed that *pylori* infection resulted in a remodeling of gastric glycophenotype with increased expression of sLe^a and sLe^x. These epitopes are SabA ligands and their binding is used to colonize gastric mucosa during inflammation⁶⁸.

Helicobacter pylori infection results in a multistep process which includes a chronic inflammatory or chronic gastritis background, followed by intestinal metaplasia and dysplasia resulting in the final step of the cascade, gastric cancer⁶⁹.

Thus, glycosylation is involved in gastric cancer carcinogenesis and additionally, 'sweet epitopes have been correlated with several processes in different types of cancer.

3.2 Glycosylation in cancer

As previously described, glycans are involved in cell-cell communication, signal transduction, protein folding and stability^{55,70}. So, it is comprehensible that aberrant glycosylation is one of the frequent features during carcinogenesis and glycans are involved in many pathophysiological steps such as regulating tumor cell proliferation, invasion, metastasis, angiogenesis^{13,15,71,72}.

Nowadays, new technologies have been developed to deeper study glycans and glycoproteins in cancer, since aberrant glycosylation is considered a hallmark of cancer. Most glycan alterations are, apart from several others, result of up or down regulation of enzymes like glycosyltransferases or mutations in Cosmc chaperone protein^{10,22,73,74}. These modifications are regulated at both posttranscriptional and posttranslational level due to epigenetic and genetic changes which result in expression of cancer associated antigens^{75,76}.

Among some of the alterations observed between normal and tumor tissues, the most relevant ones include increased of β 1-6GlcNAc branching in N-glycans due to overexpression of GlcNAc transferase V (GnT V), truncation of O-glycans leading to the appearance of Tn and sialyl Tn antigens and overexpression of sialylated Lewis structures on O- or N-linked glycoproteins¹⁰.

Many studies showed that high expression of these antigens correlate with cell growth, enhanced capacity of proliferation, invasion and metastasis of tumor cells. Regarding O-glycosylation, sialyl Tn appears to be overexpressed in many malignancies including gastric⁷⁷⁻⁷⁹, liver⁸⁰, pancreas⁸¹, bladder⁸², ovarian⁸³, breast⁸⁴, melanoma and colorectal cancers, being most of the times associated with a metastatic stage²⁸. One example for the overexpression of Tn and sialyl Tn was found to be a result of the inactivation of Cosmc gene in several cancers⁷⁶.

Sialyl Tn antigen have been implicated in different mechanisms in cancer such as angiogenesis and invasion^{22,28,76}. The expression of sTn is also observed in gastric precancerous lesions such as gastritis and intestinal metaplasia, suggesting a relation between this antigen and the development of cancer⁸⁵.

N-glycan alterations were also observed in several types of cancer and cancer cell lines⁸⁶⁻⁹⁰. For instance, some reports from our group described the importance of N-glycosylation of E-cadherin and cancer. It was demonstrated that overexpression of GnT V leads to the biosynthesis of highly branched N-glycans which play an important role in tumor invasion and metastasis through decreased cell-cell adhesion and increased cellular motility⁹¹⁻⁹⁴. In opposite, there is also evidence that GnT III and its bisecting GlcNAc structures act as metastases suppressors^{93,94}.

In addition, N-glycans have been correlated with mechanisms of progression such as proliferation⁹⁵, invasion^{92,94}, angiogenesis⁹⁶ and metastasis⁹⁷ in several types of cancer.

Sialylation present both in O- and N-linked glycans appears to be a major contributor to the metastatic cell phenotype and according to recent studies this alteration is a result of unbalance enzymes involved in the biosynthesis or removal of the sialic acid (sialyltransferases and sialidases, respectively)^{15,98-101} (fig. 5). Thus, overexpression of sialyltransferases in tumor tissues results in increased sialylated epitopes such as sLe^x, sLe^a, STn, and is generally associated with tumor progression and poor prognosis^{39,40,102}. Another evidence of the correlation between sialyltransferases and cancer is the fact that proto-oncogenes as Ras and c-myc have been shown to control transcription of these enzymes (such as ST6Gal I, ST3Gal I, II and IV)^{29,103,104}.

Overexpression of sialoglycans has been correlated with metastatic and invasive phenotype underlying tumor aggressiveness and poor prognosis of cancer patients¹⁰⁵. Some examples are the overexpression of sTn detected in different types of cancer^{106,107}, as well as sialylated Lewis antigens, sLe^a and sLe^x which have been correlated with metastatic phenotype in a wide range of cancers, including gastric, colon, renal, prostate, lung and pancreatic cancer¹⁰⁸⁻¹¹⁴. High levels of α 2,3- linked sialic acid residues were demonstrated to be associated with metastatic potential of gastric cancer cells¹¹⁵. In a previous study, our group demonstrated that overexpression of ST3Gal IV in a gastric cancer cell line induces sLe^x expression and was associated with increased invasive capacity of cancer cells both in vitro and in vivo³⁹.

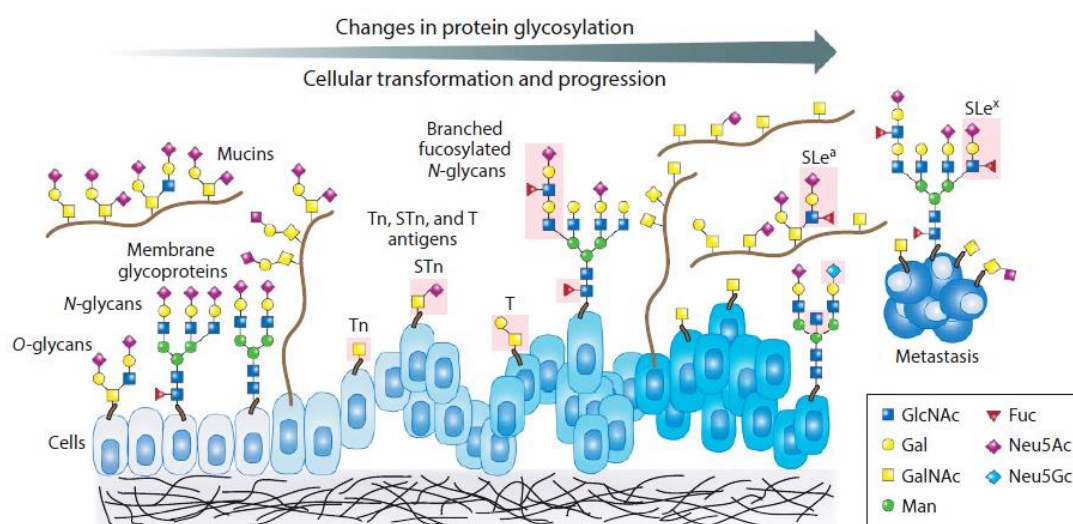


Fig. 5 – Changes on protein glycosylation observed during cancer progression¹¹⁶. Appearance of aberrant truncated glycans T, Tn and STn antigens observed in O-glycans, and increased expression of branched, sialylated and fucosylated glycans on N-glycans. Special emphasis to sLe^a and sLe^x sialoglycans present in both N-and O-glycans. These glycosylation alterations play important role in cancer cell dissemination and metastization.

During the metastatic process (fig. 6), cancer cells are able to modulate the expression of many genes in order to interact with other cell types such as fibroblasts, platelets, leukocytes and endothelial cells¹¹⁷. These alterations are involved in a process called epithelial-mesenchymal transition (EMT) and some reports also described alterations in glycosylation during this process^{92,118-120}. For instance, along with loss of epithelial markers towards a mesenchymal phenotype, overexpression of sialoglycans have been described as playing a role in cell detachment from the primary tumor mass through charge repulsion¹²¹⁻¹²³ (fig. 6A). For instance, in the particular case

of breast cancer, the appearance of sLe^a and sLe^x is associated with lymph node metastases and a decrease in the expression of the epithelial marker, E-cadherin, a glycoprotein associated with EMT¹²⁴. Since malignant transformation is usually accompanied by the expression of oncofetal antigens as sLe^x, it is thought that the observed glycosylation alterations contribute to tumor cells to exploit some events that naturally occur during development. These events include modulation of receptor activation, cell interactions, adhesion and motility all correlated with increased invasive properties of tumor cells^{22,125}.

Sialyl lewis antigens like sLe^x seem to be one of the tumor cell's secret 'weapons'. Studies suggest that sLe^x and sLe^a expression in cancer cells, mimick their normal expression on blood cells, simulating the defensive process of migration through binding to endothelial selectins during the inflammatory process^{6,126,127}. As described in chapter 2, ST3Gal IV is critical for the biosynthesis of E-selectin ligands such as sLe^x. Some reports have been demonstrating CD44 as a major E-selectin ligand¹²⁸, more specifically a CD44 glycoform containing sialylated, fucosylated binding determinants on *N*-glycans¹²⁹. Furthermore, some reports claim that CD44 variants and also carcinoembryonic antigen (CEA) express sialofucosylated structures ligands of E- and P-selectin in colon, prostate and pancreatic metastatic tumor cells¹³⁰⁻¹³² (fig. 6B). This binding facilitate extravasation of cancer cells from blood vessels to the bordered tissue contributing for metastatic spread, correlating with tumor aggressiveness and poor patient survival^{6,21,28,102,109,122,133,134}. Altogether these reports lead to consider that ST3Gal IV can play a role in the synthesis of *N*-glycan sialylation in these glycoproteins leading to the synthesis of E-selectin ligands.

In conclusion, all of these evidences described the important role of sialylated glycoproteins during cancer progression.

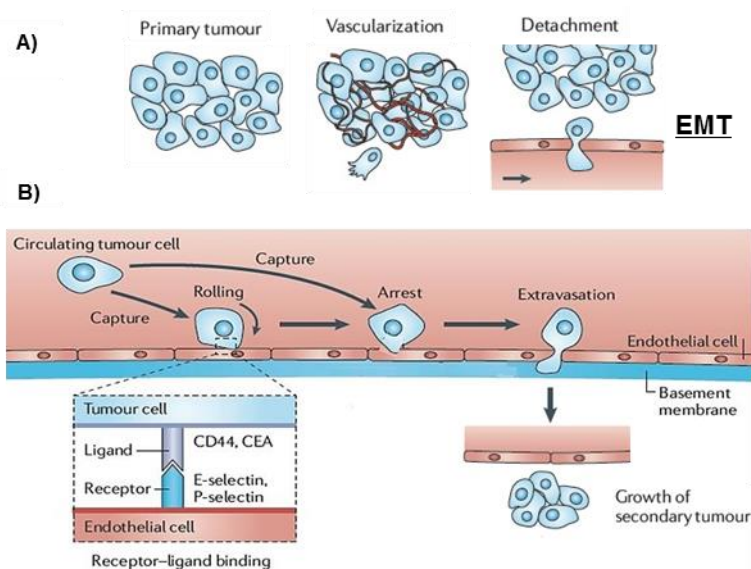


Fig. 6 – Metastatic process of cancer cells. (A) In a vascularized primary tumor, many alterations occur such as expression of sialoglycans and loss of epithelial markers resulting in increased motility of cells. This additional characteristic modulates the invasive potential of tumor cells and contributes for intravasation in the vascular system. (B) Upon entering in the circulation, tumor cells can adhere to endothelial cells on blood vessel, role and extravasate through binding of sialoglycans as sLe^x present on cancer cell's surface glycoproteins such as CD44 and CEA. This

process allows cancer cells to migrate through blood circulation reaching target tissues where they can colonize forming secondary tumors (modified from ¹²²)

3.2.1 Glycoconjugates in cancer

As described above, alterations on glycosylation are a feature of cancer, and the study of the glycoconjugates carriers of these alterations can bring important information regarding cancer development and progression.

CD44 is a highly glycosylated protein that has been described as playing an important role in many cancers¹³². CD44 represents a polymorphic group of proteins all encoded by a single and highly conserved gene where the transcripts are subject to alternative splicing¹³⁵. This regulation consists on the selection of exons or combinations of exons from a panel of ten variant ('v') exons (v1-10) of the primary transcript and final inclusion in the CD44 mRNA (representation in material and methods chapter, fig. 8). The complexity of the CD44 protein products is a result of the alternative splicing, producing a wide range of variant combinations. In addition, these different CD44 variants can be further altered with post-translational modifications, like glycosylation, given rise to glycoforms that can vary from 80 to 200 kDa^{136,137}.

Alterations on CD44 expression have been well-documented and associated to poor prognosis in many cancers¹³⁸⁻¹⁴¹. In gastric cancer, studies based on tissues expression revealed differences in standard CD44 and its variants, namely CD44v6, and these expression profiles are associated with tumor differentiation, progression and metastases¹⁴²⁻¹⁴⁵. Regarding CD44v6, it has been described as a major variant present in cancer. Its expression has been described as playing a role in many different signaling pathways that confer aggressive properties to tumor cells. For instance, it has been shown that CD44v6 can act as a co-receptor for the activation of Met, a receptor involved in proliferation and invasion^{118,146}. Thus, as described above, not only CD44 specific glycoforms can favor EMT by interacting with selectins (fig. 6B), but also can be involved in signaling pathways involved in tumor progression, survival and invasion^{118,146}.

In order to tumor cells metastasize, they need first to local invade the surrounding tissue to reach and entry the vasculature. During this process of local tissue invasion, tumor cells need to interact with extracellular matrix (ECM) components. In cancer, ECM is commonly deregulated and becomes disorganized contributing to tumor cell progression¹⁴⁷. One of the major components of extracellular matrix is perlecan, a proteoglycan also widely expressed in all basement membranes^{148,149}.

Perlecan is a heparan sulfate proteoglycan (also known as HSPG2) which consists in a core protein (>400 kDa) divided in five different domains where three long chains of heparan sulfate (glycosaminoglycan) can be attached to N-terminal (domain I). In addition, four domains except the domain II, can also undergo glycosylation showing multiple places with N- and O-glycans^{150,151}.

Perlecan is overexpressed in several carcinomas, including melanoma, breast, colon and prostate¹⁵²⁻¹⁵⁵. This proteoglycan is known to be involved in cell migration, adhesion and proliferation of cancer cells¹⁵². In addition, there is a dual role of perlecan in angiogenesis, where the different interactions can result in an inhibition or promotion of this process.

Thus, alterations in perlecan glycans during oncogenesis and tumor progression could regulate interactions between this protein and a variety of growth factors, regulating signaling pathways (leading to different functions)^{152,156}.

4. Tumor biomarker

Nowadays, most of the serological assays used in the clinic rely in the detection of glycoproteins or glycans expressed and shed by tumor cells into circulation⁷⁴.

In gastric cancer, some serological markers are used in the clinic such as carbohydrate antigen 19-9 (CA19-9), carbohydrate antigen 125 (CA125), carbohydrate antigen 72-4 and CEA, but their detection is mainly used for tumor recurrence and patients follow up¹⁵⁷⁻¹⁶¹. The potential of many other biomarkers have been claimed, however their implementation in the clinical settings have not been successful. As example, increased levels of sLe^x have been reported as a useful tumor marker in gastrointestinal carcinoma, but it is not used in clinical practice¹⁶². Despite the usefulness of the biomarkers described above, clinically there are no reliable tumor biomarkers for gastric cancer; for instance, increased levels of CEA in serum can also have non-cancer related causes¹⁶³. Besides this, patients are currently diagnosed in an advanced stage of the disease and can't undergo surgery which is the best way to treat gastric cancer. Thus, there is a need for the development of new sensible and specific biomarkers that allow early detection of gastric cancer and also new targets for therapy¹⁶⁴.

4.1 Carcinoembryonic antigen (CEA)

Carcinoembryonic antigen is a 170-180 kDa epithelial glycoprotein, which is expressed during the development of gastrointestinal tissues and also in cells of normal colon mucosa¹⁶⁵. In cancer, tumor cells usually secrete CEA into circulation allowing its detection. Thus, in clinical practice, CEA serum level is used as a biomarker for many cancers such as colon and gastric¹⁵⁷⁻¹⁶⁰.

CEA is a highly heterogeneous glycoprotein which contains around 60% carbohydrate, with 28 possible N-glycosylation sites^{166,167}. Thus, most of the molecular weight of the CEA molecule (more than 50%) consists in branched oligosaccharides containing sialic acid, fucose, galactose, mannose and N-acetylglucosamine¹⁶⁸.

CEA belongs to a group of 12 carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) which is included into the immunoglobulin superfamily. For historical reasons, the term CEA is used when it refers to the CEACAM5 gene and it is also known as CEACAM5 or CD66e^{169,170}. Epithelial CEACAMs have been implicated in cell adhesion, regulation of cell matrix attachment¹⁷¹, cell polarization¹⁷² and in intra and intercellular signaling¹⁷³. In cancer, CEACAMs have been described as playing a

role in many pathological functions like angiogenesis¹⁷⁴, cell proliferation¹⁷⁵, cell motility^{176,177}. For instance, it was reported that CEA overexpression could protect tumor cells from apoptosis¹⁷⁸. In addition, CEA was also correlated with an enhanced metastatic potential, as a carrier of E- and L-selectin ligands¹⁷⁹ and a recent study revealed cooperation with CD44 variant isoforms in colon carcinoma cell adhesion to the same selectins¹³¹.

In 1965, Gold and Freedman reported CEA as a novel tumor marker¹⁸⁰. In fact, CEA is overexpressed in nearly all colorectal cancers in comparison with low-levels of expression in gastrointestinal normal mucosas. It is also overexpressed in lung, breast and gastric cancer and correlated with a poor clinical outcome^{165,169,181}. So, in case of cancer, CEA is expressed at high levels and due to loss of cell polarity and this glycoprotein is then shed into the circulation¹⁸². Thus, the identification of CEA serum levels is used for evaluation of prognosis and postoperative surveillance of cancer patients¹⁵⁷⁻¹⁶⁰.

4.1.1 hnRNPM-CEA, a key molecule for the metastatic process?

The main cause of cancer-related death of both gastric and colon cancer is the presence of distant metastasis, mainly in the liver, at the time of diagnosis¹⁸³. High levels of CEA in serum have been associated with advanced stages of these cancer types where metastasis can be found^{183,184}.

Serum CEA levels are controlled through its production rate by the tumor and also elimination which is carried out by Kupffer cells in the liver^{185,186}. The interaction between CEA and Kupffer cells is possible through specialized receptor, a 80 kDa cell surface protein.

In fact, it has been suggested that CEA molecules released from colorectal tumor cells can enter the blood stream and interact with Kupffer cells' receptor, resulting in the release of cytokines such as IL-10 or tumor necrosis factor (TNF- α). This release results in of adhesion molecules such as E-selectin which favors the adhesion and invasion of tumor cells¹⁸⁷ (fig. 7).

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

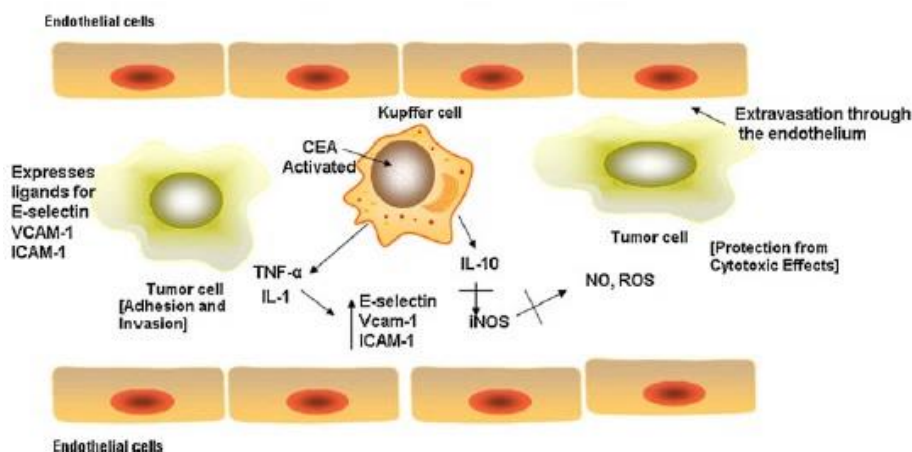


Fig. 7 – Interaction of tumor cells and Kupfer cells in the liver¹⁸⁸. Interaction of CEA with receptor present in Kupfer cells in the liver, leads to the release of cytokines such as IL-1, IL-6, IL-10 and TNF- α resulting in the up-regulation of adhesion molecules such as E-selectin which favors the adhesion and invasion of tumor cells.

Recently, *Bajenova et al* identified this receptor as hnRNPM4¹⁸⁹. HnRNPM4 is related to the hnRNP family of RNA-binding proteins with important roles in alternative splicing and transcriptional regulation that can be found in both epithelial and mesenchymal cells¹⁹⁰. These proteins are mostly expressed in the nucleus, but hnRNPM4 is one of the few exceptions which can be found as a membrane bound-receptor¹⁸⁸.

Besides CEA-hnRNPM4 interaction has been described in the activation of signaling pathways and associated with tumor invasion, a recent study demonstrated that hnRNPM is also involved in the alternative splicing of CD44 during EMT resulting in the promotion of tumor metastasis¹⁹¹. These reports point for an important role of CEA-hnRNPM interactions in the establishment of liver metastasis.

Taking this literature review in consideration, Glycobiology seems to be a piece of the mazy puzzle called ‘therapy of cancer’. Glycosylation alterations should be explored aiming the development of new serological biomarker assays for cancer detection.

AIM

Previously, our group demonstrated that a gastric carcinoma cell line transfected with a gene encoded for the enzyme ST3Gal IV leads to the expression of sLe^x glycan structures when compared to Mock cells. The expression of sLe^x resulted in an increased invasive capacity both *in vitro* and *in vivo* (CAM model) and was associated to increased activation of c-met signaling pathway³⁹. In addition, it was observed that expression of ST3Gal IV also leads to alterations in the expression of some proteins that could be involve in the invasive phenotype. Furthermore it was identified, using mass spectrometry, CEA as the main carrier of that glycosylated structure.

Therefore, this study aims to understand the role of CEA glycosylation, more specifically the presence of sLe^x structures in gastric carcinoma. In addition, this study also aims to the identification and glycan characterization of the altered expressed proteins present in sLe^x expressing cells.

To accomplish the general aim, this study will focus in the following specific objectives:

- To characterize the glycan structures present in CEA molecules from sLe^x expressing cells compared to Mock cells;
- To assess the expression of CEA, sLe^x (*in vitro*) and some evidences *in vivo*;
- To assess the role of secreted sLe^x expressing proteins in the activation of c-Met receptor;
- To identify the altered expressed proteins present in sLe^x expressing cells compared to Mock cells;
- To evaluate the presence of sLe^x structures in the identified proteins from sLe^x expressing cells;
- To evaluate the tumorigenic capacity of the sLe^x expressing cells in an *in vivo* mice model;
- To assess the expression of the CEA receptor, hnRNP M4, in sLe^x expressing cells and Mock cells.

