U.PORTO

DOUTORAMENTO
CIÊNCIAS BIOMÉDICAS

Circulating tumour cells glycosylation for improved esophageal and colorectal cancer care: a step towards precision oncology

Sofia Ribeiro Cotton

D 2024

U.PORTO

Sofia Ribeiro Cotton. Circulating tumour cells glycosylation for improved esophageal and colorectal cancer care: a step towards precision oncology

D.ICBAS 2024

Circulating tumour cells glycosylation for improved esophageal and colorectal cancer care: a step towards precision oncology

Sofia Ribeiro Cotton

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

SOFIA RIBEIRO COTTON

CIRCULATING TUMOUR CELLS GLYCOSYLATION FOR IMPROVED ESOPHAGEAL AND COLORECTAL CANCER CARE: A STEP TOWARDS PRECISION ONCOLOGY

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

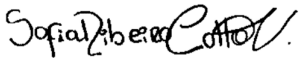
Orientador: José Alexandre Ribeiro de Castro Ferreira
Categoria: Investigador Principal
Afiliação: Instituto Português de Oncologia do Porto.

Co-orientador: Lúcio José de Lara Santos
Categoria: Professor Catedrático Convidado
Afiliação: Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Co-orientador: Luís Oliveira Lima
Categoria: Investigador Júnior
Afiliação: Instituto Português de Oncologia do Porto.

DECLARAÇÃO DE HONRA

Declaro que a presente tese é de minha autoria e não foi utilizada previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referenciação. Tenho consciência de que a prática de plágio e autoplágio constitui um ilícito académico.



Sofia Ribeiro Cotton

***Ao meu avô**.*

NOTA PRELIMINAR

Declaro que este trabalho é integralmente da minha autoria, estando devidamente referenciadas e identificadas as fontes de informação e as obras consultadas. Não contém, por isso, qualquer tipo de plágio de textos publicados ou trabalhos académicos, qualquer que seja o meio dessas publicações.

Os trabalhos de investigação conducentes a esta tese foram realizados no Grupo de Patologia e Terapêutica Experimental do Centro de Investigação do Instituto Português de Oncologia do Porto, EPE.

Declaro ainda que o presente trabalho foi financiado pela Fundação para a Ciência e a Tecnologia, pela atribuição de uma bolsa individual de doutoramento, com a referência 2020. 09394.BD. A Fundação para a Ciência e Tecnologia é co-financiada pelo Fundo Europeu de Desenvolvimento Regional da União Europeia, através do Programa Operacional Fatores de Competividade do Quadro de Referência Estratégico Nacional. Reconheço também o financiamento, pelo Norte 2020 com o apoio da união europeia do programa ESTIMA com o projeto: Early-stage cancer treatment in the context of molecular imaging, do Centro de Investigação do Instituto Português de Oncologia do Porto (PEst-OE/SAU/UI0776/201; CI-IPOP-29-2014; CI-IPOP-58-2015). Por último, delaro que o presente trabalho teve ainda o suporte financeiro do Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

De acordo com o disposto no ponto 2 do artigo 31 do Decreto-Lei n.º 65/2018, de 16 de agosto, o autor declara que na elaboração desta tese foram incluídos dados das publicações abaixo indicadas. O autor participou ativamente na conceção e execução dos trabalhos que estiveram na origem dos mesmos, assim como na sua interpretação, discussão e redação.

According to the relevant national legislation, the author declares that this thesis includes data from the publications indicated below. The author participated actively in the conception and execution of the work that originated that data, as well as in their interpretation, discussion, and writing.

The following articles were used to support this thesis:

Journal articles

a. **Cotton, Sofia**; Ferreira, Dylan; Soares, Janine; Peixoto, Andreia; Relvas-Santos, Marta; Azevedo, Rita; Piairo, Paulina; et al. "Target Score—A Proteomics Data Selection Tool Applied to Esophageal Cancer Identifies GLUT1-Sialyl Tn Glycoforms as Biomarkers of Cancer Aggressiveness". International Journal of Molecular Sciences 22 4 (2021): 1664. http://dx.doi.org/10.3390/ijms22041664.
   Authorship credit: conceptualization, methodology, formal analysis, investigation, writing – original draft
   This publication was not and will not use partially or totally in any other thesis or dissertation. This publication was integrally included in the present thesis (page 74-114). Non-commercial uses of the work are permitted without any further permission from Creative Commons atributions (CC BY), license (https://creativecommons.org/licenses/by/ 4.0/).

b. **Cotton, Sofia**; Ferreira, Dylan; Relvas-Santos, Marta; Brandão, Andreia; Afonso, Luís Pedro; Miranda, Andreia; Ferreira, Eduardo; et al. “E-selectin Affinity Glycoproteomics Reveals Neuroendocrine Proteins and the Secretin Receptor as a Poor Prognosis Signature in Colorectal Cancer”. Molecular Oncology (submitted).
   Authorship credit: conceptualization, methodology, formal analysis, investigation, writing – original draft
   This publication was not and will not use partially or totally in any other thesis or dissertation. This publication was integrally included in the present thesis (pages 116-152).

AGRADECIMENTOS

A minha tese de doutoramento que é agora um trabalho finalizado, foi apenas possível de ser concretizado pela colaboração de inúmeras pessoas e entidades às quais não posso deixar de agradecer.

Por isso, em primeiro lugar agradeço à Universidade do Porto, Faculdades de Ciências e Instituto de Ciências Biomédicas Abel Salazar por me ter aceitado como aluna integrante do plano doutoral de Ciências Biomédicas. Não menos importante foi a instituição IPO do Porto que me acolheu durante estes sete anos desde o meu mestrado: um muito obrigado ao Centro de Investigação do IPO pelo suporte prático de equipamentos fornecidos para o desenvolvimento do meu trabalho.

Ao instituto *International Iberian Nanotechnology Laboratory* (INL) por me ter acolhido durante oito meses e em especial à Doutora Lorena Diéguez e ao seu grupo *Medical Devices Research Group* por toda a dedicação e aprendizagem proporcionados durante todo o tempo integrado. Aos membros do grupo: Paulina Piairo, Krishna Kant, Lei Wu, Sara Abalde-Cela, Adriana Carneiro, e ao Celso e à Joana pela amizade, companheirismo e as lições de vida que me salientaram a importância em aprender e abraçar novas culturas.

Ao Instituto de Investigação e Inovação em Saúde (I3S), particularmente à Doutora Maria Oliveira group leader do Tumour and Microenvironment Interactions com o qual colaborei um ano, à Doutora Paula Sampaio e à Maria Azevedo, pelo acesso a tecnologia de ponta disponibilizada e todos os ensinamentos prestados.

Ao Centro De Materiais Da Universidade Do Porto (CEMUP), nomeadamente ao Professor Doutor André Silva, por todo o apoio à investigação e trabalhos técnicos concedidos.

Na minha instituição, ao serviço de Imunologia e ao Professor Doutor Carlos Palmeira pela disponibilidade prestada para a utilização dos equipamentos. Bem como à Professora Doutora Carmen Jerónimo diretora do Centro de Investigação pela integração e disponibilidade de utilização de todos os recursos alocados aos grupos de investigação.

Ao meu orientador, Doutor Alexandre Ferreira por me ter dado a oportunidade única de integrar a sua equipa, de aprender tudo o que sei hoje de investigação clínica, por acreditar em mim e por partilhar e me confiar todos os seus projetos procedentes da sua mente brilhante. Por ser compreensivo, humano e inteiramente preocupado e dedicado às teses dos seus alunos. Muitíssimo Obrigada por tudo.

Ao Professor Doutor Lúcio Lara Santos, por apoiar sempre a investigação clínica de translação e incentivar novos e diferentes trabalhos neste âmbito. A minha admiração pela sua mente brilhante e o seu trabalho é infinita.

Ao Doutor Luís Lima por ter possibilitado e apoiado o meu crescimento numa instituição diferente, o INL, e pela confiança depositada no meu trabalho. O qual agradeço muito a oportunidade.

Aos meus colegas e amigos do grupo de Patologia e Terapêutica Experimental: ao Dylan, à Marta, à Rafaela, à Rita, à Elisabete, à Andreia, à Janine, ao Rui, à Cristiana, ao Eduardo e à Carolina, pelo esforço partilhado, por todos os ensinamentos, pela ética de trabalho, pela amizade, por acreditarem em mim.

Ao ESTIMA e à FCT pelo apoio financeiro e por apoiarem a investigação básica tão importante para estabelecer a ponte para a clínica.

A elaboração deste trabalho também não teria sido possível sem o estímulo dos meus amigos e da minha família. À Susana, ao João Henrique, ao Júlio, à Renata, à Inês, ao João Luís, ao Daniel e à Cátia. Obrigada por acreditarem em mim e me motivarem sempre.

À minha família, ao meu irmão, à minha mãe, à minha avó, ao meu avô, aos meus tios, Fátima, António Manuel e Mário, e à Sara, por serem os meus fans número um, pelo apoio e dedicação incondicional, incentivo a nunca desistir e pelo amor que permitiu que alcançasse e terminasse esta etapa da minha vida. Sem vocês não teria sido possível. Estão sempre no meu coração.

Muito Obrigada.

RESUMO

O cancro representa um significativo encargo para a saúde e a sociedade, sendo imperativo avaliar e desbloquear novas ferramentas de diagnóstico e terapêuticas que melhorem a gestão do doente. Recidivas recorrentes e disseminação tumoral continuam a ser as principais causas de resultados insatisfatórios para os doentes, exigindo a identificação de assinaturas específicas associadas a esses fenómenos. Avanços recentes na oncologia de precisão revelaram os papéis das glicoproteínas que transportam alterações na glicosilação na progressão tumoral, recidiva e metástase, e seu impacto nas características do cancro. Nesse contexto, esta tese foi dedicada ao estabelecimento de marcos necessários e da fundamentação para avaliar o impacto putativo das glicosignaturas associadas ao cancro na progressão e disseminação do cancro esofágico (EC) e do cancro colorretal (CRC), visando o direcionamento preciso do cancro.

No contexto do EC, o foco foi colocado na exploração do antigénio sialil-Tn (STn), um glicano curto resultante de uma paragem prematura na O-glicosilação da proteína. O antigénio STn é prevalente em tumores de EC e células tumorais circulantes (CTCs) derivadas de EC, correlacionando-se com recidiva e redução da sobrevivência. Adicionalmente, células de EC glicosiladas que expressam STn mostraram o processo de invasão aumentado, mas o de proliferação reduzido. O O-glicoproteoma identificou o transportador de glicose tipo 1 (GLUT1) como um biomarcador de mau prognóstico, com glicoformas GLUT1-STn evidenciando a especificidade do cancro, sugerindo assim o seu potencial para estratificação e direcionamento de formas agressivas de EC.

No CRC, o foco foi nos antigénios Lewis silialados (sLe), que são mediadores bem conhecidos de metástases de células cancerígenas, através da ligação à E-selectina. Descobrimos que mais de 70% dos tumores colorretais sobrexpressam sLeA/X, incentivando uma exploração mais profunda do sialoglicoproteoma de tumores metastáticos. O enriquecimento de afinidade para E-selectin e a espectrometria de massa permitiu a identificação de quase 600 glicoproteínas, ampliando a nossa compreensão das potenciais glicoproteínas relacionadas com metástases. Notavelmente, o recetor da secretina (SCTR) emergiu como uma glicoproteína de classificação superior, mostrando associações com prognóstico desfavorável e metástases, sugerindo um papel inesperado na metástase e potencial como glicobiomarcador e ligante de E-selectina.

Estas descobertas sublinham a ligação entre as glicoformas GLUT1-STn e SCTR-sLe e a agressividade do EC e do CRC, respetivamente. Adicionalmente, estabelecem as bases para o futuro desenvolvimento de terapias direcionadas contra essas malignidades. Finalmente, demonstram o potencial dos glicoproteomas de EC e CRC como fonte de novos biomarcadores na oncologia de precisão.

ABSTRACT

Cancer poses as a significant healthcare burden on society, being imperative to assess and unlock novel diagnosis and therapeutics tools that improve patient management. Recurrent relapses and tumour dissemination remain the main causes for poor patient outcomes, urging the identification of specific signatures linked to these phenomena. Recent advances in precision oncology have unveiled the roles of glycoproteins carrying changes in glycosylation in tumour progression, relapse and metastasis, and their impact on cancer hallmarks. On this scope, this thesis was dedicated to establishing the necessary milestones and rationale to assess the putative impact of cancer-associated glycosignatures on esophageal (EC) and colorectal cancer (CRC) progression and dissemination, envisaging precise cancer targeting.

In the context of EC, focus was set on exploring the sialyl-Tn (STn) antigen, a short glycan resulting from a premature stop in protein O-glycosylation.The STn antigen was prevalent in EC tumours and EC-derived circulating tumour cells (CTCs), correlating with recurrence and reduced survival. Additionally, glycoengineered EC cells expressing STn displayed increased invasion but reduced proliferation. O-glycoproteomics identified glucose transporter type 1 (GLUT1) as a poor prognosis biomarker, with GLUT1-STn glycoforms enhancing cancer specificity, suggesting their potential for stratification, and targeting of aggressive EC forms.

In CRC, the focus was on sialylated Lewis (SLe) antigens, which are well known mediators of cancer cells metastasis through binding to E-selectin. We found that over 70% of colorectal tumours overexpress sLeA/X, prompting deeper exploration of the sialoglycoproteome of metastatic tumours. Employing E-selectin affinity enrichment and mass spectrometry, nearly 600 glycoproteins were identified, broadening our understanding of potential metastasis-related glycoproteins. Notably, the secretin receptor (SCTR) emerged as a top-ranked glycoprotein, showing associations with poor prognosis and metastasis, suggesting an unforeseen role in metastasis and potential as a biomarker and E-selectin ligand.

These findings underscore the connection between GLUT1-STn and SCTR-SLe glycoproteoforms and the aggressiveness of EC and CRC, respectively. Furthermore, they lay the groundwork for future development of targeted therapies against these life-threatening malignancies. Finally, they shed light on the potential of the EC and CRC glycoproteomes as a source of novel biomarkers precision oncology.

GRAPHICAL ABSTRACT

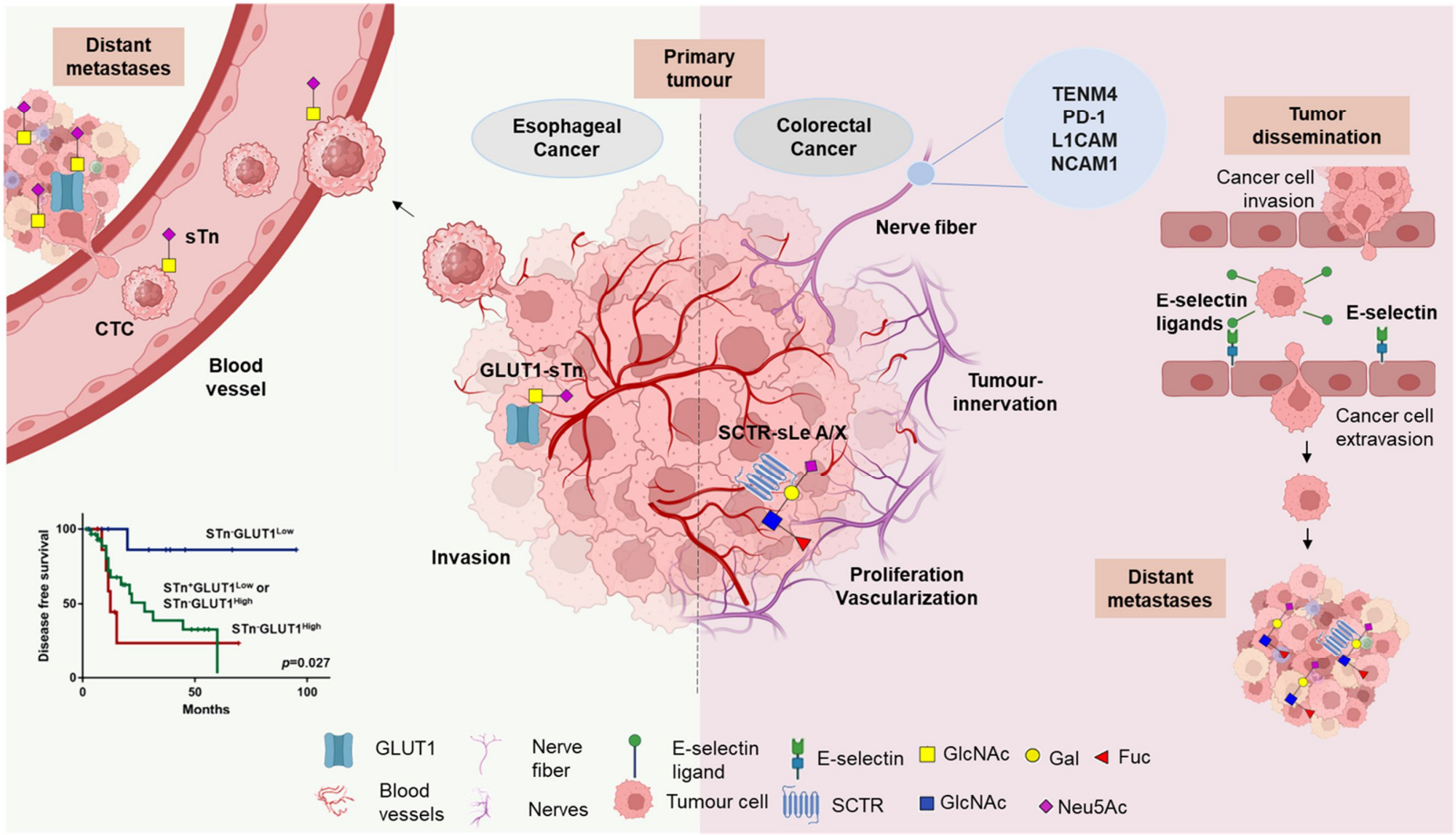


TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

CHAPTER I.

CHAPTER II.

LIST OF TABLES

CHAPTER I.

ABBREVIATIONS & ACRONYMS

AA-PEG: Aminoethylanisamide-polyethylene glycol
ADCC: Antibody-dependent cellular cytotoxicity
ADCC: Antibody-dependent cellular cytotoxicity
BBN: N-butyl-N-(4-hydroxybutyl)nitrosamine
BBN: N-butyl-N-(4-hydroxybutyl)nitrosamine
BMP7 : Bone morphogenetic protein 7
BrdU: 5-bromo-20-deoxyuridine
BSA: Bovine Serum Albumin;
CA19-9: Carbohydrate antigen 19-9
CA9: Carbonic anhydrase 9
CAFs: Cancer-associated fibroblasts
CAR: Chimeric antigen receptor
Cas: Carbonic anhydrases
CCL2: Chemokine C-C motif ligand 2
CCR2: CCL2 receptor
CD34: Cluster of differention 34
CDK4: Cyclin-dependent kinase 4
Chemoexosomes: Chemotherapy-induced exosomes
Col-I: Type I collagen
CRC: Colorectal cancer
CSC: Cancer stem cell
CSF1: Colony-stimulating factor-1
CTC: Circulating tumour cells
ctDNA: Circulating nucleic acids
DAPI: 4',6-diamidino-2-phenylindol
DC: Dendritic cell
DTC: Disseminated tumour cells
DTT: Dithiothreitol
EC: Esophageal cancer
ECM: Extracellular matrix
EGF: Epidermal growth factor
EMT: Epithelial-to-mesenchymal transition
ERK: Extracellular signal-regulated kinase
ESCC: Esophageal squamous cell carcinoma

Evs: Extracellular vesicles

FBS: Fetal bovine serum

FFPE: Formalin-fixed paraffin-embedded

FSCN1: Fascin 1

FUT: Fucosyltransferase

G-CSF: Granulocyte colony stimulating factor

G-CSF: Granulocyte colony stimulating factor

GDC: Genomic Data Common

GlcNAc: N-Acetylglucosamine

GLUT1: Glucose transporter type 1

GO: Gene Ontology

GPCR: G-protein coupled receptor

GT: Glycosyltransferases

GuHCl: Guanidine hydrochloride

HIF: Hypoxia-inducible transcription factor

HR: Hazard ratio

IgG: Immunoglobulin G

IL: Interleukin

IP: Immunoprecipitation

KOH: Potassium hydroxide

LGR5: Leucine-rich repeat-containing G protein-coupled receptor 5

MAPK: Mitogen-Activated Protein Kinase

MCTs: Monocarboxylate transporters

MCTS: Multicellular tumor spheroids

MDSC: Myeloid-derived suppressor cells

MHC-1: Histocompatibility complex I

MHC-I: Major histocompatibility complex I

MMP1: Matrix metalloproteinase 1

MMP-11: Matrix metalloproteinase 11

MS: Mass spectrometry

MSI1: Musashi homolog 1

MUC: Mucin

NDRG1 : N-myc downstream-regulated gene 1

NeuAse: Neuraminidase

NHE: Na+/H+ exchanger

NHE: Na+/H+ exchanger

NO: Nitric oxide

OS: Overall survival

Pan-CK: Pan-Cytokeratin

PARs: Protease-activated receptors

PDGF: Platelet-derived growth factor

PEDF: Pigment epithelium-derived factor

PEP: Posterior error probability

PFS: Progression-free survival

PLA: Proximity ligation assay

PMN: Pre metastatic niches

PNGase F: Peptide - N -Glycosidase F

POSTN: Periostin

RPS: Reverse phase sulfonate

RT: Room temperature

RTK: Receptor tyrosine kinase

RTK: Tyrosine kinase receptors

SCT: Secretin

SCTR: Secretin receptor

SCX: Strong cation exchange

sE-cad: Soluble E-cadherin

SIGLEC: with sialic acid-binding immunoglobulin-type lectins

sLe: Sialylated Lewis antigens

STC1: stanniocalcin-1

STR: Short tandem repeat

TAM: tumor-associated macrophages

TCGA:The Cancer Genome Atlas

TGF-β2: transforming growth factor-β2

TGF-β-RI: TGF-β receptor-I

TLR: Toll-like receptor

TNC: tenascin C

TNF: Tumor necrosis factor

TNFA: tumor necrosis factor alpha

V-ATPase: Vacuolar ATPase

VCAM-1: vascular cell adhesion molecule 1

VEGFA: vascular endothelial growth factor A

VEGFR: vascular endothelial growth factor receptor 1

INTRODUCTION

OUTLINE

The present thesis aims to showcase the relevance of glycans and protein glycoproteoforms as biomarkers of metastasis for oesophageal and colorectal cancers, given their relevance in terms of mortality and incidence. Therefore, the introduction has been organized in two sections:

- Introduction part I. The Tumour Microenvironment and Circulating Tumour Cells: A partnership driving metastasis and glycan-based opportunities for cancer control
- Introduction – Part II. The Role of Glycosylation in Esophageal and Colorectal Cancer: Clinical Implications and Biomarker Discovery.

Part I provides a brief overview of the role played by the tumour microenvironment (TME) in metastasis promotion, supported by precise alterations in glycosylation. Emphasis is also devoted to the mechanisms leading to the formation of circulating tumour cells (CTCs), which play key roles in disease dissimination. It will highlight the significance of these interactions in cancer progression and introduce glycan-based therapeutic opportunities as a novel approach for cancer control. Briefly, this chapter describes of the the TME and CTCs form a crucial partnership in driving metastasis, the spread of cancer to distant organs. The TME promotes epithelial-mesenchymal transition (EMT), enabling tumour cells to invade surrounding tissues. It also supports hematogenous metastasis by facilitating tumour cells' entry into the bloodstream. Once in circulation, CTCs rely on the TME for successful colonization at new sites, adapting to foreign microenvironments. Tumour secreted soluble factors, such as cytokines and growth factors, play a significant role in modulating the TME, promoting tumour progression and metastasis. Tumour-derived exosomes are critical in cell-to-cell communication, transferring oncogenic molecules and remodelling distant microenvironments. Glycosylation, a key post-translational modification, drives metastasis by regulating cell surface molecules and signaling pathways, presenting opportunities for targeted therapeutics. Understanding the TME-CTC partnership provides insights into metastasis mechanisms and highlights glycan-based therapeutic approaches. Ongoing challenges include the complexity of TME-CTC interactions and tumour heterogeneity, but future research promises novel strategies for cancer control, emphasizing the need for continued investigation into this dynamic and influential relationship.

Introduction – Part II explores the critical role of glycosylation in esophageal and colorectal cancers, emphasizing its biological implications and potential for biomarker discovery. It begins with an overview on these cancers, focusing on diagnosis, treatment, risk factors, and histological features. The section then discusses general glycosylation patterns in these cancers, highlighting their influence on cell signaling, immune

response, and tumour progression. Specific attention is given to the STn and sLeA antigens, which are prominently expressed in esophageal and colorectal cancers. These antigens play crucial roles in tumour progression and metastasis, presenting potential as diagnostic and prognostic markers. The clinical relevance of STn and sLeA in CTCs is also examined, emphasizing their utility in non-invasive cancer detection and monitoring. Finally, the integration of multiomics approaches, particularly glycoproteomics, is addressed. This comprehensive approach combines genomics, proteomics, and glycomics to enhance our understanding of cancer biology, leading to the discovery of novel biomarkers and the development of targeted therapies. The section underscores the promise of precision oncology in improving cancer diagnosis, monitoring, and treatment outcomes.

INTRODUCTION

PART I

The information provided in this section is based in the following publications:

a. Peixoto, Andreia; **Cotton, Sofia**; Santos, Lúcio Lara; Ferreira, José Alexandre. "The Tumour Microenvironment and Circulating Tumour Cells: A Partnership Driving Metastasis and Glycan-Based Opportunities for Cancer Control". In Advances in Experimental Medicine and Biology, 1-33. Springer International Publishing, 2021. doi:10.1007/978-3-030-73119-9_1

The Tumour Microenvironment and Circulating Tumour Cells: A partnership driving metastasis and glycan-based opportunities for cancer control

Andreia Peixoto[a,b,c,d], Sofia Cotton[a,b,d], Lúcio Lara Santos[a,b,d,e], José Alexandre Ferreira[a,b,d]

[a]Experimental Pathology and Therapeutics Group, Portuguese Institute of Oncology, 4200-162 Porto, Portugal; [b]Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-013 Porto, Portugal; [c]Institute for Research and Innovation in Health (i3s), University of Porto, 4200-135 Porto, Portugal; [d]Institute for Biomedical Engineering (INEB), Porto, Portugal, 4200-135 Porto, Portugal; [d]Porto Comprehensive Cancer Centre (P.ccc), 4200-162 Porto, Portugal; [e]Department of Surgical Oncology, Portuguese Institute of Oncology of Porto, 4200-162 Porto, Portugal;

Abstract

Circulating tumour cells (CTC) are rare cells that actively detach or are shed from primary tumours into the lymph and blood. Some CTC subpopulations gain the capacity to survive, home and colonize distant locations, forming metastasis. This results from a multifactorial process by which cancer cells acquire several molecular traits that enhance motility, invasion, immune escape in response to microenvironmental cues. Here we present evidences regarding the existence of a self-fuelling molecular crosstalk between cancer cells and the tumour stroma supporting the main milestones leading to metastasis. We discuss how the tumour microenvironment supports pre-metastatic niches, CTC development and ultimately dictates CTC fate in targeted organs. Finally, we highlight the key role played by protein glycosylation in metastasis development, its prompt response to microenvironmental stimuli and the tremendous potential of glycan-based molecular signatures for liquid biopsies and targeted therapeutics.

Introduction

The detachment of cancer cells from the primary tumour and its spread to distant organs is a critical turning point towards unfavourable outcomes. In solid tumors, sub-populations of cancer cells may acquire the capacity to invade the extracellular matrix

and surrounding anatomic structures reaching the vascular and lymphatic systems. In 80% of solid tumors, metastasis via the lymphatic system precedes metastasis via the vascular system (1). Thereby, the assessment of the metastatic involvement of sentinel lymph nodes (*i.e.* the tumour-draining lymph nodes) is considered one of the most powerful prognostication tools. Even though still challenging, the identification of metastasised nodules and its removal is often critical for disease management. On the other hand, bloodborne circulating tumour cells (CTC) can reach the vascular system by lymphatic system drainage into the blood or by directly intravasating into the primary tumour capillary. An alternative route of dissemination involves tumour cells invasion of adjacent organs as a consequence of tumour progression. Several studies have demonstrated that cells that metastasize to lymph nodes present distinct molecular features in comparison to those that directly reach the blood vessels (1, 2), providing important blueprints for intervention. However, irrespectively of the metastatic route, the acquisition of more motile and invasive traits is crucial for cancer cells to overcome a dense stromal barrier composed by the extracellular matrix (ECM; proteoglycans, hyaluronic acid, and fibrous proteins such as collagen, fibronectin, and laminin), mesenchymal supporting cells (fibroblasts, adipocytes, immune cells) and several other biomolecules (growth factors, chemokines, cytokines, metabolites, lectins, glycans, etc). This occurs by the activation of a cellular transdifferentiation program termed epithelial-to-mesenchymal transition (EMT) often accompanied by the acquisition of cancer stem cell (CSC)-like properties, notable motility, invasiveness, and increased resistance to apoptosis (3-6). Such events are in many ways triggered by the tumour microenvironment, which co-evolves with tumours cells to support disease progression. The tumour microenvironment is also known to release into circulation relevant molecular signals, including nano- and microvesicles (exosomes) containing a wide array of biomolecules that decisively contribute to establish pre-metastatic niches. Once in circulation, tumour cells capable of overcoming sheer stress and evade the immune system may dock and extravasate to distant organs. Again, depending on the microenvironment at the homing organ, these cells may experience expansion and differentiation or remain quiescence, sometimes for large periods of time, originating late relapse (7).

Despite being the driving force of metastasis, CTC are rare cells accounting for less than 0.004% of all mononucleated blood cells (8). However, elevated CTC counts have been proven to consistently predict patient survival, while being an effective marker of treatment response, holding tremendous potential to aid early interventions and

therapeutic decision (9, 10). CTC may also better reflect the molecular nature of the metastasis in comparison to biomolecules in circulation. This paves the way for liquid biopsies in alternative to metastasis biopsies that are often difficult or even impossible to perform in clinical settings. As such, the field of CTC research has experienced tremendous advances over the last five years mostly supported by micro and nanofluidics lab-on-a-chips. These technological breakthroughs have been recently reviewed by other authors and have greatly expanded our view on CTC molecular nature (11-13). Notably, several reports have already demonstrated that CTC and the metastasis present significant genetic similarities that many times differ from those found in primary tumours (14-16). Such observations further reinforce the relevance of exploiting CTC as a tool to address the metastasis in the context of liquid biopsies. Lab-on-a-chips have also allowed to recover viable cells for expansion *in vitro* enabling downstream molecular and functional studies and the adoption of more rationale and effective therapeutic strategies (17, 18). These innovative solutions are rapidly progressing towards point-of-care settings making CTC counts and molecular characterization possible at the bed side table in the near future. All these aspects are bringing CTC-based liquid biopsies one step closer to personalized oncology and will certainly constitute a turning point in cancer patients' management. In summary, CTC molecular characterization may grant insights on primary tumour most aggressive cells and will certainly open novel avenues to target metastasis. The translation of current innovations into the clinical practice is now warranted to improve on the detection of disseminated disease at early stages, increasing the odds of favourable outcome.

Even though outstanding advances have been made foreseeing CTC-based liquid biopsies, there is still an unmet need for therapeutic strategies able to contain disease dissemination and target metastatic sites. Envisaging this goal, the present chapter revises the main molecular mechanisms leading to the development of metastasis. Particular emphasis is given to the role played by the tumor microenvironment at critical stages of disease progression (formation of pre-metastatic niches; cancer cells invasion of surrounding tissues; intravasation to lymphatic and blood systems; extravasation from the bloodstream to distant tissues; adaptation to foreign microenvironments and colonization; proliferation and macrometastasis formation). Finally, we discuss the functional impact of protein glycosylation as a driving force of metastasis and the glycoproteome as a potential source of relevant biomarkers and therapeutic targets. Understanding these mechanisms should greatly facilitate cancer

patient's management and generate novel targeted therapeutic interventions able to control metastatic spread.

Tumour microenvironment promotion of EMT and tumour invasion

Throughout the course of disease, cancer cells may acquire the capability to invade surrounding tissues and adjacent organs. In this context, the tumour microenvironment provides a selective pressure towards more invasive and aggressive phenotypes. Moreover, the synergic and self-fuelling molecular interplays established by cancer cells and the tumour microenvironment are responsible by the activation and maintenance of relevant oncogenic traits that support disease progression and dissemination. It is well described that tumour infiltrating immune cells, including macrophages and mast cells, produce and secrete proteolytic enzymes responsible for extracellular matrix (ECM) remodelling. The released peptides, glycans and glycosaminoglycans are known to act as pro-invasive signalling molecules (19-24). Immune cells also release cytokines that induce an invasive behaviour in adjacent tumour cells (25), while inhibiting the expression of known metastasis suppressor genes (26). Concomitantly, immune cells create concentration gradients of growth factors that chemoattract tumour cells towards the tumour vasculature (27). Overall, the pro-inflammatory microenvironment mediated by cytokines and chemokines secretion leads to an increase in the density of immune cell infiltrates, epithelial and stromal growth factors, and matrix-remodelling enzymes needed for neoplastic progression (28, 29). Adipocytes and cancer-associated fibroblasts (CAFs) also play a critical role in the acquisition of more aggressive phenotypes **(Figure 1A)**. Mature cancer-associated adipocytes increase the expression of matrix metalloproteinase 11 (MMP-11) and pro-inflammatory cytokines IL-6 and IL-1β necessary to support pro-invasive traits in tumor cells (30). CAFs also promote sustained inflammation by producing large amounts of pro-inflammatory IL-6, which drives proliferation and migration of cancer cells towards CAFs to form heterocellular contacts for paracrine stimulation. This has been considered of key relevance for triggering EMT activation and invasion in breast, prostate and bladder carcinomas (31-33). CAFs and tumour cells also maintain a self-fuelling paracrine signalling mediated by chemokines that enhances tumor cell invasion and chemoresistance. CXCL12-CXCR4 and to less extent CCL2, CCL5, CCL7, CXCL1, CXCL8, and CXCL14 are amongst the main CAF-secreted chemokines responsible by inducing proliferation, migration, and invasion in tumour cells (34, 35). In addition, co-

culture experiments have highlighted the decisive role of fibroblasts as drivers of cancer cells invasion by creating force-mediated matrix tracks to orient the invasion front (36). CAFs also supply the tumour microenvironment with high TNF-alpha and IL-1β levels responsible by recruiting pro-cancerous monocytes to tumour sites and stimulate tumor neovascularization by inducing the production of IL by tumour cells (37, 38). The production of IL-1β also induces the activation of PI3K and NF-kB pathways and, consequently, enhances tumor angiogenesis and growth (39). Concomitantly, CAFs produce TGFβ, which causes the neighbouring cancer cells to undergo EMT and acquire cancer stem cell (CSC)-like characteristics (40). Notably, studies *in vitro* suggest that stromal derived TGFβ-induced EMT may alter glycogenes expression and consequently promote the remodelling of *N*-glycans on membrane glycoproteins towards more invasive phenotypes (41). In addition, platelet-tumor cell interactions are sufficient to prime tumor cells for metastasis. Namely, platelet-derived TGFβ and direct platelet-tumor cell contact synergistically activate TGFβ/Smad and NF-κB pathways in cancer cells, resulting in invasive mesenchymal-like phenotypes and enhanced metastasis *in vivo* (42). Altogether, the complex and synergic interplays established by tumour cells, immune cells of different linages, adipocytes, CAFs and platelets enmeshed in an evolving ECM is still far from being completely understood but decisively favours EMT and invasion (**Figure 1A**).

Hypoxia is a prominent feature of advanced stage tumours, defined as low oxygen tension within the tumour bulk as a result of sustained proliferation and defective neovascularization. It is well postulated that oxygen shortage decisively contributes to the acquisition of more motile and invasive tumour cell phenotypes (43). The absence of oxygen promotes the stabilization of hypoxia-inducible transcription factors (HIF) that mediate an array of adaptive molecular alterations that confer to cancer cells the capacity to endure environmental stress. In addition, it induces the overexpression of several genes related with angiogenesis, EMT, and dedifferentiation of tumour cells towards stem-like phenotypes as *VEGF* (44), *Sox9 (45)* and *SIRT1 (46).* Moreover, HIF-1• significantly alters the transcription of several glycogenes resulting in profound alterations in the glycosylation of key glycoproteins involved in cell-cell and cell-matrix adhesion. Namely, hypoxia significantly reduces the extension of protein *O*-glycosylation (glycosylation of Ser/Thr residues), favouring the accumulation of sialylated short-chain glycans at the cell membrane (47). These post-translational alterations decisively favour motile and invasive phenotypes, while providing an immunosuppressive microenvironment (43, 47) (**Figure 1A**). In addition, HIFs control several mediators of

vascular response, including angiogenic factors and metabolic pathways that result in the production of small molecules such as nitric oxide (NO). Loss of HIF-1α in endothelial cells reduces NO synthesis, retards tumor cell migration through endothelial layers and delays tumor cell metastasis. The loss of HIF-2α has the opposite effect (48), suggesting that HIF-mediated cellular responses are complex and can either promote or hinder metastasis depending on the HIF isoforms present in a given biological milieu. Finally, the formation of tortuous and leaky neovasculature due to the low number of vessel-associated pericytes also facilitates the extravasation of more motile tumour cells (49).

In summary, tumour stromal cells and hypoxia-mediated factors aid tumour cell invasion by supplying pro-metastatic factors, compromising vasculature and stromal barrier integrity, and by promoting EMT and the acquisition of stem-like properties.

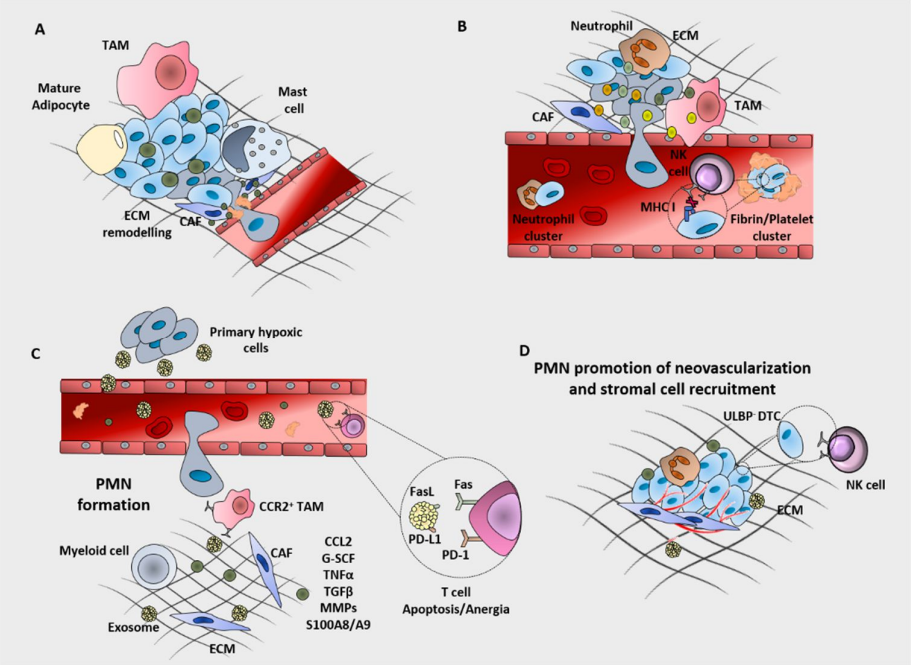


Figure 1. A) Tumour infiltrating immune cells, including macrophages and mast cells as well as mature cancer-associated adipocytes secrete proteolytic enzymes responsible for extracellular matrix (ECM) remodelling. Moreover, these pro-tumoral immune infiltrates secrete increasing amounts of growth factors and cytokines, ultimately creating a concentration gradient that chemoattracts tumour cells towards the tortuous and leaky tumour neovasculature. Hypoxic tumour cells undergo epithelial-to-mesenchymal transition and dedifferentiation towards stem-like phenotypes, acquiring more motile and invasive phenotypes. Concomitantly, CAFs form heterocellular contacts with tumour

cells inducing migration, while platelets facilitate metastasis. **B)** Perivascular macrophages CAFs and neutrophils stablish paracrine loops with tumour cells to stimulate tumor cell invasion towards blood vessels. CTC also cluster with platelets and form fibrin crosslinks to overcome sheer stress and avoid immune recognition. Of note, platelets transfer major histocompatibility complex I (MHC-I) molecules to CTC promoting NK cell surveillance escape. **C)** Primary tumour and stoma cells produce chemokines, soluble factors and exosomes capable of attracting immune cells to distant PMN, while upregulating pro-inflammatory and angiogenic molecules. For instance, granulocyte colony stimulating factor (G-CSF) promotes myeloid cell mobilization to the PMN, while CCL2 chemoattracts several CCR2 expressing immune precursors. Tumour hypoxia represents one of the main microenvironmental factors busting exosome release. Besides contributing to PMN formation, exosomes contribute to systemic immunosuppressive effects. **D)** To overcome cell dormancy and achieve successful colonization DTC require mitogenic stimulation by growth factors and cytokines that may be either pre-existent in the PMN or synthetized de novo. Subsequent recruitment of supporting stroma cells, ECM remodelling and neovascularization are also essential.

Tumour microenvironment promotion of hematogenous metastasis

Cells endowed with the capacity to migrate and/or invade the tumour stroma and ultimately reach the microvasculature may ultimately reach hematogenous circulation originating CTC. The intravasation across the endothelial barrier can be facilitated by several stromal cells, as tumour associated macrophages, fibroblasts, neutrophils, and platelets (**Figure 1B**). Particularly, tumour-associated perivascular macrophages establish a colony-stimulating factor-1 (CSF1)/epidermal growth factor (EGF) paracrine loop with tumour cells that facilitate this process. Specifically, tumor cells secrete CSF1 to recruit macrophages and macrophages reciprocally secrete EGF to stimulate tumor cell invasion towards blood vessels (50). Both cell types also secrete proteases to facilitate intravasation. Namely, macrophages induce RhoA activity in tumor cells, triggering the formation of actin-rich degradative protrusions (invadopodia) that allow tumour cells to penetrate endothelial tight junctions into the bloodstream (51, 52). Concomitantly, macrophage promoted vascular endothelial growth factor A (VEGFA), tumor necrosis factor alpha (TNFA) and IL-6 signalling induce local loss of vascular junctions and transient vascular permeability towards tumor cell intravasation (53-55). In addition, platelet-derived growth factor (PDGF)-activated CAFs secrete glycoprotein

stanniocalcin-1 (STC1) to aid tumour cell intravasation. Supporting this hypothesis, mice with STC1-deficient CAFs showed decreased tumour cell embolism, strongly suggesting decreased or no intravasation (56). Moreover, tumours showing elevated numbers of infiltrating MMP-9-positive neutrophils concomitantly exhibit enhanced levels of angiogenesis and intravasation (57). These findings directly link tumor-associated neutrophils and their TIMP-free proMMP-9 phenotype to the formation of new blood vessels that serve as conduits for tumor cell dissemination.

Once in the bloodstream, single CTC or small cell groups experience sheer stress, which can eventually disrupt cellular integrity. CTC clustering with different types of blood cells is well described and regarded as a key survival strategy (**Figure 1B**). CTC may cluster with neutrophils through VCAM1-mediated anchorage, inducing neutrophils to produce inflammatory IL-6 and IL-1β (58). This primes CTC for proliferation, as translated by the upregulation of genes associated with cell-cycle progression and increased Ki67 expression (58, 59). The acquisition of a proliferative phenotype significantly increases the probability of successful metastasis upon arrival to secondary tumour sites. CTC have also been recognized to secrete thrombin, a potent platelet activator responsible by the initiation of the coagulation cascade (60). This promotes clustering with platelets and the activation of protease-activated receptors (PARs) on both cell types to form fibrin crosslinks (**Figure 1B**). These molecular assemblies protect CTC against sheer stress and immune recognition. Namely, platelets transfer major histocompatibility complex I (MHC-I) molecules to CTC providing them with a pseudo-normal phenotype and promoting NK cell surveillance escape (61) (**Figure 1B**). This transient molecular mimicry simultaneously allows CTC to downregulate endogenous MHC-I to evade T-cell-mediated immunity (62). In summary, platelets shield CTC from physical injury and provide molecular camouflage to circumvent immune surveillance. On the same note, CTC may travel with stromal components, including activated fibroblasts, providing them an upfront colonization advantage (63).

After overcoming several physiological and anatomical challenges CTC may eventually home to distant organs by different mechanisms. CTC may become lodged in capillary beds at a given location due to size, display specific adhesion molecules that favour adhesion to microvessels in targeted sites or respond to chemoattractive cues generated by a particular tissue (64-68). Regardless, CTC may preferentially home to organs presenting a suitable for colonization and often induced by the primary tumour. Posteriorly, CTC may extravasate and invade the foreign parenchyma; nevertheless, most cancer cells undergo apoptosis within 72h after extravasation (69, 70). The fittest

cancer cells may remain in a quiescent state as micrometastases or experience clonal evolution and expansion to macrometastatic lesions, depending on the foreign microenvironment.

Tumour microenvironment promotion of CTC colonization

Only a small percentage of CTC are able to successfully colonize distant locations, mostly due to tight vascular wall barriers and unfavourable conditions for survival in circulation and at homing distant organs (71). However, these hurdles are significantly less stringent when CTC are able to re-infiltrate the tumor of origin in a process called "tumor self-seeding" (72). This has been observed for breast, colon and melanoma tumors, as well as osteosarcomas mediated by the release of IL-6 and IL-8 by the primary tumours (72). Self-seeded CTC infiltrate the primary tumour in a *collagenase 1* (*matrix metalloproteinase 1*, *MMP1*), *fascin 1* (*FSCN1*) and *CXCL1* dependent manner *(72-75)* in a process facilitated by tumour vascular permeability (76). This CTC subset often secrete stroma-recruiting signals such as the CXCL1 chemokine that mediate the repopulation of the tumour with macrophages, neutrophils, and other CD45+ cells (72). These cells contribute to increase IL-6 levels that recruit more CTC to the primary tumour in a positive signalling feedback loop. Self-seeded CTC further interacts with the tumor stroma to enhance the release of signals that foster tumor growth, angiogenesis, invasion, and metastasis; thereby increasing the aggressivity of primary tumours (72). Recent research has shed light on a novel cytokine-independent self-seeding mechanism that relies on exosome-mediated delivery of specific miRNAs to CTC. Hepatocellular carcinoma cells have shown to horizontally transfer exosome-derived microRNA-25-5p to CTC, inducing leucine-rich repeat containing 7 (LRRC7) downregulation and enhancing their migratory and invasive capacity towards the primary tumour (77). As such the miR-25-5p/LRRC7 axis may constitute an interesting novel target to reduce the multidirectional trafficking of cancer cells during cancer progression.

Regarding distant metastasis, it is well stablished that CTC organotropism is not random; instead, distant sites are selectively and actively modified by the primary tumour to favour colonization (78). The establishment of these PMN results from the combined systemic effects of tumour-secreted factors and tumour-shed extracellular vesicles (EVs) (79). Vascular leakiness is the earliest event in PMN formation, followed by the alteration of local resident cells, such as fibroblasts, and the recruitment of non-resident cells, such as bone marrow-derived cells, subsequently attracting CTC (79-81) (**Figure 1C**).