MATERIAL AND METHODS

Cell culture

The gastric cancer cell line MKN45 was obtained from the Japanese Cancer Research Bank (Tsukuba, Japan) and was previously stably transfected with a full length human gene for ST3GAL4 (MST3Gal IV) and the empty vector pcDNA3.1 (Mock) (according with *Carvalho et al*⁴⁰)

Cells were grown in monolayer in T75 cm² culture flasks and cultured in Roswell Park Memorial Institute (RPMI) 1640 GlutaMAX, HEPES medium, supplemented with 10% fetal bovine serum (FBS) (Biowest), 1% penicillin-streptomycin (P/S) (Gibco) in the presence of 0.5 mg/mL of the selective antibiotic, G418 (Invitrogen). Cell culture was maintained at 37°C in a 5% CO₂ atmosphere and culture medium was replaced every two days. Whenever cells reach 90 to 100% confluence they were trypsinized and replaced in new culture flasks to maintain the culture. Cells were also stored at -80°C in storage medium, 90% FBS and 10% Dimethyl sulfoxide (DMSO) (from Applichem) after trypsin.

Secretome collection

In order to obtain secretome of the cells, selective medium was replaced with RPMI without FBS (conditioned medium) when cells were at high confluence. After 24 hours, the conditioned medium was collected and centrifuged at 1500 rpm for 10 minutes in order to pellets cells that could be present in the medium. Then, 10 kDa cutoff filters (Millipore) were used to concentrate secretome samples, and samples were stored at -20°C.

Protein extraction

Cells at high confluence were washed with PBS and incubated with NP40 lysis buffer (1% NP40), 20mM Tris-HCl pH 8.0, 137mM NaCl, 10% glycerol, 2mM EDTA (all from Sigma) containing a cocktail of inhibitors: 1 mM sodium orthovanadate, 1 mM phenylmethylsulfonyl fluoride and protease inhibitor cocktail tablet (Roche)) . After 10 minutes of incubation, cells were scrapped, centrifuged at 13200 rpm 10 minutes at 4°C, supernatant collected and stored at -20°C.

Protein extraction from tissues was performed using RIPA lysis buffer (150mM NaCl, 1% NP40, 0.5% sodium deoxycholate, 0,1% sodium dodecyl sulfate (SDS), 50 mM Tris pH 8.0 (all from Sigma) with addition of a cocktail of inhibitors). After several

washes with cold PBS 1x, three gastric tumors (between 144-166 mg) and one colon biopsy (about 18 mg) were incubated with lysis buffer, the mixture was homogenized with a pistol and left in ice for 30 minutes. After this, the tubes were centrifuged at maximal rotation for 10 minutes at 4°C and the supernatants collected were stored at -20°C.

SDS-PAGE and Western blotting analysis

Protein concentration from cell extracts and secretome was determined accordingly to Protein Assay (BIORAD), and 50 µg of total proteins were loaded on 7.5% tris-glycine SDS-polyacrylamide gel (table 1). Gels were run at a constant voltage of 125V.

Table 1 – Composition of 7.5% tris-glycine SDS-polyacrylamide gel.

	Resolving Gel 7.5% (V _r =10 mL)	Stacking Gel 4% (V _r =5 mL)
MilliQ H ₂ O	4.8 mL	2.95 mL
30% Acrylamide (37.5;1) (SERVA)	2.5 mL	0.67 mL
1.5M Tris (pH 8.8 with HCl) (BIORAD)	2.5 mL	1.28 mL
10% SDS (National diagnostics)	100 µL	50 µL
10% Ammonium Persulfate (APS) (Sigma)	100 µL	50 µL
TEMED (SIGMA)	10 µL	5 µL

After electrophoresis, gels were stained with Coomassie blue or transferred to nitrocellulose membranes (Amersham) using a semi-dry system.

The efficiency of protein transfer was monitored using Ponceaux Red (Sigma) and membranes were washed with PBS 0.05% Tween 20 (PBST), followed by blocking with 5% BSA (Sigma-Aldrich) or 5% milk in PBST (Table 2). After blocking, primary antibodies were incubated overnight at 4°C (table 2).

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Table 2 – List of primary antibodies used with the respective dilutions.

Primary antibodies	Incubation	Dilution
CEA (Cell Signaling)	BSA 5% PBST 0.05%	1:3000
CD44v6 (Invitrogen)		1:1000
sLe ^x (KM93) (Millipore)		1:500
hnRNPM4 (OriGene)		1:3000
p-Met (Invitrogen)		1:1000
p-Src (Cell Signaling)		1:1000
Perlecan (Santa Cruz)	Milk 5% PBST 0.05%	1:2000
Src (Cell Signaling)		1:1000
c-Met (Invitrogen)		1:3000
Actin (Santa Cruz)		1:4000
CD44 (Cell Signaling)		1:1000

Followed antibody incubation, membranes were washed three times with PBST and incubated with secondary antibodies conjugated with horseradish peroxidase in 1% milk or BSA for 1 hour. Membranes were washed three times with PBST and analysis was done by chemiluminescence using ECL Western Blotting detection reagent and films (GE Healthcare).

The secondary antibodies used were: goat anti-mouse IgG₁ antibody 1:25000 (Jackson ImmunoResearch); anti-mouse IgG_{2a} antibody (Jackson ImmunoResearch) 1:25000; anti-mouse IgM antibody (Jackson ImmunoResearch) 1:10000; anti-rabbit antibody (Santa Cruz), all conjugated with horseradish peroxidase.

Immunoprecipitation

Immunoprecipitation was performed in Mock and MST3Gal IV total cell lysates and secretome. Approximately 60 µL of ProteinG Sepharose beads (Sigma) were coupled with the anti-CEA antibody (2µL), anti-CD44 (8 µL), anti-CD44v6 (4 µL) or anti-Perlecan (5 µL) during 2h at 4 °C. Beads were then washed three times with PBS1x and in some cases BS³ crosslink was performed to avoid co-elution of antibody heavy- and light chains. In brief, after beads-antibody incubation, beads were equilibrated three times with BS3 conjugation buffer (20mM of sodium phosphate dibasic dihydrate and 0,15mM of NaCl, pH 7-9, all from Sigma) followed by 30 minutes of BS³ crosslinker incubation. The reaction was stopped by incubation of 12.5 µL of quenching buffer (1M Tris pH 7.5, from Sigma). Beads were washed three times with PBS1x and pre-

-clear protein samples incubated overnight at 4 °C. Pre-clear of protein samples were performed by incubation of 30 µL of protein G sepharose beads with 800 µg of protein samples at 4 °C.

Supernatants of the pre-clear were transferred to the tubes with the antibodies and incubated overnight at 4 °C. In the following day, supernatants were collected (input) and the pellet (IP) was washed three times with PBS1x with 1% Triton X-100 and PBS1x. The IP and input samples were analyzed through western blotting.

Protein N-deglycosylation

Digestion of N-glycans was performed in CEA immunoprecipitated molecules using PNGase F, Endo H and combination of both enzymes (New England Biolabs). After immunoprecipitation, IP samples (beads) were washed five times with PBS1x followed by the addition of 2,5 µL of 10x denaturing buffer to a final volume of 25 µL using deionized water (dH₂O). After denaturation at 100°C for 10 minutes, the reaction mixtures containing the endoglycosidases and respective buffers were added according to table 3 (total final volume approximately of 30 µL). Control samples were performed by adding all the mixture constituents except the enzymes into the CEA immunoprecipitated molecules. Samples were incubated overnight at 37°C with gentle agitation. The results were analyzed through western blotting.

Table 3 – PNGaseF and EndoH enzymatic reaction mixtures for CEA immunoprecipitated samples.

PNGase F	Endo H
3 µL 10x NP40 buffer	-
3 µL 10x G7 buffer	3 µL 10x G5 buffer
1.0 µL PNGase F	1.5 µL Endo H
-	3 µL dH ₂ O

A similar protocol was also applied to 40 µg of total cell lysates and secretome samples. After denaturation, reaction mixtures were added according to table 4 in a final volume of 40 µL, followed by the same procedure as described above.

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Table 4 – PNGase F and Endo H enzymatic reaction mixtures for cell lysates and secretome samples.

PNGase F	Endo H
4 μ L 10x NP40 buffer	-
4 μ L 10x G7 buffer	4 μ L 10x G5 buffer
2 μ L PNGase F	5 μ L Endo H

Immunofluorescence staining

MKN45 transfected cells were seeded and cultured on glass coverslips in 24-well plates (Orange Scientific). After approximately 80% confluency, cells were washed with PBS1x and fixed during 30 min with methanol. After fixation, cells were washed with PBS 1x and blocked using goat serum (Dako) diluted 1:5 in 10% BSA PBS for 30 min at room temperature, followed by overnight primary antibody incubation at 4°C. Then, cells were washed with PBS1x and incubated for 30 min with secondary antibodies goat anti-mouse IgG Alexa 594 and goat anti-mouse IgM Alexa 594 (from Invitrogen) diluted 1:500 in 5% BSA PBS.

Nuclear counter staining was done with DAPI and coverslips were mounted in a microscope slide with vectashield (Vector Labs). Fluorescence was examined in a fluorescence microscopy and images were acquired using a Zeiss Axio cam MRm and the AxioVision Rel. 4.8 software (Carl ZEISS) and in some cases immersion oil was used.

Monoclonal antibodies used: mouse monoclonal KM93 1:60 directed against sLe^x structures and mouse monoclonal antibody anti-hnRNPM4 1:1000.

Condition medium assay

Mock and MST3Gal IV cells were grown in 6-well plate until reached 80-100% confluence and medium was replaced and leaved for 24 hours. After 24h incubation, condition medium from MST3GalIV was collected and incubated in MST3 Mock cells for 24h. After incubation cells were lysate and total cell lysate collected. After protein extraction and quantification, the analysis was possible through western blotting.

RNA isolation

Transfected cells at high confluence were trypsinized and centrifuged at 1200 rpm for 5 minutes. The resultant pellets were washed two to three times with PBS 1x and

transferred to eppendorfs, followed by addition of tri-reagent (Sigma). Cells were then resuspended until the formation of a viscous solution and incubated for 5 minutes at room temperature. For each mL of tri-reagent 200 μ L of chloroform (Sigma) was added and samples were vortexed and leaved at room temperature for 15 minutes. The mixtures were then centrifuged at 12 000 g for 5 minutes at 4°C resulting in the formation of three phases, and the supernatant (upper layer) was collected to a new tube. Approximately 500 μ L of isopropanol (Sigma) was added and incubated for 10 minutes. After incubation, the tubes were centrifuged at 12 000 g during 10 minutes at 4°C, and the resultant pellets were washed two times with 75% ethanol. After air drying, pellets were resuspended in 40 μ L of dH₂O and RNA concentration was measured using nanodrop system.

cDNA synthesis and PCR analysis

RNA was converted to cDNA using the SuperScript® II Reverse Transcriptase (Invitrogen). Reverse transcription was performed in two steps: 1) a mixture of 3 μ g of total RNA, 1 μ L of 40mM dNTPs, 1 μ L of nonspecific random primers and pure dH₂O in a total volume of 12 μ L was incubated at 65°C for 10 minutes; 2) after incubation, 4 μ L of 10x buffer, 2 μ L of dithiothreitol (DTT), 0.3 μ L of RNase out, 0.75 μ L of reverse transcriptase (all from Invitrogen) and 0.95 μ L pure water was added into a final volume of 20 μ L and incubated at 25°C for 10 minutes followed by a cycle of 50 minutes at 42°C and 15 minutes at 70 °C.

For all the different PCR reactions, 1 μ L of cDNA was used. In the case of Perlecan's PCR, six different PCR reactions were done covering 3 different protein domains (domain I, III and V). PCR mixtures contain 1 μ L of cDNA, 2.5 μ L 10x PCR buffer, 0.5 μ L of 10mM dNTPs (Thermo Scientific), 0.5 μ L of both For and Rev specific primers (Sigma) (according with table 5), 0.2 μ L of Taq DNA polymerase (Invitrogen), 0.75 μ L of 50mM magnesium chloride (MgCl₂) and dH₂O into a final volume of 25 μ L. PCR conditions were: hold denaturation cycle of 2 minutes at 95°C, followed by 40 cycles of 1minute at 95°C denaturation step, 1 minute at 60°C annealing step and 90 seconds at 72°C extension step and a final extension cycle of 10 minutes at 72°C.

Perlecan PCR products were separated on a 1% agarose gel.

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

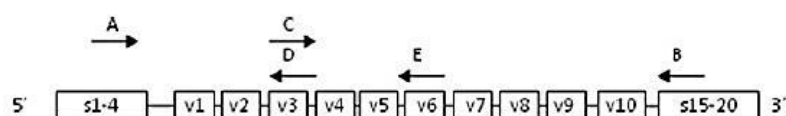
Table 5 – Primers for PCR amplification of Perlecan cDNA (adapted from¹⁹²)

Domain	Exons	Primer sequences
I	3-7	F: 5'-ACCTGGGCAGTGGGGACCTG-3' R: 5'-GCCTCCGTGCAGGCTCTTGG-3'
	2-7	F: 5'-CCATGGGCTGAGGGCATACTG-3' R: 5'-GGCACTGTGCCCAGGCGTCGGAAC-3'
III	29-36	F: 5'-ACCCACCTGTGATGCGTGCTC-3' R: 5'-GTGGCCCGGATCAGGAGCTCAT-3'
	35-37	F: 5'-CGATGTGCAGATCAGGGC-3' R: 5'-CTGGCATTGCGAGCAGG-3'
V	87-97	F: 5'-GTGTGACAGTGACCACCCCT-3' R: 5'-CTACGAGGGGCAGGGGCGT-3'
	87-94	F: 5'-GTGTGACAGTGACCACCCCT-3' R: 5'-CTGAAGACAAGGTGCCCCG-3'

For CD44, five different PCR were performed using primers flanking specific variants within CD44 transcripts (Figure 8). PCR mixture were composed of 1 μ L of cDNA, 2.5 μ L of 10xPCR buffer, 1 μ L 10mM dNTPs, 0.5 μ L of both For and Rev primers (Sigma) (according with table 6), 0.2 μ L of Taq DNA polymerase (Invitrogen), 0.75 μ L of $MgCl_2$ and dH_2O into a final volume of 25 μ L. Primer combinations were the follow: A-B, A-D, A-E, C-B, C-E (see fig. 8). PCR conditions were: hold denaturation cycle of 2 minutes at 94°C, followed by 40 cycles of 1 minute at 95°C, 1 minute at 60°C, 90 seconds at 72°C and a final extension cycle at 72°C for 10 minutes.

Table 6 – Primers for PCR amplification of CD44 cDNA (adapted from ¹⁹³)

Primer	Primer sequences
A	Standard 5'-AGTCACAGACCTGCCCAATGCCTTT
B	Standard 3'-TTTGCTCCACCTTCTTGACTCCCATG
C	v3 exon 5'-GGGAGCCAAATGAAGAAAATGAAGATGAAAG
D	v3 exon 3'-GGTGCCTGTCTCTTTCATCTTCATTTTCTTCATTT
E	v6 exon 3'-TCTGTTGCCAAACCACTGTTTCCTTCTG

Fig. 8 – CD44 specific primers (adapted from ¹⁹³).

CD44 PCR products were separated on 3%, 2% and 1% agarose gel and a control cell line (colo 205) was also used. PCR products from CD44 and Perlecan were resolved in agarose gel with Invitrogen 1Kb ladder and the detection was done using nucleic acid staining solution RedSafe™ (iNtRON Biotechnology). The results were observed using a BIORAD Molecular Imager® ChemiDoc™ XRS.

In vivo tumorigenic mice model

Male athymic nude mice (N:NIH(S)II-nu/nu) were reproduced, maintained, and housed at IPATIMUP Animal House facility at the Medical Faculty of the University of Porto. Animal experiments were carried out in accordance with the European Guidelines for the Care and Use of Laboratory Animals, Directive 2010/63/UE and the National Regulation published in 2013 (Diário da República, 1.^a série — N.º 151 — 7 de agosto de 2013). In all experiments we established the following Humane Endpoints for euthanasia: any signals of distress, suffering or pain, body weight loss greater than 20-25% of the body mass, anorexia and moribund state, related or not to the experimental procedure.

Athymic nude mice, aged 9 weeks, were orthotopically inoculated with 1×10^7 viable Mock cells (n=5) and MST3Gal IV cells (n=6) using a 25 gauge needle. Mice were monitored for development of primary xenograf tumours: mice were weighed, tumour size was measured using callipers, and tumour volumes (mm^3) were estimated using the formula: $W \times L^2 \times \frac{1}{2}$, where W is the width and L is the length of the tumour. In order to extend mice life and promote metastization, tumours of the two mice groups were surgically removed when tumors reached a mean volume of $\sim 2000\text{-}2500\text{mm}^3$. Mice were anesthetized intraperitoneal with 100 μL of a mixture containing 0.5 mg/kg of ketamin and 1 mg/kg of medetomidine hydrochloride to remove the tumour tissue, trying, when possible, to excise all tumour. To revert the effect of anesthesia, 2.5 mg/kg of Revertor® were used per mice. Mice were treated with an oral solution of 10 mg/kg Tramal, every 8h for 24-48h to avoid pain. At the time of surgeries, tumours were collected and fixed 10% neutral buffered formalin. All mice were euthanized according to the established Humane Endpoints and necropsies were made in all animals. Tumor masses and the organs (lymph nodes, lung and liver) were removed and fixed in 10% neutral buffered formalin. After dehydration and embedding in paraffin wax, sequential 3 μm sections were cut from each block. One section was stained with haematoxylin and eosin (H&E) for routine histological examination and subsequent sections were used for immunohistochemical studies.

RESULTS

1. Characterization of CEA glycosylation *in vitro* and in clinical samples

1.1 Characterization of CEA glycosylation in glycoengineered gastric carcinoma cells

Previously, our group demonstrated that the expression of a specific sialyltransferase, ST3Gal IV, leads to the expression of sLe^x glycan structures in the gastric carcinoma cell line MKN45.

In this work, the expression of sLe^x was assessed in the different cell lines, MST3Mock and MST3GalIV. Western blot analysis confirmed that only MST3GalIV cell line expresses sLe^x glycan structures (fig. 9A). Further, it was evaluated the expression of sLe^x in cells cultured in conditioned medium (medium without Fetal Bovine Serum) and in secreted proteins. With this, it was observed that MST3GalIV cultured in condition medium as well as MST3Gal IV secreted protein express sLe^x glycan structures (fig. 9B).

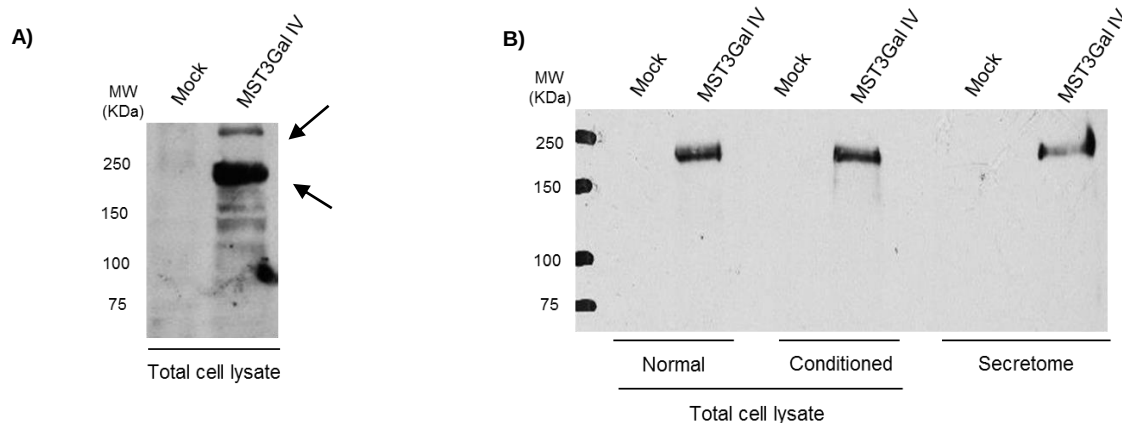


Fig. 9 – SLe^x expression in gastric carcinoma cell line stably transfected with ST3Gal IV. A) SLe^x expression was observed in MST3Gal IV cells compared with Mock cells. Western blot analysis show the presence of two major bands (black arrows) for sLe^x expression when membranes were exposure during 20 minutes. B) MST3Gal IV cells express sLe^x antigen in total cell lysate cultured both in the presence and absence of fetal bovine serum, and in secreted proteins.

In order to disclose the nature of glycan type of sLe^x structures, it was performed enzymatic digestion using PNGase F and Endo H treatment in total cell lysates and secretome from both Mock and MST3GalIV. After enzymatic treatment sLe^x expression was assessed by western blot analysis. PNGase F and Endo H treatment resulted both in a shift in the bands positive for sLe^x when compared to no treatment (fig. 10). Very

similar results were obtained for both total cell lysates and secretome samples, with PNGase F treatment leading to a higher difference in the molecular mass shift.

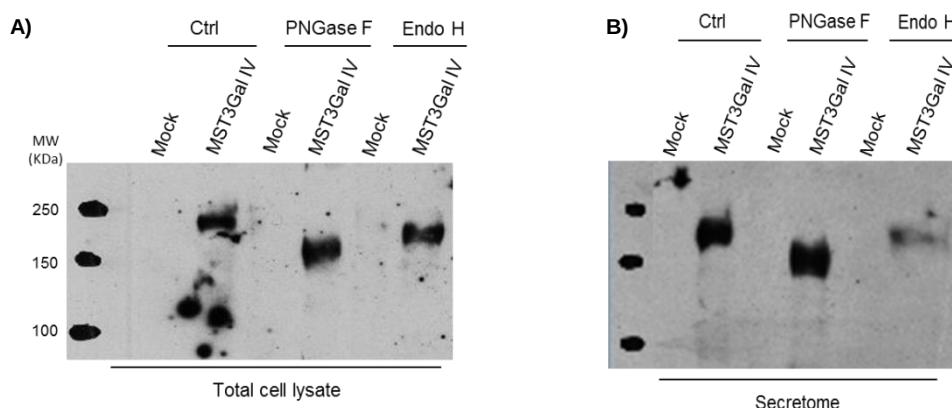


Fig. 10 – SLe^x expression upon PNGase F and Endo H enzymatic digestion. A) SLe^x expression in total cell lysates upon enzymatic treatment showing a shift in the molecular weight for sLe^x bands. B) SLe^x expression in secretome samples upon enzymatic treatment showing also a shift in the molecular weight for sLe^x bands.

Previously, our group observed, by western blot, that sLe^x protein carriers are high molecular weight proteins. In order to identify the putative sLe^x protein carriers, total cell lysates were separated in SDS PAGE and glycoproteins were stained using a periodic acid-Schiff (PAS) (fig. 11A). This method specifically detects glycosylated proteins which contain sialic acid in its composition.