Tumour secreted soluble factors

It is well established that primary tumours contribute in an orchestrated and stepwise manner to the formation and evolution of PMN. The first evidences include the demonstration that bone marrow-derived haematopoietic progenitor cells expressing vascular endothelial growth factor receptor 1 (VEGFR1; also known as Flt1) home to pre-metastatic sites and form cellular clusters before the arrival of CTC (80). $VEGFR1^+$ cells
expressed integrin α4β1 (also known as VLA-4) and upregulate fibronectin, an integrin α4β1 ligand in resident fibroblasts, providing a permissive niche for incoming tumour cells (80). VEGFR1 blockage with monoclonal antibodies or removal of $VEGFR1^+$ cells from the bone marrow of wild-type mice has been shown to abrogate the formation of pre-metastatic niches and prevent tumour metastasis, reinforcing the role of this mechanism in metastasis development. Altogether, these findings highlight the key role played by $VEGFR1^+$ haematopoietic progenitors in the metastasis process and suggest that expression patterns of fibronectin and VEGFR1+ integrin α4β1+ clusters dictate organ-specific tumour spread. Another well-known mechanism underlying PMN formation may be mediated by granulocyte colony stimulating factor (G-CSF) and prokinecitin 2, also known as Bv8. Both G-CSF and the G-CSF-induced Bv8 protein have preferential expression in refractory tumors, initiating myeloid cell mobilization and angiogenesis that mediate tumor refractoriness to anti-VEGF therapy in mouse models (82). Anti-G-CSF treatment dramatically suppressed circulating or tumor-associated CD11b(+)Gr1(+) cells, reduced Bv8 levels, and affected tumor vasculature (82). Accordingly, G-CSF expression by tumor or stromal cells is a determinant of refractoriness to anti-VEGF-A treatment, being sufficient to mimic lung PMN and to enhance metastatic properties. Another example relates with the role played by cytokines and chemokines produced by both tumour and stoma cells in PMN development. Amongst the most studied is chemokine C-C motif ligand 2 (CCL2), a potent chemoattractant of monocytes, macrophages, memory T lymphocytes and NK cells expressing CCR2 (CCL2 receptor) to distant PMN *(83)* (**Figure 1C**). The role of CCL2-CCCR2 axis in PMN development and metastasis has been elegantly demonstrated for breast cancer. In these tumours, cancer cell and stroma-derived CCL2 recruits Gr1-positive inflammatory monocytes that facilitate CTC extravasation in a VEGFA-dependent manner, thus effectively promoting lung micrometastasis (84). The subsequent recruitment of metastasis-associated macrophages and their interaction with

metastatic breast tumour cells was also dependent on CCL2-CCR2 interactions. The inhibition of CCL2-CCR2 signalling blocked the recruitment of inflammatory monocytes, inhibited metastasis *in vivo* and prolonged the survival of tumour-bearing mice. Furthermore, the depletion of tumour-cell-derived CCL2 inhibited metastatic seeding (84). On the same note, MDA-MB-231 human breast cancer cells overexpressing CCL2 prompted $CCR2^+$ macrophage infiltration to the lungs and increased $CCR2^+$ osteoclast differentiation in the bone, thereby facilitating colonization to the lung and bone (85, 86). Contrastingly, in mouse models of breast cancer, primary tumour-derived CCL2 induced anti-tumoral responses in lung neutrophils to inhibit metastatic seeding, suggesting a novel inhibitory role of CCL2 in PMN formation *(87)*. In addition, injection of mice with hypoxic mammary tumor cells conditioned medium containing CCL2 resulted in PMN formation in the lungs of immune-competent mice in the absence of a primary tumor. Furthermore, an increased metastatic burden in mammary and melanoma experimental metastasis models was observed due to reduced antitumor responses of PMN immune cells (88). Alternate mechanisms for PMN formation involve TNFα, TGFβ, lysyl-oxidase (LOX), MMP2, MMP9, CXCR4, and SDF1α, among others, all of which direct or indirect HIF targets and generally important in the metastatic process (47, 79, 89, 90), pointing tumour hypoxia as a key player in PMN formation (**Figure 1C**). These findings demonstrate that hypoxia in the primary tumor can promote PMN formation in secondary organs by providing cytokines and growth factors capable of creating PMN through recruitment of myeloid cells and reduction of the cytotoxic effector functions of NK cell populations.

The family of S100 Ca2+-binding proteins has been implicated in many aspects of the interaction between cancer cells and stromal cells, contributing to the establishment of pro-inflammatory tumour microenvironments and PMN, mostly by inducing cytokine production, immune cell recruitment and infiltration (91, 92) (**Figure 1C**). S100A8 and S100A9 have been detected in pre-metastatic lungs in response to the production of TNF, TGFβ and VEGFA by primary lesions (93, 94). S100A8/A9 are responsible by recruiting macrophages and haematopoietic progenitor cells and inducing the production of serum amyloid A 3 (SAA3) by lung endothelial and myeloid cells. In addition, SAA3 interacts with the Toll-like receptor (TLR) 4 of lung endothelial cells and macrophages in the pre-metastatic niche in a positive feedback signalling loop that fuels the production of recruiting signals. Moreover, SAA3 stimulates NF-kappaB signalling in a TLR4-dependent manner and facilitates metastasis. This inflammation-like state accelerates the migration of primary tumour cells to lung tissues, which is suppressed

by the inhibition of either TLR4 or SAA3 (95). Thus, blocking SAA3-TLR4 function in the pre-metastatic phase may provide a strategy to impede pulmonary metastasis.

Tumour-derived exosomes

Metastatic tumour cells present an increased capability of releasing exosomes, which have been extensively involved in the preparation of premetastatic niches (96). Exosomes are nanoscale tumour-derived vesicles carrying proteins, lipids, DNA, mRNA and microRNAs that mediate intercellular communication within the primary tumour and at distant locations (97, 98). Furthermore, exosome subpopulations are highly heterogeneous and demonstrate diverse organ biodistribution patterns, suggesting distinct biological functions (99). Particularly, exosomes from mouse and human lung, liver and brain tumour cells fuse preferentially with resident lung fibroblasts and epithelial cells, liver Kupffer cells and brain endothelial cells at their predicted destination (100). Moreover, proteomics studies revealed distinct integrin expression patterns that may decisively contribute target these vesicles to different organs. Namely, exosomal integrins α6β4 and α6β1 were associated with lung metastasis, while exosomal integrin αvβ5 was linked to liver metastasis. Targeting integrins α6β4 and αvβ5 decreased exosome uptake, as well as lung and liver metastasis, respectively, indicating that exosomal integrins could be used to predict organ-specific metastasis (100). Furthermore, exosome integrin uptake by resident cells activated Src phosphorylation and pro-inflammatory *S100* gene expression to induce PMN formation (100), reinforcing the role of integrins in this process.

Tumour hypoxia represents one of the main microenvironmental factors busting exosome release by cancer cells in a HIF-1α-dependent manner (101, 102) (**Figure 1C**). Moreover, several studies demonstrate that exosomes produced by cancer cells were able to promote angiogenic signalling and vascular leakage (103, 104). As example, hypoxic renal cancer cells release exosomes containing carbonic anhydrase 9 (CA9). The uptake of CA9-exosomes by endothelial cells induces migration and tube formation, increasing neovascularization (105).

In addition, exosomes derived from stem-like and mesenchymal cancer cells contribute to trigger the angiogenic switch and coordinate metastatic diffusion during tumor progression (106, 107). Soluble E-cadherin (sE-cad), abundantly released in the form of exosomes by the primary tumour, also plays a relevant role in the definition of PMN. sE-cad-positive exosomes heterodimerize with VE-cadherin on endothelial cells

and transduce the sequential activation of β-catenin and NFκB signalling (108). Furthermore, microparticles derived from mesenchymal-like tumour cells contain increased levels of angiogenic molecules including PDGF, IL8 and angiogenin. These molecules promote the activation of endothelial cells through Akt phosphorylation and *Arf6* upregulation to constitute a pro-metastatic vascular niche (107). In addition, in endothelial monolayers, exosome-mediated transfer of cancer-secreted miR-105 efficiently abrogates tight junctions' integrity. Overexpression of miR-105 in non-metastatic cancer cells induced metastasis and vascular permeability in distant organs, whereas inhibition of miR-105 in highly metastatic tumors eases these effects (109). Similarly, exosomal miR-25-3p regulates the expression of VEGFR2, ZO-1, occludin and Claudin-5 in endothelial cells by targeting KLF2 and KLF4, consequently promoting vascular permeability and angiogenesis (110). These findings suggest that exosomal miRNAs are involved in pre-metastatic niche formation and may be used as a blood-based biomarker for metastasis.

Exosomes also participate in all known mechanisms by which cancer cells evade the immune system. They influence the differentiation and activation of immune suppressor cells, modulate antigen presentation, and are able to induce T-cell apoptosis (111). It has been reported that metastatic melanomas release exosomes carrying surface PD-L1 (**Figure 1C**). This decreases the proportion of proliferating PD-1^{+} CD8 T cells in both spleen and lymph nodes, suggesting that these vesicles may decisively contribute to suppresses anti-tumour immunity systemically (112). The evaluation of PD-L1 exosomes in the blood of melanoma patients also holds tremendous potential for liquid biopsies. PD-L1 exosome levels at early stages of anti-PD-1 therapy, indicating adaptive responses of the tumour cells to T cell reinvigoration, may be used to stratify clinical responders from non-responders (112). Other exosome proteins such as FasL may also contribute to immunosuppressive effects (113-115) (**Figure 1C**). Particularly, plasma-derived exosomes from advanced head-and-neck cancer patients are often enriched in proteins such as COX-2, TGFβ-LAP, PD-1, CTLA-4 and TRAIL, which ultimately induce apoptosis of CD8+ T cells, suppression of CD4+ T cell proliferation, up-regulation of regulatory T cell suppressor functions and down-regulation of NKG2D expression levels in NK cells, compromising anti-tumor immune responses (113). Similarly, melanoma exosomes presenting S100A8 and S100A9 as part of its protein cargo were capable of compromising dendritic cell (DC) maturation, highlighting a potential mechanism for PMN induction in tumor-draining sentinel lymph nodes (116). Interestingly, there are evidences supporting the existence of protective exosomes

produced by non-metastatic cancer cells. Namely, exosomes isolated from patients with non-metastatic primary melanomas have the ability to suppress lung metastasis (117). The mechanism appears to be mediated by exosomes capable of stimulating an innate immune response by inducing the expansion of Ly6Clow patrolling monocytes in the bone marrow. This positively contributes to cause cancer cell clearance at the PMN *via* the recruitment of NK cells and TRAIL-dependent killing of cancer cells by macrophages. These events require the induction of the Nr4a1 transcription factor and are dependent on pigment epithelium-derived factor (PEDF) on the outer surface of exosomes (117). In summary, exosomes influence immune response in PMN either by promoting immune surveillance through increase of mononuclear cells, recruitment of NK cells and induction of macrophage phagocytosis or by aiding immune escape through recruitment of suppressive immune cells and/or inducing anergia in CD8 T-cells through PD-L1 signalling.

Tumour derived exosomes also contribute to ECM remodelling towards enhanced metastatic potential of cancer cells. Showcasing this aspect, tumour-derived exosomes carrying proteases capable of degrading collagen, laminin and fibronectin are capable of binding ECM components contributing to disrupt its integrity and favouring cell motility and invasiveness (118). Proteomics studies have also demonstrated that chemotherapy-induced exosomes (chemoexosomes) have high heparanase levels in comparison to exosomes from non-treated cells. Chemoexosomes are able to transfer this cargo to tumour cells, enhancing their heparan sulfate degrading activity, ERK signalling and increasing syndecan-1 proteoglycan shedding, ultimately shaping tumour cells behaviour (119).

In summary, exosomes are molecular messengers that decisively contribute to PMN formation. The main exosome-mediated mechanisms include the upregulation of pro-inflammatory and angiogenic molecules, promotion of immune escape as well as extracellular matrix remodelling. Exosomes of distinct sizes and molecular natures have been found in the blood of cancer patients in early stages of disease, associated with the risk of metastasis, response to treatment and poor prognosis (120). These observations reinforce the potential of exosomes for liquid biopsies and the importance of engaging in comprehensive large-scale translational studies in this area. Moreover, the targeted nature of exosomes for metastatic sites is currently being explored as a strategy to deliver therapy to more aggressive cancer cells. Namely, paclitaxel-loaded macrophage-derived exosomes were bioengineered to incorporate an aminoethylanisamide-polyethylene glycol (AA-PEG) vector moiety to target the sigma

receptor, which is overexpressed by metastatic lung cancer cells. This strategy resulted in increased exosome loading capacity, profound ability to accumulate in cancer cells upon systemic administration, and improved therapeutic outcomes *in vivo (121)*. These findings showcase an example in which the combination of targeting ability with the biocompatibility of exosome-based drug formulations offers a powerful and novel delivery platform for anticancer therapy.

Adaptation of CTC to foreign site microenvironments

CTC that successfully attach to metastatic sites are generically termed as disseminated tumour cells (DTC). However, most DTC are unable to overcome the microenviromental challenges posed by the homing site and undergo apoptosis within 72h after extravasation into a secondary location (122). Nevertheless, more fit DTC may prevail in metastatic sites as single quiescent cells, as micrometastasis or as actively growing macrometastasis (123). DTC may become dormant as a result of inhibitory stimuli encountered in distant microenvironments or the absence of supporting cues from primary tumours (124). Some dormancy-inducing cues include strong and specific transforming growth factor-β2 (TGF-β2) signalling, resulting in the induction of DEC2/SHARP1 and p27, in a TGF-β receptor-I (TGF-β-RI), TGF-β-RIII and SMAD1/5 dependent manner, and downregulation of cyclin-dependent kinase 4 (CDK4). Importantly, systemic inhibition of TGF-β-RI or p38α/β activities awakened dormant DTC and fuelled multiorgan metastasis in HNSCC mouse models (125). Similarly, BMP7 (bone morphogenetic protein 7) secreted from bone stromal cells also induced senescence in prostate cancer stem-like cells (CSCs) by activating p38 mitogen-activated protein kinase, and increasing the expression of the cell cycle inhibitor p21 and metastasis suppressor NDRG1 (N-myc downstream-regulated gene 1) (126). These findings strongly suggest that the BMP7-BMPR2-p38-NDRG1 axis plays a critical role in dormancy and recurrence of prostate CSCs in bone. Furthermore, dormant DTC might inhabit specific niches supporting their survival while restraining proliferative signalling. Namely, prostate cancer cells that metastasize to the bone were found to compete with hematopoietic stem cells for the occupancy of the endosteal stem cell niche in a CXCL12-CXCR4 signalling axis dependent manner (127). Furthermore, dormant DTC are frequently localized in perivascular niches, probably due to endothelial-derived thrombospondin-1 dormancy signalling (128). However, this suppressive cue was lost in sprouting neovasculature where endothelial tip cells not only permitted but also

accelerated breast DTC outgrowth in an active TGF-β1 and periostin dependent manner (128) (**Figure 1D**). Consistent with this idea, dormant DTC activation in the brain also seems to be dependent on angiogenesis, since vascular endothelial growth factor-A (VEGF-A) inhibition induced long-term dormancy of lung cancer micrometastases by preventing angiogenic growth to macrometastases (129).

The immune system is also key to define DTC fate in metastatic niches where DTC are no longer shielded by the immunosuppressive environment of the primary tumours. To overcome immune surveillance, DTC downregulate ULBP ligands for NK cells, thereby evading NK-cell-mediated clearance (130) (**Figure 1D**). On the other hand, adaptive immunity may contribute to induce cancer dormancy for extended periods, maintaining cancer cells occult (131). Contrastingly, localized inflammatory states trigger DTC activation and progression to macroscopic metastasis by activating EMT programs in DTC in a *ZEB1*-dependent manner (132).

To overcome cell dormancy and achieve successful colonization DTC require mitogenic stimulation by growth factors and cytokines that may be either pre-existent in the PMN or synthetized *de novo*. Subsequent recruitment of supporting stroma cells, ECM remodelling, the establishment of signalling cues able to sustain key oncogenic pathways and neovascularization are also essential (**Figure 1D**). It has been demonstrated that dormant bone metastasis overexpressing vascular cell adhesion molecule 1 (VCAM-1) recruit monocytic osteoclast progenitors by interacting with the cognate receptor integrin α4β1, ultimately promoting the transition of indolent micrometastasis to overt metastasis (133). On the same note, Coco, a secreted antagonist of TGF-β ligands, reactivates dormant breast cancer cells in lung metastatic sites by blocking lung-derived BMP ligands in orthotopic mouse models. Furthermore, depletion of Coco suppresses lung colonization by preventing the reactivation of dormant cells (134). These findings reveal that metastasis-initiating cells need to overcome organ-specific antimetastatic signals in order to undergo reactivation. Other key molecular determinants of cellular re-initiation were identified through *in vivo* selection of multiple ER-negative human breast cancer populations. Namely, tumorigenic-enriched subpopulations exhibited increased expression of *LAMA4, FOXQ1* and *NAP1L3* genes. *FOXQ1* is known to promote *LAMA4* expression, ultimately leading to enhanced clonal expansion. This activation cascade was followed by substratum detachment *in vitro*, tumour re-initiation in multiple organs, and DTC proliferation and colonization (135). Other studies point a non-canonical mechanism driving breast cancer metastatic reactivation in the lung, bone, and brain. Namely, the atypical tetraspanin TM4SF1

couples the collagen receptor tyrosine kinase DDR1 to the cortical adaptor syntenin 2 and, hence, to PKCα kinase which phosphorylates and activates JAK2, leading to the activation of STAT3. This non-canonical mechanism induces the expression of SOX2 and NANOG stem-like cells markers and drives metastasis reactivation (136). Such finding may account for the simultaneous appearance of multiorgan metastasis observed in advanced breast cancer patients.

Successful metastatic colonization is also dependent on the formation of supportive ECM. Amongst several examples, cancer cell-derived tenascin C (TNC) promotes the survival and outgrowth of pulmonary micrometastasis until the tumor stroma takes over as a source of TNC. Mechanistically, TNC enhances the expression of stem cell signalling components musashi homolog 1 (MSI1) and leucine-rich repeat-containing G protein-coupled receptor 5 (LGR5), a positive regulator of NOTCH signalling and a target gene of the WNT pathway respectively, supporting the fitness of metastasis-initiating breast cancer cells (137). Similarly, DTC induce periostin (POSTN) expression by fibroblasts in secondary target organs, which recruits Wnt ligands and increases Wnt signalling in stem-like DTC, allowing colonization in a ECM-dependent manner (138). The induction of fibrosis, associated with the deposition of type I collagen (Col-I) in metastatic microenvironments *in vivo*, induces dormant breast cancer cells proliferation through beta1-integrin activation. Namely, beta1-integrin activation of Src and focal adhesion kinase leads to extracellular signal-regulated kinase (ERK)-dependent myosin light chain phosphorylation by myosin light chain kinase and actin stress fibber formation, leading to breast cancer transition from dormancy to metastatic growth (139).

In summary, the last stages of colonization are strongly dependent on DTC overcoming cell dormancy. We are still beginning to understand the intricate mechanisms that dictate DTC fate and the role played by the microenvironment in this context. However, effective colonization largely relies on supportive microenvironment cues created during PMN formation or recruited upon DTC arrival.

Glycosylation: A key post-translational modification driving metastasis and its potential for targeted therapeutics

Most events underlying metastasis development require significant molecular and functional adaptations, starting with the membrane proteome. In fact, cell surface proteins need to endorse higher motility and plasticity to invade, attach to distinct and

evolving ECM, promote immune escape, mediate vascular intravasation and allow extravasation to targeted organs. The most accepted paradigm is that these phenotypic changes are accomplished at the expenses of significant proteome remodelling in response to microenvironmental stimuli. Moreover, membrane proteins assimilate extensive post-translational modifications mediated by sugars, termed glycosylation, which decisively contribute to protein function, stability and recognition by external ligands such as lectins and antibodies (140). Furthermore, glycans decisively contribute to increase the dimension, functional diversity and plasticity of the proteome. Two main classes of glycans can be found at the cell membrane, *O*-GalNAc glycans in serine and threonine residues and *N*-glycans at asparagine residues of polypeptide chains. Even though human glycan chains are built from a rather restricted number of sugar donors, the possibility of establishing different combinations of glycosidic linkages significantly amplifies structural diversity. Moreover, glycan chains do not present templated structures as observed for other post-translational modifications. Instead, the length, composition and structural organization of glycans are the result of tightly regulated processes mediated by an array of glycosyltransferases, glycosidases, sugar donors and other co-factors available in a given biological milieu (141). Alterations in the phosphorylation, sulfation, methylation and acetylation of sugar residues have also been described, which further diversifies the glycome, ultimately impacting on the glycoproteome (142). Protein glycosylation patterns evolve throughout the course of disease, mostly due to alterations in the transcription of glycosyltransferases (GT), mutations in key GT chaperone molecules, and miss-localization of GT in secretory organelles (143). Notwithstanding, little is known about functional mutations in glycogenes driving glycan alterations and glycogene mutations are thought to be rare (144). Notably, the levels and activity of these molecules rapidly change in response to environmental cues, endowing glycans with the necessary structural diversity to dynamically shape protein functions according to incoming molecular signals (43). This may include reductions in glycans length, increased branching or modification of terminal glycoepitopes driven by fucosylation and/or sialylation (43, 145).

On a functional standpoint, altering the glycosylation landscape of mucins, integrins, CD44 and cadherins severely compromises cell-cell and cell-ECM adhesion, frequently favouring cancer cells detachment and metastasis development (145-147) (**Figure 2**). On the same note, the more aggressive tumour cells experience a premature stop in *O*-GalNAc glycosylation, frequently driven by early sialylation in glycosylation pathways. From this hallmark event, a main cancer-associated glycoform arises, namely

the sialyl-Tn (STn) antigen. This known pancarcinoma antigen has been largely associated with invasion, metastasis development and poor prognosis in most advanced solid tumours (148). In addition, the STn antigen is absent from most healthy tissues and presents a very restricted pattern of expression in the respiratory and gastrointestinal epithelia, mostly associated with goblet cells. Inducing STn expression in cancer cells of distinct natures has resulted in increased motility and invasion capacity *in vitro* and *in vivo*, irrespectively of the used cancer model (149-153). Furthermore, mice xenografted with gastric cancer cells overexpressing the antigen presented less cohesive tumours and significantly decreased survival (152). The fact that STn expression may dramatically change the glycosylation landscape of many relevant adhesion proteins, including mucins, cadherins, integrins and CD44 may account for these findings (47). Moreover, the presence of short-chain *O*-GalNAc glycans induced the activation of EGFR and ErbB2 receptors and a transcriptomic signature switch of gastric cancer cells towards a more aggressive phenotype associated with patients' poor-survival (152) (**Figure 2**). Taken together these findings place the STn antigen as part of an array of molecular events that alter protein functionality towards more metastatic cell phenotypes. The STn has also impact on cell recognition by the immune system, namely it has been described to induce a tolerogenic phenotype in innate and adaptive immune cells against cancer cells, which was regarded as a crucial hallmark driving tumour cell survival (154). Furthermore, it has been recently demonstrated that STn and other sialylated short-chain *O*-glycans are recognized by immune cell sialic acid-binding immunoglobulin-type lectins, such as siglec 15, 7, and 9, in antigen presenting cells and NK cells to modulate inhibitory immune signalling (43) (**Figure 2**). Reinforcing the association between STn and metastasis, we have also recently described that the majority of CTC in the blood of metastatic gastric, colorectal and bladder cancer patients present this antigen, mimicking its expression in the primary tumours and corresponding metastasis (155). Of note, STn was not present in other circulating blood cells, reinforcing its cancer-associated nature (155). This provides the missing molecular evidence linking altered glycosylation of primary tumours with the patterns found in disseminated disease, setting STn as a potential target for clinical intervention. Interestingly, the studies that have addressed the functional impact of STn in cancer cells were almost exclusively performed using glycoengineered cell lines. In fact, in opposition to advanced solid tumours, cells grown in *in vitro* settings tend to lack or modestly express STn, denoting its strong dependence on microenvironment stimuli (149, 151, 154). However, the particular molecular cues supporting STn expression were not yet fully disclosed. Notwithstanding, there are

suggestions that low oxygen levels experienced by cancer cells under hypoxia may be amongst the contributing factors (47). Studies involving bladder cancer cell models demonstrated that hypoxia induces the downregulation of several glycosyltransferases involved in *O*-GalNAc glycans extension in an HIF-1• dependent manner, leading to an accumulation of short-chain glycan precursors (47). This is accompanied by the activation of transcription programs associated with EMT as well as increased motility and invasion; thus, in agreement with the functional role played by STn in bladder cancer and other solid tumours (47).

Two other well-known cancer-associated glycans are sialyl Lewis A (sLe^A) and its isomer sialyl-Lewis X (sLe^X), commonly found as terminal epitopes in both *O*- and *N*-glycans in glycoproteins and glycolipids. Similarly to the STn antigen, both have been associated with metastasis development and poor prognosis in a wide number of solid tumours (156, 157). However, while the STn antigen appears to be driving cancer cells to metastasis by inducing higher motility and invasion, $sLe^{A/X}$ epitopes appear to play a role as promoters of haematogenous dissemination (**Figure 2**). Tumour cell $sLe^{A/X}$ are specific ligands for E- and P-selectins on endothelial cells, ultimately mediating intravasation of tumour cells into circulation and extravasation to distant organs (158, 159) (**Figure 2**). Moreover, these glycoepitopes may decisively contribute to CTC survival in circulation by enabling attachment to P-selectin in activated platelets, forming protective clusters (160) (**Figure 2)**. Of note, inflammatory tumor microenvironments characterized by increases levels of pro-inflammatory cytokines, as IL-1β and IL-6, were found to regulate the expression of biosynthetic glycosyltransferases increasing the expression of these sialylated antigens in tumour cells (161, 162). These findings highlight the microenvironmental regulation of cancer cell glycosylation.

In addition, the role of glycans in cancer dissemination is not limited to the cancer cell. It is now known that cancer-derived extracellular vesicles display particular glycan coats that may differently interplay with other cells within the tumour microenvironment as well as at distant sites (163). Namely, studies using extracellular vesicles (EV) derived from liver cells have demonstrated that the modification of EV glycosylation could affect its biodistribution, since the enzymatic removal of sialic acids resulted in EV bioaccumulation into the lungs of mouse models (164). Nevertheless, these are still preliminary observations and more comprehensive approaches are required to fully disclose the role of glycans in intercellular communication under physiological and pathological settings. More detailed information on the structural and functional role of glycans, its alterations in cancer and potential therapeutics opportunities are beyond the

scope of this chapter; however, these topics have been comprehensively explored in several recent publications (43, 140, 145).

In summary, glycans directly mediate a wide array of relevant molecular and biological functions spanning all stages of metastasis development. Moreover, changes in glycans length, composition, density and distribution of glycosites decisively alter protein functions and immunogenicity patterns, contributing to more aggressive phenotypes. These events are triggered and sustained by pro-metastatic microenvironments, as comprehensively reviewed recently (43). The sweet side to this sour end resides on the fact that glycans present a cell-surface nature, easily accessible to several ligands. Moreover, exploiting the cancer-associated nature of certain known antigens such as STn and $sLe^{A/X}$ offers tremendous potential for designing novel detection methods and targeted therapeutics.

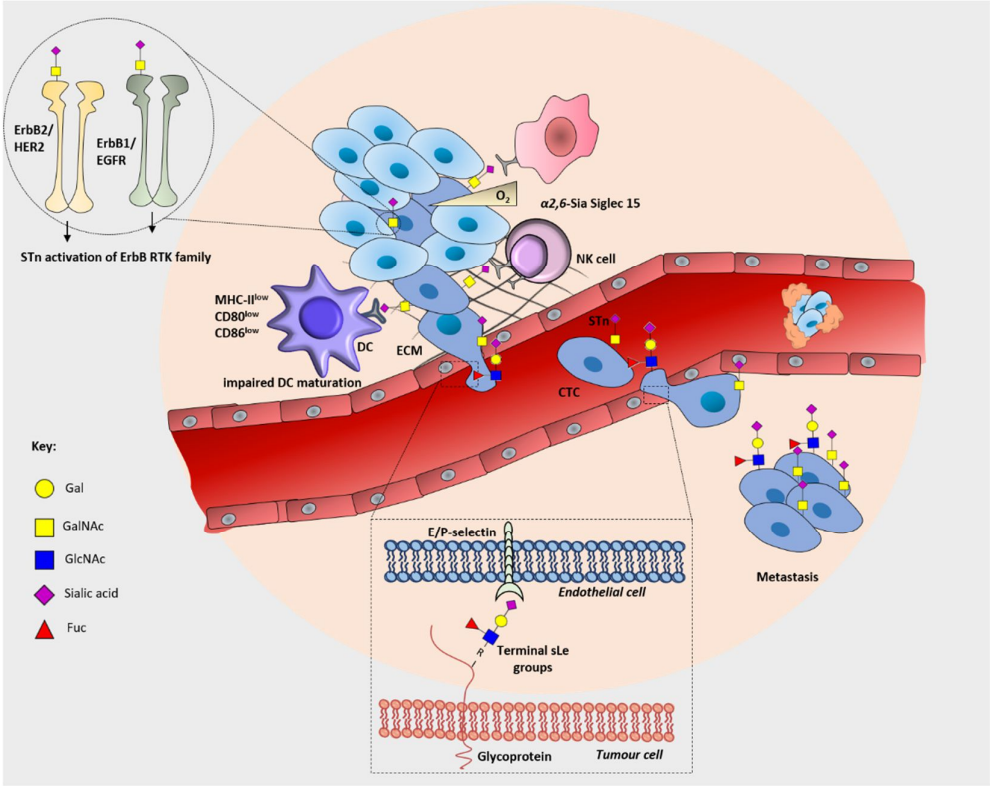


Figure 2. Sialyl-Tn overexpressing cells frequently acquire more motile and invasive phenotypes by weakly interacting with ECM components and stablishing poorly cohesive tumours. STn overexpression itself is though to be largely dependent on

microenvironmental cues as low oxygen tension within the tumour bulk. Furthermore, STn overexpression induces activation of EGFR and ErbB2 tyrosine kinase receptors (RTK), consequently initiating oncogenic pathways. STn also has impact on cell recognition by the immune system, namely by inducing a tolerogenic and immature phenotype of dendritic cells and by interacting with sialic acid-binding immunoglobulin-type lectins (SIGLEC), such as siglec 15 in antigen presenting cells and NK cells to modulate inhibitory responses. In turn, tumour cell $sLe^{A/X}$ are specific ligands for E- and P-selectins on endothelial cells, ultimately mediating intravasation of tumour cells into circulation and extravasation to distant organs. Moreover, these glycoepitopes may decisively contribute to CTC survival in circulation by enabling attachment to P-selectin in activated platelets, forming protective clusters.

Concluding Remarks

CTC are amongst the driving forces of metastasis and may be released from primary tumours and secondary lesions at all stages of disease. Moreover, elevated CTC counts in the blood of cancer patients has been associated not only with the presence of metastasis but also with higher risk of disease progression (165). It may also be indicative of occult micrometastasis, providing a good complement to imageology. Elevated CTC counts are associated with decreased response to treatment for a wide array of solid tumours and therapeutic schemes and may be used, in conjugation with other molecular markers, for follow-up (166, 167). CTC better reflect the molecular nature of metastasis and can be explored in non-invasive settings, making its interrogation more attractive than diagnostic tumour biopsies. This enables a more educated election of therapeutic schemes and may constitute a critical milestone towards novel therapeutic approaches. Such striking observations have revolutionized the concept of liquid biopsies and made CTC analysis an attractive tool for early detection of disseminated disease, enabling the adoption of more effective interventions, and a decisive option for treatment follow-up. The exploitation of CTC has been accompanied by ground-breaking observations regarding the role of exosomes in cell-cell communication and contribution to PMN formation. As such, exosome analysis with emphasis on circulating exosomes may complement CTC analysis and provide relevant blueprints on disease status. Technical advances and ongoing clinical studies are rapidly refining the use of CTC and exosome-based liquid biopsies for cancer control and these solutions are likely to reach clinical practice in a near future.

In parallel with the advances in liquid biopsies technology, the last ten years have allowed a more in-depth understanding of the critical milestones leading to metastasis. The focus has been progressively moving from looking at aggressive subpopulations of cancer cells has isolated entities to a more comprehensive vision in which cancer cells stablish supportive cross-talks with the tumour microenvironment. In the early steps of metastasis, primary stromal cells and hypoxia-mediated factors support tumour cell invasion by supplying pro-metastatic factors, compromising vasculature and stromal barrier integrity, endorsing EMT and promoting acquisition of stem-like properties. Subsequently, active stromal cells as tumour-associated macrophages, fibroblasts, neutrophils and platelets aid tumour cell intravasation into the tumour microvasculature by providing soluble factors, promoting the formation of invadopodia and ECM degradation. Once in the bloodstream, single CTC or small CTC groups form protective clusters with neutrophils, platelets and ECM components. In this process, CTC are protected from sheer stress in the bloodstream, avoid immune recognition and are primed to proliferate upon secondary site arrival. Furthermore, CTC are able to travel with activated stromal components as CAFs, determining a more favourable colonization step **(Figure 3)**. Interestingly, CTC may also return to their primary tumour in a process called tumour self-seeding. The interaction between self-seeded CTC and the tumor stroma promotes the release of signals that foster tumor growth, angiogenesis, invasion, and metastasis; thereby increasing the aggressivity of primary tumours. Distant metastasis organotropism is not random and depends on the formation of PMN capable of attracting CTC. PMN formation results from the combined systemic effects of tumour-secreted factors and tumour-shed extracellular vesicles (EV). Specifically, tumour and stroma-derived growth factors and chemokines are potent chemoattractants of immune cells to PMN, ultimately aiding CTC colonization. On the same note, tumour-derived exosomes contribute to the formation of PMN by upregulating pro-inflammatory and angiogenic molecules, promoting ECM and endorsing immune escape. Once on the secondary site, cancer cells may lay quiescent or progress to actively growing metastasis in a way that is dependent on the quality of the PMN and on the genetic background of homing cells. Even though these findings support the decisive role of the microenvironment on metastasis development, the intricate molecular network supporting these events remains to be fully elucidated. Nevertheless, several therapeutic strategies targeting the tumour microenvironment have already been proposed and progressed to human clinical trials and clinical implementation (**Figure 4C**). Some of this approaches include: i) agents that degrade and/or deconstruct the supportive tumour

ECM; as heparanases iii) MMP inhibitors; iii) HIF-1α and HIF-1α targets inhibitors; iv) proton exchanger and carbonic anhydrase inhibitors; v) antiangiogenic drugs; vi) multireceptor tyrosine kinase inhibitors; vii) chemical agents for the inhibition of the recruitment and/or proliferation of myeloid-derived suppressor cells (MDSC) and tumor-associated macrophages (TAM), viii) chemical agents for the selective depletion of TAM and MDSC; ix) chemical agents for the re-education of TAMs and MDSCs anti-tumor functions; x) chemical agents for the differentiation of MDSC; xi) chemical agents targeting the pro-tumoral mediators of TAM and MDSC; xii) interleukin or interleukin-receptor inhibitors; xiii) immune checkpoint inhibitors; xiv) CAF targeting agents; xv) exosome biogenesis inhibitors; and xvi) exosomes for the delivery of drugs and siRNA. All these therapeutic approaches have been extensively described in recent publications (168, 169), including combinatory therapeutic strategies.

Targeting cancer cells glycosylation may also offer novel strategies to control disease dissemination and ultimately provide curative treatments. In this context, it is well described that the microenvironment may induce critical alterations in the glycosylation of membrane proteins that favour disease dissemination (43). Namely, glycans like STn and $sLe^{A/X}$ are absent or modestly expressed in healthy adult organs but significantly overexpressed in more aggressive cancer subpopulations, opening an avenue for targeted therapeutics. Furthermore, several studies support the functional impact of these glycans in CTC development and survival, setting an important rationale for intervention (155, 170). A more comprehensive interrogation of the glycoproteome for cancer-associated protein species carrying these post-translational modifications is now required to gain more functional insights and design bi-specific ligands.

In summary, this chapter highlights the multifactorial molecular nature underlying metastasis development and its strong dependence on the tumour microenvironment. Therefore, disease control will most likely require the introduction of multiplex biomarker panels for patient stratification, enabling early intervention and adoption of more effective therapeutic schemes according to the potential for metastasis development. Many studies also indicate that patients will most likely benefit from combinatory therapies capable of supressing relevant microenvironmental cues responsible for CTC development, maintenance in circulation and distant location colonization. Exploiting membrane glycosylation may provide highly specific means to target these cells while reducing off-target effects.

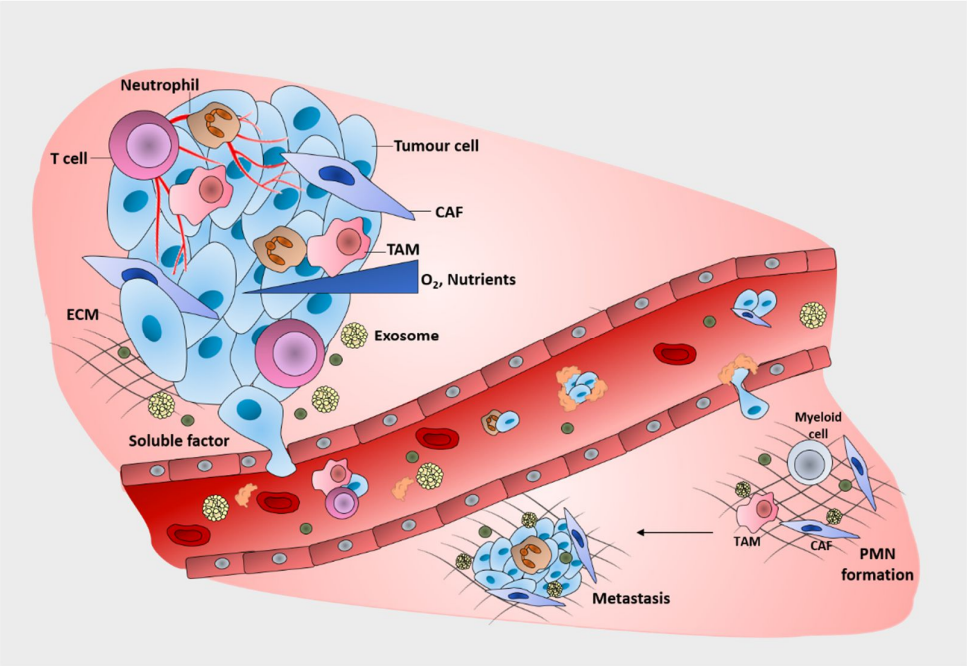


Figure 3. In the first steps of the metastatic cascade, stromal cells and hypoxia-mediated factors promote tumour cell invasion by supplying the tumour microenvironment with pro-metastatic factors, favouring EMT, promoting ECM remodelling and vascular leakage. Subsequently, tumour cell intravasation into the tumour microvasculature is supported by soluble factors provided by active stromal cells as tumour-associated macrophages, fibroblasts, neutrophils and platelets. Once in the bloodstream, single CTC or small CTC groups form protective clusters with neutrophils, platelets and ECM components. These cooperative crosslinks protect CTC from sheer stress and immune recognition, while priming them for proliferation upon successful dissemination. CTC also stablish clusters with activated CAFs, determining a more favourable colonization step by bringing their own supporting stroma to distant locations.

Ongoing challenges and future perspectives

Controlling cancer progression and spread is a tremendously challenging task due to significant molecular heterogeneity and distinct pathological nature presented by different types of human tumours. The multifactorial nature of the tumour-stroma cross-talk, the wide array of microenvironmental factors driving metastasis and its constantly adapting nature during the course of disease management further adds to these difficulties. High throughput technologies and mature bioinformatics tools now enable the

exploitation of cancer systems biology involving a comprehensive integration of multi-omics approaches (genomics, transcriptomics, proteomics, metabolomics and others). This is expected to provide important clues on the molecular and functional mechanisms adopted by cancer systems at multiple biological scales and ultimately pinpoint relevant biological networks nodes for intervention.

Another defying topic relates with the adoption of relevant models for biomedical research. Most available studies have focused on the cross-talk between cancer cells and selected stromal elements exploiting the potential of 2D co-cultures *in vitro* and mouse xenograft cancer models *in vivo*. However, adequate preclinical models that mimic the tumour microenvironment are warranted. In this context, currently available 3D culture models, including tumor tissue explants, “tumor on a chip”, and multicellular tumor spheroids (MCTS) may be useful tools to achieve increasingly translational research. Furthermore, efforts should be conducted to progressively increase the complexity of these models by taking advantage of novel microfluidics systems, allowing mimicking the cellular vasculature and the movement of biological fluids; thereby, improving similarities with real tumour microenvironments. Current *in vivo* studies also pose some limitations that should be addressed in the near future envisaging translatable results. For example, the maintenance of more aggressive cancer cells subpopulations is significantly dependent on the capacity to evade the immune system. In particular, CTC employ several mechanisms of immune escape and even immune cooperation to avoid clearance. However, most studies employ athymic mouse models that fail to reflect the contribution of the immune system to tumour progression at different levels. As such, it becomes pressing to introduce engineered animal models capable of reflecting the stepwise co-evolution of cancer and immunity. Other potentially interesting approaches may contemplate inducing tumours in relevant animal models that reflect the molecular nature of human tumours. For example, urothelial tumours induced by exposure to N-butyl-N-(4-hydroxybutyl)nitrosamine (BBN) have been shown to reflect the molecular nature of human cancer and have been extensively used for molecular and functional studies and therapy development (171). Synergistic models xenografted with relevant cell models may also constitute a valuable alternative.

In addition, little is known about the molecular nature of CTC, in particular the mechanisms that define their capacity to resist microenvironmental challenges and the molecular features that enable them to target different organs. The extent of CTC self-seeding events and its contribution to late-stage disease is also poorly understood and may be critical to comprehend the spatial-temporal evolution of disease progression.

There are also many unanswered questions regarding the resemblances between CTC, the parental lesions and the metastasis at different levels. It is also becoming clear that CTC of the same patient may present significant genotypic and phenotypic differences, even though the functional and clinical significance of these observations remain to be fully understood. The response to these outstanding questions will certainly aid more adequate therapy selection and the design of more effective approaches to treat cancer patients.

In parallel, the liquid biopsies field is being propelled by technological breakthroughs in micro- and nano-sensors. These advances are reaching higher technology readiness levels and will most likely be decisive tools for CTC-based biomedical research and clinical practice. Furthermore, liquid biopsies are starting to evolve to multiparametric settings, exploiting CTC, exosomes, circulating nucleic acids and other biomolecules to provide more comprehensive models for patient stratification and to guide clinical decisions (**Figure 4A**). We believe that future studies should continue to refine multiplex settings using comprehensive panomics approaches backed by artificial intelligence. We anticipate that this will be the cornerstone towards progressively more personalized solutions. However, the expansion of CTC outside the human body remains yet a critical milestone. Its limited success is many times accomplished at the expenses of losing CTC original molecular phenotypes. A deeper understanding of the microenvironmental cues faced by CTC in different physiological and clinicopathological contexts is required to overcome these limitations. Successful and robust expansion of CTC will be instrumental to evaluate drug susceptibility, disclose relevant driver mutations, CTC immunogenicity and mechanisms of immune escape as well as therapy design.

Another critical challenge relates with the adoption of preventive measures for metastasis development at early stages of disease. This will strongly benefit from a better understanding of the events underlying PMN formation. Most promising strategies may include blocking molecular messengers such as exosomes and pro-inflammatory factors, hampering angiogenesis and ECM interactions, preventing the recruitment of anti-inflammatory and pro-tumour immune cells and employing strategies to revive anti-tumour immune responses (168). In addition, it will be crucial to improve therapeutic options against more aggressive cancer cells in primary tumours, CTC and the metastasis. In this context, the current therapeutics tool box is well equipped with relevant molecular tools to stimulate the immune system against more aggressive cancer cells, including immune check-point inhibitors, vaccines based on cancer neoantigens

and chimeric antigen receptor (CAR) T-cell therapies. Moreover, there is a wide panoply of novel therapeutic antibodies targeting relevant cell-surface receptors and capable of inhibiting key oncogenic pathways, while cross priming antigen-presenting cells for tumour destruction. Tremendous advances in genetic therapies further increase the number of current options. Exciting progresses have also been made concerning intelligent nanoparticles specifically bioengineered to enhance drug delivery to more aggressive cells. However, the success of these approaches against disseminated disease is still heterogeneous and will most likely benefit from a deeper understanding of the events underlying metastasis formation. It is also crucial to identify cell-surface molecules that may be used to specifically target these therapies to cancer cells while limiting off-target effects. A comprehensive glycoproteomics screening of human tumours offers an exciting opportunity for identification of unique molecular signatures. Notably, there are still few reports on the nature of the glycoproteome of primary lesions and none on metastasis, leaving still a long ground to cover envisaging clinically useful glycobiomarkers. However, in the past five years many of the limitations related with the annotation of minute amounts of glycospecies by mass spectrometry have been overcome, laying the foundations for comprehensive glycoproteomics studies. As the sensitivity of modern mass spectrometers improves, it will also be possible to perform analysis at near single cell level, a crucial milestone when addressing minor amounts of biological material generally recovered from biopsied metastasis or CTC isolated from cancer patients. Taking advantage of these novel technological advancements and looking at highly context-specific glycosignatures as attractive targets for personalized medicine, several glycan-based therapeutic strategies and immunotherapies have emerged. Namely, inhibitors of glycosyltransferases catalytic activity (172) and glycoengineered antibodies capable of inducing increased antibody-dependent cellular cytotoxicity (173). Furthermore, glycan-based antibodies may be used to guide nanotherapies (174, 175) or improve the cancer specificity of chimeric antigen receptors (CAR) of genetically modified T cells (176).On another note, tumour specific glycoliposomes can be used to prime DC through DC-SIGN engagement to stimulate tumour-specific $CD4^+$ and $CD8^+$ T cell anti-tumour responses, constituting a promising glycan-based immunotherapy (177). Finally, anticancer multicomponent glycoconjugate vaccines, based on glycan antigens coupled to T-cell peptide epitopes or immunostimulant epitopes, have been demonstrated effective in circumventing cancer immunotolerance (178, 179), providing an appealing option for the much-awaited development of new glycan-based therapeutic agents (**Figure 4B**). In summary, to

effectively treat metastasis, we must inhibit fundamental metastatic processes and develop specific preclinical and clinical strategies that excel the classical view focused on cancer cells, setting the focus on the intense crosstalk between different types of primary tumour cells, CTC, and DTC. Nevertheless, current technological advances ensure capacity to support these challenges and the transition to molecular-based precision oncology.

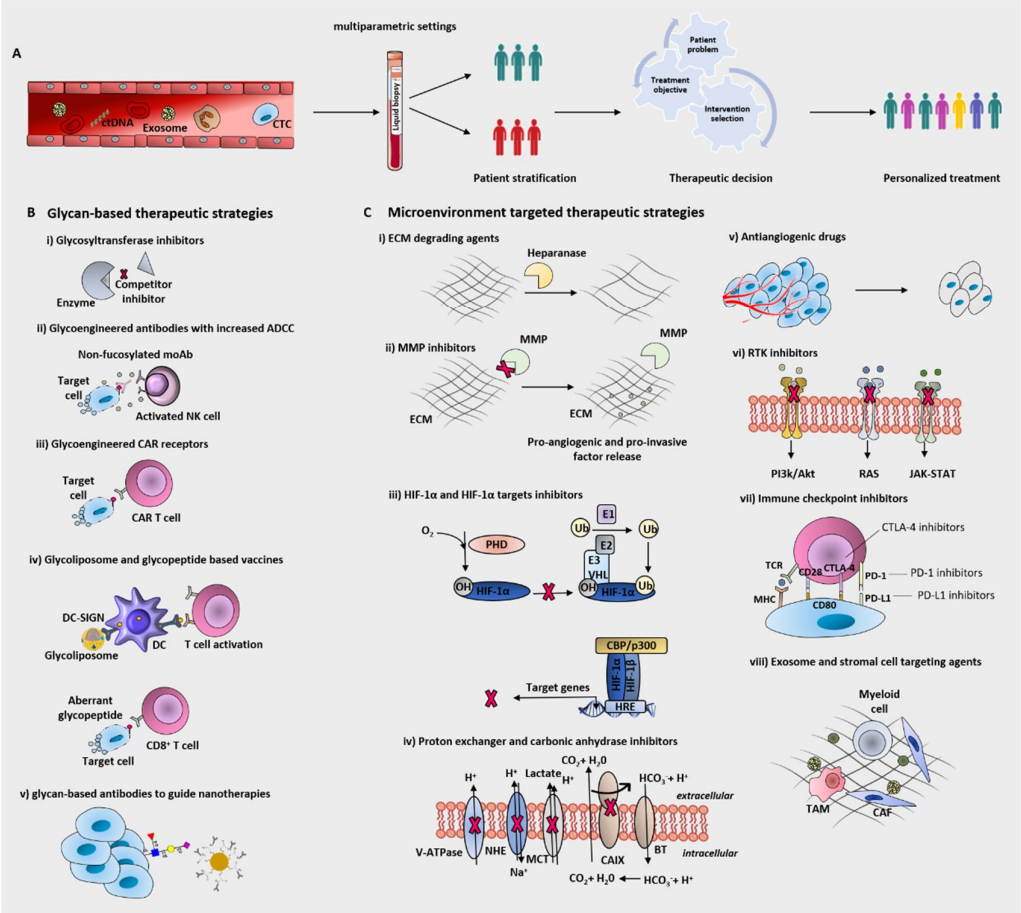


Figure 4. A) Liquid biopsies are starting to evolve to multiparametric settings, exploiting CTC, exosomes, circulating nucleic acids (ctDNA) and other biomolecules to provide more comprehensive models for patient stratification and to guide clinical decisions towards personalized medicine. **B)** Several glycan-based therapeutic strategies and immunotherapies have emerged. Namely, inhibitors of glycosyltransferases catalytic activity, glycoengineered antibodies capable of inducing increased antibody-dependent cellular cytotoxicity (ADCC), glycan-based antibodies to guide nanotherapies or improve

the cancer specificity of chimeric antigen receptors (CAR) of genetically modified T cells, as well as anticancer multicomponent glycoconjugate vaccines. **C)** Several therapeutic strategies targeting the tumour microenvironment have already been proposed. These include: i) ECM deconstructing agents as heparanases iii) MMP inhibitors; iii) HIF-1α and HIF-1α targets inhibitors; iv) proton exchanger and carbonic anhydrase inhibitors (mainly vacuolar ATPase (V-ATPase) Na+/H+ exchanger (NHE), monocarboxylate transporters (MCTs), and carbonic anhydrases (CAs); v) antiangiogenic drugs; vi) multireceptor tyrosine kinase (RTK) inhibitors; vii) immune checkpoint inhibitors; viii) chemical agents targeting exosome biosynthesis and specific stromal cells as myeloid-derived suppressor cells (MDSC), tumor-associated macrophages (TAM) and CAF targeting agents.