In parallel western blot analysis of sLe^x was performed and the highlighted bands (fig. 11B) were matched to the stained gel. Two

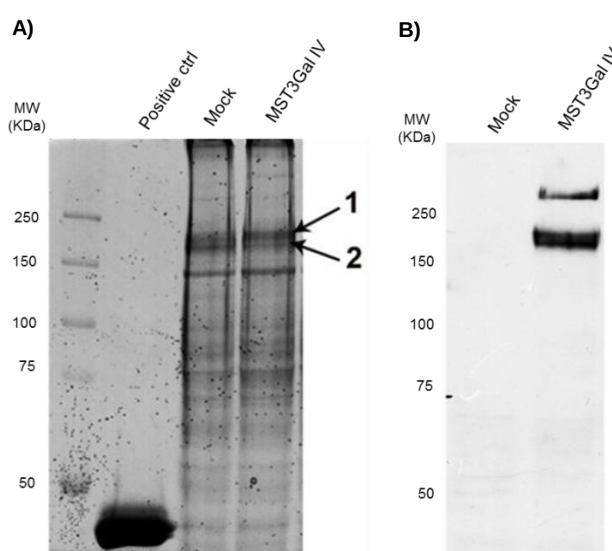


Fig. 11 – Glycoprotein analysis in Mock and MST3Gal IV cells. A) PAS stained gel with arrows 1 and 2 which represent the selected proteins for mass spectrometry analysis; Horseradish peroxidase was used as positive staining control. B) SLe^x expression exclusively on MST3Gal IV cells.

bands were removed from the gel (arrow 1 and 2 figure 11A) and protein identification was performed using MALDI-TOF mass spectrometry analysis.

Mass spectrometry analysis identified two proteins as putative sLe^x carriers in MST3Gal IV cells (table 7). From the two identified proteins, Carcinoembryonic antigen (CEA or CEACAM5) was the only protein that has been described has being glycosylated and the most promising protein regarding cancer biomarkers.

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Table 7 – Proteins identified by mass spectrometry regarding sLe^x antigen detection by western blot.

Spot ID	Protein description	Accession number	MASCOT Protein C.I. %	Peptide count	% cov	Peaks matched	MOWSE score
1	Carcinoembryonic antigen-related cell adhesion molecule 5	CEAM5_HUMAN	99	4	8	5	92
	Ras GTPase-activating-like protein IQGAP1	IQGA1_HUMAN	100	39	21	40	268
2	Carcinoembryonic antigen-related cell adhesion molecule 5	CEAM5_HUMAN	100	5	10	7	234

The presence of CEA molecules in both cell lines was then assessed by western blot analysis in different conditions: total cell lysates both in normal and conditioned medium, and secretome. According to the results, CEA is expressed in the analyzed conditions in both cell lines (fig. 12). In addition, it was possible to notice a slightly difference in the molecular weight of CEA in the different cell lines, with MST3GalIV presenting a slight increase in the molecular weight that can be due to the expression of sLe^x.

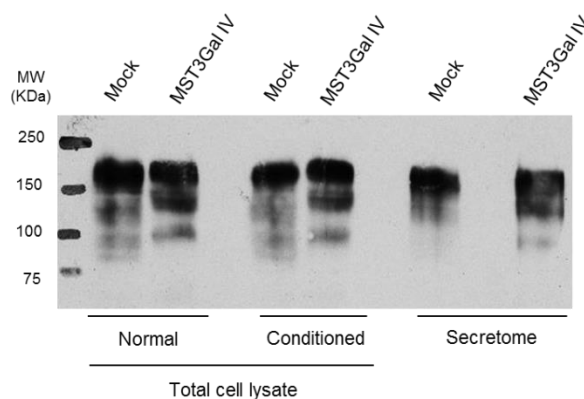


Fig. 12 – Expression of CEA glycoprotein in both Mock and MST3Gal IV cells in different conditions. CEA is expressed in both total cell lysates in normal and conditioned medium as well as in secretome.

In order to validate that CEA is a carrier of sLe^x, it was performed immunoprecipitation of total cell lysates from Mock and MST3Gal IV cells followed by western blot analysis for sLe^x antigen. Figure 13A shows the expression of CEA upon immunoprecipitation demonstrating the efficiency in the immunocapturing of CEA molecules from total cell lysates. In addition, sLe^x western blot analysis of CEA immunoprecipitates demonstrated the presence of a band in MST3Gal IV immunoprecipitation (fig. 13B). This band appeared to be in the same molecular weight as sLe^x in total cell lysates (fig. 13B).

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

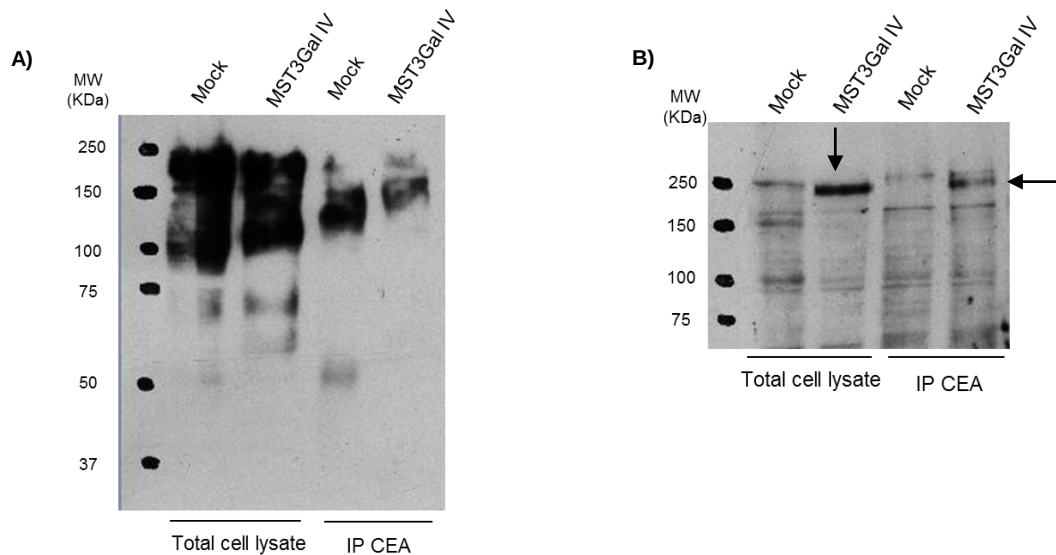


Fig. 13 – Western blot analysis of CEA and sLe^x in CEA immunoprecipitates. A) CEA expression upon its immunoprecipitation of both MST3Mock and MST3Gal IV cell lysates, demonstrating that CEA molecules were successfully immunoprecipitated, B) SLe^x expression of CEA immunoprecipitates from Mock and MST3Gal IV cell lysates, confirming the presence of sLe^x in CEA molecules from MST3Gal IV cells. Arrows indicate the sLe^x band corresponding to a CEA glycoform in MST3Gal IV lysates.

Furthermore, and in order to characterize the sLe^x structural glycan type within CEA molecules, it was performed two different enzymatic glycan digestions: 1) PNGase F which hydrolyzes the amide bond of the asparagine residue to which the glycan is attached, removing all types of N-glycans from peptides and proteins, and 2) Endo H which cleaves the bond connecting the two GlcNAc group in asparagine-linked mannose rich oligosaccharides, but not complex oligosaccharides.

CEA immunoprecipitates were enzymatically digested with PNGase F and Endo H followed by western blot analysis of CEA and sLe^x. Regarding CEA expression, after PNGase F treatment, it was observed a significant reduction in CEA molecular weight from both cell lines characterized by a large shift in the western blot bands (fig. 14A). The same result was also observed when CEA immunoprecipitates were treated with both PNGase F and Endo H enzymes. In samples treated with Endo H enzyme, it was not observed any differences in CEA molecular weight in both cell lines when compared with the control immunoprecipitates (fig. 14A).

Regarding sLe^x expression, it was observed the expression of sLe^x in cell lysate (full arrow in fig. 14B), in CEA immunoprecipitates from MST3Gal IV (arrow 1 in fig. 14B) and no expression in Mock lysates. Furthermore, it was observed, in CEA immunoprecipitates from MST3Gal IV treated with PNGase F, a shift in the sLe^x band to a lower molecular weight (arrow 2 in fig. 14B). This result correlates with the CEA shift after PNGase F and the two endoglycosidases treatment in CEA western blot (figure 14A). In immunoprecipitates treated with Endo H enzyme, it was not found any

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

differences in sLe^x expression when compared with control immunoprecipitates (arrow 3 in fig. 14B). Moreover, a similar result to PNGase F treated immunoprecipitates was observed in immunoprecipitates treated with the two enzymes showing the expression of sLe^x at a lower molecular weight in MST3Gal IV (arrow 4 in fig. 14B).

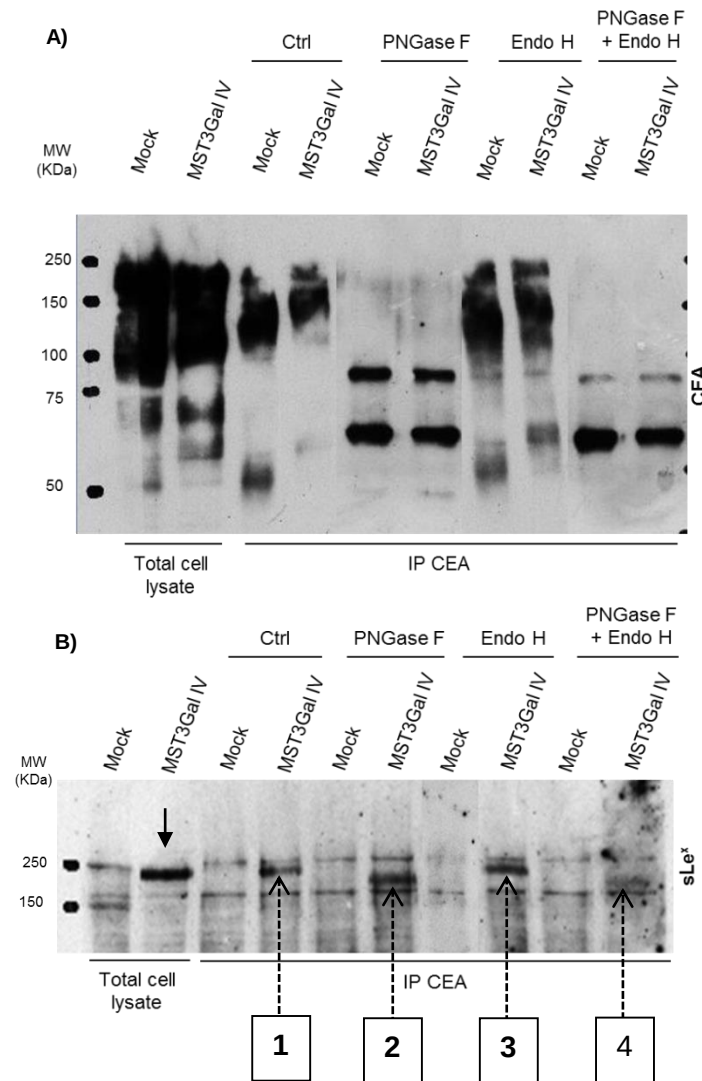


Fig. 14 – Glycan characterization of CEA immunoprecipitated molecules using PNGase F and Endo H enzymatic digestion. A) CEA western blot analysis upon enzymatic digestion. Evidence of a large shift on CEA molecular weights upon PNGase F treatment in immunoprecipitates from both cell lines. B) Western blot analysis of sLe^x on CEA immunoprecipitated samples upon enzymatic digestion. Full arrow indicates the presence of sLe^x on total cell lysate from MST3Gal IV. Arrow 1 shows the presence of sLe^x in CEA immunoprecipitates, demonstrating the presence of CEA/sLe^x molecular complex in MST3Gal IV cells; arrow 2 indicates the sLe^x shift to a lower molecular weight in MST3Gal IV after PNGase F treatment; arrow 3 demonstrate no differences in the molecular weight of sLe^x expression upon Endo H treatment; arrow 4 indicates a sLe^x shift to a lower molecular weight in CEA immunoprecipitates from MST3Gal IV upon treatment with the two enzymes, showing a similar result to PNGase F treatment.

CEA western blot analysis upon PNGase F treatment in immunoprecipitates from both cell lines: incubation of the membranes with ECL PLUS showed a slight band for MST3Gal IV cells which could correlate with the shift in the sLe^x on CEA immunoprecipitated samples upon PNGase F treatment to a lower molecular weight.

1.2 Characterization of CEA glycosylation in clinical samples

Previously, our group screened 31 gastric carcinoma tissues for the expression of sLe^x, CEA and the presence of the CEA/sLe^x molecular complex. In that screening it was observed that all (100%) gastric carcinoma tissues showed CEA expression and 26 (83.9%) showed sLe^x expression. The evaluation of CEA/sLe^x molecular complex assessed by proximity ligation assay (PLA) demonstrated all cases (100%) that were positive for sLe^x were also positive for the CEA/sLe^x molecular complex. Very interestingly, the expression of CEA/sLe^x molecular complex was statistical significantly associated with venous invasion of the tumors (supplementary materials and data).

Taking all these findings in consideration, it was hypothesized that CEA molecules in gastric carcinoma tissues also present sLe^x structures in their composition. In order to prove this observation, it was performed protein extraction from frozen gastric carcinoma (GC) tissues and assessed CEA and sLe^x expression by western blot analysis. Colon biopsy was used as control for the CEA expression. The western blot analysis demonstrated that all the analyzed tissues express CEA molecules (fig. 15A). In addition, CEA pattern of expression vary in the different analyzed samples as can be seen by a different band patterns in CEA western blot from gastric carcinoma lysates.

Regarding sLe^x, western blot analysis showed expression of sLe^x in the gastric carcinoma cases (fig. 15B). The sLe^x immunoreactivity (arrows in fig. 15B) seems to co-localize with CEA immunoreactivity observed in the CEA western blot.

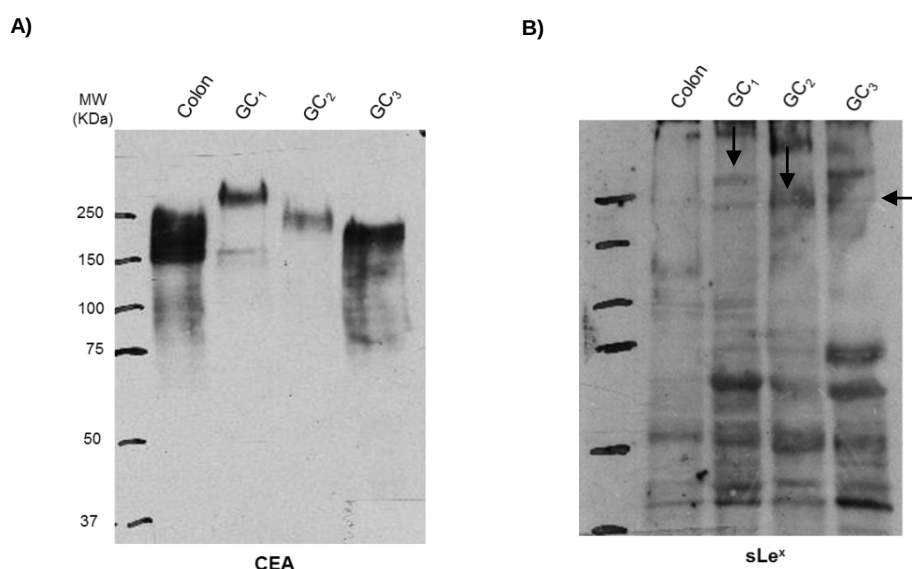


Fig. 15 – CEA and sLe^x expression in gastric carcinoma tumor samples and colon biopsy. A) Western blot analysis of CEA expression in gastric carcinoma samples and colon biopsy. B) Western blot analysis of sLe^x expression in gastric carcinoma samples and colon biopsy.

2. Identification of putative new protein targets which can be involved in MST3Gal IV invasive phenotype

2.1 Role of secreted sLe^x expressing proteins on c-Met and Src activation

Previously, our group showed that MST3Gal IV cells present increased activation of hepatocyte growth factor receptor (HGFR/c-Met) resulting in the activation of survival signaling pathways.

In this work, it was confirmed the activation of the described signaling pathway. By western blot analysis it was observed the increased phosphorylation of c-Met as well as the downstream effector Src in MST3Gal IV cells when compared to Mock cells (fig. 16).

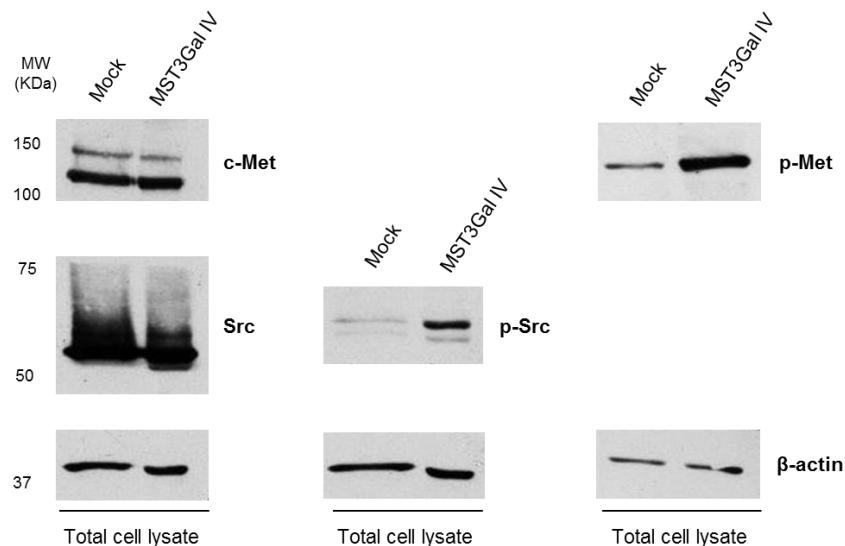


Fig.16 – Western blot analysis of total and phosphorylated c-Met and Src expression levels in gastric carcinoma cells. MST3Gal IV cells reveal an increased activation of c-Met and Src when compared to Mock cells, with no alterations in the total protein expression.

In order to disclose a possible role of secreted proteins expressing sLe^x in the activation of c-Met, it was performed a condition medium assay. Briefly, Mock and MST3Gal IV cells were incubated with condition medium (medium from MST3Gal IV or Mock, respectively) during 24h. After incubation, the activation of c-Met and Src was evaluated by western blot analysis as well as sLe^x expression.

Results show no differences in the expression of sLe^x in Mock and MST3Gal IV cells cultured in condition medium (fig. 17A). Regarding activation of c-Met and Src proteins, it was possible to confirm the increased activation of c-Met and Src in MST3Gal IV cells when compared to Mock cells cultured in normal conditions (fig. 17B). Also, it was

observed that Mock cells cultured in condition medium present increased activation of c-Met and Src (increased phosphorylation of c-Met and Src) mimic the differences observe between Mock and MST3Gal IV in normal conditions (fig. 17B). Although these differences were clearly seen by western blot, the statistical analysis of different experiments didn't reach a statistical significant difference.

Nevertheless these results point to a role of sLe^x secreted proteins in the activation of c-Met and a downstream target Src.

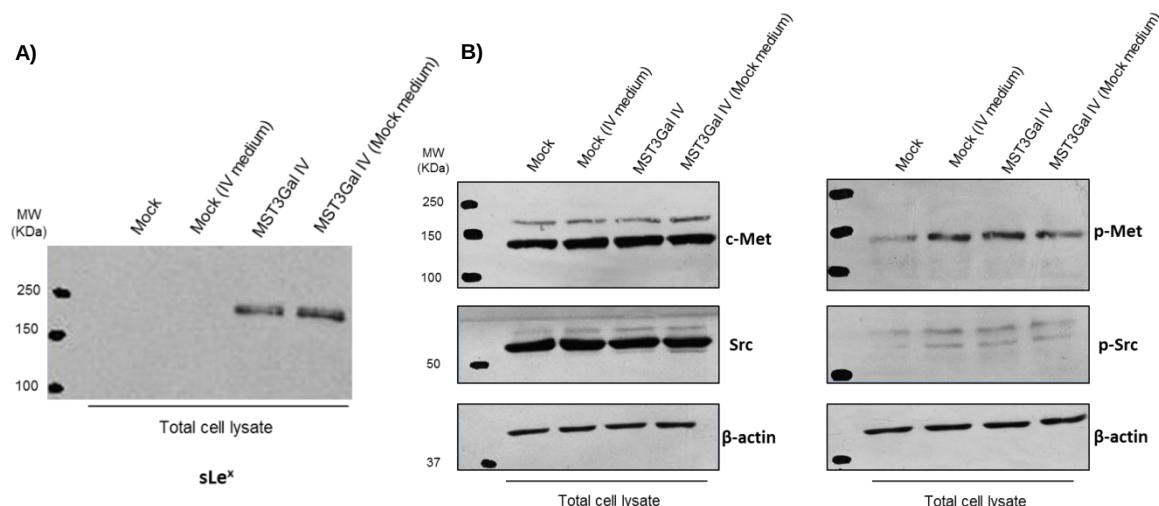


Fig. 17 – Role of sLe^x secreted proteins in the activation of c-Met and Src proteins. Evaluation of c-Met and Src activation on gastric carcinoma cells cultured in condition medium. A) Expression of sLe^x in Mock and MST3Gal IV cells cultured in normal and condition medium. Expression of sLe^x is only detected in MST3Gal IV cells in both conditions. B) Expression of total and phosphorylated forms of c-Met and Src in Mock and MST3Gal IV cells cultured in normal and condition medium. Western blot analysis show increased activation of c-Met and Src in Mock cells cultured in condition medium, in comparison to Mock cells in normal conditions. No difference observed in the total levels of c-Met and Src.

2.2 Identification and glycan characterization of differential expressed proteins in MST3Gal IV cells

In a previous work, our group observed that coomassie blue stained secretome samples from Mock and MST3Gal IV cells displayed few differences in their protein content (fig. 18). A clear band at high molecular weight was exclusively present in MST3Gal IV and identified by mass spectrometry, as Perlecan.

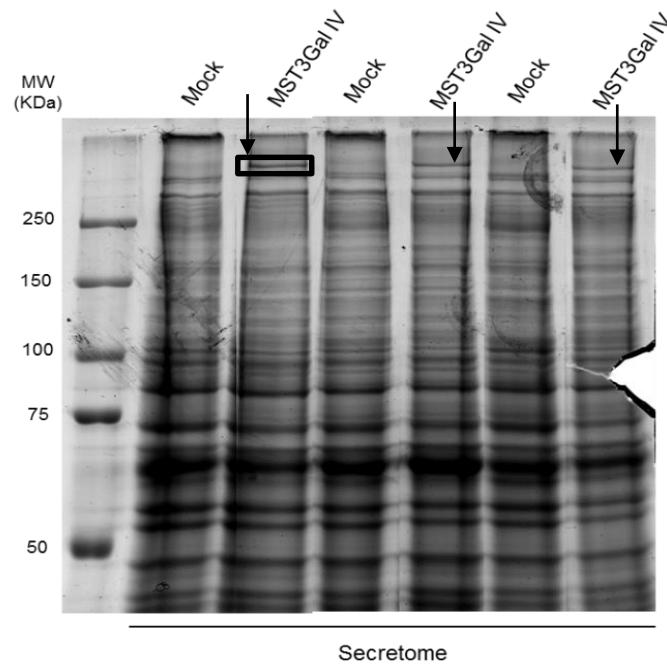


Fig. 18 – Coomassie Blue staining of secretome samples from MST3Mock and MST3Gal IV. Black arrows indicate a high molecular weight protein band that is only present in MST3Gal IV secretome.