Acknowledgments

The authors wish to acknowledge the Portuguese Foundation for Science and Technology (FCT) for the human resources grants: PhD grant SFRH/BD/111242/2015 (AP) and FCT auxiliary researcher grant CEECIND/03186/2017 (JAF). FCT is co-financed by European Social Fund (ESF) under Human Potential Operation Programme (POPH) from National Strategic Reference Framework (NSRF). The authors also acknowledge FCT the funding for CI-IPOP research unit (PEst-OE/SAU/UI0776/201), the Portuguese Oncology Institute of Porto Research Centre (CI-IPOP-29-2014; CI-IPOP-58-2015) and PhD Programs in Biomedicine and Pathology and Molecular Genetics of ICBAS-University of Porto. The authors EF, DF and MR also acknowledge the financial support of the Research Group of Digestive Cancers (GICD) and the “Early stage cancer treatment, driven by context of molecular imaging (ESTIMA)” framework (NORTE-01-0145-FEDER-000027) and IPO-Score (DSAIPA/DS/0042/2018) for financial support.

References

1. Mohammed SI, Torres-Luquis O, Walls E, Lloyd F. Lymph-circulating tumor cells show distinct properties to blood-circulating tumor cells and are efficient metastatic precursors. Mol Oncol. 2019;13(6):1400-18.
2. Wong SY, Hynes RO. Lymphatic or hematogenous dissemination: how does a metastatic tumor cell decide? Cell Cycle. 2006;5(8):812-7.
3. Charafe-Jauffret E, Ginestier C, Iovino F, Wicinski J, Cervera N, Finetti P, et al. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. Cancer Res. 2009;69(4):1302-13.
4. Pang R, Law WL, Chu AC, Poon JT, Lam CS, Chow AK, et al. A subpopulation of CD26+ cancer stem cells with metastatic capacity in human colorectal cancer. Cell Stem Cell. 2010;6(6):603-15.
5. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704-15.
6. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. Cell. 2009;139(5):871-90.
7. Maheswaran S, Haber DA. Circulating tumor cells: a window into cancer biology and metastasis. Curr Opin Genet Dev. 2010;20(1):96-9.
8. Dong Y, Skelley AM, Merdek KD, Sprott KM, Jiang C, Pierceall WE, et al. Microfluidics and circulating tumor cells. J Mol Diagn. 2013;15(2):149-57.
9. Yan W-T, Cui X, Chen Q, Li Y-F, Cui Y-H, Wang Y, et al. Circulating tumor cell status monitors the treatment responses in breast cancer patients: a meta-analysis. Scientific Reports. 2017;7:43464.
10. Li J, Fu W, Zhang W, Li P. High Number of Circulating Tumor Cells Predicts Poor Survival of Cutaneous Melanoma Patients in China. Med Sci Monit. 2018;24:324-31.
11. Azevedo R, Soares J, Peixoto A, Cotton S, Lima L, Santos LL, et al. Circulating tumor cells in bladder cancer: Emerging technologies and clinical implications foreseeing precision oncology. Urol Oncol. 2018;36(5):221-36.
12. Huang Q, Wang Y, Chen X, Wang Y, Li Z, Du S, et al. Nanotechnology-Based Strategies for Early Cancer Diagnosis Using Circulating Tumor Cells as a Liquid Biopsy. Nanotheranostics. 2018;2(1):21-41.

13. Ma Y-HV, Middleton K, You L, Sun Y. A review of microfluidic approaches for investigating cancer extravasation during metastasis. Microsystems &Amp; Nanoengineering. 2018;4:17104.

14. Josefsson A, Larsson K, Mansson M, Bjorkman J, Rohlova E, Ahs D, et al. Circulating tumor cells mirror bone metastatic phenotype in prostate cancer. Oncotarget. 2018;9(50):29403-13.

15. Jiang R, Lu YT, Ho H, Li B, Chen JF, Lin M, et al. A comparison of isolated circulating tumor cells and tissue biopsies using whole-genome sequencing in prostate cancer. Oncotarget. 2015;6(42):44781-93.

16. Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. Nat Med. 2014;20(8):897-903.

17. Khoo BL, Grenci G, Lim YB, Lee SC, Han J, Lim CT. Expansion of patient-derived circulating tumor cells from liquid biopsies using a CTC microfluidic culture device. Nature Protocols. 2017;13:34.

18. Tang M, Wen CY, Wu LL, Hong SL, Hu J, Xu CM, et al. A chip assisted immunomagnetic separation system for the efficient capture and in situ identification of circulating tumor cells. Lab Chip. 2016;16(7):1214-23.

19. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. Cell. 2010;141(1):52-67.

20. Giannelli G, Falk-Marzillier J, Schiraldi O, Stetler-Stevenson WG, Quaranta V. Induction of cell migration by matrix metalloprotease-2 cleavage of laminin-5. Science. 1997;277(5323):225-8.

21. Cheng K, Xie G, Raufman JP. Matrix metalloproteinase-7-catalyzed release of HB-EGF mediates deoxycholyltaurine-induced proliferation of a human colon cancer cell line. Biochem Pharmacol. 2007;73(7):1001-12.

22. Gocheva V, Zeng W, Ke D, Klimstra D, Reinheckel T, Peters C, et al. Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. Genes Dev. 2006;20(5):543-56.

23. Vasiljeva O, Papazoglou A, Kruger A, Brodoefel H, Korovin M, Deussing J, et al. Tumor cell-derived and macrophage-derived cathepsin B promotes progression and lung metastasis of mammary cancer. Cancer Res. 2006;66(10):5242-50.

24. Gorden DL, Fingleton B, Crawford HC, Jansen DE, Lepage M, Matrisian LM. Resident stromal cell-derived MMP-9 promotes the growth of colorectal metastases in

the liver microenvironment. International journal of cancer Journal international du cancer. 2007;121(3):495-500.

25. Oguma K, Oshima H, Aoki M, Uchio R, Naka K, Nakamura S, et al. Activated macrophages promote Wnt signalling through tumour necrosis factor-alpha in gastric tumour cells. EMBO J. 2008;27(12):1671-81.

26. Sager R, Sheng S, Pemberton P, Hendrix MJ. Maspin. A tumor suppressing serpin. Adv Exp Med Biol. 1997;425:77-88.

27. Ojalvo LS, Whittaker CA, Condeelis JS, Pollard JW. Gene expression analysis of macrophages that facilitate tumor invasion supports a role for Wnt-signaling in mediating their activity in primary mammary tumors. J Immunol. 2010;184(2):702-12.

28. Mantovani A. Molecular pathways linking inflammation and cancer. Curr Mol Med. 2010;10(4):369-73.

29. Dube DH, Bertozzi CR. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. Nat Rev Drug Discov. 2005;4(6):477-88.

30. Dirat B, Bochet L, Dabek M, Daviaud D, Dauvillier S, Majed B, et al. Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion. Cancer Res. 2011;71(7):2455-65.

31. Osuala KO, Sameni M, Shah S, Aggarwal N, Simonait ML, Franco OE, et al. Il-6 signaling between ductal carcinoma in situ cells and carcinoma-associated fibroblasts mediates tumor cell growth and migration. BMC Cancer. 2015;15:584.

32. Giannoni E, Bianchini F, Masieri L, Serni S, Torre E, Calorini L, et al. Reciprocal activation of prostate cancer cells and cancer-associated fibroblasts stimulates epithelial-mesenchymal transition and cancer stemness. Cancer Res. 2010;70(17):6945-56.

33. Yeh CR, Hsu I, Song W, Chang H, Miyamoto H, Xiao GQ, et al. Fibroblast ERalpha promotes bladder cancer invasion via increasing the CCL1 and IL-6 signals in the tumor microenvironment. Am J Cancer Res. 2015;5(3):1146-57.

34. Fang WB, Mafuvadze B, Yao M, Zou A, Portsche M, Cheng N. TGF-beta Negatively Regulates CXCL1 Chemokine Expression in Mammary Fibroblasts through Enhancement of Smad2/3 and Suppression of HGF/c-Met Signaling Mechanisms. PLoS One. 2015;10(8):e0135063.

35. Mishra P, Banerjee D, Ben-Baruch A. Chemokines at the crossroads of tumor-fibroblast interactions that promote malignancy. J Leukoc Biol. 2011;89(1):31-9.

36. Gaggioli C, Hooper S, Hidalgo-Carcedo C, Grosse R, Marshall JF, Harrington K, et al. Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. Nat Cell Biol. 2007;9(12):1392-400.

37. Katanov C, Lerrer S, Liubomirski Y, Leider-Trejo L, Meshel T, Bar J, et al. Regulation of the inflammatory profile of stromal cells in human breast cancer: prominent roles for TNF-alpha and the NF-kappaB pathway. Stem Cell Res Ther. 2015;6:87.

38. Limoge M, Safina A, Beattie A, Kapus L, Truskinovsky AM, Bakin AV. Tumor-fibroblast interactions stimulate tumor vascularization by enhancing cytokine-driven production of MMP9 by tumor cells. Oncotarget. 2017;8(22):35592-608.

39. Erez N, Truitt M, Olson P, Arron ST, Hanahan D. Cancer-Associated Fibroblasts Are Activated in Incipient Neoplasia to Orchestrate Tumor-Promoting Inflammation in an NF-kappaB-Dependent Manner. Cancer Cell. 2010;17(2):135-47.

40. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. Science. 2011;331(6024):1559-64.

41. Guo J, Li X, Tan Z, Lu W, Yang G, Guan F. Alteration of N-glycans and expression of their related glycogenes in the epithelial-mesenchymal transition of HCV29 bladder epithelial cells. Molecules. 2014;19(12):20073-90.

42. Labelle M, Begum S, Hynes RO. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. Cancer Cell. 2011;20(5):576-90.

43. Peixoto A, Relvas-Santos M, Azevedo R, Santos LL, Ferreira JA. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Frontiers in Oncology. 2019;9(380).

44. Morfoisse F, Kuchnio A, Frainay C, Gomez-Brouchet A, Delisle MB, Marzi S, et al. Hypoxia induces VEGF-C expression in metastatic tumor cells via a HIF-1alpha-independent translation-mediated mechanism. Cell Rep. 2014;6(1):155-67.

45. Choi B-J, Park S-A, Lee S-Y, Cha YN, Surh Y-J. Hypoxia induces epithelial-mesenchymal transition in colorectal cancer cells through ubiquitin-specific protease 47-mediated stabilization of Snail: A potential role of Sox9. Scientific Reports. 2017;7(1):15918.

46. Qin J, Liu Y, Lu Y, Liu M, Li M, Li J, et al. Hypoxia-inducible factor 1 alpha promotes cancer stem cells-like properties in human ovarian cancer cells by upregulating SIRT1 expression. Scientific Reports. 2017;7(1):10592.

47. Peixoto A, Fernandes E, Gaiteiro C, Lima L, Azevedo R, Soares J, et al. Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. Oncotarget. 2016;7(39):63138-57.

48. Branco-Price C, Zhang N, Schnelle M, Evans C, Katschinski DM, Liao D, et al. Endothelial cell HIF-1alpha and HIF-2alpha differentially regulate metastatic success. Cancer Cell. 2012;21(1):52-65.

49. Cooke VG, LeBleu VS, Keskin D, Khan Z, O'Connell JT, Teng Y, et al. Pericyte depletion results in hypoxia-associated epithelial-to-mesenchymal transition and metastasis mediated by met signaling pathway. Cancer Cell. 2012;21(1):66-81.

50. Goswami S, Sahai E, Wyckoff JB, Cammer M, Cox D, Pixley FJ, et al. Macrophages promote the invasion of breast carcinoma cells via a colony-stimulating factor-1/epidermal growth factor paracrine loop. Cancer Res. 2005;65(12):5278-83.

51. Roh-Johnson M, Bravo-Cordero JJ, Patsialou A, Sharma VP, Guo P, Liu H, et al. Macrophage contact induces RhoA GTPase signaling to trigger tumor cell intravasation. Oncogene. 2014;33(33):4203-12.

52. Wyckoff JB, Wang Y, Lin EY, Li JF, Goswami S, Stanley ER, et al. Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. Cancer Res. 2007;67(6):2649-56.

53. Harney AS, Arwert EN, Entenberg D, Wang Y, Guo P, Qian BZ, et al. Real-Time Imaging Reveals Local, Transient Vascular Permeability, and Tumor Cell Intravasation Stimulated by TIE2hi Macrophage-Derived VEGFA. Cancer Discov. 2015;5(9):932-43.

54. Zervantonakis IK, Hughes-Alford SK, Charest JL, Condeelis JS, Gertler FB, Kamm RD. Three-dimensional microfluidic model for tumor cell intravasation and endothelial barrier function. Proc Natl Acad Sci U S A. 2012;109(34):13515-20.

55. Wang J, Cao Z, Zhang XM, Nakamura M, Sun M, Hartman J, et al. Novel mechanism of macrophage-mediated metastasis revealed in a zebrafish model of tumor development. Cancer Res. 2015;75(2):306-15.

56. Pena C, Cespedes MV, Lindh MB, Kiflemariam S, Mezheyeuski A, Edqvist PH, et al. STC1 expression by cancer-associated fibroblasts drives metastasis of colorectal cancer. Cancer Res. 2013;73(4):1287-97.

57. Bekes EM, Schweighofer B, Kupriyanova TA, Zajac E, Ardi VC, Quigley JP, et al. Tumor-recruited neutrophils and neutrophil TIMP-free MMP-9 regulate coordinately the levels of tumor angiogenesis and efficiency of malignant cell intravasation. Am J Pathol. 2011;179(3):1455-70.

58. Szczerba BM, Castro-Giner F, Vetter M, Krol I, Gkountela S, Landin J, et al. Neutrophils escort circulating tumour cells to enable cell cycle progression. Nature. 2019;566(7745):553-7.

59. Jansson S, Bendahl PO, Larsson AM, Aaltonen KE, Ryden L. Prognostic impact of circulating tumor cell apoptosis and clusters in serial blood samples from patients with metastatic breast cancer in a prospective observational cohort. BMC Cancer. 2016;16:433.

60. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. 2011;11(2):123-34.

61. Placke T, Orgel M, Schaller M, Jung G, Rammensee HG, Kopp HG, et al. Platelet-derived MHC class I confers a pseudonormal phenotype to cancer cells that subverts the antitumor reactivity of natural killer immune cells. Cancer Res. 2012;72(2):440-8.

62. Zitvogel L, Tesniere A, Kroemer G. Cancer despite immunosurveillance: immunoselection and immunosubversion. Nat Rev Immunol. 2006;6(10):715-27.

63. Duda DG, Duyverman AM, Kohno M, Snuderl M, Steller EJ, Fukumura D, et al. Malignant cells facilitate lung metastasis by bringing their own soil. Proc Natl Acad Sci U S A. 2010;107(50):21677-82.

64. Alix-Panabieres C, Riethdorf S, Pantel K. Circulating tumor cells and bone marrow micrometastasis. Clin Cancer Res. 2008;14(16):5013-21.

65. Abdel-Ghany M, Cheng HC, Elble RC, Pauli BU. The breast cancer beta 4 integrin and endothelial human CLCA2 mediate lung metastasis. J Biol Chem. 2001;276(27):25438-46.

66. Wang H, Fu W, Im JH, Zhou Z, Santoro SA, Iyer V, et al. Tumor cell alpha3beta1 integrin and vascular laminin-5 mediate pulmonary arrest and metastasis. J Cell Biol. 2004;164(6):935-41.

67. Brown DM, Ruoslahti E. Metadherin, a cell surface protein in breast tumors that mediates lung metastasis. Cancer Cell. 2004;5(4):365-74.

68. Hess KR, Varadhachary GR, Taylor SH, Wei W, Raber MN, Lenzi R, et al. Metastatic patterns in adenocarcinoma. Cancer. 2006;106(7):1624-33.

69. Kim JW, Wong CW, Goldsmith JD, Song C, Fu W, Allion MB, et al. Rapid apoptosis in the pulmonary vasculature distinguishes non-metastatic from metastatic melanoma cells. Cancer Lett. 2004;213(2):203-12.

70. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful

extravasation and limited survival of early micrometastases. Am J Pathol. 1998;153(3):865-73.

71. Nguyen DX, Bos PD, Massague J. Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer. 2009;9(4):274-84.

72. Kim MY, Oskarsson T, Acharyya S, Nguyen DX, Zhang XH, Norton L, et al. Tumor self-seeding by circulating cancer cells. Cell. 2009;139(7):1315-26.

73. Arihiro K, Oda H, Kaneko M, Inai K. Cytokines facilitate chemotactic motility of breast carcinoma cells. Breast Cancer. 2000;7(3):221-30.

74. Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, et al. Lung metastasis genes couple breast tumor size and metastatic spread. Proc Natl Acad Sci U S A. 2007;104(16):6740-5.

75. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature. 2005;436(7050):518-24.

76. Dondossola E, Crippa L, Colombo B, Ferrero E, Corti A. Chromogranin A regulates tumor self-seeding and dissemination. Cancer Res. 2012;72(2):449-59.

77. Liu H, Chen W, Zhi X, Chen EJ, Wei T, Zhang J, et al. Tumor-derived exosomes promote tumor self-seeding in hepatocellular carcinoma by transferring miRNA-25-5p to enhance cell motility. Oncogene. 2018;37(36):4964-78.

78. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, et al. Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer. 2017;17(5):302-17.

79. Psaila B, Lyden D. The metastatic niche: adapting the foreign soil. Nat Rev Cancer. 2009;9(4):285-93.

80. Kaplan RN, Riba RD, Zacharoulis S, Bramley AH, Vincent L, Costa C, et al. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. Nature. 2005;438(7069):820-7.

81. Joyce JA, Pollard JW. Microenvironmental regulation of metastasis. Nat Rev Cancer. 2009;9(4):239-52.

82. Shojaei F, Wu X, Qu X, Kowanetz M, Yu L, Tan M, et al. G-CSF-initiated myeloid cell mobilization and angiogenesis mediate tumor refractoriness to anti-VEGF therapy in mouse models. Proc Natl Acad Sci U S A. 2009;106(16):6742-7.

83. Melgarejo E, Medina MA, Sanchez-Jimenez F, Urdiales JL. Monocyte chemoattractant protein-1: a key mediator in inflammatory processes. Int J Biochem Cell Biol. 2009;41(5):998-1001.

84. Qian BZ, Li J, Zhang H, Kitamura T, Zhang J, Campion LR, et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature. 2011;475(7355):222-5.

85. Lu X, Kang Y. Chemokine (C-C motif) ligand 2 engages CCR2+ stromal cells of monocytic origin to promote breast cancer metastasis to lung and bone. J Biol Chem. 2009;284(42):29087-96.

86. Chen W, Hoffmann AD, Liu H, Liu X. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. NPJ Precis Oncol. 2018;2(1):4.

87. Granot Z, Henke E, Comen EA, King TA, Norton L, Benezra R. Tumor entrained neutrophils inhibit seeding in the premetastatic lung. Cancer Cell. 2011;20(3):300-14.

88. Sceneay J, Chow MT, Chen A, Halse HM, Wong CS, Andrews DM, et al. Primary tumor hypoxia recruits CD11b+/Ly6Cmed/Ly6G+ immune suppressor cells and compromises NK cell cytotoxicity in the premetastatic niche. Cancer Res. 2012;72(16):3906-11.

89. Semenza GL. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. Oncogene. 2010;29(5):625-34.

90. Rankin EB, Giaccia AJ. Hypoxic control of metastasis. Science. 2016;352(6282):175-80.

91. Lukanidin E, Sleeman JP. Building the niche: the role of the S100 proteins in metastatic growth. Semin Cancer Biol. 2012;22(3):216-25.

92. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of S100 proteins. Curr Mol Med. 2013;13(1):24-57.

93. Hiratsuka S, Nakamura K, Iwai S, Murakami M, Itoh T, Kijima H, et al. MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. Cancer Cell. 2002;2(4):289-300.

94. Hiratsuka S, Watanabe A, Aburatani H, Maru Y. Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. Nat Cell Biol. 2006;8(12):1369-75.

95. Hiratsuka S, Watanabe A, Sakurai Y, Akashi-Takamura S, Ishibashi S, Miyake K, et al. The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase. Nat Cell Biol. 2008;10(11):1349-55.

96. Guo Y, Ji X, Liu J, Fan D, Zhou Q, Chen C, et al. Effects of exosomes on pre-metastatic niche formation in tumors. Mol Cancer. 2019;18(1):39.

97. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015;17(6):816-26.

98. Ostenfeld MS, Jeppesen DK, Laurberg JR, Boysen AT, Bramsen JB, Primdal-Bengtson B, et al. Cellular disposal of miR23b by RAB27-dependent exosome release is linked to acquisition of metastatic properties. Cancer Res. 2014;74(20):5758-71.

99. Zhang H, Freitas D, Kim HS, Fabijanic K, Li Z, Chen H, et al. Identification of distinct nanoparticles and subsets of extracellular vesicles by asymmetric flow field-flow fractionation. Nat Cell Biol. 2018;20(3):332-43.

100. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, et al. Tumour exosome integrins determine organotropic metastasis. Nature. 2015;527(7578):329-35.

101. King HW, Michael MZ, Gleadle JM. Hypoxic enhancement of exosome release by breast cancer cells. BMC Cancer. 2012;12:421.

102. Wysoczynski M, Ratajczak MZ. Lung cancer secreted microvesicles: underappreciated modulators of microenvironment in expanding tumors. Int J Cancer. 2009;125(7):1595-603.

103. Park JE, Tan HS, Datta A, Lai RC, Zhang H, Meng W, et al. Hypoxic tumor cell modulates its microenvironment to enhance angiogenic and metastatic potential by secretion of proteins and exosomes. Mol Cell Proteomics. 2010;9(6):1085-99.

104. Svensson KJ, Kucharzewska P, Christianson HC, Skold S, Lofstedt T, Johansson MC, et al. Hypoxia triggers a proangiogenic pathway involving cancer cell microvesicles and PAR-2-mediated heparin-binding EGF signaling in endothelial cells. Proc Natl Acad Sci U S A. 2011;108(32):13147-52.

105. Horie K, Kawakami K, Fujita Y, Sugaya M, Kameyama K, Mizutani K, et al. Exosomes expressing carbonic anhydrase 9 promote angiogenesis. Biochem Biophys Res Commun. 2017;492(3):356-61.

106. Grange C, Tapparo M, Collino F, Vitillo L, Damasco C, Deregibus MC, et al. Microvesicles released from human renal cancer stem cells stimulate angiogenesis and formation of lung premetastatic niche. Cancer Res. 2011;71(15):5346-56.

107. Pasquier J, Thawadi HA, Ghiabi P, Abu-Kaoud N, Maleki M, Guerrouahen BS, et al. Microparticles mediated cross-talk between tumoral and endothelial cells promote the constitution of a pro-metastatic vascular niche through Arf6 up regulation. Cancer Microenviron. 2014;7(1-2):41-59.

108. Tang MKS, Yue PYK, Ip PP, Huang RL, Lai HC, Cheung ANY, et al. Soluble E-cadherin promotes tumor angiogenesis and localizes to exosome surface. Nat Commun. 2018;9(1):2270.

109. Zhou W, Fong MY, Min Y, Somlo G, Liu L, Palomares MR, et al. Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. Cancer Cell. 2014;25(4):501-15.

110. Zeng Z, Li Y, Pan Y, Lan X, Song F, Sun J, et al. Cancer-derived exosomal miR-25-3p promotes pre-metastatic niche formation by inducing vascular permeability and angiogenesis. Nat Commun. 2018;9(1):5395.

111. Czernek L, Duchler M. Functions of Cancer-Derived Extracellular Vesicles in Immunosuppression. Arch Immunol Ther Exp (Warsz). 2017;65(4):311-23.

112. Chen G, Huang AC, Zhang W, Zhang G, Wu M, Xu W, et al. Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. Nature. 2018;560(7718):382-6.

113. Ludwig S, Floros T, Theodoraki MN, Hong CS, Jackson EK, Lang S, et al. Suppression of Lymphocyte Functions by Plasma Exosomes Correlates with Disease Activity in Patients with Head and Neck Cancer. Clin Cancer Res. 2017;23(16):4843-54.

114. Kim JW, Wieckowski E, Taylor DD, Reichert TE, Watkins S, Whiteside TL. Fas ligand-positive membranous vesicles isolated from sera of patients with oral cancer induce apoptosis of activated T lymphocytes. Clin Cancer Res. 2005;11(3):1010-20.

115. Wieckowski EU, Visus C, Szajnik M, Szczepanski MJ, Storkus WJ, Whiteside TL. Tumor-derived microvesicles promote regulatory T cell expansion and induce apoptosis in tumor-reactive activated CD8+ T lymphocytes. J Immunol. 2009;183(6):3720-30.

116. Maus RLG, Jakub JW, Nevala WK, Christensen TA, Noble-Orcutt K, Sachs Z, et al. Human Melanoma-Derived Extracellular Vesicles Regulate Dendritic Cell Maturation. Front Immunol. 2017;8:358.

117. Plebanek MP, Angeloni NL, Vinokour E, Li J, Henkin A, Martinez-Marin D, et al. Pre-metastatic cancer exosomes induce immune surveillance by patrolling monocytes at the metastatic niche. Nat Commun. 2017;8(1):1319.

118. Mu W, Rana S, Zoller M. Host matrix modulation by tumor exosomes promotes motility and invasiveness. Neoplasia. 2013;15(8):875-87.

119. Bandari SK, Purushothaman A, Ramani VC, Brinkley GJ, Chandrashekar DS, Varambally S, et al. Chemotherapy induces secretion of exosomes loaded with heparanase that degrades extracellular matrix and impacts tumor and host cell behavior. Matrix Biol. 2018;65:104-18.

120. Huang T, Deng CX. Current Progresses of Exosomes as Cancer Diagnostic and Prognostic Biomarkers. Int J Biol Sci. 2019;15(1):1-11.

121. Kim MS, Haney MJ, Zhao Y, Yuan D, Deygen I, Klyachko NL, et al. Engineering macrophage-derived exosomes for targeted paclitaxel delivery to pulmonary metastases: in vitro and in vivo evaluations. Nanomedicine. 2018;14(1):195-204.

122. Massague J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature. 2016;529(7586):298-306.

123. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging Biological Principles of Metastasis. Cell. 2017;168(4):670-91.

124. Sosa MS, Bragado P, Aguirre-Ghiso JA. Mechanisms of disseminated cancer cell dormancy: an awakening field. Nat Rev Cancer. 2014;14(9):611-22.

125. Bragado P, Estrada Y, Parikh F, Krause S, Capobianco C, Farina HG, et al. TGF-beta2 dictates disseminated tumour cell fate in target organs through TGF-beta-RIII and p38alpha/beta signalling. Nat Cell Biol. 2013;15(11):1351-61.

126. Kobayashi A, Okuda H, Xing F, Pandey PR, Watabe M, Hirota S, et al. Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone. J Exp Med. 2011;208(13):2641-55.

127. Shiozawa Y, Pedersen EA, Havens AM, Jung Y, Mishra A, Joseph J, et al. Human prostate cancer metastases target the hematopoietic stem cell niche to establish footholds in mouse bone marrow. J Clin Invest. 2011;121(4):1298-312.

128. Ghajar CM, Peinado H, Mori H, Matei IR, Evason KJ, Brazier H, et al. The perivascular niche regulates breast tumour dormancy. Nat Cell Biol. 2013;15(7):807-17.

129. Kienast Y, von Baumgarten L, Fuhrmann M, Klinkert WE, Goldbrunner R, Herms J, et al. Real-time imaging reveals the single steps of brain metastasis formation. Nat Med. 2010;16(1):116-22.

130. Malladi S, Macalinao DG, Jin X, He L, Basnet H, Zou Y, et al. Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT. Cell. 2016;165(1):45-60.

131. Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, et al. Adaptive immunity maintains occult cancer in an equilibrium state. Nature. 2007;450(7171):903-7.

132. De Cock JM, Shibue T, Dongre A, Keckesova Z, Reinhardt F, Weinberg RA. Inflammation Triggers Zeb1-Dependent Escape from Tumor Latency. Cancer Res. 2016;76(23):6778-84.

133. Lu X, Mu E, Wei Y, Riethdorf S, Yang Q, Yuan M, et al. VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging alpha4beta1-positive osteoclast progenitors. Cancer Cell. 2011;20(6):701-14.

134. Gao H, Chakraborty G, Lee-Lim AP, Mo Q, Decker M, Vonica A, et al. The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. Cell. 2012;150(4):764-79.

135. Ross JB, Huh D, Noble LB, Tavazoie SF. Identification of molecular determinants of primary and metastatic tumour re-initiation in breast cancer. Nat Cell Biol. 2015;17(5):651-64.

136. Gao H, Chakraborty G, Zhang Z, Akalay I, Gadiya M, Gao Y, et al. Multi-organ Site Metastatic Reactivation Mediated by Non-canonical Discoidin Domain Receptor 1 Signaling. Cell. 2016;166(1):47-62.

137. Oskarsson T, Acharyya S, Zhang XH, Vanharanta S, Tavazoie SF, Morris PG, et al. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. Nat Med. 2011;17(7):867-74.

138. Malanchi I, Santamaria-Martinez A, Susanto E, Peng H, Lehr HA, Delaloye JF, et al. Interactions between cancer stem cells and their niche govern metastatic colonization. Nature. 2011;481(7379):85-9.

139. Barkan D, El Touny LH, Michalowski AM, Smith JA, Chu I, Davis AS, et al. Metastatic growth from dormant cells induced by a col-I-enriched fibrotic environment. Cancer Res. 2010;70(14):5706-16.

140. Azevedo R, Peixoto A, Gaiteiro C, Fernandes E, Neves M, Lima L, et al. Over forty years of bladder cancer glycobiology: Where do glycans stand facing precision oncology? Oncotarget. 2017;8(53):91734-64.

141. Rini JM, Esko JD. Glycosyltransferases and Glycan-Processing Enzymes. In: rd, Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, et al., editors. Essentials of Glycobiology. Cold Spring Harbor (NY)2015. p. 65-75.

142. Muthana SM, Campbell CT, Gildersleeve JC. Modifications of glycans: biological significance and therapeutic opportunities. ACS Chem Biol. 2012;7(1):31-43.

143. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol. 2019;15(6):346-66.

144. Hansen L, Lind-Thomsen A, Joshi HJ, Pedersen NB, Have CT, Kong Y, et al. A glycogene mutation map for discovery of diseases of glycosylation. Glycobiology. 2015;25(2):211-24.

145. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015;15(9):540-55.

146. Singh C, Shyanti RK, Singh V, Kale RK, Mishra JPN, Singh RP. Integrin expression and glycosylation patterns regulate cell-matrix adhesion and alter with breast cancer progression. Biochem Biophys Res Commun. 2018;499(2):374-80.

147. Taniguchi T, Woodward AM, Magnelli P, McColgan NM, Lehoux S, Jacobo SMP, et al. N-Glycosylation affects the stability and barrier function of the MUC16 mucin. J Biol Chem. 2017;292(26):11079-90.

148. Julien S, Videira PA, Delannoy P. Sialyl-tn in cancer: (how) did we miss the target? Biomolecules. 2012;2(4):435-66.

149. Ferreira JA, Videira PA, Lima L, Pereira S, Silva M, Carrascal M, et al. Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. Mol Oncol. 2013;7(3):719-31.

150. Wang WY, Cao YX, Zhou X, Wei B, Zhan L, Sun SY. Stimulative role of ST6GALNAC1 in proliferation, migration and invasion of ovarian cancer stem cells via the Akt signaling pathway. Cancer Cell Int. 2019;19:86.

151. Pinho S, Marcos NT, Ferreira B, Carvalho AS, Oliveira MJ, Santos-Silva F, et al. Biological significance of cancer-associated sialyl-Tn antigen: modulation of malignant phenotype in gastric carcinoma cells. Cancer Lett. 2007;249(2):157-70.

152. Freitas D, Campos D, Gomes J, Pinto F, Macedo JA, Matos R, et al. O-glycans truncation modulates gastric cancer cell signaling and transcription leading to a more aggressive phenotype. EBioMedicine. 2019;40:349-62.

153. Davidson B, Gotlieb WH, Ben-Baruch G, Kopolovic J, Goldberg I, Nesland JM, et al. Expression of carbohydrate antigens in advanced-stage ovarian carcinomas and their metastases-A clinicopathologic study. Gynecol Oncol. 2000;77(1):35-43.

154. Carrascal MA, Severino PF, Guadalupe Cabral M, Silva M, Ferreira JA, Calais F, et al. Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. Mol Oncol. 2014;8(3):753-65.

155. Neves M, Azevedo R, Lima L, Oliveira MI, Peixoto A, Ferreira D, et al. Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: A novel biomarker and an analytical tool for precision oncology applications. N Biotechnol. 2019;49:77-87.

156. Trinchera M, Aronica A, Dall'Olio F. Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. Biology (Basel). 2017;6(1).

157. Liang JX, Liang Y, Gao W. Clinicopathological and prognostic significance of sialyl Lewis X overexpression in patients with cancer: a meta-analysis. Onco Targets Ther. 2016;9:3113-25.

158. Natoni A, Macauley MS, O'Dwyer ME. Targeting Selectins and Their Ligands in Cancer. Front Oncol. 2016;6:93.

159. Li J, King MR. Adhesion receptors as therapeutic targets for circulating tumor cells. Front Oncol. 2012;2:79.

160. Stone JP, Wagner DD. P-selectin mediates adhesion of platelets to neuroblastoma and small cell lung cancer. J Clin Invest. 1993;92(2):804-13.

161. Padro M, Mejias-Luque R, Cobler L, Garrido M, Perez-Garay M, Puig S, et al. Regulation of glycosyltransferases and Lewis antigens expression by IL-1beta and IL-6 in human gastric cancer cells. Glycoconj J. 2011;28(2):99-110.

162. Bassaganas S, Allende H, Cobler L, Ortiz MR, Llop E, de Bolos C, et al. Inflammatory cytokines regulate the expression of glycosyltransferases involved in the biosynthesis of tumor-associated sialylated glycans in pancreatic cancer cell lines. Cytokine. 2015;75(1):197-206.

163. Freitas D, Balmana M, Pocas J, Campos D, Osorio H, Konstantinidi A, et al. Different isolation approaches lead to diverse glycosylated extracellular vesicle populations. J Extracell Vesicles. 2019;8(1):1621131.

164. Royo F, Cossio U, Ruiz de Angulo A, Llop J, Falcon-Perez JM. Modification of the glycosylation of extracellular vesicles alters their biodistribution in mice. Nanoscale. 2019;11(4):1531-7.

165. Giuliano M, Giordano A, Jackson S, De Giorgi U, Mego M, Cohen EN, et al. Circulating tumor cells as early predictors of metastatic spread in breast cancer patients with limited metastatic dissemination. Breast Cancer Res. 2014;16(5):440.

166. Yan WT, Cui X, Chen Q, Li YF, Cui YH, Wang Y, et al. Circulating tumor cell status monitors the treatment responses in breast cancer patients: a meta-analysis. Sci Rep. 2017;7:43464.

167. Krebs MG, Hou JM, Ward TH, Blackhall FH, Dive C. Circulating tumour cells: their utility in cancer management and predicting outcomes. Ther Adv Med Oncol. 2010;2(6):351-65.

168. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting Tumor Microenvironment for Cancer Therapy. Int J Mol Sci. 2019;20(4).

169. Szebeni GJ, Vizler C, Nagy LI, Kitajka K, Puskas LG. Pro-Tumoral Inflammatory Myeloid Cells as Emerging Therapeutic Targets. Int J Mol Sci. 2016;17(11).

170. Kannagi R. Carbohydrate antigen sialyl Lewis a--its pathophysiological significance and induction mechanism in cancer progression. Chang Gung Med J. 2007;30(3):189-209.

171. Fantini D, Glaser AP, Rimar KJ, Wang Y, Schipma M, Varghese N, et al. A Carcinogen-induced mouse model recapitulates the molecular alterations of human muscle invasive bladder cancer. Oncogene. 2018;37(14):1911-25.

172. Compain P, Martin OR. Carbohydrate mimetics-based glycosyltransferase inhibitors. Bioorg Med Chem. 2001;9(12):3077-92.

173. Liu SD, Chalouni C, Young JC, Junttila TT, Sliwkowski MX, Lowe JB. Afucosylated antibodies increase activation of FcgammaRIIIa-dependent signaling components to intensify processes promoting ADCC. Cancer Immunol Res. 2015;3(2):173-83.

174. Azevedo R, Ferreira JA, Peixoto A, Neves M, Sousa N, Lima A, et al. Emerging antibody-based therapeutic strategies for bladder cancer: A systematic review. J Control Release. 2015;214:40-61.

175. Fernandes E, Ferreira JA, Andreia P, Luís L, Barroso S, Sarmento B, et al. New trends in guided nanotherapies for digestive cancers: A systematic review. J Control Release. 2015;209:288-307.

176. Posey AD, Jr., Schwab RD, Boesteanu AC, Steentoft C, Mandel U, Engels B, et al. Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. Immunity. 2016;44(6):1444-54.

177. Boks MA, Ambrosini M, Bruijns SC, Kalay H, van Bloois L, Storm G, et al. MPLA incorporation into DC-targeting glycoliposomes favours anti-tumour T cell responses. J Control Release. 2015;216:37-46.

178. Lakshminarayanan V, Thompson P, Wolfert MA, Buskas T, Bradley JM, Pathangey LB, et al. Immune recognition of tumor-associated mucin MUC1 is achieved by a fully synthetic aberrantly glycosylated MUC1 tripartite vaccine. Proc Natl Acad Sci U S A. 2012;109(1):261-6.

179. Abdel-Aal AB, Lakshminarayanan V, Thompson P, Supekar N, Bradley JM, Wolfert MA, et al. Immune and anticancer responses elicited by fully synthetic aberrantly glycosylated MUC1 tripartite vaccines modified by a TLR2 or TLR9 agonist. Chembiochem. 2014;15(10):1508-13.

INTRODUCTION

PART II

The Role of Glycosylation in Esophageal and Colorectal Cancer: Clinical Implications and Biomarker Discovery

Decoding Esophageal and Colorectal Cancer: Insights into Diagnosis, Treatment, Risk Factors, and Histology

Esophageal (EC) and colorectal (CRC) cancers are among the most prevalent and fatal malignancies worldwide, accounting for nearly one-third of all cancer-related deaths (1, 2). In 2020, CRC held the third in terms of incidence (1.9 million new cases) and the second in terms of mortality (935,000 deaths). On the other hand, EC ranks seventh in terms of incidence (604,000 new cases) and sixth in terms of mortality (544,000 deaths) (2). Furthermore, most cases of CRC are detected in Western countries and high developed countries, while EC was much higher in low- and middle-income countries (1, 3, 4).

EC and CRC arise from different tissues and exhibit distinct molecular characteristics. Still, they share common diagnostic and treatment challenges, often combined to late-stage diagnoses and limited effectiveness of current treatments. Current diagnostic procedures, including upper digestive tract endoscopy and colonoscopy respectively, though effective, are invasive and therefore unsuitable for large-scale screening, hence creating a necessity for more non-invasive diagnostic tools (5, 6).

The histological features of esophageal lesions are divided into the following terms: squamous cell carcinoma (ESCC), which represents most of the cases, resulting from the abnormal transformation of the stratified squamous epithelial layer; or adenocarcinoma (EAC) in which glandular cells exchange the squamous epithelium (7, 8). According to the World Health Organization, EC grade 1 represents well-differentiated tumour, grade 2 corresponds to moderately differentiated, and grade 3 to poorly differentiated. Importantly, poorer differentiated carcinomas present a decreased prognosis compared to well-differentiated esophageal cancers (9, 10). The predominant histological types of colorectal carcinomas consist largely of adenocarcinomas or mucinous adenocarcinomas (11, 12). CRC is also categorized in three major groups according to histological grade staging, similarly to the EC.

Several risk factors contribute to the onset of these cancers, including tobacco, alcohol consumption, poor diet, physical inactivity, obesity, and exposure to environmental carcinogens, mostly nitrosamines (13, 14). Furthermore, chronic inflammation in the esophagus can promote the growth of esophageal carcinoma by disrupting normal cell signalling and development (15). The correlation between tumour growth and lymph node metastases is positively associated with tumour aggressiveness, being the bases for the

development of the TNM system (16). Despite these findings, it is crucial to highlight the role of the intratumoral molecular heterogeneity that exists in EC and CRC, which adds a significant layer of complexity to patient stratification and treatment decision-making (5, 6, 17). High-throughput genomics and transcriptomics studies have begun to uncover the potential molecular markers for CRC and EC, emphasising promising steps towards revolutionizing future personalized patient care(18-22).

Treatments for EC and CRC include a range of options from surgical resection, chemotherapy, radiotherapy, to immunotherapy, all of which come with their limitations and side effects (22-24). While targeted therapeutics such as trastuzumab (anti-HER2) and cetuximab (anti-EGFR) have shown promises, their effectiveness requires further validation (25, 26). The introduction of immune checkpoint inhibitors targeting PD-1/PD-L1 represent a significant advancement in cancer immunotherapy, yet the need for accurate molecular models to predict treatment responders remains a significant barrier (27, 28).

Research has directed its focus towards the development of new and effective therapies, namely combined therapies guided by patient responses (29). Furthermore, oncoproteogenomics, which links mass spectrometry (MS) with next-generation sequencing (NGS), is seen as the next frontier in cancer research. This multidisciplinary approach has the potential to provide deeper insights into the role of protein variants in several biological processes and diseases and identify new cancer antigens for therapy development.

In conclusion, despite significant progress in understanding EC and CRC molecular landscape, these malignancies continue to present substantial challenges due to their complexity and heterogeneity. These issues underscore the need for a concerted effort to develop more effective diagnostic tools, therapeutic strategies, and predictive models to improve patient outcomes, pushing the boundaries of personalized medicine.

Glycosylation patterns in esophageal and colorectal cancers: exploring general features.

Glycosylation, the most frequent posttranslational modification (PTM) of proteins, prominently occurs at multiple sites of the extracellular domains of membrane-anchored and secreted proteins. This complex process, dictated by the assembled action of nucleotide sugar transporters, glycosyltransferases, and glycosidases, results in a carbohydrate-dense glycocalyx on the cell surface. The complex nature of glycosylation significantly contributes to protein folding, trafficking, stability, cell-cell adhesion, cell differentiation, and migration, further regulating cell signalling pathways, host-pathogen interaction, and immune escape (30-34). The glycosylation landscape encompasses three

major classes of glycoconjugates - proteoglycans, glycosphingolipids and glycoproteins, that modulate these biological features. Importantly, variations in glycosidic linkages, including *N*-, *O*-, and *C*-linked glycosylation, glypiation (GPI anchor attachment), and phosphoglycosylation, depend on multiple factors such as enzyme availability and macromolecular conformation of the glycoconjugates (35-40).

In cancer, alterations in the glycoproteome, including changes in glycosite density and distribution and/or glycan chains structure are observed, impacting the cellular homeostasis (41-43). Predominant changes comprise altered fucose and sialic acid residue substitution patterns, *O*-GalNAc glycans oversialylation resulting in their dramatic reduction, and alterations in *N*-glycans branching, core fucosylation, and sialylation (37, 44-46). Additionally, terminal glycan chain epitopes, such as Lewis's blood group patterns, often exhibit variations (47). These aberrant glycans modulate intracellular pathways, influencing cell proliferation, survival, and adhesion dynamics, often promoting more invasive phenotypes (41, 43, 48). Glycans also manage immune response by mediating immune cell lectin recognition, often leading towards cancer tolerogenic phenotypes (49, 50). Specific changes in tumour environment that promote these glycophenotypes, frequently leads to disease development by undisclosed mechanisms (48, 51-53). Given their role as critical mediators of cellular adaptation and disease progression, glycans are considered interesting theragnostic targets. Focus on *O*-GalNAc and *N*-GalNAc glycosylation and particularly, on two clinically relevant glycans - the sialyl-Tn (STn) and the blood group related sialyl lewis A/X (sLeA/X) antigens, should provide the ground to disclose the impact of protein glycosylation in the development of EC and CRC.

O-GalNAc and N-GalNAc glycosylation

Glycosylation gives rise to two central categories of glycans that constitute the glycocalyx present on the cell's surface: *N*-glycans and *O*-glycans (54). Cancer progression in esophageal and colorectal malignancies is deeply influenced by changes in glycosylation patterns (29). *O*-glycosylation is an essential post-translational modification, a crucial player in cellular function and integrity. This vital layer significantly impacts cellular functions, thus maintaining homeostasis (30, 32, 55). *O*-glycosylation, the second most prevalent form of glycosylation, initiates by adding a *N*-Acetylgalactosamine (*GalNAc*), a carbohydrate fragment, to protein serine or threonine residues. This complex process is controlled by 20 membrane-bound UDP-GalNAc enzymes: polypeptide glycosyltransferases (ppGalNAc-Ts). Each enzyme in this family has unique yet overlapping specificities, modifying the initiation of *O*-glycosylation. This process lacks a consensus sequence for the activity of *N*-acetylgalactosamine transferase. Following the initial sugar addition, a varied number of

sugars are subsequently added to the growing glycan chain. The range includes galactose, *N*-acetylglucosamine, fucose, or sialic acid (56-58).

The first step of *O*-glycosylation results in the formation of the Tn antigen, by adding a GalNAc, in the Golgi apparatus. This is followed by the synthesis of the core 1 structure through the action of Gal-transferase, leading to the formation of structures referred to Thomsen-Friedenreich antigen (T antigen). Sialyltransferases can also sialylate the Tn and T antigens to form sialyl-Tn, sialyl-T, and disialyl-T antigens. Notably, the formation of these sialylated structures stops further processing of the oligosaccharide chain. This core 1 structure can be further modified to other core structures by adding different monosaccharides, such as galactose, GalNAc, *N*-acetylglucosamine, and sialic acids. The most common structures in humans are cores 1-4. The extension of these core units is catalyzed by *N*-*β*3/4-acetylglucosaminyltransferases and *β*3/4-galactosyltransferases, leading to the formation of side chains termed type-1 and type-2 chains. These chains are widely expressed among epithelial tissues (56-58).

O-glycosylation is particularly significant in modulating the biological role of mucins, a high molecular weight glycoproteins rich in serine and threonine residues. Mucins are ubiquitously present in mucous secretions, cell surfaces, as transmembrane glycoproteins, and body fluids. They are known to be involved in signal transduction, mediating cell-cell adhesion, or performing anti-adhesive functions. Mucins also play roles in fertilization and immune responses. Their presence on epithelial surfaces serves to protect against physical and chemical damage, and against pathogen infection (59). Alterations on the glycosylation pattern of these glycoproteins could determine its biological function.

Conversely, in *N*-glycosylation, *N*-glycans are transferred to protein asparagine (Asn) residues via *N*-glycosidic bonds, forming the most frequent constituent, GlcNAcβ1-Asn (60). Their synthesis initiates on the cytosolic face of the endoplasmic reticulum (ER) and the structure extends into the ER lumen (61).

Glycoproteins primed to undergo *N*-glycosylation exhibit marked motifs known as Asn-X-Ser/Thr "sequons", where the symbol "X" denotes any amino acid residue, excluding proline (62). The oligosaccharide transferase (OSTase) surveys nascent protein polypeptides for this consensus sequence, proceeding to shift the precursor glycan (Glc3Man9GlcNac2-) from dolichol pyrophosphate to the Asn residue (63). All *N*-linked glycoproteins share this precursor glycan structure.

The Golgi apparatus initiates glycan processing, introducing diversity by sequentially trimming and adding sugars (64). Mature *N*-glycans can undergo further modifications such as fucosylation and sialylation, generating terminal structures such as Lewis's blood group related antigens (Lea, SLea, Lex, SLex, Leb, Ley), or ABO(H) blood group determinants (65). Additionally, other modifications like phosphorylation, *O*-acetylation of sialic acids, and

O-sulfation of galactose and N-acetylglucosamine residues amplify the glycome's structural complexity (66).

Shifting focus to cancer, cell-surface glycoconjugates in tumoral cells often exhibit an overexpression of sialylated Lewis-type blood group antigens such as sLeA and its isomer sLeX as terminal epitopes of *O*-glycans, *N*-glycans, and glycolipids (67). The intricate process of sialyl Lewis antigen biosynthesis involves the synchronized expression of multiple glycosyltransferases, which may differ based on the glycoconjugate carrying the antigen (68). Increases in sLeA/X levels in *O*-glycans correlate mostly with C2GnT overexpression (69), while aberrant glycolipid sLeA/X is linked to *β*1,3-GlcNAc transferase activation (70). The competition between type 1 (sLeA) and type 2 (sLeX) chain biosynthesis varies by tissue and is majorly controlled by the basal levels of *β*-1,3 and *β*-1,4-galactosyltransferases (71).

In conclusion, *O*- and *N*-glycosylation and the resulting intricate array of glycan structures plays an essential role in normal cellular functions. Alterations in this process, particularly in the context of cancer, have been linked to tumour progression, thereby suggesting potential avenues for targeted therapeutics. Continued exploration of the implications of glycosylation changes in cancer is crucial for research.

Unravelling the significance of STn and sLeA antigen expression in esophageal and colorectal cancers

Malignant transformation of esophageal and colorectal epithelium is often accompanied by a profound dysregulation of the normal glycosylation pattern.

This alteration stems from disruptions in typical glycan extension processes, leading to the neo- and/or overexpression of several glycans, such as STn and SLe (72, 73). These changes on the glycosylation pattern across the carcinogenic cascade, highlight the potential role of glycans in tumorigenesis.