In this work, expression of Perlecan was evaluated, by western blot analysis, in normal and condition medium total cell lysates as well as secretome from Mock and MST3Gal IV cells (fig. 19). The results showed the expression of similar bands for Perlecan in total cell lysates from Mock and MST3Gal IV (fig. 19A). However, secretome samples displayed a shift on the molecular weights of Perlecan protein expression in Mock when compared to MST3Gal IV. Secretome samples from MST3Gal IV presented a reduction in the molecular weight of some Perlecan isoforms when compared to Mock secretome (fig. 19B).

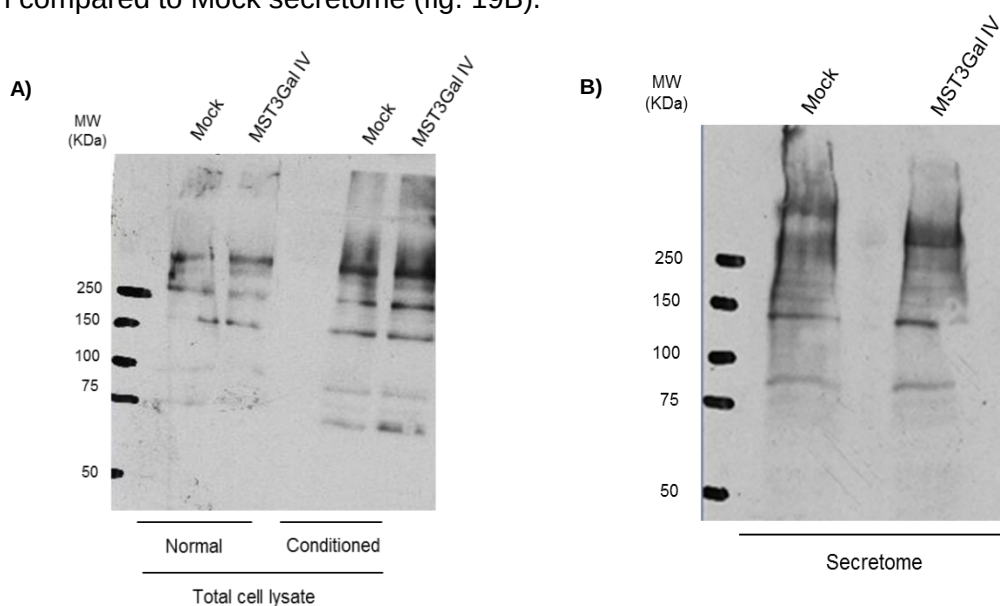


Fig. 19 – Expression of Perlecan in different conditions in Mock and MST3Gal IV cells. Western blot analysis showed that secretome samples present differences in the molecular weight of Perlecan in the different cell lines.

In order to evaluate if these differences in the protein expression could be due to alternative splicing in Perlecan, it was performed PCR analysis using different primer pairs flanking specific exons within its domains. PCR results showed no major differences in the transcription of the targeted exons between the different cell lines (figure 20). The analysis of exon 87-97 within domain V showed a reduction in the expression of a large transcript (approximately 1000bp) in MST3Gal IV when compared with Mock cells.

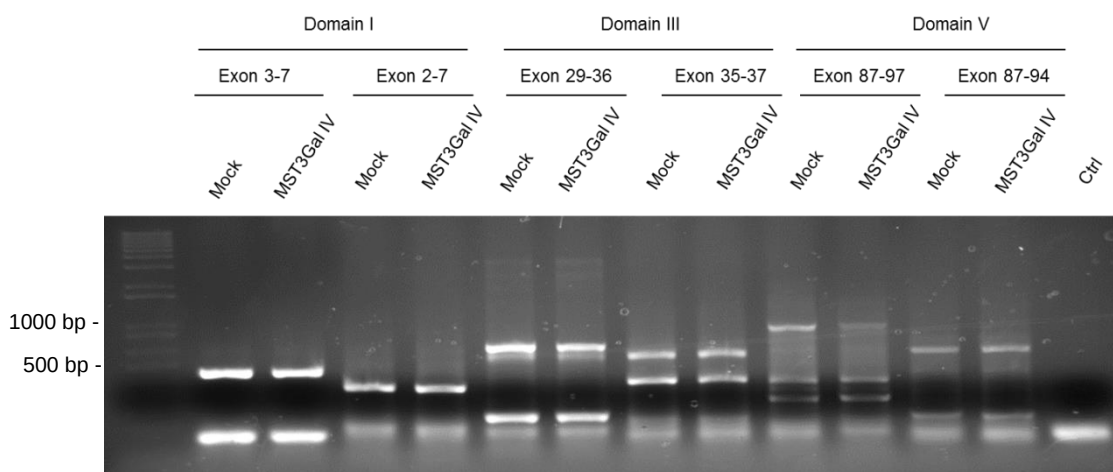


Fig. 20 – mRNA expression of Perlecan from Mock and MST3Gal IV using domain specific primers. mRNA derived from both cell lines was isolated and used to generate cDNA which was amplified using domain-specific primers and electrophoresed on 1% (w/v) agarose gels. Deionized water was used as a control of the PCR.

Perlecan is a proteoglycan which contains several different glycans attached to the protein backbone. Knowing this, and in order to evaluate if the differences observed in the molecular weight could be due to the presence of N-glycans, it was performed an enzymatic digestion using PNGase F and Endo H in total cell lysates and secretome followed by Perlecan western blot analysis.

Results from total cell lysates showed that treatment with both enzymes lead to a decrease in the molecular weight of the different isoforms in both cell lines. The observed shift in the molecular weight was equal in both cell lines (figure 21A).

Regarding secretome samples, it was also observed a decreased in the molecular weight of the different isoforms after the two types of enzymatic treatment in both cell lines (fig. 21B). It was found a slightly decrease of a band in MST3Gal IV secretome treated with PNGase F when compared to Mock (fig. 21B, full square). In addition, PNGase F treatment almost abrogates the difference observed previously in secreted Perlecan between Mock and MST3Gal IV cells (fig. 21B dotted square).

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

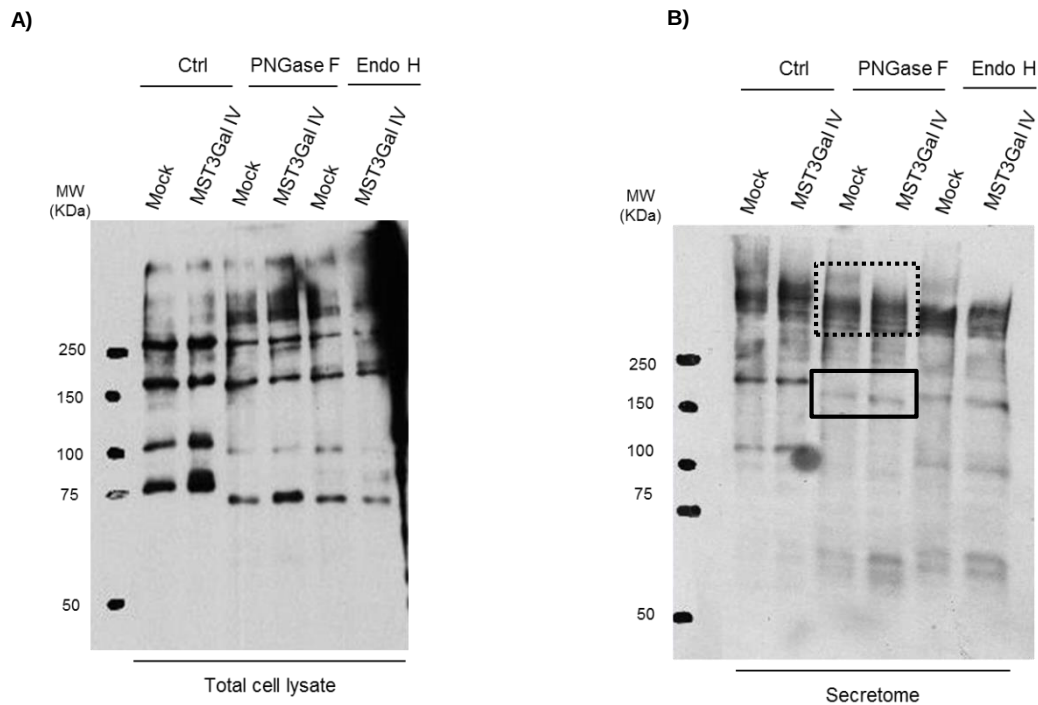


Fig. 21 – Effect of enzymatic digestion, using PNGase F and Endo H in Perlecan isoforms. Enzymatic digestion was performed in both total cell lysates (A) and secretome (B) from Mock and MST3Gal IV cells. Differences were observed in total cell lysates and in secretome from both cell lines after enzymatic treatment.

In order to evaluate the expression of sLe^x in Perlecan isoforms, it was performed Perlecan immunoprecipitation followed by sLe^x western blot analysis. Perlecan immunoprecipitation (IP) was performed using total cell lysates and secretome samples from Mock and MST3Gal IV cells (fig. 22).

Perlecan IP of total cell lysates didn't show any differences in the molecular weight from the different cell lines. In addition, sLe^x expression was only detected in total cell lysates from MST3Gal IV cells (fig. 22B black arrows), with no expression of sLe^x on Perlecan IP from both cell lines (fig. 22B).

Regarding Perlecan IP of secretome samples, it was observed differences in the molecular weight of Perlecan from the different cell lines (fig. 22C) similarly to the previous perlecan western blot (fig. 19B). In addition, we observed a band for sLe^x expression at high molecular weight isoform of Perlecan IP in MST3Gal IV (fig. 22D dotted arrow).

Furthermore, in these experiments it was also performed negative control IP using mouse immunoglobulins. However, in these negative controls IP it was possible to detect Perlecan expression in both samples from the different cell lines despite many

optimizations procedures (fig. 22A, 22C). These results can be explained by the presence of IgG like domains in Perlecan protein that can be recognized by the immunoglobulins used in the control IP.

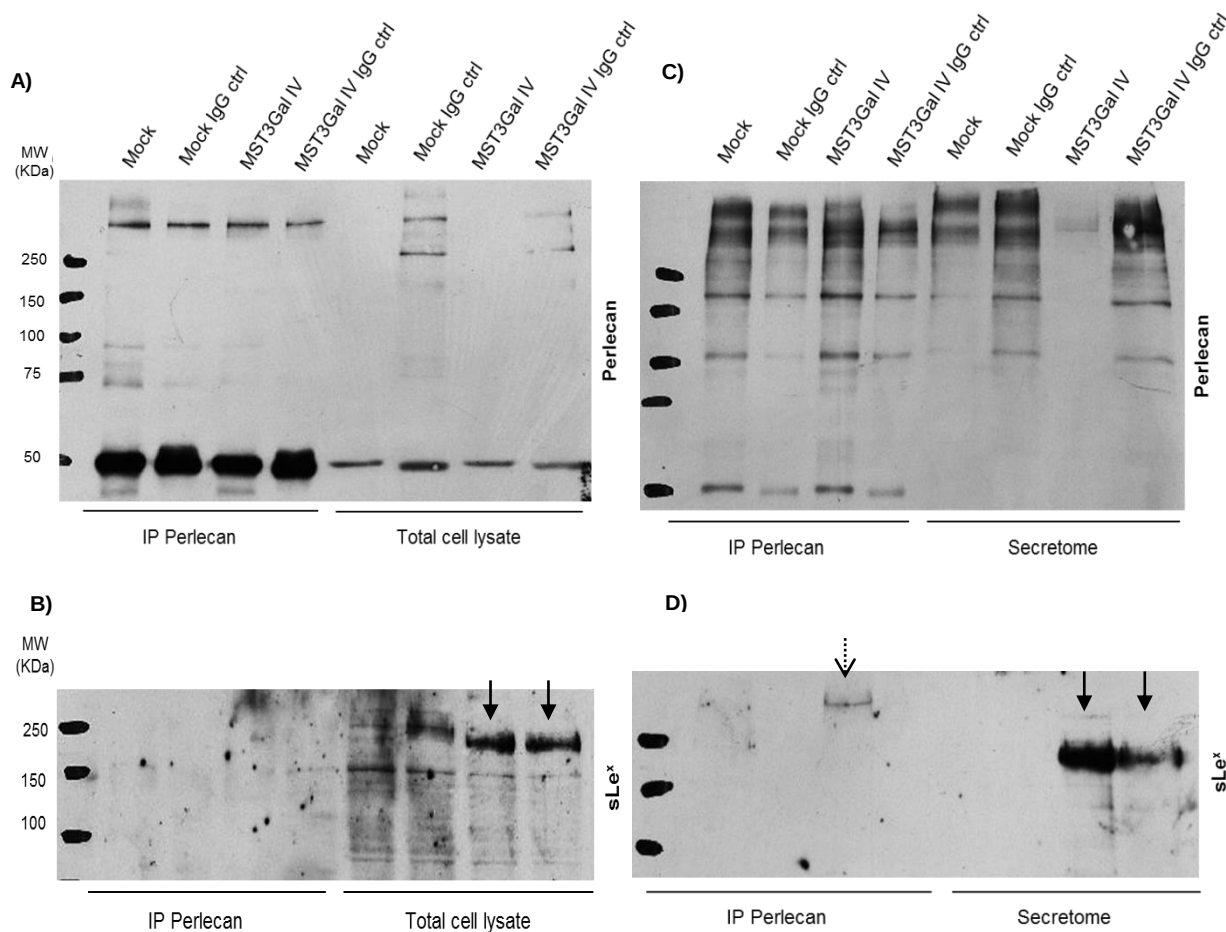


Fig. 22 – Western blot analysis of Perlecan immunoprecipitation in total cell lysates and secretome from Mock and MST3Gal IV cells. A) Perlecan expression in IP from total cell lysates. B) sLe^x expression in Perlecan IP from total cell lysates; C) Perlecan expression in IP from secretome; D) sLe^x expression in Perlecan IP from secretome.

According to the literature, CD44 is a glycoprotein described as a selectin ligand which can interact with CEA cooperating and mediating tumor cell adhesion to endothelial cells during tumor dissemination. Thus, it was interesting to evaluate the expression and glycan characterization of CD44 in our cell lines. Therefore, it was assessed the expression of CD44 in Mock and MST3Gal IV using western blot analysis. The results showed a pronounce shift in the molecular weight of MST3Gal IV total cell lysates when compared to Mock (fig. 23A).

To follow up this result, the expression of a specific and cancer-related CD44 variant – CD44v6 – was assessed in both cell lines by western blot. The results showed that MST3Gal IV cells display a different pattern of CD44v6 expression when compared to Mock cells, characterized by a shift to lower molecular weights. This phenomenon was observed in total cell lysates from cells cultured both in normal and conditioned medium as well as in the secretome (fig. 23B).

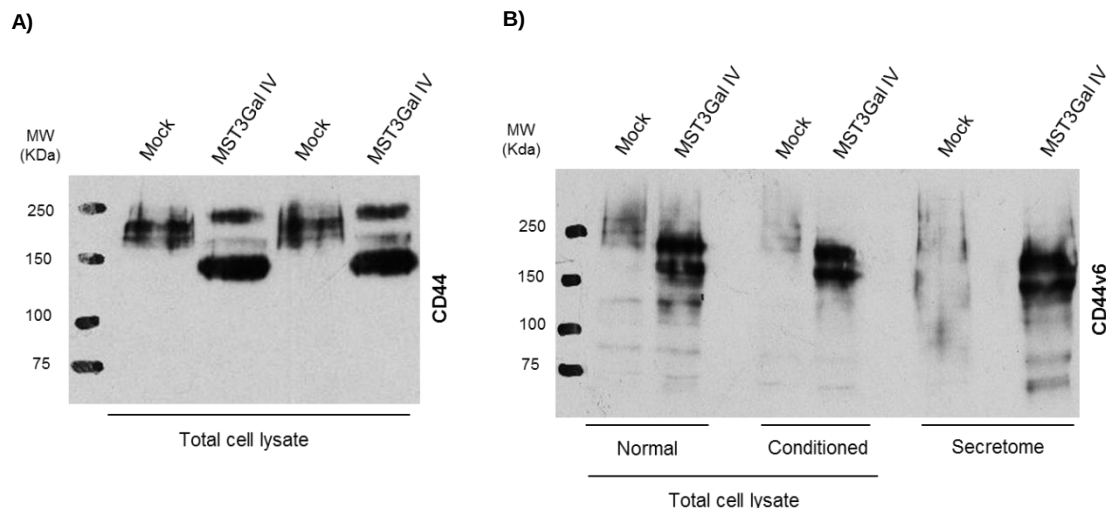
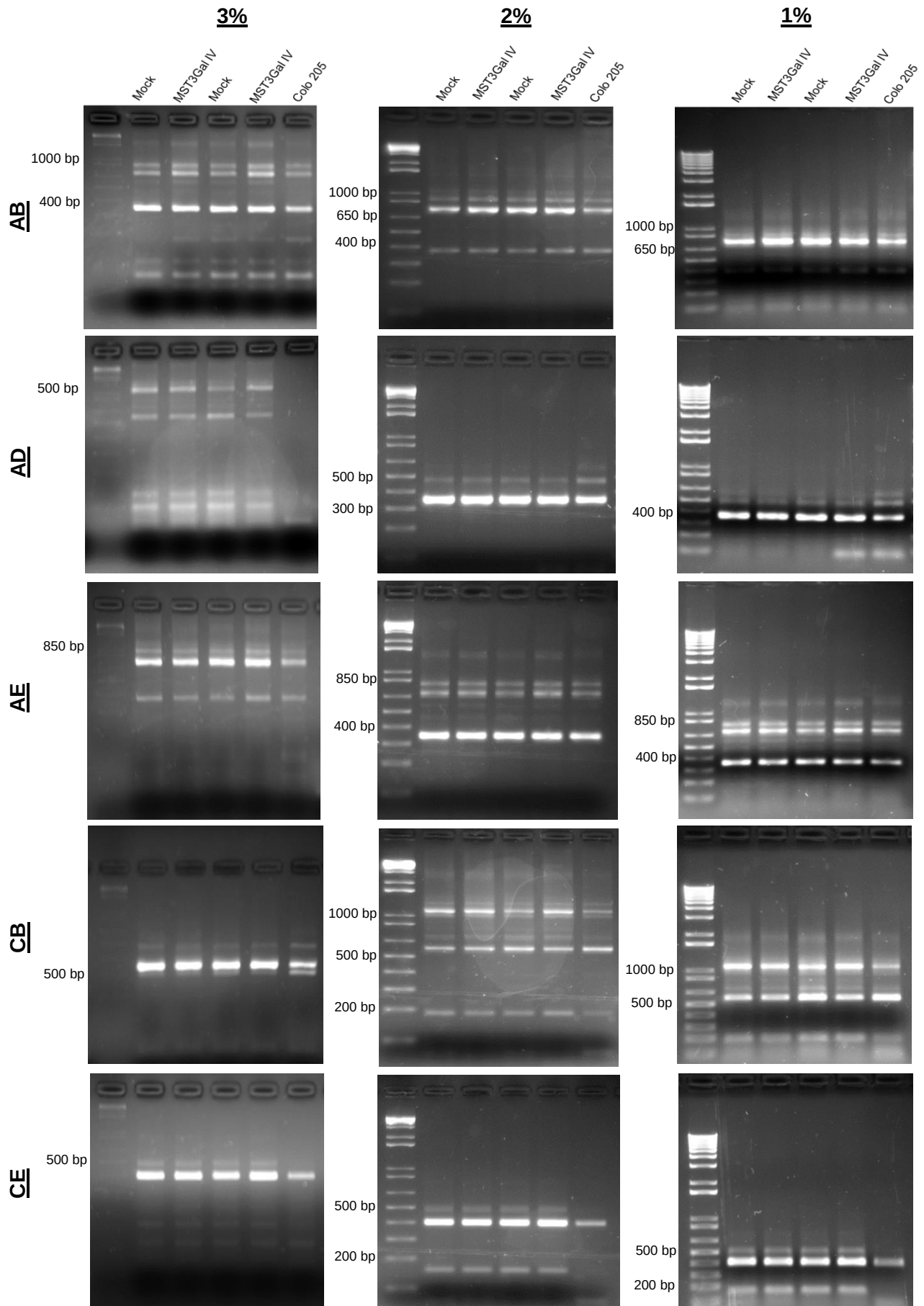


Fig. 23 – Expression of CD44 and CD44v6 containing isoforms in Mock and MST3Gal IV cells. A) Western blot analysis of CD44 in total cell lysates. B) Western blot analysis of CD44v6 containing isoforms in total cell lysates from normal, conditioned medium (without FBS) and in secretome.

Taking these results in consideration, we evaluated mRNA expression of CD44 transcripts in order to characterize the presence of specific variants in both cell lines (fig. 24). In this experiment it was selected a panel of different primers flanking different CD44 variants (AB – flanking CD44s 5' to CD44s 3' reaching all variants; AD – flanking CD44s 5' to v3 exon 3'; AE – flanking CD44s 5' to v6 exon 3'; CB – flanking CD44s 3' to v3 exon 5'; and CE – flanking CD44s 3' to v6 exon 3') for PCR analysis, and gel electrophoresis was performed using different percentages of agarose (3%, 2% and 1%). In general, the results showed the expression of different transcripts for the different primer pairs used. In addition, within each primer pairs it was not observed

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

any difference in CD44 transcript expression for MST3Gal IV when compared to Mock cells.



Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Fig. 24 – PCR analysis of CD44 transcripts in Mock and MST3Gal IV cells. PCR analysis was performed in both cell lines using primers flanking different variant regions within CD44 transcripts, and gel electrophoresis was performed using different percentages of agarose. mRNA from Colo 205 cell line was used as a control of CD44 expression. (AB – flanking CD44s 5' to CD44s 3' reaching all variants; AD – flanking CD44s 5' to v3 exon 3'; AE – flanking CD44s 5' to v6 exon 3'; CB – flanking CD44s 3' to v3 exon 5'; and CE – flanking CD44s 3' to v6 exon 3')

Then, we went for the analysis of C44v6 containing isoforms glycosylation using PNGase F and Endo H enzymatic digestion in total cell lysates and secretome, followed by western blot analysis (fig. 25). Enzymatic treatment of Mock total cell lysates and secretome using both enzymes didn't present any shift in the molecular weight of CD44v6. Regarding MST3Gal IV samples, it was observed that both total cell lysates and secretome treated with PNGase F presented a shift in the CD44v6 molecular weight (fig. 25 dotted and full arrows). In addition, MST3Gal IV secretome samples enzymatically treated with Endo H also showed a shift in the CD44v6 molecular weight, although smaller than the one observed in PNGase F treatment (fig. 25B full arrows).

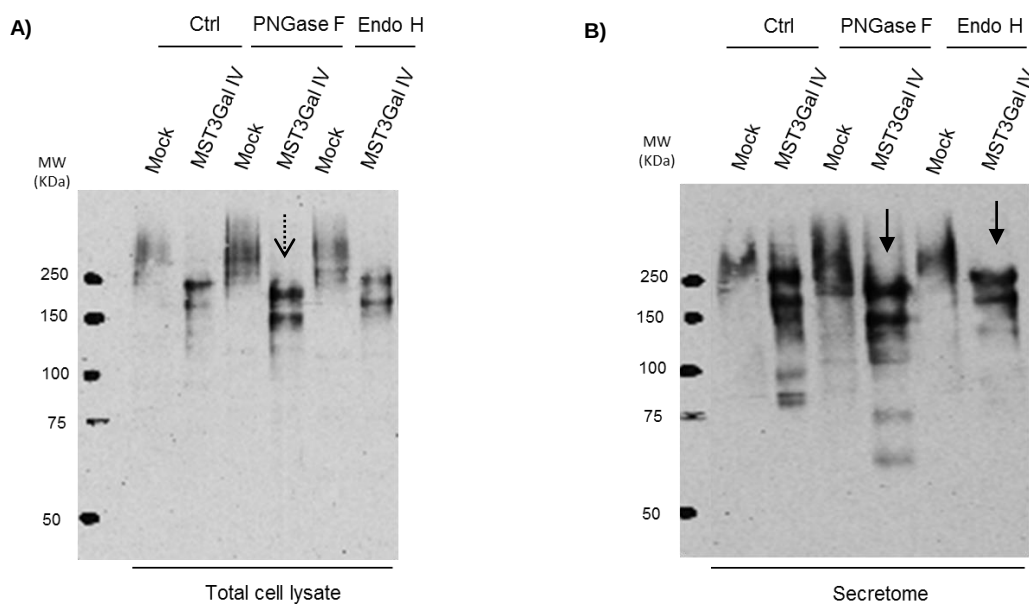


Fig. 25 – CD44v6 western blot analysis after treatment with PNGase F and Endo H in Mock and MST3Gal IV total cell lysates (A) and secretome (B). Arrows indicate the shift in the molecular weight of CD44v6 observed in MST3Gal IV samples after enzymatic treatment.

In order to evaluate if CD44v6 is a putative sLe^x carrier, we performed CD44v6 immunoprecipitation (IP) in total cell lysates and in secretome from both cell lines, followed by western blot analysis (fig. 26).

Regarding sLe^x western blot analysis of total cell lysates CD44v6 immunoprecipitates, it was observed a thin band in MST3Gal IV (dotted arrow in fig. 26A) which seems to

be at the same molecular weight as the sLe^x bands in the inputs (full arrows in fig. 26A).

About CD44v6 Immunoprecipitation of secretome samples, this result did not show any evidence of sLe^x expression on CD44v6 glycoprotein (fig. 26B).

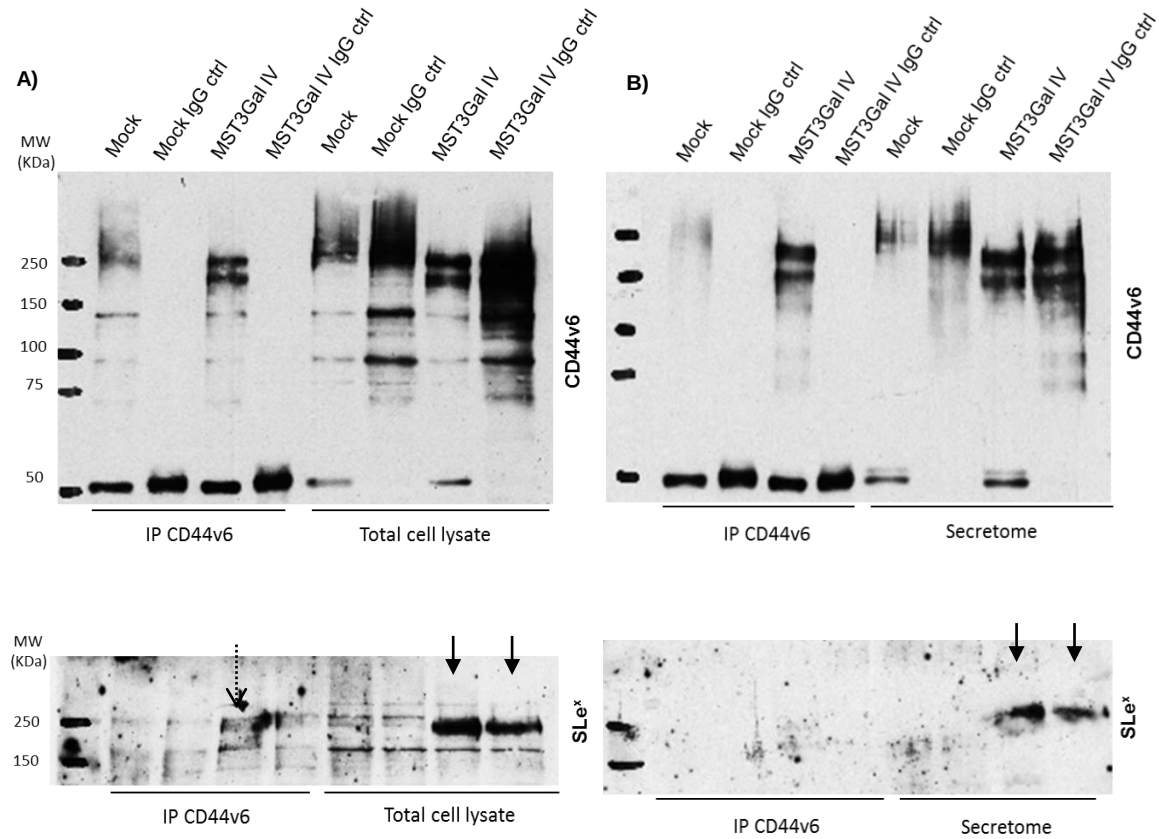


Fig. 26 – Evaluation of sLe^x expression in CD44v6 from Mock and MST3Gal IV total cell lysates and secretome. A) Dotted arrow shows sLe^x expression in CD44v6 immunoprecipitated proteins from MST3Gal IV total cell lysates. B) SLe^x western blot analysis of CD44v6 immunoprecipitation from Mock and MST3Gal IV secretome, with no expression of this glycan structure.

3. Evaluation of tumorigenic potential of sLe^x expressing cells

In order to disclose a possible mechanism of invasion mediated by CEA/sLe^x molecular complex, it was designed a mice model of orthotropic injection with both cell lines in athymic nude mice. In this study, it was possible to evaluate the tumorigenic potential (capacity to form tumors in mice) of the cells.

The evaluation of tumor formation of the injected cells showed that both cell lines present capacity to form tumors in mice (fig. 27). By measuring tumor size within 20 days it was possible to see a slightly increased in MST3Gal IV injected mice when compared to tumors of Mock injected mice. This tendency was observed at 8th day after injection and it was more evident after 13 days of injection (fig. 28). However these results didn't reach statistical significance.

In addition, MST3Gal IV injected mice developed signs of disease, such as weight loss, earlier than Mock injected mice. Moreover after tumor removal both types of injected mice evidenced tumor relapse.

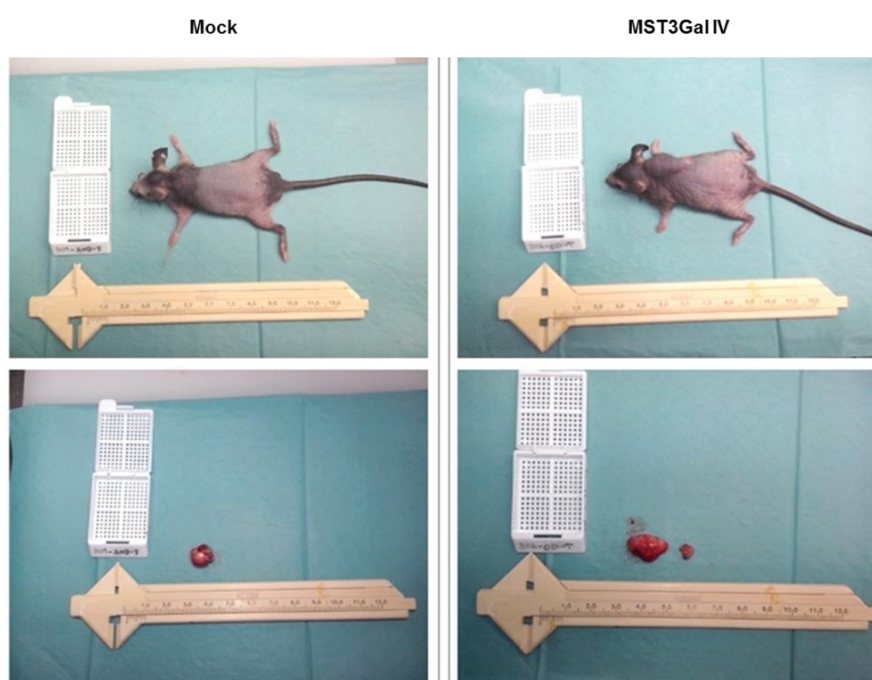


Fig. 27 – Tumorigenic potential of Mock and MST3Gal IV cells in athymic nude mice. SLe^x expressing (MST3Gal IV) and Mock cells were subcutaneously injected in athymic nude mice. SLe^x expressing mice group showed formation of larger tumors when compared with Mock control mice group.

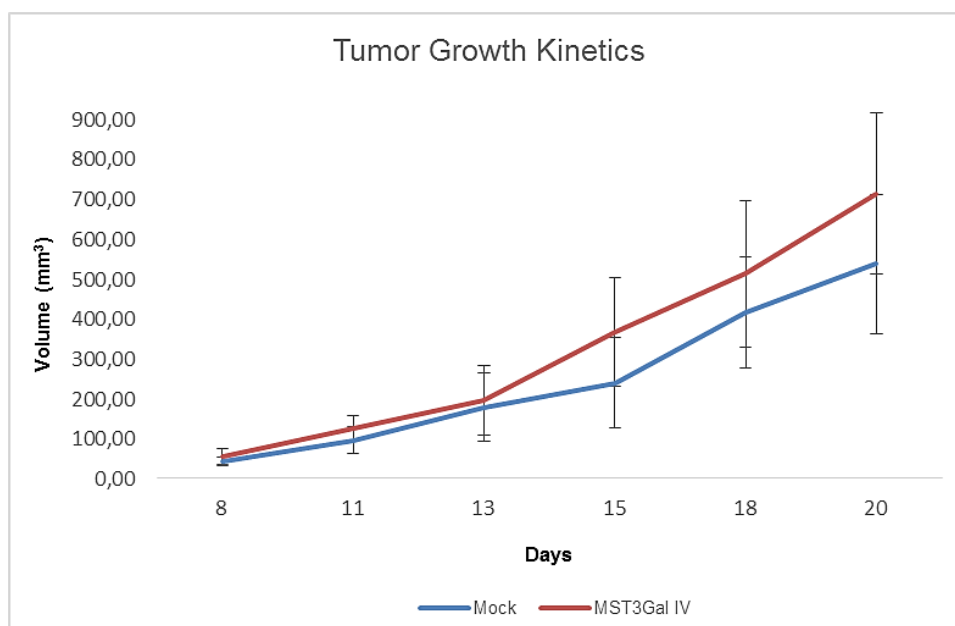


Fig. 28 – Tumor growth kinetics of Mock and MST3Gal IV injected mice. MST3Gal IV injected mice have the capacity to promote larger tumors when compared with Mock situation. Blue line represents the tumor growth of Mock injected mice and red line represents MST3Gal IV injected mice.

According to the literature, some studies suggest CEA-hnRNPM4 interaction and consequent activation of signaling pathways, as a hypothesis for the metastization of tumor cells. Thus, in order to evaluate if Mock and MST3Gal IV cells express hnRNPM4, it was performed western blot analysis of both cell lines. The results show that both cell lines express hnRNPM4, however any differences were observed between Mock and MST3Gal IV cells (fig. 29).

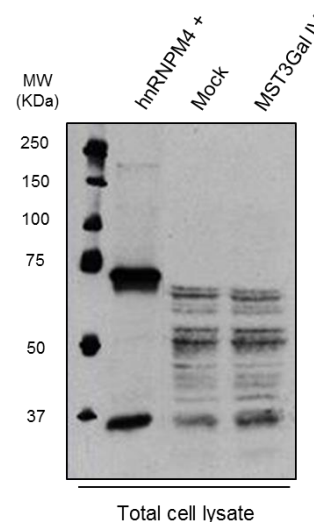


Fig. 29 – HnRNPM4 expression in Mock and MST3Gal IV cell lysates. HnRNPM4 + correspond to HEK293T cell line overexpressing hnRNPM4 and was used as a western blot positive control.

To further confirm the expression and localization of hnRNPM4 in Mock and MST3Gal IV cells it was performed immunofluorescence analysis. In addition, it was also evaluated the expression of sLe^x in both cell lines. SLe^x expression was only detected in MST3Gal IV with no expression in Mock cells (fig. 30).

Immunofluorescence for hnRNPM4 showed expression in Mock and MST3Gal IV both at the cytoplasm and nucleus (fig. 30).

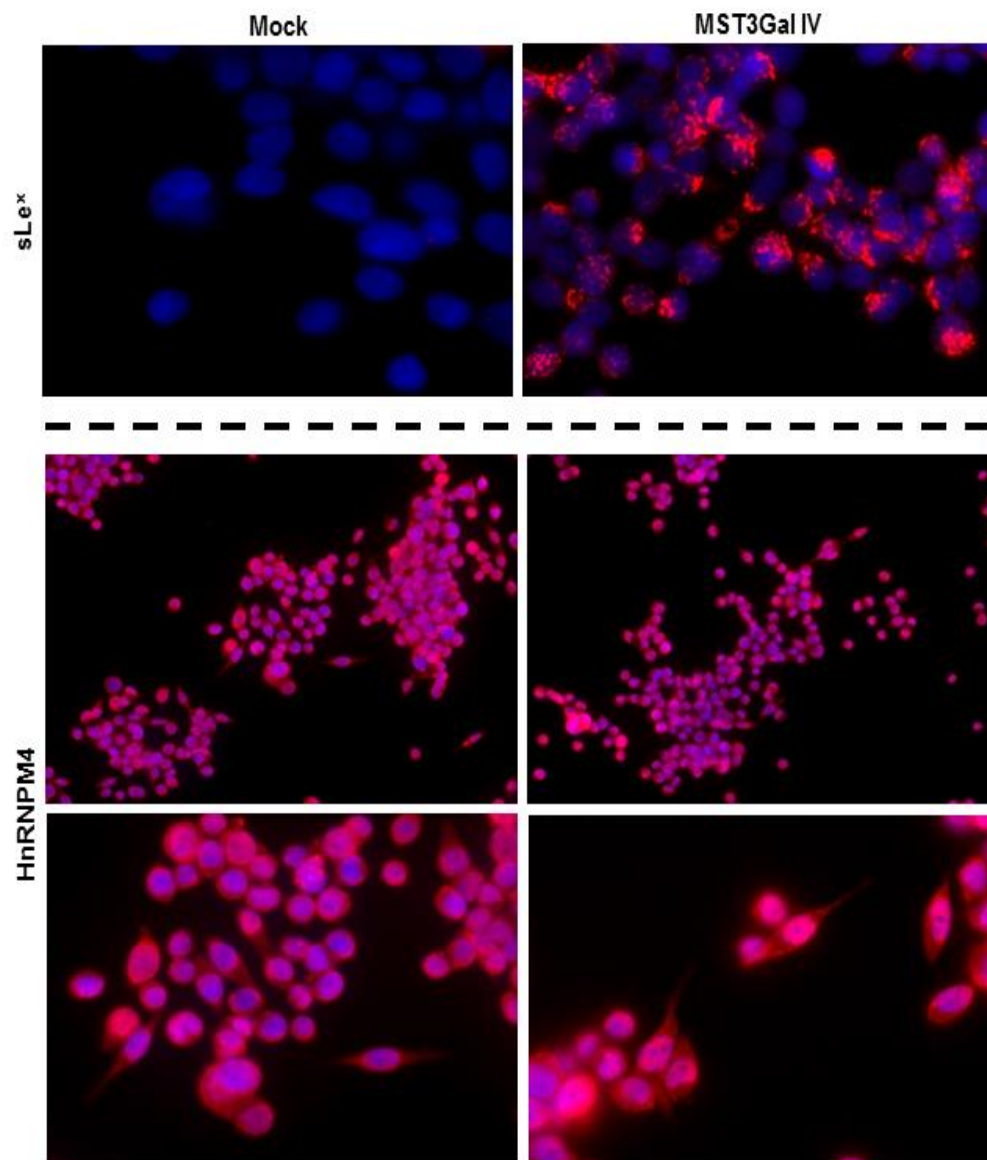


Fig. 30 – Immunofluorescence analysis to evaluate the expression sLe^x and hnRNPM4 in Mock and MST3Gal IV gastric carcinoma cells. SLe^x is only expressed in MST3Gal IV and both cell lines expressed hnRNPM4. DAPI was used for nuclear counter staining.

DISCUSSION

Gastric cancer is one of the most common cancers worldwide, presenting also a high mortality rate. Alterations of glycosylation have been strongly associated with cancer. Deregulation of glycosylation is a feature in many types of cancer and leads to the expression of distinct carbohydrates profiles in malignant tissues¹⁰⁸⁻¹¹⁴. Moreover, aberrant glycosylation has been considered a hallmark of cancer and one of the most common alterations observed is the abnormal expression of sialyltransferases. These enzymes are responsible for the synthesis of sialylated antigens added to cell surface molecules and to secreted proteins. SLe^x is one of these examples and its expression has been associated with cancer progression and aggressiveness, such as invasive gastric carcinomas^{102,114}.

In the first part of this work, it was confirmed the expression of sLe^x in MKN45 previously transfected with a gene encoded for the enzyme ST3Gal IV^{39,40}. The results confirmed the expression of sLe^x glycan structures on glycoproteins from total cell lysates and secreted proteins in MST3Gal IV cells when compared to Mock cells (figure 9). In addition, this result was further confirmed by immunofluorescence, where it was possible to see sLe^x antigen expression only in MST3Gal IV cells (figure 30). Altogether, these observations confirm previous work of our group³⁹ and other studies that described the importance of ST3Gal IV enzyme in the synthesis of sLe^x structures⁴⁰.

CEA is a glycoprotein that shows limited expression in normal adult tissues, most restricted to colon epithelial cells¹⁶⁹. During carcinogenesis, cells loss their normal polarity and tumor cells usually secreted CEA into circulation allowing its detection in serum. CEA is commonly used in the clinic as a biomarker of several cancers such as gastric cancer. However its detection is usually limited to monitor disease recurrence and for patients follow up¹⁵⁷⁻¹⁶¹. Previous results of our group identified, by mass spectrometry, CEA as major protein carrier of sLe^x (table 7, unpublished results). In this work, it was possible to confirm the presence of sLe^x in CEA glycoprotein by immunoprecipitation (fig. 13). In addition, the CEA immunoprecipitates demonstrated that CEA molecules from MST3Gal IV cells present a slightly higher molecular weight when compared to Mock immunoprecipitates.

CEA family members have been described to carry Lewis x structures^{194,195}. Differences in the glycosylation pattern of CEA have been described in colonic tissues, where normal mucosa and pre-tumor lesions showed the presence of CEA molecules with different molecular weights when compared to colon tumor cells¹⁹⁶. Taking these facts in consideration, we hypothesized that the difference observed in CEA molecular

weights from MST3Gal IV cells could be associated to differences in the glycosylation pattern, more specifically the expression of sLe^x.

CEA is a heavily glycosylated protein with 28 putative N-glycosylation sites¹⁹⁷. Thus, in order to characterize CEA N-glycans, we performed enzymatic deglycosylation using PNGase F and Endo H in immunoprecipitated CEA molecules (fig. 14A). The results showed that PNGase F enzymatic digestion leads to a large reduction in CEA molecular weight, with no alterations observed using Endo H enzymatic digestion. PNGase F is an enzyme which hydrolyzes the amide bond of the asparagine residue to which the glycan is attached, removing all N-glycan types from peptides and proteins. On the other hand, Endo H cleaves N-glycans between the two N-acetylglucosamine (GlcNAc) subunits directly proximal to the asparagine residue of high mannose and hybrid N-glycan types. Taking this into consideration, our result leads us to conclude that CEA from both cell lines contains large amounts of complex N-glycans. The complexity of N-glycans in CEA molecules have been largely investigated, and a recent study also described that treating MKN45 cells with PNGase F leads to a pronounce reduction in the molecular weight of CEA molecules¹⁹⁸. Moreover, we still observe sLe^x expression in CEA immunoprecipitates from MST3Gal IV after N-glycans digestion (fig. 14B). This result might suggest: 1) sLe^x antigen might be present in O-glycans or 2) the deglycosylation process was incomplete and some CEA molecules might have N-glycan attached. Indeed, when we treated total cell lysates and secretome directly with both enzymes, it was still possible to see sLe^x expression, strongly indicating the presence of sLe^x O-glycans. However, this result doesn't clear confirm that sLe^x is present in CEA molecules after enzymatic treatment (fig. 10).

CEA is considered to be highly expressed in gastric and colorectal carcinoma tissues¹⁹⁹⁻²⁰² (review in¹⁶⁹). In this study we went to evaluate the expression of CEA and sLe^x in proteins extracted from three gastric carcinoma tissues. All the samples present expression of CEA and sLe^x and it looks like we have co-localization of the two molecules in gastric carcinoma lysates (fig. 15). These results are in agreement with previous observations from our group where it was possible to show CEA/sLe^x molecular complex using PLA in gastric carcinoma tissues (unpublished; supplementary data). Interestingly, the pattern of CEA expression was very different in the three gastric carcinoma sample tissues analyzed. We hypothesized that these variations might be explained by different glycosylation patterns in the different carcinoma tissues.