In colorectal cancer stem cells (CRC-CSC), STn overexpression is implicated in promoting chemoresistance and sphere formation, thereby enhancing the survival and proliferation of these cells [85]. The alteration is attributed to Akt pathway activation in conjunction with Galectin-3, indicating potential therapeutic targets [85]. STn overexpression is central to the progression and aggressiveness of various cancers, supporting its role as a therapeutic target. Further research is needed to translate these findings into treatments, exploring the pathways and molecular mechanisms involved. STn's contribution to oncogenic traits across tumour types, including breast, ovarian, prostate, and bladder cancers, affirms its pancarcinomic nature [43, 86, 87]. The pursuit of strategies to inhibit STn expression or function may provide a promising direction for future cancer therapy development.

In several malignant forms, including EC (74), and CRC (75, 76), the glycoepitopes sLeA and sLeX have been reported to be significantly overexpressed. The increased presence of these glycoepitopes appears to be strongly linked to increased cancer cell malignancy. Accordingly, several reports claim that these molecules serve as E-selectin ligands on vascular endothelial cells, aiding tumour cell invasion into blood vessels and facilitating the arrest of circulating tumour cells at distant sites during hematogenous metastasis (77, 78). In addition, these ligands can enhance the adhesive interactions between tumour cells and the extracellular matrix proteins of vessel walls, promoting extravasation (79). Furthermore, it has been proposed that sLeA/X, in colon cancer cells, interact with P-selectin on platelets, thereby aiding hematogenous dissemination by protecting tumour cells from shear stress in circulation and preventing immune recognition by natural killer (NK) cells (80).Through this interaction, activated platelets may also release growth factors that stimulate the development of highly metastatic colon cancer cells (81). Moreover, colon cancer cells undergoing epithelial-to-mesenchymal transition (EMT) have been observed to upregulate sLeA/X levels and E-selectin binding capacity, while also enhancing *vascular endothelial growth factor* (VEGF) production (82). These discoveries suggest a potential pro-angiogenic feedback loop initiated by the interaction between sLeA/X and E-selectin. The molecular mechanism associated with EMT-related sLeA/X expression in CRC cells seems to involve transcriptomic regulation of glycogenes in a manner dependent on c-Myc and CDX2. Specifically, EMT-associated increases in c-Myc transcriptional activity result in the upregulation of ST3GAL1/3/4 and FUT3, while CDX2 downregulation suppresses FUT2 transcription, leading to an overexpression of sLeA/X (82).

In sum, STn and sLe antigens overexpression seems to accompany the carcinogenesis process and aggressiveness of several cancers, rendering it a promising therapeutic target for aggressive cancer phenotypes [43, 86, 87]. These findings underscore the importance of targeted therapeutic interventions for cancer-specific post-translational modifications (PTMs), however further research may devote to uncovering the underlying molecular pathways leading to these glycophenotypes.

Clinical relevance of STn and sLe^A in esophageal and colorectal cancers: emerging potential in the circulating tumour cells.

Late diagnosis of CRC and EC, associated with clinical evidence of metastases, poses significant healthcare burden, being critical to address the following clinical problems: i) Identification of patients at higher risk of metastasis development, including those with radiographically occult lesions; ii) Implementation of non-invasive molecular-based patient stratification methods for disease follow-up and early intervention; iii) Tailoring of

therapeutic schemes to improve efficacy while reducing toxicity; iv) Development of novel targeted therapeutics. Recent studies are focusing these clinical challenges through the analysis of aberrant glycosylation markers, such as STn and sLeA. The clinical assessment of STn and sLeA levels on cancer remains circumscribed to certain tumours, mainly from the gastrointestinal tract, however without a significant clinical value (83).

In EC few studies have been made in approaching the expression of STn. Was reported that the overexpression of this antigen is correlated to the poorer survival. There is evidence of the appearance of STn in healthy esophagus cells (83), also agglomerates linked to serine and cross reaction to Tn (84, 85).

SLeA in ESCC indicate a cancer specific nature occurring in membrane and cytoplasm cancer cells in opposition to stromal and normal tissues (86). This antigen was associated with poor prognosis, tumour differentiation and metastasis (87). However, clinical pathological variables were not connected to sleA expression (88).

In CRC, the expression of STn antigen provides insights into the malignancy's biological behaviour, with its absence in normal mucosa adjacent to tumours—except in goblet cells—underscoring its association with cancer (89-93).

STn's detection in colonocytes post de-acetylation and the reduction of sialic acids *O*-acetylation on mucins may serve as sensitive markers for malignant transformation (94-96). The antigen's localization in tumour cells, notably in apical cytoplasm or at cell membranes, intensifies with the invasion extent, aligning with histological findings of poor differentiation and perineural invasion (89, 93). STn expression is not only detected in CRC tumours but also in transitional tissue and adenomas, where it correlates significantly with tumour stage, suggesting its utility as a marker for early carcinogenic stages (91-93). On the other hand, in CRC, sLeA expression has been significantly associated with a range of clinicopathological factors including lymph node metastasis, recurrence, survival rates, and tumour stage (97, 98). Yet, some investigations have not found a definitive clinical significance for sLeA in CRC (99, 100). Despite these mixed results, the majority of studies endorse sLeA as a marker indicative of tumour aggressiveness and patient prognosis in advanced CRC cases. The exploration of sLeA levels in both tumour tissues and serum has suggested that their presence could identify patients at an elevated risk of recurrence and mortality, potentially guiding decisions for adjuvant therapy (69, 101).

The presence of specific tumour-associated antigens, such as STn and sLeA, in the bloodstream of patients diagnosed with EC and CRC facilitates their detection. These antigens, often found on proteins secreted or shed by tumours, serve as biomarkers in non-invasive diagnostic procedures. Serological assays are the gold standard method to assess the levels of these two cancer-associated glycans, mainly on colorectal and pancreatic

cancer. No implementation of these strategies was amplified to others digestive tumours, such as EC. The CA72-4 assay, which targets the tumour-associated glycoprotein 72 (TAG-72) carrying the STn antigen, is frequently utilized in prognosis of esophageal and colorectal cancer patients (103, 104). Despite STn's presence in CRC tissues, its levels do not consistently reflect in serum measurements by the CA72-4 test, indicating a disconnection between tissue and serum antigen levels (101). Meanwhile, the CA19-9 assay detects the sLeA antigen, primarily on mucins in patient sera, but can also identify monosialogangliosides with sLeA in gastrointestinal tumour tissues (105). SLeA antigen, when secreted in small quantities by normal tissues, is upregulated in benign gastrointestinal disorders, potentially leading to false-positive results. Individuals with the Lewis (a-b-) genotype present a challenge for accurate cancer discrimination due to false negatives (106, 107). In sum, serum sLeA detection is impeded by low sensitivity and the genetic variability in the population concerning the LewisA blood group antigen, which influences the synthesis of the sLeA antigen (108). A substantial meta-analysis encompassing 6434 CRC patients revealed a significant correlation between elevated pre-treatment serum CA19-9 levels and reduced survival outcomes, highlighting its prognostic value independent of variables such as region, analysis type, sample size, and treatment modalities (109). Although fewer studies are available for CA72-4, initial results suggest a similar trend in CRC, yet the conclusions are currently limited due to insufficient statistical power (110, 111).

The clinical assessment of the STn and sLeA antigens in tumour tissues, unlike serological assays, is not commonly practiced. However, these antigens' expression in the serum of CRC and EC patients open an avenue for the development of non-invasive detection tools. On this context, emerging research emphasizes the clinical potential of circulating tumour cells (CTCs) in cancer prognostication and the assessment of micrometastasis. These cells have been identified even in patients without radiological evidence of metastasis, suggesting their utility in early detection and intervention (112-114). Recent studies on EC have indicated a potential correlation between CTC in peripheral blood and distant metastasis (115-117). Even though metastasis appear at a later stage of the disease, several studies suggest a mechanism that occurs at an early stage of tumour development, in which a subset of tumour cells remain quiescent and can be reactivated after a long period of apparent freedom from the disease (118). Therefore, it becomes critical to develop analytical tools capable of identify the cells involved in the metastatic process in a non-invasive manner (119). Once, CTC detached from the tumour niche and go into the bloodstream, it became a relevant source of information. CTC provide a fingerprint for metastasis molecular heterogeneity and its temporal evolution through the course of disease management (119, 120). Conventionally, isolated CTC, defined based on surface

expression of epithelial adhesion molecule (EpCAM) or cytokeratin (CK), and the absence of pan-leukocyte marker CD45, have shown to correlate with progression-free survival and overall survival in patients with CRC (121-123). Recent investigations indicate that a majority of CTC, including those from CRC, express STn on their surface (124). This discovery not only improves the sensitivity of CTC detection techniques but also correlates with the progression of disease in patients, linking STn expression with increased cell motility, invasion, and immune evasion (124). Considering the role of sLeA in metastasis and its potential expression on CTCs, there is a need for further studies to validate its clinical significance (71). The insights gained from studying STn and sLeA in CTCs can significantly enhance liquid biopsy techniques and support the development of glycan-based therapeutic strategies aimed at managing disseminated disease.

Multiomics on the Precision Oncology Era: Glycoproteomic Insights on the Cancer Biomarker Discovery

Advanced mass spectrometry (MS) technologies, complemented by spectral data from tandem MS experiments, facilitate the annotation of glycopeptides and glycosites (125). These techniques are now robust enough for clinical research, supported by comprehensive guidelines (126).

Glycoproteomic analysis is critical for advancing precision oncology, particularly in EC and CRC. It focuses on identifying specific glycosylation patterns in serum proteins that reflect disease presence and progression. For instance, the carcinoembryonic antigen (CEA), a well-known marker in colorectal cancer, presents unique glycoforms in tumours that could improve its diagnostic utility (127, 128). Overall, the serum glycoproteome's alterations, largely driven by systemic inflammatory responses, hold the key to enhancing the sensitivity of glycan-based diagnostics in oncology (42, 129).

In the context of precision oncology, explorations into the glycoproteome have unveiled alterations in the glycosylation patterns of serum proteins, potentially linked to inflammatory diseases and colorectal cancer (CRC) (130). This has been supported by studies identifying changes in the abundance of specific glycoproteins, such as haptoglobulin and *α*-1-acid glycoprotein, which are known to engage in local and systemic inflammatory responses, as well as in extracellular matrix remodelling – key processes in the tumour niche (131, 132).

Particularly in CRC, research using lectin-enrichment techniques has spotlighted the rise in sialylation and fucosylation of proteins like complement C3, pointing towards a significant divergence from normal and adenomatous tissues (133). In terms of clinical diagnostics, carcinoembryonic antigen (CEA) has been a cornerstone marker. Despite its widespread clinical use for monitoring CRC post-therapy and aiding pre-operative assessments, its

sensitivity for early disease detection is limited by its elevation in non-cancerous conditions (134, 135). Recent glycomic research suggests that tumour-specific glycoforms of CEA could be harnessed to enhance the assay's diagnostic accuracy (136, 137). However, it's becoming clear that the serum glycoproteome is largely influenced by systemic inflammatory responses rather than the tumour cells themselves, which may explain the limited sensitivity of glycan-based serological tests and underscores the critical need for more targeted glycoproteomic analyses (42, 129). These insights pave the way for refined detection methods and underscore the potential of glycoproteomics in identifying disease-associated protein alterations.

In colorectal cancer (CRC), serological assessments reveal humoral responses against atypical glycosylation of cancer-associated proteins. Notably, Pedersen et al. discovered autoantibodies against unusual glycopeptides from MUC1 and MUC5AC, suggesting glycan changes as potential biomarkers for CRC, achieving a 79% sensitivity and 92% specificity in detection (138). These glycan alterations may represent cancer neoantigens, which could elicit immune responses, though their effectiveness is often suppressed by the immunomodulatory actions of short-chain O-glycans. The detection of such autoantibodies could lead to the identification of distinct cancer-associated glycan signatures, which may enhance diagnostic precision and therapeutic targeting in oncology.

The knowledge about STn and sLeA as potential biomarkers in CTCs of CRC and EC is still limited. However, the importance of these glycans for cancer control is undoubtful as they are predominantly overexpressed in CTCs. It is also important to consider the microenvironment as an impacting factor of cancer control actions, as it impacts the presence of CTCs and their fate, and the metastasis development.

The expansive landscape of glycoproteomics is yet to be fully explored, particularly concerning the glycoproteome associated with esophageal and colorectal cancers. Current research has been predominantly exploratory, focusing on cell lines or serum from patients and often featuring limited patient cohorts with varied clinical backgrounds. Despite these limitations, there's a strong indication that cancer provokes significant shifts in serum glycoproteomics not mirrored in non-cancerous conditions. Identified changes are largely linked to systemic inflammation and extracellular matrix alterations. These insights underscore the necessity of focus on cancer-associated glycan markers to refine non-invasive diagnostic techniques for CRC and EC. Surprisingly, the potent clinical and functional insights provided by STn and sLeA antigens have not yet been fully leveraged in a targeted examination of the glycoproteomes in both primary and metastatic lesions within these cancers. This gap hinders a deeper understanding of oncogenesis and delays the development of precision bi-specific interventions that could exploit cancer-specific glycan profiles.

The surge in biomarker research is largely due to the integration of omics technologies into the field of precision oncology. This has led to the transition from traditional single-target investigations to a holistic systems biology approach. Next-generation sequencing, both genomic and transcriptomic, has been pivotal in this shift, facilitated by the vast computational resources and sophisticated bioinformatics. This advance is exemplified by the availability of vast datasets in repositories like GEO (139), ArrayExpress (140), and Oncomine (141), which have become instrumental in data mining and the cross-examination of hypotheses. Specialized databases such as MSigDB (142) and the Cancer Genome Project (143) further enrich the pool of cancer-related molecular data (144). Transcriptomic studies redefine colorectal cancer classifications by identifying distinct molecular subtypes with prognostic value, such as CMS1 with its immunogenic profile and CMS2 with its activation of the WNT-β-catenin pathway, indicative of better survival outcomes (145). These subtypes, identified through rigorous data analysis, are continuously refined by incorporating clinical findings, underscoring the symbiotic relationship between molecular data and patient-specific information in enhancing predictive models within oncology.

In the pursuit of precision oncology, panomics approaches have notably shifted focus from genomic and transcriptomic analyses to include the proteome, which often presents inconsistencies when mirrored against gene expression levels (146). The recognition of the proteome's distinct nature has driven the development of oncoproteomics, aiming to unify genomic alterations with protein function, yet databases often lack comprehensive genomic variant representation. Emerging oncoproteomics strategies are seeking to reconcile these disparities through custom databases derived from genomic data (147, 148), thus enhancing the identification of cancer-specific protein signatures for therapeutic targeting. This integration has refined the molecular models of esophageal and colorectal cancers, identifying critical targets for precise interventions. The synthesis of multi-level molecular data underscores its indispensability in formulating targeted cancer treatments.

References

1. Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, et al. Global Burden of 5 Major Types of Gastrointestinal Cancer. Gastroenterology. 2020;159(1):335-49 e15.

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.

3. Oesophago-Gastric Anastomotic Audit Collaborative: Writing C, Steering C, National L, Site L, Collaborators. Mortality from esophagectomy for esophageal cancer across low, middle, and high-income countries: An international cohort study. Eur J Surg Oncol. 2021;47(6):1481-8.

4. Marmol I, Sanchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. Int J Mol Sci. 2017;18(1).

5. Bijlsma MF, Sadanandam A, Tan P, Vermeulen L. Molecular subtypes in cancers of the gastrointestinal tract. Nat Rev Gastroenterol Hepatol. 2017;14(6):333-42.

6. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? Int J Mol Sci. 2018;19(12).

7. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. Gut. 2015;64(3):381-7.

8. Blot WJ, Devesa SS, Fraumeni JF, Jr. Continuing climb in rates of esophageal adenocarcinoma: an update. JAMA. 1993;270(11):1320.

9. Wang Y, Bai G, Guo L, Chen W. Associations Between Apparent Diffusion Coefficient Value With Pathological Type, Histologic Grade, and Presence of Lymph Node Metastases of Esophageal Carcinoma. Technol Cancer Res Treat. 2019;18:1533033819892254.

10. Lam AK. Histopathological Assessment for Esophageal Squamous Cell Carcinoma. Methods Mol Biol. 2020;2129:7-18.

11. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.

12. Hermanek P. Colorectal carcinoma: histopathological diagnosis and staging. Baillieres Clin Gastroenterol. 1989;3(3):511-29.

13. Uhlenhopp DJ, Then EO, Sunkara T, Gaduputi V. Epidemiology of esophageal cancer: update in global trends, etiology and risk factors. Clin J Gastroenterol. 2020;13(6):1010-21.

14. Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16(12):713-32.

15. Huang FL, Yu SJ. Esophageal cancer: Risk factors, genetic association, and treatment. Asian J Surg. 2018;41(3):210-5.

16. Marzouk O, Schofield J. Review of histopathological and molecular prognostic features in colorectal cancer. Cancers (Basel). 2011;3(2):2767-810.

17. Ross-Innes CS, Becq J, Warren A, Cheetham RK, Northen H, O'Donovan M, et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. Nat Genet. 2015;47(9):1038-46.

18. Lee BH. Commentary on: "Comprehensive molecular characterization of papillary renal-cell carcinoma." Cancer Genome Atlas Research Network.: N Engl J Med. 2016 Jan 14;374(2):135-45. Urol Oncol. 2017;35(9):578-9.

19. Vasaikar S, Huang C, Wang X, Petyuk VA, Savage SR, Wen B, et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. Cell. 2019;177(4):1035-49 e19.

20. Rho JH, Ladd JJ, Li CI, Potter JD, Zhang Y, Shelley D, et al. Protein and glycomic plasma markers for early detection of adenoma and colon cancer. Gut. 2018;67(3):473-84.

21. Visnovska T, Biggs PJ, Schmeier S, Frizelle FA, Purcell RV. Metagenomics and transcriptomics data from human colorectal cancer. Sci Data. 2019;6(1):116.

22. Rustgi A, El-Serag HB. Esophageal carcinoma. N Engl J Med. 2015;372(15):1472-3.

23. Thanikachalam K, Khan G. Colorectal Cancer and Nutrition. Nutrients. 2019;11(1).

24. Ohashi S, Miyamoto S, Kikuchi O, Goto T, Amanuma Y, Muto M. Recent Advances From Basic and Clinical Studies of Esophageal Squamous Cell Carcinoma. Gastroenterology. 2015;149(7):1700-15.

25. Samson P, Lockhart AC. Biologic therapy in esophageal and gastric malignancies: current therapies and future directions. J Gastrointest Oncol. 2017;8(3):418-29.

26. Kirstein MM, Lange A, Prenzler A, Manns MP, Kubicka S, Vogel A. Targeted therapies in metastatic colorectal cancer: a systematic review and assessment of currently available data. Oncologist. 2014;19(11):1156-68.

27. Yaghoubi N, Soltani A, Ghazvini K, Hassanian SM, Hashemy SI. PD-1/ PD-L1 blockade as a novel treatment for colorectal cancer. Biomed Pharmacother. 2019;110:312-8.

28. Janjigian YY, Shitara K, Moehler M, Garrido M, Salman P, Shen L, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. Lancet. 2021;398(10294):27-40.

29. Fernandes E, Sores J, Cotton S, Peixoto A, Ferreira D, Freitas R, et al. Esophageal, gastric and colorectal cancers: Looking beyond classical serological biomarkers towards glycoproteomics-assisted precision oncology. Theranostics. 2020;10(11):4903-28.

30. Chandler KB, Costello CE. Glycomics and glycoproteomics of membrane proteins and cell-surface receptors: Present trends and future opportunities. Electrophoresis. 2016;37(11):1407-19.

31. Loureiro LR, Sousa DP, Ferreira D, Chai W, Lima L, Pereira C, et al. Novel monoclonal antibody L2A5 specifically targeting sialyl-Tn and short glycans terminated by alpha-2-6 sialic acids. Sci Rep. 2018;8(1):12196.

32. Christiansen MN, Chik J, Lee L, Anugraham M, Abrahams JL, Packer NH. Cell surface protein glycosylation in cancer. Proteomics. 2014;14(4-5):525-46.

33. Brockhausen I, Schachter H, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

34. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. 3rd ed. Cold Spring Harbor (NY)2015.

35. Haltiwanger RS, Lowe JB. Role of glycosylation in development. Annu Rev Biochem. 2004;73:491-537.

36. Hart GW, Copeland RJ. Glycomics hits the big time. Cell. 2010;143(5):672-6.

37. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer. 2015;15(9):540-55.

38. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J. 2013;32(10):1478-88.

39. Taniguchi N, Kizuka Y. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. Adv Cancer Res. 2015;126:11-51.

40. Varki A. Biological roles of oligosaccharides: all of the theories are correct. Glycobiology. 1993;3(2):97-130.

41. Azevedo R, Peixoto A, Gaiteiro C, Fernandes E, Neves M, Lima L, et al. Over forty years of bladder cancer glycobiology: Where do glycans stand facing precision oncology? Oncotarget. 2017;8(53):91734-64.

42. Chen L, Deng H, Cui H, Fang J, Zuo Z, Deng J, et al. Inflammatory responses and inflammation-associated diseases in organs. Oncotarget. 2018;9(6):7204-18.

43. Ferreira IG, Pucci M, Venturi G, Malagolini N, Chiricolo M, Dall'Olio F. Glycosylation as a Main Regulator of Growth and Death Factor Receptors Signaling. Int J Mol Sci. 2018;19(2).

44. Fernandez-Rodriguez J, Feijoo-Carnero C, Merino-Trigo A, Paez de la Cadena M, Rodriguez-Berrocal FJ, de Carlos A, et al. Immunohistochemical analysis of sialic acid and fucose composition in human colorectal adenocarcinoma. Tumour Biol. 2000;21(3):153-64.

45. Padler-Karavani V. Aiming at the sweet side of cancer: aberrant glycosylation as possible target for personalized-medicine. Cancer Lett. 2014;352(1):102-12.

46. Zhao H, Chen W, Wu J, Wang L, Mao W. Clinical significance of preoperative serum tumor markers in esophageal squamous cell carcinoma. J Cancer Res Ther. 2014;10 Suppl:C179-85.

47. Blanas A, Sahasrabudhe NM, Rodriguez E, van Kooyk Y, van Vliet SJ. Fucosylated Antigens in Cancer: An Alliance toward Tumor Progression, Metastasis, and Resistance to Chemotherapy. Front Oncol. 2018;8:39.

48. Peixoto A, Relvas-Santos M, Azevedo R, Santos LL, Ferreira JA. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front Oncol. 2019;9:380.

49. Amon R, Reuven EM, Leviatan Ben-Arye S, Padler-Karavani V. Glycans in immune recognition and response. Carbohydr Res. 2014;389:115-22.

50. RodrIguez E, Schetters STT, van Kooyk Y. The tumour glyco-code as a novel immune checkpoint for immunotherapy. Nat Rev Immunol. 2018;18(3):204-11.

51. Peixoto A, Fernandes E, Gaiteiro C, Lima L, Azevedo R, Soares J, et al. Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. Oncotarget. 2016;7(39):63138-57.

52. Chandler KB, Costello CE, Rahimi N. Glycosylation in the Tumor Microenvironment: Implications for Tumor Angiogenesis and Metastasis. Cells. 2019;8(6).

53. Silva-Filho AF, Sena WLB, Lima LRA, Carvalho LVN, Pereira MC, Santos LGS, et al. Glycobiology Modifications in Intratumoral Hypoxia: The Breathless Side of Glycans Interaction. Cell Physiol Biochem. 2017;41(5):1801-29.

54. Smith DE, Fisher PA. Identification, developmental regulation, and response to heat shock of two antigenically related forms of a major nuclear envelope protein in Drosophila embryos: application of an improved method for affinity purification of antibodies using polypeptides immobilized on nitrocellulose blots. J Cell Biol. 1984;99(1 Pt 1):20-8.

55. Brockhausen I, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of Glycobiology. 3rd ed. Cold Spring Harbor (NY)2015. p. 113-23.

56. Ferreira JA, Magalhaes A, Gomes J, Peixoto A, Gaiteiro C, Fernandes E, et al. Protein glycosylation in gastric and colorectal cancers: Toward cancer detection and targeted therapeutics. Cancer Lett. 2017;387:32-45.

57. Gill DJ, Tham KM, Chia J, Wang SC, Steentoft C, Clausen H, et al. Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. Proc Natl Acad Sci U S A. 2013;110(34):E3152-61.

58. Gill DJ, Clausen H, Bard F. Location, location, location: new insights into O-GalNAc protein glycosylation. Trends Cell Biol. 2011;21(3):149-58.

59. . !!! INVALID CITATION !!! [58, 59].

60. Aebi M. N-linked protein glycosylation in the ER. Biochim Biophys Acta. 2013;1833(11):2430-7.

61. Pfeffer S, Dudek J, Schaffer M, Ng BG, Albert S, Plitzko JM, et al. Dissecting the molecular organization of the translocon-associated protein complex. Nat Commun. 2017;8:14516.

62. Nilsson IM, von Heijne G. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J Biol Chem. 1993;268(8):5798-801.

63. Chavan M, Yan A, Lennarz WJ. Subunits of the translocon interact with components of the oligosaccharyl transferase complex. J Biol Chem. 2005;280(24):22917-24.