Concerning sLe^x expression, previous studies have shown that increased cellular sialylation results in the activation of receptors and signaling pathways which contribute

to cancer progression^{203,204}. In fact, previous results from our group demonstrated that sLe^x expressing cells present increased activation of c-Met as well as downstream effectors such as Src. In addition, sLe^x expressing cells exhibited increased invasive capacity mediated by c-Met activation³⁹. Some reports showed that overexpression of c-Met has been correlated with poor prognosis in gastrointestinal, hepatocellular, breast, endometrial and nasopharyngeal carcinomas²⁰⁵⁻²⁰⁷. In this work we also confirmed the increased activation of c-Met and Src in MST3Gal IV cells when compared to Mock cells (fig. 16). Also, when Mock cells were cultured in condition medium, we observed an increase in c-Met activation with no alterations in sLe^x expression (fig. 17). Thus, we can conclude that the noticeable change on the receptor activation in Mock cells cultured in condition medium might be due to the interaction of c-Met with sLe^x secreted proteins.

Taking into consideration this result, we went to identify and characterize the MST3Gal IV secreted proteins that can be mediators of c-Met activation, playing a role in cancer cell invasion. Previous results from our group showed the presence of a specific band at high molecular weight in secretome samples from MST3Gal IV cells when compared to Mock cells (fig. 18). This band was further identified, by mass spectrometry, as Perlecan (unpublished results). Using western blot analysis, we confirmed the presence of this proteoglycan in Mock and MST3Gal IV cells. However, differences in the molecular weight were only detected in secretome with MST3Gal IV samples expressing a band at higher molecular weight (fig. 19B). This result was further confirmed by Perlecan immunoprecipitation in secretome samples, characterized by a different pattern of Perlecan expression in Mock when compare to MST3Gal IV (fig 22C). Follow this result, we went to analyze the expression of possible splicing variants within Perlecan transcripts. In this experiment we used primers flanking specific exons within domains I, III and V of Perlecan. We couldn't find major differences in the transcription of the targeted exons in MST3Gal IV compared to Mock cells (fig. 20). Only the PCR product for exon 87-97 in domain V showed a reduce intensity band at approximately 1000 bp in MST3Gal IV when compared to Mock. Since we didn't perform a loading control we cannot claim that MST3Gal IV present reduced expression of this transcript. However, the role of different Perlecan fragment including differences in exon 87-97 transcripts, have been described in mast cells and associated to regulation of angiogenesis and matrix turnover, which are both key events in wound healing²⁰⁸. Taking this into consideration, future studies on Perlecan transcripts including analysis of exon 87-97, can give further insights on MST3Gal IV increased cell invasive capacity.

Perlecan is a proteoglycan of approximately 400kDa (core protein) which is highly glycosylated with both O-linked oligosaccharides and several heparan sulfate chains^{150,151}. In addition, Perlecan contains 10 potential N-glycosylation sites distributed along the protein core²⁰⁹. In order to evaluate if the differences observed in the molecular weight of Perlecan could be due to alterations in glycosylation, we started by performing N-glycan enzymatic digestion followed by Perlecan western blot analysis. Our results showed a shift in Perlecan molecular weight from both Mock and MST3Gal IV after enzymatic treatment (fig. 21). This observation suggests the presence of all types of N-glycans in Perlecan from Mock and MST3Gal IV in total cell lysates and secretome. Very interesting, in secretome samples treated with PNGase F, we observed that previous differences in the molecular weight from MST3Gal IV compared to Mock were almost totally abolished (fig. 21B). This result leads us to conclude that part of this difference is possibly due to the presence of complex type N-glycans in MST3Gal IV which aren't in Mock secreted Perlecan.

Then, in order to assess if Perlecan could contain sLe^x, we performed Perlecan immunoprecipitation in total cell lysates and secretome and looked for sLe^x expression. The immunoprecipitation results suggest that Perlecan in secretome from MST3Gal IV cells contains sLe^x (fig. 22D).

Several studies suggest secreted Perlecan as potential biomarker for the detection of advanced prostate cancer and could also provide a drug target for cancers such as gastric cancer^{152,156}. In addition, one study that focused on the identification of proteins carriers of sLe^x in plasma from breast cancer patients showed that sLe^x could be found in several proteins such as proteoglycan 4²¹⁰. For our knowledge, these results seem to be the first description of Perlecan carrying sLe^x in gastric cancer cell lines.

In this study, we also focus our attention in CD44 glycoprotein. CD44 is a highly glycosylated protein which has been reported to have important role in many cancers. Thus, it was suggested that CD44 can function as a selectin ligand and cooperate with CEA mediating tumor cell adhesion to the ECM facilitating tumor cell dissemination¹³¹. Besides CD44 standard form, this glycoprotein can express different variants as a result of alternative splicing and post-translational modifications^{132,135-137}. A study conducted by *Yokozaki et al* the expression of different CD44 splice variants in gastric carcinoma tissues and also in metastatic foci²¹¹. Furthermore, CD44v6 has been described as a major variant expressed in cancer and found as a co-receptor for the activation of c-Met^{146,212,213}. In addition, it was shown that the extracellular domain of CD44v6 can efficiently bind to HGF and present it to c-Met leading to activation of signaling pathways²¹².

In our study we confirmed the expression of CD44 and CD44v6 in both Mock and MST3Gal IV cell lines, and interestingly we saw a shift in the molecular weight for the sLe^x expressing cells (fig. 23). This result lead us to hypothesize that CD44 from MST3Gal IV cells could be expressing different variants when compared to Mock cells. To validate this hypothesis, we analyze the CD44 mRNA expression by PCR using primers flanking different exons in order to amplify different variant combinations. The PCR results didn't show differences between Mock and MST3Gal IV (fig. 24).

Then we looked for a possible alteration in glycosylation of CD44 using treatment with PNGase F and Endo H. The results showed that treatment with both enzymes in Mock total cell lysates and secretome don't alter the molecular weight of the protein, indicating that CD44v6 from Mock cells doesn't contain any N-glycan (fig. 25). On the other hand, treatment with both enzymes in MST3Gal IV leads to a reduction in the molecular weight indicating the presence of N-glycan structures. Specifically, in total cell lysates, it was observed a shift in the molecular weight when samples were treated with PNGase F and no alteration with Endo H treatment, indicating the presence of complex type N-glycans (fig. 25A). In secretome samples from the same cell line, observed different shifts using different enzymatic treatment indicating the presence of all type of N-glycans (fig. 25B).

Then we looked for sLe^x in CD44v6 containing isoforms from total cell lysates and secretome samples, and found a weak expression of this antigen in MST3Gal IV lysates samples (fig. 26B). Specific CD44 variants in cancer cells were identified as carriers of sialylated antigens, such as sLe^x, which facilitate selectin binding²¹⁴. *Lim et al* also showed that CD44 variants expressing cancer-associated antigens such as sLe^x were frequently increased in the serum of cancer patients²¹⁵. As described above, CD44v6 has been described as a c-Met co-receptor being involved in its activation. Taking all these facts and findings into consideration, we hypothesized that CD44v6/sLe^x molecular complex might be involved in the increased activation of c-Met in MST3Gal IV cells. As future analysis it would be interesting to see CD44v6 and c-Met using proximity ligation assay (PLA) as well as immunoprecipitation.

As main conclusions, we can say that ST3Gal IV expression influence and alter the glycosylation pattern of CD44, and Perlecan. Further analysis using mass spectrometry analysis will be of extremely importance to exactly determine the glycan structure in the identified proteins.

Apart from these in vitro studies using our cell model, we also evaluate the biological impact of sLe^x expressing cells *in vivo* in a mice model. We subcutaneously injected

both cell lines in athymic nude mice and our preliminary results for tumor growth kinetics showed a tendency in MST3Gal IV injected mice to form larger tumors (fig. 27). However, this result didn't reach a statistical significant difference (fig. 28). Regarding previous work from our group, this result is in agreement with other *in vivo* model using chicken chorioallantoic membrane which also shows no statistical difference in tumor size from both cell lines³⁹. It is important to reemphasize that these are preliminary result and this assay should be repeated in an extended timeline with the evaluation of other conditions.

Regarding tumor cells dissemination, some studies suggest CEA-hnRNPM4 interaction as an important event leading to the activation of signaling pathways, as a involved in the metastatic process¹⁸⁷⁻¹⁸⁹. HnRNPM4 apart from being described as CEA receptor at the cell membrane, it can also be found in the cytoplasm and nucleus mediating alternative splicing of several proteins, such as CD44^{188,189,191}. Taking these facts into consideration, we also started a study to evaluate the capacity of CEA/sLe^x in the activation of hnRNPM4 therefore mediating alternative splicing of target mRNAs. To achieve this, we started by assessing hnRNPM4 expression in both cell lines using western blot (fig. 29) and immunofluorescence (fig. 30) analysis. The results demonstrated that both cell lines express hnRNPM4, and immunofluorescence analysis showed its expression both at the cytoplasm and nucleus. However, membranous expression was not clear evident using this technique so further analysis using confocal microscopy are needed.

In conclusion, we reported the expression of sLe^x in different proteins of ST3Gal IV transfected cells, in opposite to the absence of this antigen in proteins from control cells. We confirmed CEA as a sLe^x protein carrier in MST3Gal IV cells and a possible presence of this antigen in CEA O-glycans. Also, we identified Perlecan and CD44v6 as putative carriers of sLe^x in different conditions. Moreover, we observed increase activation of c-Met and Src in MST3Gal IV cells, as well as in Mock cells treated with condition medium suggesting a possible role of secreted sLe^x protein carriers in c-Met activation. Preliminary *in vivo* results suggest that mice injected with sLe^x expressing cells present larger tumors when compared with Mock injected mice.

FUTURE PERSPECTIVES

Alterations in glycosylation have been largely described in cancer and most of the times associated to more aggressive tumors. At the moment, gastric cancer patients are currently diagnosed in an advanced stage of the disease and can't undergo surgery. For this reason, there is a need in the development of new specific biomarkers which allow early detection of gastric cancer. In addition, the identification of new therapeutic targets is also an important feature to improve cancer treatment. Taking these in consideration, we found very important the identification and characterization of glycosylation alterations in cancer, specifically in gastric carcinoma.

In this work we were able to identify alteration in the glycosylation of carcinoembryonic antigen (CEA), more specifically expression of sLe^x in gastric carcinoma which seems to be associated to a more aggressive phenotype of cancer cells. As future work and in order to validate this result, it would be crucial to characterize the structure of the N- and O-glycans within CEA molecules using cutting edge mass spectrometry technologies. Ultimately, we will try to develop a method to identify CEA/sLe^x in serum from gastric cancer patients. In addition, we also found differences in the regulation and glycan composition of other proteins in sLe^x expressing cells. According to these results, it would be also interesting to validate Perlecan and CD44v6 as putative carriers of sLe^x in MST3Gal IV cells using mass spectrometry analysis.

Previous results from our group showed that sLe^x expressing cells have an increased invasive capacity associated to the activation of c-Met³⁹. In this work, a role of secreted sLe^x protein carriers in the activation of c-Met was also described. In order to follow this result, it would be interesting to perform an invasion assay, using condition medium, to disclose the role of MST3Gal IV secreted proteins in the invasive capacity of cancer cells. Furthermore, to clearly validate the role of sLe^x structures in the invasive phenotype, a specific blocking of sLe^x epitopes or removal of the sialic acids using, respectively, specific antibodies and neuraminidase treatment should also be performed. In addition, the activation of c-Met and downstream effectors should also be evaluated upon these treatments.

Regarding our *in vivo* mice model injected with Mock or MST3Gal IV cells, it was already collected specific organs in order to assess tumor invasion. Tissues were fixed using formaldehyde and paraffin embedded, and slide sections performed. As future work it will be evaluated, by immunohistochemistry, the presence of metastasis in ganglia, lung, and liver paraffin sections. In addition, and if micrometastasis were identified, we will also evaluation the expression of sLe^x antigen, CEA, CD44v6, Perlecan, p-Met and hnRNPM4 in these sections.

Finally, and knowing that hnRNPM4 can act as a CEA receptor activating signaling pathways, also it will be evaluated the activation of hnRNPM4 *in vitro* and in our mice model in order to disclose a possible role sLe^x expression in this process.

Our results and future outlooks can bring important contributions to the study and the development of novel serological assays for gastric cancer detection and treatment.

BIBLIOGRAPHY

- 1 Varki, A. Sialic acids in human health and disease. *Trends in molecular medicine* **14**, 351-360, doi:10.1016/j.molmed.2008.06.002 (2008).
- 2 Landsteiner, K. Individual Differences in Human Blood. *Science* **73**, 403-409, doi:10.1126/science.73.1894.403 (1931).
- 3 Varki, A. & Sharon, N. in *Essentials of Glycobiology* (eds A. Varki et al.) (2009).
- 4 Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nature reviews. Cancer* **15**, 540-555, doi:10.1038/nrc3982 (2015).
- 5 Bertozzi, C. R. & Sasisekharan, R. in *Essentials of Glycobiology* (eds A. Varki et al.) (2009).
- 6 Reis, C. A., Osorio, H., Silva, L., Gomes, C. & David, L. Alterations in glycosylation as biomarkers for cancer detection. *Journal of clinical pathology* **63**, 322-329, doi:10.1136/jcp.2009.071035 (2010).
- 7 Varki, A., Esko, J. D. & Colley, K. J. in *Essentials of Glycobiology* (eds A. Varki et al.) (2009).
- 8 Tuccillo, F. M. et al. Aberrant glycosylation as biomarker for cancer: focus on CD43. *BioMed research international* **2014**, 742831, doi:10.1155/2014/742831 (2014).
- 9 Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et biophysica acta* **1473**, 4-8 (1999).
- 10 Varki, A. & Freeze, H. H. in *Essentials of Glycobiology* (eds A. Varki et al.) (2009).
- 11 Kang, J. G., Ko, J. H. & Kim, Y. S. Pros and cons of using aberrant glycosylation as companion biomarkers for therapeutics in cancer. *BMB reports* **44**, 765-771 (2011).
- 12 Suji, G. & Sivakami, S. Glucose, glycation and aging. *Biogerontology* **5**, 365-373, doi:10.1007/s10522-004-3189-0 (2004).
- 13 Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867, doi:10.1016/j.cell.2006.08.019 (2006).
- 14 Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143**, 672-676, doi:10.1016/j.cell.2010.11.008 (2010).
- 15 Schultz, M. J., Swindall, A. F. & Bellis, S. L. Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer metastasis reviews* **31**, 501-518, doi:10.1007/s10555-012-9359-7 (2012).
- 16 Crocker, P. R. & Feizi, T. Carbohydrate recognition systems: functional triads in cell-cell interactions. *Current opinion in structural biology* **6**, 679-691 (1996).
- 17 Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* **291**, 2364-2369 (2001).
- 18 Karlsson, K. A. Meaning and therapeutic potential of microbial recognition of host glycoconjugates. *Molecular microbiology* **29**, 1-11 (1998).
- 19 Varki, A., Kannagi, R. & Toole, B. P. in *Essentials of Glycobiology* (eds A. Varki et al.) (2009).
- 20 Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A. & Dwek, R. A. Glycosylation and the immune system. *Science* **291**, 2370-2376 (2001).
- 21 Li, M., Song, L. & Qin, X. Glycan changes: cancer metastasis and anti-cancer vaccines. *Journal of biosciences* **35**, 665-673 (2010).
- 22 Fuster, M. M. & Esko, J. D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature reviews. Cancer* **5**, 526-542, doi:10.1038/nrc1649 (2005).
- 23 Clausen, H. & Bennett, E. P. A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyl-transferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* **6**, 635-646 (1996).
- 24 Brockhausen, I., Yang, J., Dickinson, N., Ogata, S. & Itzkowitz, S. H. Enzymatic basis for sialyl-Tn expression in human colon cancer cells. *Glycoconjugate journal* **15**, 595-603 (1998).

- 25 Brockhausen, I., Yang, J., Lehotay, M., Ogata, S. & Itzkowitz, S. Pathways of mucin O-glycosylation in normal and malignant rat colonic epithelial cells reveal a mechanism for cancer-associated Sialyl-Tn antigen expression. *Biological chemistry* **382**, 219-232, doi:10.1515/BC.2001.029 (2001).
- 26 Marcos, N. T. *et al.* Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated sialyl-Tn antigen. *Cancer research* **64**, 7050-7057, doi:10.1158/0008-5472.CAN-04-1921 (2004).
- 27 Brockhausen, I., Schachter, H. & Stanley, P. in *Essentials of Glycobiology* (eds A. Varki *et al.*) (2009).
- 28 Dall'Olio, F., Malagolini, N., Trinchera, M. & Chiricolo, M. Mechanisms of cancer-associated glycosylation changes. *Frontiers in bioscience* **17**, 670-699 (2012).
- 29 Hauselmann, I. & Borsig, L. Altered tumor-cell glycosylation promotes metastasis. *Frontiers in oncology* **4**, 28, doi:10.3389/fonc.2014.00028 (2014).
- 30 Mellquist, J. L., Kasturi, L., Spitalnik, S. L. & Shakin-Eshleman, S. H. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry* **37**, 6833-6837, doi:10.1021/bi972217k (1998).
- 31 Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annual review of biochemistry* **73**, 1019-1049, doi:10.1146/annurev.biochem.73.011303.073752 (2004).
- 32 Stanley, P., Schachter, H. & Taniguchi, N. in *Essentials of Glycobiology* (eds A. Varki *et al.*) (2009).
- 33 Johnson, J. L., Jones, M. B., Ryan, S. O. & Cobb, B. A. The regulatory power of glycans and their binding partners in immunity. *Trends in immunology* **34**, 290-298, doi:10.1016/j.it.2013.01.006 (2013).
- 34 Pinho, S. S. *et al.* Gastric cancer: adding glycosylation to the equation. *Trends in molecular medicine* **19**, 664-676, doi:10.1016/j.molmed.2013.07.003 (2013).
- 35 Schauer, R. Sialic acids as regulators of molecular and cellular interactions. *Current opinion in structural biology* **19**, 507-514, doi:10.1016/j.sbi.2009.06.003 (2009).
- 36 Tsuji, S. Molecular cloning and functional analysis of sialyltransferases. *Journal of biochemistry* **120**, 1-13 (1996).
- 37 Christiansen, M. N. *et al.* Cell surface protein glycosylation in cancer. *Proteomics* **14**, 525-546, doi:10.1002/pmic.201300387 (2014).
- 38 Mondal, N. *et al.* ST3Gal-4 is the primary sialyltransferase regulating the synthesis of E-, P-, and L-selectin ligands on human myeloid leukocytes. *Blood* **125**, 687-696, doi:10.1182/blood-2014-07-588590 (2015).
- 39 Gomes, C. *et al.* Expression of ST3GAL4 leads to SLe(x) expression and induces c-Met activation and an invasive phenotype in gastric carcinoma cells. *PloS one* **8**, e66737, doi:10.1371/journal.pone.0066737 (2013).
- 40 Carvalho, A. S. *et al.* 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. *The international journal of biochemistry & cell biology* **42**, 80-89, doi:10.1016/j.biocel.2009.09.010 (2010).
- 41 Fukushi, Y. *et al.* Location and distribution of difucoganglioside (VI3NeuAcV3III3Fuc2nLc6) in normal and tumor tissues defined by its monoclonal antibody FH6. *Cancer research* **45**, 3711-3717 (1985).
- 42 Tsuchida, A. *et al.* Synthesis of disialyl Lewis a (Le(a)) structure in colon cancer cell lines by a sialyltransferase, ST6GalNAc VI, responsible for the synthesis of alpha-series gangliosides. *The Journal of biological chemistry* **278**, 22787-22794, doi:10.1074/jbc.M211034200 (2003).
- 43 Miyazaki, K. *et al.* Loss of disialyl Lewis(a), the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with

- increased sialyl Lewis(a) expression on human colon cancers. *Cancer research* **64**, 4498-4505, doi:10.1158/0008-5472.CAN-03-3614 (2004).
- 44 Delmotte, P. *et al.* Tumor necrosis factor alpha increases the expression of glycosyltransferases and sulfotransferases responsible for the biosynthesis of sialylated and/or sulfated Lewis x epitopes in the human bronchial mucosa. *The Journal of biological chemistry* **277**, 424-431, doi:10.1074/jbc.M109958200 (2002).
 - 45 Colomb, F. *et al.* TNF regulates sialyl-Lewisx and 6-sulfo-sialyl-Lewisx expression in human lung through up-regulation of ST3GAL4 transcript isoform BX. *Biochimie* **94**, 2045-2053, doi:10.1016/j.biochi.2012.05.030 (2012).
 - 46 Higai, K., Miyazaki, N., Azuma, Y. & Matsumoto, K. Interleukin-1beta induces sialyl Lewis X on hepatocellular carcinoma HuH-7 cells via enhanced expression of ST3Gal IV and FUT VI gene. *FEBS letters* **580**, 6069-6075, doi:10.1016/j.febslet.2006.09.073 (2006).
 - 47 Jun, L. *et al.* Altered mRNA expressions of sialyltransferases in human gastric cancer tissues. *Medical oncology* **29**, 84-90, doi:10.1007/s12032-010-9771-1 (2012).
 - 48 Cazet, A., Julien, S., Bobowski, M., Burchell, J. & Delannoy, P. Tumour-associated carbohydrate antigens in breast cancer. *Breast cancer research : BCR* **12**, 204, doi:10.1186/bcr2577 (2010).
 - 49 Croce, M. V. *et al.* Lewis x is highly expressed in normal tissues: a comparative immunohistochemical study and literature revision. *Pathology oncology research : POR* **13**, 130-138, doi:PAOR.2007.13.2.0130 (2007).
 - 50 Ilver, D. *et al.* Bacterium-host protein-carbohydrate interactions. *Methods in enzymology* **363**, 134-157, doi:10.1016/S0076-6879(03)01049-8 (2003).
 - 51 Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochimica et biophysica acta* **1760**, 527-537, doi:10.1016/j.bbagen.2005.12.008 (2006).
 - 52 Lehmann, F., Tiralongo, E. & Tiralongo, J. Sialic acid-specific lectins: occurrence, specificity and function. *Cellular and molecular life sciences : CMLS* **63**, 1331-1354, doi:10.1007/s00018-005-5589-y (2006).
 - 53 Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **446**, 1023-1029, doi:10.1038/nature05816 (2007).
 - 54 Dias, A. M. *et al.* Dysregulation of T cell receptor N-glycosylation: a molecular mechanism involved in ulcerative colitis. *Human molecular genetics* **23**, 2416-2427, doi:10.1093/hmg/ddt632 (2014).
 - 55 Varki, A., Freeze, H. H. & Vacquier, V. D. in *Essentials of Glycobiology* (eds A. Varki *et al.*) (2009).
 - 56 Bray, F., Jemal, A., Grey, N., Ferlay, J. & Forman, D. Global cancer transitions according to the Human Development Index (2008-2030): a population-based study. *The Lancet. Oncology* **13**, 790-801, doi:10.1016/S1470-2045(12)70211-5 (2012).
 - 57 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).
 - 58 Yao, D., Dai, C. & Peng, S. Mechanism of the mesenchymal-epithelial transition and its relationship with metastatic tumor formation. *Molecular cancer research : MCR* **9**, 1608-1620, doi:10.1158/1541-7786.MCR-10-0568 (2011).
 - 59 Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* **119**, 1420-1428, doi:10.1172/JCI39104 (2009).
 - 60 Yamaoka, M. & Nakajima, S. Prevalence of subjects at a high or very high risk of gastric cancer in Japan. *Gut and liver* **3**, 95-100, doi:10.5009/gnl.2009.3.2.95 (2009).
 - 61 La Torre, G. *et al.* Smoking status and gastric cancer risk: an updated meta-analysis of case-control studies published in the past ten years. *Tumori* **95**, 13-22 (2009).

- 62 Ramon, J. M., Serra, L., Cerdo, C. & Oromi, J. Dietary factors and gastric cancer risk. A case-control study in Spain. *Cancer* **71**, 1731-1735 (1993).
- 63 Peleteiro, B., Lopes, C., Figueiredo, C. & Lunet, N. Salt intake and gastric cancer risk according to *Helicobacter pylori* infection, smoking, tumour site and histological type. *British journal of cancer* **104**, 198-207, doi:10.1038/sj.bjc.6605993 (2011).
- 64 Correa, P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer research* **52**, 6735-6740 (1992).
- 65 D'Elia, L., Rossi, G., Ippolito, R., Cappuccio, F. P. & Strazzullo, P. Habitual salt intake and risk of gastric cancer: a meta-analysis of prospective studies. *Clinical nutrition* **31**, 489-498, doi:10.1016/j.clnu.2012.01.003 (2012).
- 66 Peek, R. M., Jr. *Helicobacter pylori* strain-specific modulation of gastric mucosal cellular turnover: implications for carcinogenesis. *Journal of gastroenterology* **37 Suppl 13**, 10-16 (2002).
- 67 Magalhaes, A., Ismail, M. N. & Reis, C. A. Sweet receptors mediate the adhesion of the gastric pathogen *Helicobacter pylori*: glycoproteomic strategies. *Expert review of proteomics* **7**, 307-310, doi:10.1586/epr.10.18 (2010).
- 68 Magalhaes, A. *et al.* *Helicobacter pylori* chronic infection and mucosal inflammation switches the human gastric glycosylation pathways. *Biochimica et biophysica acta* **1852**, 1928-1939, doi:10.1016/j.bbadis.2015.07.001 (2015).
- 69 Correa, P. & Houghton, J. Carcinogenesis of *Helicobacter pylori*. *Gastroenterology* **133**, 659-672, doi:10.1053/j.gastro.2007.06.026 (2007).
- 70 Varki, A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97-130 (1993).
- 71 Kannagi, R. Carbohydrate-mediated cell adhesion involved in hematogenous metastasis of cancer. *Glycoconjugate journal* **14**, 577-584 (1997).
- 72 Hakomori, S. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10231-10233, doi:10.1073/pnas.172380699 (2002).
- 73 Dennis, J. W., Granovsky, M. & Warren, C. E. Protein glycosylation in development and disease. *BioEssays : news and reviews in molecular, cellular and developmental biology* **21**, 412-421, doi:10.1002/(SICI)1521-1878(199905)21:5<412::AID-BIES8>3.0.CO;2-5 (1999).
- 74 Drake, P. M. *et al.* Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clinical chemistry* **56**, 223-236, doi:10.1373/clinchem.2009.136333 (2010).
- 75 Wu, X. *et al.* Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nature medicine* **10**, 518-523, doi:10.1038/nm1041 (2004).
- 76 Ju, T. *et al.* Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer research* **68**, 1636-1646, doi:10.1158/0008-5472.CAN-07-2345 (2008).
- 77 Conze, T. *et al.* MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. *Glycobiology* **20**, 199-206, doi:10.1093/glycob/cwp161 (2010).
- 78 Marcos, N. T. *et al.* ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front Biosci (Elite Ed)* **3**, 1443-1455 (2011).
- 79 Campos, D. *et al.* Probing the O-glycoproteome of gastric cancer cell lines for biomarker discovery. *Molecular & cellular proteomics : MCP* **14**, 1616-1629, doi:10.1074/mcp.M114.046862 (2015).
- 80 Cao, Y., Karsten, U., Otto, G. & Bannasch, P. Expression of MUC1, Thomsen-Friedenreich antigen, Tn, sialosyl-Tn, and alpha2,6-linked sialic acid in hepatocellular carcinomas and preneoplastic hepatocellular lesions. *Virchows Archiv : an international journal of pathology* **434**, 503-509 (1999).

- 81 Itzkowitz, S. *et al.* Expression of Tn, sialosyl Tn, and T antigens in human pancreas. *Gastroenterology* **100**, 1691-1700 (1991).
- 82 Ferreira, J. A. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Molecular oncology* **7**, 719-731, doi:10.1016/j.molonc.2013.03.001 (2013).
- 83 Kobayashi, H., Terao, T. & Kawashima, Y. Sialyl Tn as a prognostic marker in epithelial ovarian cancer. *British journal of cancer* **66**, 984-985 (1992).
- 84 Cho, S. H., Sahin, A., Hortobagyi, G. N., Hittelman, W. N. & Dhingra, K. Sialyl-Tn antigen expression occurs early during human mammary carcinogenesis and is associated with high nuclear grade and aneuploidy. *Cancer research* **54**, 6302-6305 (1994).
- 85 Gomes, C. *et al.* Glycoproteomic analysis of serum from patients with gastric precancerous lesions. *Journal of proteome research* **12**, 1454-1466, doi:10.1021/pr301112x (2013).
- 86 Nakano, M. *et al.* Site-specific analysis of N-glycans on haptoglobin in sera of patients with pancreatic cancer: a novel approach for the development of tumor markers. *International journal of cancer. Journal international du cancer* **122**, 2301-2309, doi:10.1002/ijc.23364 (2008).
- 87 Saldova, R., Reuben, J. M., Abd Hamid, U. M., Rudd, P. M. & Cristofanilli, M. Levels of specific serum N-glycans identify breast cancer patients with higher circulating tumor cell counts. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **22**, 1113-1119, doi:10.1093/annonc/mdq570 (2011).
- 88 de Leoz, M. L. *et al.* Glycomic approach for potential biomarkers on prostate cancer: profiling of N-linked glycans in human sera and pRNS cell lines. *Disease markers* **25**, 243-258 (2008).
- 89 Kim, Y. G. *et al.* Rapid and high-throughput analysis of N-glycans from ovarian cancer serum using a 96-well plate platform. *Analytical biochemistry* **391**, 151-153, doi:10.1016/j.ab.2009.05.015 (2009).
- 90 Lattova, E. *et al.* Mass spectrometric study of N-glycans from serum of woodchucks with liver cancer. *Rapid communications in mass spectrometry : RCM* **23**, 2983-2995, doi:10.1002/rcm.4202 (2009).
- 91 Pinho, S. S. *et al.* Role of E-cadherin N-glycosylation profile in a mammary tumor model. *Biochemical and biophysical research communications* **379**, 1091-1096, doi:10.1016/j.bbrc.2009.01.024 (2009).
- 92 Pinho, S. S. *et al.* Modulation of E-cadherin function and dysfunction by N-glycosylation. *Cellular and molecular life sciences : CMLS* **68**, 1011-1020, doi:10.1007/s00018-010-0595-0 (2011).
- 93 Pinho, S. S. *et al.* The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin. *Human molecular genetics* **18**, 2599-2608, doi:10.1093/hmg/ddp194 (2009).
- 94 Pinho, S. S. *et al.* Loss and recovery of Mgat3 and GnT-III Mediated E-cadherin N-glycosylation is a mechanism involved in epithelial-mesenchymal-epithelial transitions. *PloS one* **7**, e33191, doi:10.1371/journal.pone.0033191 (2012).
- 95 Girnita, L. *et al.* Inhibition of N-linked glycosylation down-regulates insulin-like growth factor-1 receptor at the cell surface and kills Ewing's sarcoma cells: therapeutic implications. *Anti-cancer drug design* **15**, 67-72 (2000).
- 96 Pili, R. *et al.* The alpha-glucosidase I inhibitor castanospermine alters endothelial cell glycosylation, prevents angiogenesis, and inhibits tumor growth. *Cancer research* **55**, 2920-2926 (1995).
- 97 Shimura, T. *et al.* Galectin-3, a novel binding partner of beta-catenin. *Cancer research* **64**, 6363-6367, doi:10.1158/0008-5472.CAN-04-1816 (2004).

- 98 Dall'Olio, F. & Chiricolo, M. Sialyltransferases in cancer. *Glycoconjugate journal* **18**, 841-850 (2001).
- 99 Varki, N. M. & Varki, A. Diversity in cell surface sialic acid presentations: implications for biology and disease. *Laboratory investigation; a journal of technical methods and pathology* **87**, 851-857, doi:10.1038/labinvest.3700656 (2007).
- 100 Cha, S. K. *et al.* Removal of sialic acid involving Klotho causes cell-surface retention of TRPV5 channel via binding to galectin-1. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9805-9810, doi:10.1073/pnas.0803223105 (2008).
- 101 Isaev, D. *et al.* Role of extracellular sialic acid in regulation of neuronal and network excitability in the rat hippocampus. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 11587-11594, doi:10.1523/JNEUROSCI.2033-07.2007 (2007).
- 102 Amado, M., Carneiro, F., Seixas, M., Clausen, H. & Sobrinho-Simoes, M. Dimeric sialyl-Le(x) expression in gastric carcinoma correlates with venous invasion and poor outcome. *Gastroenterology* **114**, 462-470 (1998).
- 103 Astronomo, R. D. & Burton, D. R. Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nature reviews. Drug discovery* **9**, 308-324, doi:10.1038/nrd3012 (2010).
- 104 Bull, C., Stoel, M. A., den Brok, M. H. & Adema, G. J. Sialic acids sweeten a tumor's life. *Cancer research* **74**, 3199-3204, doi:10.1158/0008-5472.CAN-14-0728 (2014).
- 105 Julien, S. *et al.* Selectin ligand sialyl-Lewis x antigen drives metastasis of hormone-dependent breast cancers. *Cancer research* **71**, 7683-7693, doi:10.1158/0008-5472.CAN-11-1139 (2011).
- 106 Kjeldsen, T. *et al.* 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 (1988).
- 107 Julien, S. *et al.* Sialyl-Tn vaccine induces antibody-mediated tumour protection in a relevant murine model. *British journal of cancer* **100**, 1746-1754, doi:10.1038/sj.bjc.6605083 (2009).
- 108 Ogawa, J., Inoue, H. & Koide, S. Expression of alpha-1,3-fucosyltransferase type IV and VII genes is related to poor prognosis in lung cancer. *Cancer research* **56**, 325-329 (1996).
- 109 Nakamori, S. *et al.* Increased expression of sialyl Lewisx antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study. *Cancer research* **53**, 3632-3637 (1993).
- 110 Jorgensen, T. *et al.* Up-regulation of the oligosaccharide sialyl LewisX: a new prognostic parameter in metastatic prostate cancer. *Cancer research* **55**, 1817-1819 (1995).
- 111 Tozawa, K. *et al.* Positive correlation between sialyl Lewis X expression and pathologic findings in renal cell carcinoma. *Kidney international* **67**, 1391-1396, doi:10.1111/j.1523-1755.2005.00216.x (2005).
- 112 Takahashi, S. *et al.* Overexpression of sialyl Lewis x antigen is associated with formation of extratumoral venous invasion and predicts postoperative development of massive hepatic metastasis in cases with pancreatic ductal adenocarcinoma. *Pathobiology : journal of immunopathology, molecular and cellular biology* **69**, 127-135, doi:48767 (2001).
- 113 Renkonen, J., Paavonen, T. & Renkonen, R. Endothelial and epithelial expression of sialyl Lewis(x) and sialyl Lewis(a) in lesions of breast carcinoma. *International journal of cancer. Journal international du cancer* **74**, 296-300 (1997).

- 114 Tatsumi, M. *et al.* Immunohistochemical expression of the sialyl Lewis x antigen on gastric cancer cells correlates with the presence of liver metastasis. *Clinical & experimental metastasis* **16**, 743-750 (1998).
- 115 Wang, F. L. *et al.* High expression of alpha 2, 3-linked sialic acid residues is associated with the metastatic potential of human gastric cancer. *Cancer detection and prevention* **32**, 437-443, doi:10.1016/j.cdp.2009.01.001 (2009).
- 116 Stowell, S. R., Ju, T. & Cummings, R. D. Protein glycosylation in cancer. *Annual review of pathology* **10**, 473-510, doi:10.1146/annurev-pathol-012414-040438 (2015).
- 117 Mueller, M. M. & Fusenig, N. E. Friends or foes - bipolar effects of the tumour stroma in cancer. *Nature reviews. Cancer* **4**, 839-849, doi:10.1038/nrc1477 (2004).
- 118 Orian-Rousseau, V. CD44, a therapeutic target for metastasising tumours. *European journal of cancer* **46**, 1271-1277, doi:10.1016/j.ejca.2010.02.024 (2010).
- 119 Freire-de-Lima, L. *et al.* Involvement of O-glycosylation defining oncofetal fibronectin in epithelial-mesenchymal transition process. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17690-17695, doi:10.1073/pnas.1115191108 (2011).
- 120 Ono, M. & Hakomori, S. Glycosylation defining cancer cell motility and invasiveness. *Glycoconjugate journal* **20**, 71-78, doi:10.1023/B:GLYC.0000018019.22070.7d (2004).
- 121 Thiery, J. P. & Sleeman, J. P. Complex networks orchestrate epithelial-mesenchymal transitions. *Nature reviews. Molecular cell biology* **7**, 131-142, doi:10.1038/nrm1835 (2006).
- 122 Wirtz, D., Konstantopoulos, K. & Searson, P. C. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. *Nature reviews. Cancer* **11**, 512-522, doi:10.1038/nrc3080 (2011).
- 123 Reymond, N., d'Agua, B. B. & Ridley, A. J. Crossing the endothelial barrier during metastasis. *Nature reviews. Cancer* **13**, 858-870, doi:10.1038/nrc3628 (2013).
- 124 Jeschke, U. *et al.* Expression of sialyl Lewis X, sialyl Lewis A, E-cadherin and cathepsin-D in human breast cancer: immunohistochemical analysis in mammary carcinoma in situ, invasive carcinomas and their lymph node metastasis. *Anticancer research* **25**, 1615-1622 (2005).
- 125 Varki, A. & Schauer, R. in *Essentials of Glycobiology* (eds A. Varki *et al.*) (2009).
- 126 McEver, R. P. Selectin-carbohydrate interactions during inflammation and metastasis. *Glycoconjugate journal* **14**, 585-591 (1997).
- 127 Ugorski, M. & Laskowska, A. Sialyl Lewis(a): a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta biochimica Polonica* **49**, 303-311 (2002).
- 128 Hidalgo, A., Peired, A. J., Wild, M. K., Vestweber, D. & Frenette, P. S. Complete identification of E-selectin ligands on neutrophils reveals distinct functions of PSGL-1, ESL-1, and CD44. *Immunity* **26**, 477-489, doi:10.1016/j.immuni.2007.03.011 (2007).
- 129 Dimitroff, C. J., Lee, J. Y., Rafii, S., Fuhlbrigge, R. C. & Sackstein, R. CD44 is a major E-selectin ligand on human hematopoietic progenitor cells. *The Journal of cell biology* **153**, 1277-1286 (2001).
- 130 Napier, S. L., Healy, Z. R., Schnaar, R. L. & Konstantopoulos, K. Selectin ligand expression regulates the initial vascular interactions of colon carcinoma cells: the roles of CD44v and alternative sialofucosylated selectin ligands. *The Journal of biological chemistry* **282**, 3433-3441, doi:10.1074/jbc.M607219200 (2007).
- 131 Thomas, S. N., Zhu, F., Schnaar, R. L., Alves, C. S. & Konstantopoulos, K. Carcinoembryonic antigen and CD44 variant isoforms cooperate to mediate colon carcinoma cell adhesion to E- and L-selectin in shear flow. *The Journal of biological chemistry* **283**, 15647-15655, doi:10.1074/jbc.M800543200 (2008).

- 132 Barthel, S. R. *et al.* Alpha 1,3 fucosyltransferases are master regulators of prostate cancer cell trafficking. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19491-19496, doi:10.1073/pnas.0906074106 (2009).
- 133 Yusa, A., Miyazaki, K., Kimura, N., Izawa, M. & Kannagi, R. Epigenetic silencing of the sulfate transporter gene DTDST induces sialyl Lewisx expression and accelerates proliferation of colon cancer cells. *Cancer research* **70**, 4064-4073, doi:10.1158/0008-5472.CAN-09-2383 (2010).
- 134 Fukuoka, K., Narita, N. & Saijo, N. Increased expression of sialyl Lewis(x) antigen is associated with distant metastasis in lung cancer patients: immunohistochemical study on bronchofiberscopic biopsy specimens. *Lung cancer* **20**, 109-116 (1998).
- 135 Sreaton, G. R. *et al.* Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 12160-12164 (1992).
- 136 Ponta, H., Sherman, L. & Herrlich, P. A. CD44: from adhesion molecules to signalling regulators. *Nature reviews. Molecular cell biology* **4**, 33-45, doi:10.1038/nrm1004 (2003).
- 137 Naor, D., Sionov, R. V. & Ish-Shalom, D. CD44: structure, function, and association with the malignant process. *Advances in cancer research* **71**, 241-319 (1997).
- 138 Muller, W. *et al.* Expression and prognostic value of the CD44 splicing variants v5 and v6 in gastric cancer. *The Journal of pathology* **183**, 222-227, doi:10.1002/(SICI)1096-9896(199710)183:2<222::AID-PATH923>3.0.CO;2-C (1997).
- 139 Wielenga, V. J. *et al.* Expression of CD44 variant proteins in human colorectal cancer is related to tumor progression. *Cancer research* **53**, 4754-4756 (1993).
- 140 Cho, S. H. *et al.* CD44 enhances the epithelial-mesenchymal transition in association with colon cancer invasion. *International journal of oncology* **41**, 211-218, doi:10.3892/ijo.2012.1453 (2012).
- 141 Gvozdenovic, A. *et al.* CD44 enhances tumor formation and lung metastasis in experimental osteosarcoma and is an additional predictor for poor patient outcome. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **28**, 838-847, doi:10.1002/jbmr.1817 (2013).
- 142 Mayer, B. *et al.* De-novo expression of CD44 and survival in gastric cancer. *Lancet* **342**, 1019-1022 (1993).
- 143 Hong, R. L., Lee, W. J., Shun, C. T., Chu, J. S. & Chen, Y. C. Expression of CD44 and its clinical implication in diffuse-type and intestinal-type gastric adenocarcinomas. *Oncology* **52**, 334-339 (1995).
- 144 Yamamichi, K., Uehara, Y., Kitamura, N., Nakane, Y. & Hioki, K. Increased expression of CD44v6 mRNA significantly correlates with distant metastasis and poor prognosis in gastric cancer. *International journal of cancer. Journal international du cancer* **79**, 256-262 (1998).
- 145 Naor, D., Nedvetzki, S., Golan, I., Melnik, L. & Faitelson, Y. CD44 in cancer. *Critical reviews in clinical laboratory sciences* **39**, 527-579, doi:10.1080/10408360290795574 (2002).
- 146 Nestl, A. *et al.* Gene expression patterns associated with the metastatic phenotype in rodent and human tumors. *Cancer research* **61**, 1569-1577 (2001).
- 147 Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. *The Journal of cell biology* **196**, 395-406, doi:10.1083/jcb.201102147 (2012).
- 148 Yurchenco, P. D. & Schittny, J. C. Molecular architecture of basement membranes. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **4**, 1577-1590 (1990).

- 149 Murdoch, A. D., Liu, B., Schwarting, R., Tuan, R. S. & Iozzo, R. V. Widespread expression of perlecan proteoglycan in basement membranes and extracellular matrices of human tissues as detected by a novel monoclonal antibody against domain III and by in situ hybridization. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **42**, 239-249 (1994).
- 150 Mishra, M., Naik, V. V., Kale, A. D., Ankola, A. V. & Pilli, G. S. Perlecan (basement membrane heparan sulfate proteoglycan) and its role in oral malignancies: an overview. *Indian journal of dental research : official publication of Indian Society for Dental Research* **22**, 823-826, doi:10.4103/0970-9290.94678 (2011).
- 151 Iozzo, R. V. Basement membrane proteoglycans: from cellar to ceiling. *Nature reviews. Molecular cell biology* **6**, 646-656, doi:10.1038/nrm1702 (2005).
- 152 Datta, S., Pierce, M. & Datta, M. W. Perlecan signaling: helping hedgehog stimulate prostate cancer growth. *The international journal of biochemistry & cell biology* **38**, 1855-1861, doi:10.1016/j.biocel.2006.03.022 (2006).
- 153 Cohen, I. R. *et al.* Abnormal expression of perlecan proteoglycan in metastatic melanomas. *Cancer research* **54**, 5771-5774 (1994).
- 154 Sharma, B. *et al.* Antisense targeting of perlecan blocks tumor growth and angiogenesis in vivo. *The Journal of clinical investigation* **102**, 1599-1608, doi:10.1172/JCI3793 (1998).
- 155 Nerlich, A. G., Wiest, I., Wagner, E., Sauer, U. & Schleicher, E. D. Gene expression and protein deposition of major basement membrane components and TGF-beta 1 in human breast cancer. *Anticancer research* **17**, 4443-4449 (1997).
- 156 Datta, M. W. *et al.* Perlecan, a candidate gene for the CAPB locus, regulates prostate cancer cell growth via the Sonic Hedgehog pathway. *Molecular cancer* **5**, 9, doi:10.1186/1476-4598-5-9 (2006).
- 157 Ychou, M., Duffour, J., Kramar, A., Gourgou, S. & Grenier, J. Clinical significance and prognostic value of CA72-4 compared with CEA and CA19-9 in patients with gastric cancer. *Disease markers* **16**, 105-110 (2000).
- 158 Tocchi, A. *et al.* The role of serum and gastric juice levels of carcinoembryonic antigen, CA19.9 and CA72.4 in patients with gastric cancer. *Journal of cancer research and clinical oncology* **124**, 450-455 (1998).
- 159 Gaspar, M. J., Arribas, I., Coca, M. C. & Diez-Alonso, M. Prognostic value of carcinoembryonic antigen, CA 19-9 and CA 72-4 in gastric carcinoma. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* **22**, 318-322, doi:50633 (2001).
- 160 Carpelan-Holmstrom, M., Louhimo, J., Stenman, U. H., Alfthan, H. & Haglund, C. CEA, CA 19-9 and CA 72-4 improve the diagnostic accuracy in gastrointestinal cancers. *Anticancer research* **22**, 2311-2316 (2002).
- 161 Haga, Y. *et al.* Clinical significance of serum CA125 values in patients with cancers of the digestive system. *The American journal of the medical sciences* **292**, 30-34 (1986).
- 162 Bones, J. *et al.* Glycomic and glycoproteomic analysis of serum from patients with stomach cancer reveals potential markers arising from host defense response mechanisms. *Journal of proteome research* **10**, 1246-1265, doi:10.1021/pr101036b (2011).
- 163 Hauck, W. & Stanners, C. P. Control of carcinoembryonic antigen gene family expression in a differentiating colon carcinoma cell line, Caco-2. *Cancer research* **51**, 3526-3533 (1991).
- 164 Liu, L. *et al.* The identification and characterization of novel N-glycan-based biomarkers in gastric cancer. *PloS one* **8**, e77821, doi:10.1371/journal.pone.0077821 (2013).

- 165 Berinstein, N. L. Carcinoembryonic antigen as a target for therapeutic anticancer vaccines: a review. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **20**, 2197-2207 (2002).
- 166 Paxton, R. J., Mooser, G., Pande, H., Lee, T. D. & Shively, J. E. Sequence analysis of carcinoembryonic antigen: identification of glycosylation sites and homology with the immunoglobulin supergene family. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 920-924 (1987).
- 167 Yamashita, K. *et al.* Structural studies of the carbohydrate moieties of carcinoembryonic antigens. *Cancer research* **47**, 3451-3459 (1987).
- 168 Westwood, J. H., Bessell, E. M., Bukhari, M. A., Thomas, P. & Walker, J. M. Studies on the structure of the carcinoembryonic antigen. I. some deductions on the basis of chemical degradations. *Immunochemistry* **11**, 811-818 (1974).
- 169 Hammarstrom, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Seminars in cancer biology* **9**, 67-81, doi:10.1006/scbi.1998.0119 (1999).
- 170 Hammarstrom, S. & Baranov, V. Is there a role for CEA in innate immunity in the colon? *Trends in microbiology* **9**, 119-125 (2001).
- 171 Muenzner, P., Rohde, M., Kneitz, S. & Hauck, C. R. CEACAM engagement by human pathogens enhances cell adhesion and counteracts bacteria-induced detachment of epithelial cells. *The Journal of cell biology* **170**, 825-836, doi:10.1083/jcb.200412151 (2005).
- 172 Benchimol, S. *et al.* Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell* **57**, 327-334 (1989).
- 173 Beauchemin, N. & Arabzadeh, A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer metastasis reviews* **32**, 643-671, doi:10.1007/s10555-013-9444-6 (2013).
- 174 Ergun, S. *et al.* CEA-related cell adhesion molecule 1: a potent angiogenic factor and a major effector of vascular endothelial growth factor. *Molecular cell* **5**, 311-320 (2000).
- 175 Singer, B. B. *et al.* Deregulation of the CEACAM expression pattern causes undifferentiated cell growth in human lung adenocarcinoma cells. *PloS one* **5**, e8747, doi:10.1371/journal.pone.0008747 (2010).
- 176 Klaile, E., Muller, M. M., Kannicht, C., Singer, B. B. & Lucka, L. CEACAM1 functionally interacts with filamin A and exerts a dual role in the regulation of cell migration. *Journal of cell science* **118**, 5513-5524, doi:10.1242/jcs.02660 (2005).
- 177 Muller, M. M., Singer, B. B., Klaile, E., Obrink, B. & Lucka, L. Transmembrane CEACAM1 affects integrin-dependent signaling and regulates extracellular matrix protein-specific morphology and migration of endothelial cells. *Blood* **105**, 3925-3934, doi:10.1182/blood-2004-09-3618 (2005).
- 178 Ordonez, C., Screaton, R. A., Ilantzis, C. & Stanners, C. P. Human carcinoembryonic antigen functions as a general inhibitor of anoikis. *Cancer research* **60**, 3419-3424 (2000).
- 179 Jessup, J. M., Samara, R., Battle, P. & Laguinge, L. M. Carcinoembryonic antigen promotes tumor cell survival in liver through an IL-10-dependent pathway. *Clinical & experimental metastasis* **21**, 709-717 (2004).
- 180 Gold, P. & Freedman, S. O. Demonstration of Tumor-Specific Antigens in Human Colonic Carcinomata by Immunological Tolerance and Absorption Techniques. *The Journal of experimental medicine* **121**, 439-462 (1965).
- 181 Nakane, Y. *et al.* Correlation of preoperative carcinoembryonic antigen levels and prognosis of gastric cancer patients. *Cancer* **73**, 2703-2708 (1994).
- 182 Laguinge, L. *et al.* Surface expression and CEA binding of hnRNP M4 protein in HT29 colon cancer cells. *Anticancer research* **25**, 23-31 (2005).

- 183 Dikken, J. L. *et al.* Treatment of resectable gastric cancer. *Therapeutic advances in gastroenterology* **5**, 49-69, doi:10.1177/1756283X11410771 (2012).
- 184 Scheele, J., Stang, R., Altendorf-Hofmann, A. & Paul, M. Resection of colorectal liver metastases. *World journal of surgery* **19**, 59-71 (1995).
- 185 Thomas, P., Toth, C. A., Saini, K. S., Jessup, J. M. & Steele, G., Jr. The structure, metabolism and function of the carcinoembryonic antigen gene family. *Biochimica et biophysica acta* **1032**, 177-189 (1990).
- 186 Thomas, P. & Zamcheck, N. Role of the liver in clearance and excretion of circulating carcinoembryonic antigen (CEA). *Digestive diseases and sciences* **28**, 216-224 (1983).
- 187 Gangopadhyay, A., Bajenova, O., Kelly, T. M. & Thomas, P. Carcinoembryonic antigen induces cytokine expression in Kupffer cells: implications for hepatic metastasis from colorectal cancer. *Cancer research* **56**, 4805-4810 (1996).
- 188 Thomas, P., Forse, R. A. & Bajenova, O. Carcinoembryonic antigen (CEA) and its receptor hnRNP M are mediators of metastasis and the inflammatory response in the liver. *Clinical & experimental metastasis* **28**, 923-932, doi:10.1007/s10585-011-9419-3 (2011).
- 189 Bajenova, O. V. *et al.* Heterogeneous RNA-binding protein M4 is a receptor for carcinoembryonic antigen in Kupffer cells. *The Journal of biological chemistry* **276**, 31067-31073, doi:10.1074/jbc.M104093200 (2001).
- 190 Bajenova, O. *et al.* Surface expression of heterogeneous nuclear RNA binding protein M4 on Kupffer cell relates to its function as a carcinoembryonic antigen receptor. *Experimental cell research* **291**, 228-241 (2003).
- 191 Xu, Y. *et al.* Cell type-restricted activity of hnRNPM promotes breast cancer metastasis via regulating alternative splicing. *Genes & development* **28**, 1191-1203, doi:10.1101/gad.241968.114 (2014).
- 192 Lord, M. S. *et al.* The role of vascular-derived perlecan in modulating cell adhesion, proliferation and growth factor signaling. *Matrix biology : journal of the International Society for Matrix Biology* **35**, 112-122, doi:10.1016/j.matbio.2014.01.016 (2014).
- 193 Banky, B. *et al.* Characteristics of CD44 alternative splice pattern in the course of human colorectal adenocarcinoma progression. *Molecular cancer* **11**, 83, doi:10.1186/1476-4598-11-83 (2012).
- 194 Stocks, S. C. & Kerr, M. A. Neutrophil NCA-160 (CD66) is the major protein carrier of selectin binding carbohydrate groups LewisX and sialyl lewisX. *Biochemical and biophysical research communications* **195**, 478-483, doi:10.1006/bbrc.1993.2068 (1993).
- 195 Lucka, L. *et al.* Identification of Lewis x structures of the cell adhesion molecule CEACAM1 from human granulocytes. *Glycobiology* **15**, 87-100, doi:10.1093/glycob/cwh139 (2005).
- 196 Garcia, M. *et al.* Carcinoembryonic antigen has a different molecular weight in normal colon and in cancer cells due to N-glycosylation differences. *Cancer research* **51**, 5679-5686 (1991).
- 197 Oikawa, S., Nakazato, H. & Kosaki, G. Primary structure of human carcinoembryonic antigen (CEA) deduced from cDNA sequence. *Biochemical and biophysical research communications* **142**, 511-518 (1987).
- 198 Hatakeyama, K. *et al.* Novel protein isoforms of carcinoembryonic antigen are secreted from pancreatic, gastric and colorectal cancer cells. *BMC research notes* **6**, 381, doi:10.1186/1756-0500-6-381 (2013).
- 199 Thompson, J. *et al.* CGM2, a member of the carcinoembryonic antigen gene family is down-regulated in colorectal carcinomas. *The Journal of biological chemistry* **269**, 32924-32931 (1994).

- 200 Baranov, V., Yeung, M. M. & Hammarstrom, S. Expression of carcinoembryonic antigen and nonspecific cross-reacting 50-kDa antigen in human normal and cancerous colon mucosa: comparative ultrastructural study with monoclonal antibodies. *Cancer research* **54**, 3305-3314 (1994).
- 201 Kinugasa, T. *et al.* Expression of four CEA family antigens (CEA, NCA, BGP and CGM2) in normal and cancerous gastric epithelial cells: up-regulation of BGP and CGM2 in carcinomas. *International journal of cancer. Journal international du cancer* **76**, 148-153 (1998).
- 202 Shi, Z. R., Tacha, D. & Itzkowitz, S. H. Monoclonal antibody COL-1 reacts with restricted epitopes on carcinoembryonic antigen: an immunohistochemical study. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **42**, 1215-1219 (1994).
- 203 Seales, E. C. *et al.* Hypersialylation of beta1 integrins, observed in colon adenocarcinoma, may contribute to cancer progression by up-regulating cell motility. *Cancer research* **65**, 4645-4652, doi:10.1158/0008-5472.CAN-04-3117 (2005).
- 204 Seales, E. C. *et al.* A protein kinase C/Ras/ERK signaling pathway activates myeloid fibronectin receptors by altering beta1 integrin sialylation. *The Journal of biological chemistry* **280**, 37610-37615, doi:10.1074/jbc.M508476200 (2005).
- 205 Danilkovitch-Miagkova, A. & Zbar, B. Dysregulation of Met receptor tyrosine kinase activity in invasive tumors. *The Journal of clinical investigation* **109**, 863-867, doi:10.1172/JCI15418 (2002).
- 206 Ma, P. C., Maulik, G., Christensen, J. & Salgia, R. c-Met: structure, functions and potential for therapeutic inhibition. *Cancer metastasis reviews* **22**, 309-325 (2003).
- 207 Takeuchi, H. *et al.* c-MET expression level in primary colon cancer: a predictor of tumor invasion and lymph node metastases. *Clinical cancer research : an official journal of the American Association for Cancer Research* **9**, 1480-1488 (2003).
- 208 Jung, M. *et al.* Mast cells produce novel shorter forms of perlecan that contain functional endorepellin: a role in angiogenesis and wound healing. *The Journal of biological chemistry* **288**, 3289-3304, doi:10.1074/jbc.M112.387811 (2013).
- 209 Noonan, D. M. *et al.* The complete sequence of perlecan, a basement membrane heparan sulfate proteoglycan, reveals extensive similarity with laminin A chain, low density lipoprotein-receptor, and the neural cell adhesion molecule. *The Journal of biological chemistry* **266**, 22939-22947 (1991).
- 210 Cho, W., Jung, K. & Regnier, F. E. Sialylated Lewis x antigen bearing glycoproteins in human plasma. *Journal of proteome research* **9**, 5960-5968, doi:10.1021/pr100747p (2010).
- 211 Yokozaki, H. *et al.* Expression of CD44 abnormal transcripts in human gastric carcinomas. *Cancer letters* **83**, 229-234 (1994).
- 212 Orian-Rousseau, V., Chen, L., Sleeman, J. P., Herrlich, P. & Ponta, H. CD44 is required for two consecutive steps in HGF/c-Met signaling. *Genes & development* **16**, 3074-3086, doi:10.1101/gad.242602 (2002).
- 213 Zoller, M. CD44: can a cancer-initiating cell profit from an abundantly expressed molecule? *Nature reviews. Cancer* **11**, 254-267, doi:10.1038/nrc3023 (2011).
- 214 Hanley, W. D. *et al.* Variant isoforms of CD44 are P- and L-selectin ligands on colon carcinoma cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **20**, 337-339, doi:10.1096/fj.05-4574fje (2006).
- 215 Lim, K. T., Miyazaki, K., Kimura, N., Izawa, M. & Kannagi, R. Clinical application of functional glycoproteomics - dissection of glycotopes carried by soluble CD44 variants in sera of patients with cancers. *Proteomics* **8**, 3263-3273, doi:10.1002/pmic.200800147 (2008).

SUPPLEMENTARY DATA

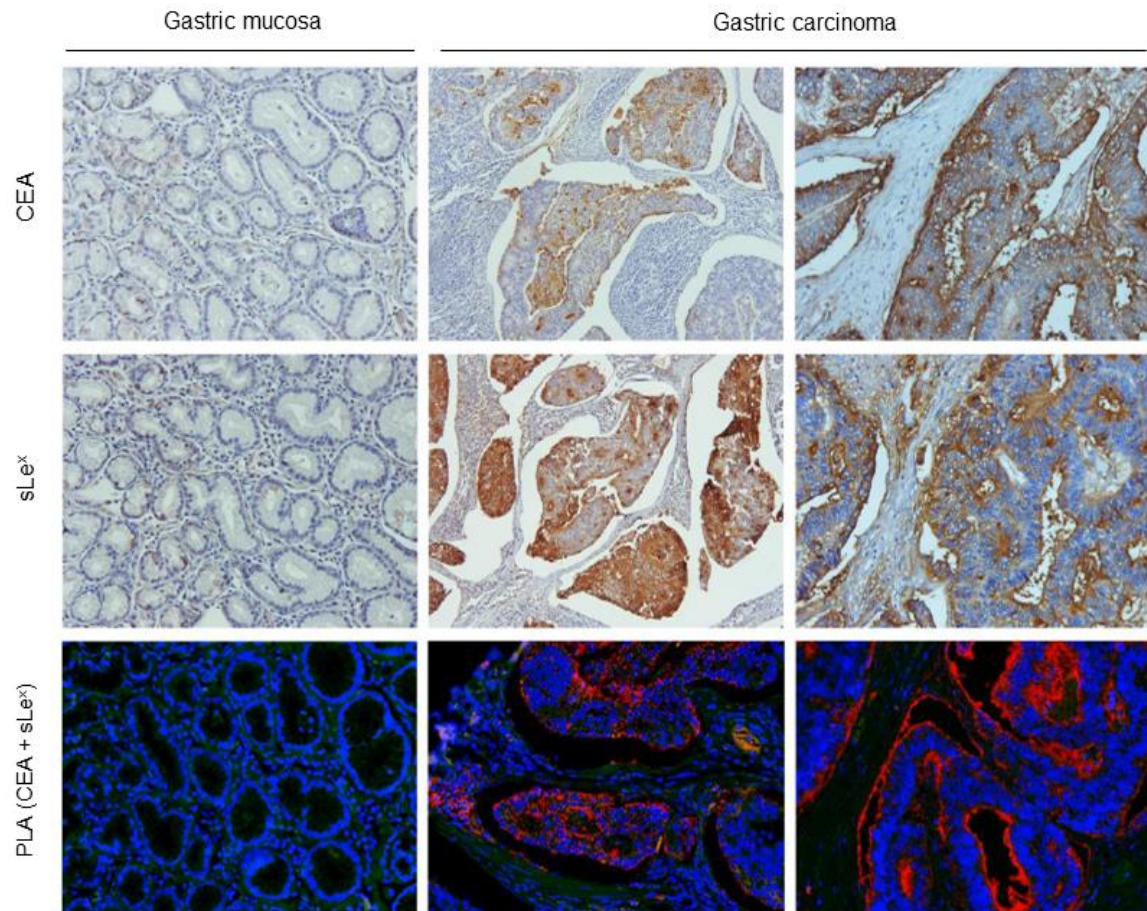
Material and methods: immunohistochemistry analysis

Gastric carcinomas and gastric mucosa adjacent to carcinomas were obtained from individuals undergoing surgery at Centro Hospitalar São João (CHSJ), University of Porto Medical Faculty (Porto, Portugal). Study was performed with the approval of the local CHSJ ethical committee. Gastric carcinoma tissues (n=31) formalin-fixed paraffin-embedded were used for slide sections.

Paraffin sections were dewaxed, rehydrated and blocked for endogenous peroxidase activity with 3% H₂O₂ in methanol for 30 min. Sections were then incubated with normal rabbit or swine serum diluted 1:5 in PBS containing 10% BSA for 30 min followed by incubation with the monoclonal antibodies overnight at 4°C. Incubation with biotinylated rabbit anti-mouse (DAKO) was done during 30 min at room temperature followed by avidin/biotin complex detection (Vectastain). Staining was performed with 3,3'-diaminobenzidine tetrahydrochloride from Sigma containing 0.02% hydrogen peroxide and counter staining of the nucleus was done with Mayer's hematoxylin. Mouse monoclonal antibodies used KM93 1:60, anti-CEA (CB30) 1:300 with antigen retrieval with citrate buffer pH 6.0.

Material and methods: Proximity ligation assay (PLA)

PLA assay was performed in cultured cells on glass coverslips and on gastric carcinoma tissue sections for detection of ligation dependent SLe^x on CEA glycoproteins. Briefly, paraffin tissue sections were dewaxed and rehydrated followed by antigen retrieval with citrate buffer pH 6.0. Tissue sections and cells on glass coverslips were then incubated with normal goat serum diluted 1:5 in PBS containing 10% BSA for 30 min followed by incubation with the two monoclonal antibodies in the same solution overnight at 4°C. Proximity ligation assays (PLA) were performed using the DuoLink® II Fluorescence Kit (Olink® Bioscience, Uppsala, Sweden) according to the manufacturer's instructions. PLA probes anti-IgG plus and anti-IgM minus were incubated 1 h at 37°C, followed by ligation step 30 min at 37°C and amplification step 100 min at 37°C. Nuclei were stained with DAPI and slides were mounted in an appropriated medium (Duolink II). PLA products are seen as fluorescent red dots. Fluorescence was examined in a fluorescence microscopy and images were acquired using a Zeiss Axio cam MRm and the AxioVision Rel. 4.8 software (Carl ZEISS) and in some cases immersion oil was used.



Supplementary figure 1 – CEA is a sLe^x carrier in gastric carcinoma tissues. Immunohistochemical evaluation of the expression of sLe^x and CEA in gastric carcinoma tissues. The figure illustrates a normal gastric mucosa (left column) and two gastric carcinomas. The result shows CEA and sLe^x expression in gastric carcinoma tissues and no expression in normal gastric mucosa. PLA signal was only observed in gastric carcinoma (200x magnification).

Role of carcinoembryonic antigen (CEA) glycosylation in gastric cancer

Supplementary table 1 – Expression of CEA and sLe^x in gastric carcinoma tissues and expression of CEA/sLe^x by PLA.

	CEA				SLe ^x				PLA signal		
	0-25%	25-75%	>75%	P*	0-25%	25-75%	>75%	P*	Neg.	Pos.	P*
Lauren classification											
Intestinal (n=13)	1 (7.7%)	2 (15.4%)	10 (76.9%)	ns	1 (7.7%)	9 (69.2%)	3 (23.1%)	ns	1 (7.7%)	12 (92.3%)	ns
Diffuse (n=2)	0 (0%)	1 (50%)	1 (50%)		0 (0%)	2 (100%)	0 (0%)		1 (50%)	1 (50%)	
Unclassified (n=16)	1 (6.2%)	5 (31.3%)	10 (62.5%)		5 (25%)	6 (37.5%)	6 (37.5%)		4 (25%)	12 (75%)	
Carneiro classification											
Glandular (n=13)	1 (7.7%)	2 (15.4%)	10 (76.9%)	ns	1 (7.7%)	9 (69.2%)	3 (23.1%)	ns	1 (7.7%)	12 (92.3%)	0.069
Mixed (n=15)	1 (6.7%)	5 (33.3%)	9 (60%)		3 (20%)	6 (40%)	6 (40%)		3 (20%)	12 (80%)	
Others (Isolated cells + solid) (n=3)	0 (0%)	1 (33.3%)	2 (66.7%)		1 (33.3%)	2 (66.7%)	0 (0%)		2 (66.7%)	1 (33.3%)	
Ming classification											
Expansive (n=8)	1 (12.5%)	2 (25%)	5 (62.5%)	ns	4 (40%)	3 (37.5%)	1 (12.5%)	0.04	4 (50%)	4 (50%)	0.038
Infiltrative (n=22)	1 (4.5%)	5 (22.8%)	16 (72.7%)		1 (4.5%)	13 (59.1%)	8 (36.4%)		2 (9.1%)	20 (90.9%)	
Unclassified (n=1)	0 (0%)	1 (100%)	0 (0%)		0 (0%)	1 (100%)	0 (0%)		0 (0%)	1 (100%)	
OMS classification											
Well-differentiated (n=15)	2 (13.3%)	2 (13.4%)	11 (73.3%)	ns	2 (13.3%)	9 (60%)	4 (26.7%)	ns	2 (13.3%)	13 (86.7%)	ns
Others (n=16)	0 (0%)	6 (37.5%)	10 (62.5%)		3 (18.75%)	8 (50%)	5 (31.25%)		4 (25%)	12 (75%)	
Lymphatic invasion											
Present (n=24)	1 (4.2%)	4 (16.6%)	19 (79.2%)	0.042	4 (16.7%)	13 (54.1%)	7 (29.2%)	ns	4 (16.7%)	20 (83.3%)	ns
Absent (n=7)	1 (14.3%)	4 (57.1%)	2 (28.6%)		1 (14.3%)	4 (57.1%)	2 (28.6%)		2 (28.6%)	5 (71.4%)	
Venous invasion											
Present (n=19)	0 (0%)	3 (15.8%)	16 (84.2%)	0.029	1 (5.3%)	12 (63.1%)	6 (31.6%)	ns	1 (5.3%)	18 (94.7%)	0.012
Absent (n=12)	2 (16.7%)	5 (41.7%)	5 (41.7%)		4 (33.3%)	5 (41.7%)	3 (25%)		5 (41.7%)	7 (58.3%)	
pTNM											
I+II (n=16)	1 (6.25%)	7 (43.75%)	8 (50%)	0.058	4 (25%)	8 (50%)	4 (25%)	ns	5 (31.25%)	11 (68.75%)	0.08
III+IV (n=15)	1 (6.7%)	1 (6.7%)	13 (86.6%)		1 (6.7%)	9 (60%)	5 (33.3%)		1 (6.7%)	14 (93.3%)	
Total	2 (6.5%)	8 (25.8%)	21 (67.7%)		5 (16.1%)	17 (54.9%)	9 (29%)		6 (19.4%)	25 (80.6%)	