64. Helenius A, Aebi M. Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem. 2004;73:1019-49.

65. Lowe JB, Marth JD. A genetic approach to Mammalian glycan function. Annu Rev Biochem. 2003;72:643-91.

66. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology. 2002;12(4):43R-56R.

67. Dall'Olio F, Malagolini N, Trinchera M, Chiricolo M. Mechanisms of cancer-associated glycosylation changes. Front Biosci (Landmark Ed). 2012;17(2):670-99.

68. Holgersson J, Lofling J. Glycosyltransferases involved in type 1 chain and Lewis antigen biosynthesis exhibit glycan and core chain specificity. Glycobiology. 2006;16(7):584-93.

69. Shimodaira K, Nakayama J, Nakamura N, Hasebe O, Katsuyama T, Fukuda M. Carcinoma-associated expression of core 2 beta-1,6-N-acetylglucosaminyltransferase gene in human colorectal cancer: role of O-glycans in tumor progression. Cancer Res. 1997;57(23):5201-6.

70. Holmes EH, Hakomori S, Ostrander GK. Synthesis of type 1 and 2 lacto series glycolipid antigens in human colonic adenocarcinoma and derived cell lines is due to activation of a normally unexpressed beta 1----3N-acetylglucosaminyltransferase. J Biol Chem. 1987;262(32):15649-58.

71. Trinchera M, Aronica A, Dall'Olio F. Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. Biology (Basel). 2017;6(1).

72. Itzkowitz SH, Yuan M, Montgomery CK, Kjeldsen T, Takahashi HK, Bigbee WL, et al. Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. Cancer Res. 1989;49(1):197-204.

73. Ikeda Y, Kuwano H, Baba K, Ikebe M, Matushima T, Adachi Y, et al. Expression of Sialyl-Tn antigens in normal squamous epithelium, dysplasia, and squamous cell carcinoma in the esophagus. Cancer Res. 1993;53(7):1706-8.

74. Zhang H, Li H, Ma Q, Yang FY, Diao TY. Predicting malignant transformation of esophageal squamous cell lesions by combined biomarkers in an endoscopic screening program. World J Gastroenterol. 2016;22(39):8770-8.

75. Ito K, Ye CL, Hibi K, Mitsuoka C, Kannagi R, Hidemura K, et al. Paired tumor marker of soluble E-selectin and its ligand sialyl Lewis A in colorectal cancer. J Gastroenterol. 2001;36(12):823-9.

76. Takada A, Ohmori K, Yoneda T, Tsuyuoka K, Hasegawa A, Kiso M, et al. Contribution of carbohydrate antigens sialyl Lewis A and sialyl Lewis X to adhesion of human cancer cells to vascular endothelium. Cancer Res. 1993;53(2):354-61.

77. Kannagi R, Izawa M, Koike T, Miyazaki K, Kimura N. Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis. Cancer Sci. 2004;95(5):377-84.

78. Shimada Y, Maeda M, Watanabe G, Imamura M. High serum soluble E-selectin levels are associated with postoperative haematogenic recurrence in esophageal squamous cell carcinoma patients. Oncol Rep. 2003;10(4):991-5.

79. McCarty OJ, Mousa SA, Bray PF, Konstantopoulos K. Immobilized platelets support human colon carcinoma cell tethering, rolling, and firm adhesion under dynamic flow conditions. Blood. 2000;96(5):1789-97.

80. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. 2011;11(2):123-34.

81. Honn KV, Tang DG, Crissman JD. Platelets and cancer metastasis: a causal relationship? Cancer Metastasis Rev. 1992;11(3-4):325-51.

82. Sakuma K, Aoki M, Kannagi R. Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF/bFGF-induced epithelial-mesenchymal transition. Proc Natl Acad Sci U S A. 2012;109(20):7776-81.

83. Thor A, Ohuchi N, Szpak CA, Johnston WW, Schlom J. Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. Cancer Res. 1986;46(6):3118-24.

84. Nuti M, Teramoto YA, Mariani-Costantini R, Hand PH, Colcher D, Schlom J. A monoclonal antibody (B72.3) defines patterns of distribution of a novel tumor-associated antigen in human mammary carcinoma cell populations. Int J Cancer. 1982;29(5):539-45.

85. Reddish MA, Jackson L, Koganty RR, Qiu D, Hong W, Longenecker BM. Specificities of anti-sialyl-Tn and anti-Tn monoclonal antibodies generated using novel clustered synthetic glycopeptide epitopes. Glycoconj J. 1997;14(5):549-60.

86. Oshiba G, Kijima H, Tanaka H, Kenmochi T, Himeno S, Kise Y, et al. Frequent expression of sialyl Le(a) in human esophageal squamous cell carcinoma. Int J Oncol. 2000;17(4):701-5.

87. Makino T, Shimada Y, Maeda M, Komoto I, Imamura M. Carbohydrate antigens as a risk factor for hematogenous recurrence of esophageal squamous cell carcinoma patients. Oncol Rep. 2001;8(5):981-5.

88. Ikeda Y, Kuwano H, Ikebe M, Baba K, Toh Y, Adachi Y, et al. Immunohistochemical detection of CEA, CA19-9, and DF3 in esophageal carcinoma limited to the submucosal layer. J Surg Oncol. 1994;56(1):7-12.

89. Yamachika T, Nakanishi H, Inada K, Kitoh K, Kato T, Irimura T, et al. Reciprocal control of colon-specific sulfomucin and sialosyl-Tn antigen expression in human colorectal neoplasia. Virchows Arch. 1997;431(1):25-30.

90. Lundin M, Nordling S, Roberts PJ, Lundin J, Carpelan-Holmstrom M, von Boguslawsky K, et al. Sialyl Tn is a frequently expressed antigen in colorectal cancer: No correlation with patient prognosis. Oncology. 1999;57(1):70-6.

91. Nanashima A, Nakagoe T, Sawai T, Nakamura S, Yamaguchi H, Yasutake T, et al. Different expressions of sialyl Tn antigen between polypoid and flat-type early colorectal cancers. Dis Colon Rectum. 1997;40(10 Suppl):S37-42.

92. Vazquez-Martin C, Cuevas E, Gil-Martin E, Fernandez-Briera A. Correlation analysis between tumor-associated antigen sialyl-Tn expression and ST6GalNAc I activity in human colon adenocarcinoma. Oncology. 2004;67(2):159-65.

93. Xu F, Fan C, Fan S, Liu F, Wen T, An G, et al. Expression profile of mucin-associated sialyl-Tn antigen in Chinese patients with different colorectal lesions (adenomas, carcinomas). Int J Clin Exp Pathol. 2015;8(9):11549-54.

94. Kjeldsen T, Clausen H, Hirohashi S, Ogawa T, Iijima H, Hakomori S. Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2----6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope. Cancer Res. 1988;48(8):2214-20.

95. Ogata S, Koganty R, Reddish M, Longenecker BM, Chen A, Perez C, et al. Different modes of sialyl-Tn expression during malignant transformation of human colonic mucosa. Glycoconj J. 1998;15(1):29-35.

96. Shen Y, Tiralongo J, Kohla G, Schauer R. Regulation of sialic acid O-acetylation in human colon mucosa. Biol Chem. 2004;385(2):145-52.

97. Shimono R, Mori M, Akazawa K, Adachi Y, Sgimachi K. Immunohistochemical expression of carbohydrate antigen 19-9 in colorectal carcinoma. Am J Gastroenterol. 1994;89(1):101-5.

98. Nakayama T, Watanabe M, Katsumata T, Teramoto T, Kitajima M. Expression of sialyl Lewis(a) as a new prognostic factor for patients with advanced colorectal carcinoma. Cancer. 1995;75(8):2051-6.

99. Lorenzi M, Vindigni C, Minacci C, Tripodi SA, Iroatulam A, Petrioli R, et al. Histopathological and prognostic evaluation of immunohistochemical findings in colorectal cancer. Int J Biol Markers. 1997;12(2):68-74.

100. Portela SV, Martin CV, Romay LM, Cuevas E, Martin EG, Briera AF. sLea and sLex expression in colorectal cancer: implications for tumourigenesis and disease prognosis. Histol Histopathol. 2011;26(10):1305-16.

101. Akamine S, Nakagoe T, Sawai T, Tsuji T, Tanaka K, Hidaka S, et al. Differences in prognosis of colorectal cancer patients based on the expression of sialyl Lewisa, sialyl Lewisx and sialyl Tn antigens in serum and tumor tissue. Anticancer Res. 2004;24(4):2541-6.

102. . !!! INVALID CITATION !!! [85]

103. Yang Y, Huang X, Zhou L, Deng T, Ning T, Liu R, et al. Clinical use of tumor biomarkers in prediction for prognosis and chemotherapeutic effect in esophageal squamous cell carcinoma. BMC Cancer. 2019;19(1):526.

104. Wu T, Mo Y, Wu C. Prognostic values of CEA, CA19-9, and CA72-4 in patients with stages I-III colorectal cancer. Int J Clin Exp Pathol. 2020;13(7):1608-14.

105. Magnani JL, Steplewski Z, Koprowski H, Ginsburg V. Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. Cancer Res. 1983;43(11):5489-92.

106. Xu Y, Chen Y, Wei L, Lai S, Zheng W, Wu F. Serum tumor-associated glycoprotein 72, a helpful predictor of lymph nodes invasion in esophagogastric junction adenocarcinoma. Biochem Biophys Res Commun. 2019;509(1):133-7.

107. Kannagi R. Carbohydrate antigen sialyl Lewis a--its pathophysiological significance and induction mechanism in cancer progression. Chang Gung Med J. 2007;30(3):189-209.

108. Magnani JL, Nilsson B, Brockhaus M, Zopf D, Steplewski Z, Koprowski H, et al. A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. J Biol Chem. 1982;257(23):14365-9.

109. Yu Z, Chen Z, Wu J, Li Z, Wu Y. Prognostic value of pretreatment serum carbohydrate antigen 19-9 level in patients with colorectal cancer: A meta-analysis. PLoS One. 2017;12(11):e0188139.

110. Gao Y, Wang J, Zhou Y, Sheng S, Qian SY, Huo X. Evaluation of Serum CEA, CA19-9, CA72-4, CA125 and Ferritin as Diagnostic Markers and Factors of Clinical Parameters for Colorectal Cancer. Sci Rep. 2018;8(1):2732.

111. Yanqing H, Cheng D, Ling X. Serum CA72-4 as a Biomarker in the Diagnosis of Colorectal Cancer: A Meta-analysis. Open Med (Wars). 2018;13:164-71.

112. Anderson RL, Balasas T, Callaghan J, Coombes RC, Evans J, Hall JA, et al. A framework for the development of effective anti-metastatic agents. Nat Rev Clin Oncol. 2019;16(3):185-204.

113. Azevedo R, Soares J, Peixoto A, Cotton S, Lima L, Santos LL, et al. Circulating tumor cells in bladder cancer: Emerging technologies and clinical implications foreseeing precision oncology. Urol Oncol. 2018;36(5):221-36.

114. Alunni-Fabbroni M, Muller V, Fehm T, Janni W, Rack B. Monitoring in metastatic breast cancer: is imaging outdated in the era of circulating tumor cells? Breast Care (Basel). 2014;9(1):16-21.

115. Li H, Song P, Zou B, Liu M, Cui K, Zhou P, et al. Circulating Tumor Cell Analyses in Patients With Esophageal Squamous Cell Carcinoma Using Epithelial Marker-Dependent and -Independent Approaches. Medicine (Baltimore). 2015;94(38):e1565.

116. Gallerani G, Fabbri F. Circulating Tumor Cells in the Adenocarcinoma of the Esophagus. Int J Mol Sci. 2016;17(8).

117. Yin XD, Yuan X, Xue JJ, Wang R, Zhang ZR, Tong JD. Clinical significance of carcinoembryonic antigen-, cytokeratin 19-, or survivin-positive circulating tumor cells in the peripheral blood of esophageal squamous cell carcinoma patients treated with radiotherapy. Dis Esophagus. 2012;25(8):750-6.

118. Schilling D, Todenhofer T, Hennenlotter J, Schwentner C, Fehm T, Stenzl A. Isolated, disseminated and circulating tumour cells in prostate cancer. Nat Rev Urol. 2012;9(8):448-63.

119. Bobek V, Matkowski R, Gurlich R, Grabowski K, Szelachowska J, Lischke R, et al. Cultivation of circulating tumor cells in esophageal cancer. Folia Histochem Cytobiol. 2014;52(3):171-7.

120. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013;501(7467):355-64.

121. Oh BY, Kim J, Lee WY, Kim HC. A New Size-based Platform for Circulating Tumor Cell Detection in Colorectal Cancer Patients. Clin Colorectal Cancer. 2017;16(3):214-9.

122. Zhang J, Chen K, Fan ZH. Circulating Tumor Cell Isolation and Analysis. Adv Clin Chem. 2016;75:1-31.

123. Yang MH, Imrali A, Heeschen C. Circulating cancer stem cells: the importance to select. Chin J Cancer Res. 2015;27(5):437-49.

124. Neves M, Azevedo R, Lima L, Oliveira MI, Peixoto A, Ferreira D, et al. Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: A novel biomarker and an analytical tool for precision oncology applications. N Biotechnol. 2019;49:77-87.

125. Mechref Y, Muddiman DC. Recent advances in glycomics, glycoproteomics and allied topics. Anal Bioanal Chem. 2017;409(2):355-7.

126. Campbell MP, Abrahams JL, Rapp E, Struwe WB, Costello CE, Novotny M, et al. The minimum information required for a glycomics experiment (MIRAGE) project: LC guidelines. Glycobiology. 2019;29(5):349-54.

127. Kodera Y, Isobe K, Yamauchi M, Satta T, Hasegawa T, Oikawa S, et al. Expression of carcinoembryonic antigen (CEA) and nonspecific crossreacting antigen (NCA) in gastrointestinal cancer; the correlation with degree of differentiation. Br J Cancer. 1993;68(1):130-6.

128. Amri R, Bordeianou LG, Sylla P, Berger DL. Preoperative carcinoembryonic antigen as an outcome predictor in colon cancer. J Surg Oncol. 2013;108(1):14-8.

129. Arora A, Patil V, Kundu P, Kondaiah P, Hegde AS, Arivazhagan A, et al. Serum biomarkers identification by iTRAQ and verification by MRM: S100A8/S100A9 levels predict tumor-stroma involvement and prognosis in Glioblastoma. Sci Rep. 2019;9(1):2749.

130. Zhang S, Shang S, Li W, Qin X, Liu Y. Insights on N-glycosylation of human haptoglobin and its association with cancers. Glycobiology. 2016;26(7):684-92.

131. Mann B, Madera M, Klouckova I, Mechref Y, Dobrolecki LE, Hickey RJ, et al. A quantitative investigation of fucosylated serum glycoproteins with application to esophageal adenocarcinoma. Electrophoresis. 2010;31(11):1833-41.

132. Song E, Zhu R, Hammoud ZT, Mechref Y. LC-MS/MS quantitation of esophagus disease blood serum glycoproteins by enrichment with hydrazide chemistry and lectin affinity chromatography. J Proteome Res. 2014;13(11):4808-20.

133. Qiu Y, Patwa TH, Xu L, Shedden K, Misek DE, Tuck M, et al. Plasma glycoprotein profiling for colorectal cancer biomarker identification by lectin glycoarray and lectin blot. J Proteome Res. 2008;7(4):1693-703.

134. Duffy MJ, Lamerz R, Haglund C, Nicolini A, Kalousova M, Holubec L, et al. Tumor markers in colorectal cancer, gastric cancer and gastrointestinal stromal cancers: European group on tumor markers 2014 guidelines update. Int J Cancer. 2014;134(11):2513-22.

135. Goldstein MJ, Mitchell EP. Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. Cancer Invest. 2005;23(4):338-51.

136. Huang C, Zhan T, Liu Y, Li Q, Wu H, Ji D, et al. Glycomic profiling of carcinoembryonic antigen isolated from human tumor tissue. Clin Proteomics. 2015;12(1):17.

137. Ju L, Wang Y, Xie Q, Xu X, Li Y, Chen Z, et al. Elevated level of serum glycoprotein bifucosylation and prognostic value in Chinese breast cancer. Glycobiology. 2016;26(5):460-71.

138. Pedersen JW, Blixt O, Bennett EP, Tarp MA, Dar I, Mandel U, et al. Seromic profiling of colorectal cancer patients with novel glycopeptide microarray. Int J Cancer. 2011;128(8):1860-71.

139. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10.
140. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. ArrayExpress--a public repository for microarray gene expression data at the EBI. Nucleic Acids Res. 2003;31(1):68-71.
141. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004;6(1):1-6.
142. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.
143. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature. 2010;463(7278):191-6.
144. Pavlopoulou A, Spandidos DA, Michalopoulos I. Human cancer databases (review). Oncol Rep. 2015;33(1):3-18.
145. Thanki K, Nicholls ME, Gajjar A, Senagore AJ, Qiu S, Szabo C, et al. Consensus Molecular Subtypes of Colorectal Cancer and their Clinical Implications. Int Biol Biomed J. 2017;3(3):105-11.
146. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Sci Rep. 2015;5:10775.
147. Sheynkman GM, Shortreed MR, Cesnik AJ, Smith LM. Proteogenomics: Integrating Next-Generation Sequencing and Mass Spectrometry to Characterize Human Proteomic Variation. Annu Rev Anal Chem (Palo Alto Calif). 2016;9(1):521-45.
148. Wang X, Zhang B. Integrating genomic, transcriptomic, and interactome data to improve Peptide and protein identification in shotgun proteomics. J Proteome Res. 2014;13(6):2715-23.

AIMS & STUDY OUTLINE

AIMS & STUDY OUTLINE

Esophageal cancer (EC) and colorectal cancer (CRC) are amongst the most common and deadliest cancers. Accurate molecular-based patient stratification remains challenging, hampering educated and potentially life-saving clinical decisions. Furthermore, despite the promising introduction of targeted therapeutics and immune checkpoint inhibitors, mainstay treatments mainly consist of surgery, chemotherapy, and radiotherapy, which often fall short in preventing tumour relapse and metastasis. These aspects underscore the urgent need for continuous biomarker discovery, leveraging innovative omics-based strategies, towards precision oncology and patient-tailored therapeutics.

This thesis explores the notion that the cancer glycocalyx undergoes significant changes with disease progression and dissemination as result of genetic as well as epigenetic events, that largely impact on most cancer hallmarks. Furthermore, it is well documented that abnormal protein glycosylation linked to cancer may originate unique glycoproteoforms holding significant potential for precise cancer targeting. Building on these observations, it hypothesizes that the EC and CRC glycoproteomes could yield clinically relevant signatures linked to cancer aggressiveness, providing a rationale for to facilitate future advancements in precision oncology and therapeutic development. To achieve this goal, targeted mass spectrometry-based glycoproteomics has been employed to interrogate the glycoproteome of EC and CRC for cancer-specific glycoproteoforms associated with aggressiveness. The focus was on exploiting glycoproteins carrying the immature O-glycan STn antigen in EC (addressed in chapter I) or terminal SLe epitopes in CRC (addressed in chapter II), both of which are linked to invasion and metastasis.

In summary, this thesis comprehends the following specific aims:

i) Interrogated the STn-targeted glycoproteomics in ESSC envisaging cancer-specific glycoproteoforms for precision oncology;

ii) Interrogate the CRC glycome and SLe-glycoproteome envisaging cancer-specific glycoproteoforms for precision oncology;

iii) Demonstrate the relevance of the glycoproteome as a source of clinically meaningful cancer biomarkers, supported by in silico approaches.

This thesis is organized in two chapters that together meet the aims. Chapter I. Glycobiomarkers discovery in EC incorporates the article entitled “Target Score—A Proteomics Data Selection Tool Applied to Esophageal Cancer Identifies GLUT1-Sialyl Tn Glycoforms as Biomarkers of Cancer Aggressiveness”. This chapter characterizes the expression of the STn antigen in EC primary tumors and metastasis, establishing a link between this glycan and poor prognosis. Furthermore, it utilizes microfluidic devices to

demonstrate its expression in CTCs recovered from patients, reinforcing its association with disease spread and poor prognosis. By exploiting glycoengineered cell lines, it further demonstrates that STn plays a functional role in driving cancer aggressiveness. Building on these findings, it employs lectin-affinity chromatography and mass spectrometry to identify glycoproteins carrying this glycan in EC. It further proposes an algorithm, drawing on information from the Human Protein Atlas (https://www.proteinatlas.org/ (accessed on 1 January 2021)) consulted in March 2020, to prioritize potentially targetable glycoproteins, which led to the identification of GLUT1. Finally, it is dedicated to supporting the potential cancer-specific GLUT1-STn glycoproteoforms, envisioning a rationale for targeting aggressive EC cells.

Chapter II. Glycobiomarkers discovery in CRC, englobes the article “E-selectin Affinity Glycoproteomics Reveals Neuroendocrine Proteins and the Secretin Receptor as a Poor Prognosis Signature in Colorectal Cancer”. This chapter provides an in-depth investigation on glycoprotein carrying the metastasis-mediator SLe glycoepitopes, in in-depth investigation on glycoproteins carrying SLe glycoepitopes, aiming to address the low cancer-specificity of these metastasis mediating glycans. It describes that more than 70% of CRC tumors overexpressed sLeA/X, yet without discernible associations with metastasis or survival. However, by integrating TCGA analysis unveiled differing expression patterns of sLeA/X-related glycogenes correlating with disease severity, indicating context dependent regulation by distinct glycosyltransferases. Deeper exploration of metastatic tumour sialoglycoproteome identified nearly 600 glycoproteins, greatly expanding the understanding on the metastasis-related glycoproteome. Using the algorithm developed in Chapter I, it further identified the secretin receptor (SCTR), a poorly studied neuroendocrine-related protein, as a top-ranked targetable signature linked to poor prognosis and metastasis. Furthermore, it highlighted the potential cancer-specificity of SCTR-SLe glycoproteoforms. Overall, Chapters I and II highlight the unexplored potential of the cancer glycoproteome, offering a framework for advancement in this field. Furthermore, they underscore potentially targetable glycoproteoforms for EC and CRC, which could support precision oncology, thereby addressing unmet clinical challenges in this context.

CHAPTER I

GLYCOBIOMARKERS DISCOVERY IN ESOPHAGEAL CANCER

Target Score—A Proteomics Data Selection Tool Applied to Esophageal Cancer Identifies GLUT1-Sialyl Tn Glycoforms as Biomarkers of Cancer Aggressiveness

Sofia Cotton[1,2,3,4], Dylan Ferreira[1,2,3,4], Janine Soares[1,2,5], Andreia Peixoto[1,3,4], Marta Relvas-Santos[1,2,3,4,6], Rita Azevedo[7], Paulina Piairo[8], Lorena Diéguez[8], Carlos Palmeira[1,9,10], Luís Lima[1,11], André M. N. Silva[6], Lúcio Lara Santos[1,2,10,11,12] and José Alexandre Ferreira[1,2,11*]

1 Experimental Pathology and Therapeutics Group, IPO Porto Research Center (CI-IPOP), Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal; sofia.ribeiro.cotton@ipoporto.min-saude.pt (S.C.); dylan.ferreira@hotmail.com (D.F.); janinepaivasoares@gmail.com (J.S.); andreia.peixoto@ipoporto.min-saude.pt (A.P.); marta_frds@hotmail.com (M.R.-S.); carlospalmeira@ipoporto.min-saude.pt (C.P.); luis.carlos.lima@ipoporto.min-saude.pt (L.L.); llarasantos@gmail.com (L.L.S.)

2 Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-313 Porto, Portugal

3 Institute for Research and Innovation in Health (i3S), University of Porto, 4200-135 Porto, Portugal

4 Institute for Biomedical Engineering (INEB), 4200-135 Porto, Portugal

5 QOPNA/LAQV, Department of Chemistry, Campus Universitário de Santiago, University of Aveiro, 3810-193 Aveiro, Portugal

6 REQUIMTE-LAQV, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal; andre.silva@fc.up.pt

7 Institute of Biomedicine, University of Turku, FI-20014 Turku, Finland; rita.pereiraazevedo@utu.fi

8 Medical Devices Research Group, International Iberian Nanotechnology Laboratory (INL), 4715-330 Braga, Portugal; paulina.piairo@inl.int (P.P.); lorena.dieguez@inl.int (L.D.)

9 Department of Immunology, Portuguese Institute of Oncology of Porto, 4200-072 Porto, Portugal

10 Health Science Faculty, University of Fernando Pessoa, 4249-004 Porto, Portugal

11 Porto Comprehensive Cancer Center (P.ccc), 4200-072 Porto, Portugal

12 Department of Surgical Oncology, Portuguese Institute of Oncology, 4200-072 Porto, Portugal

Corresponding author: José Alexandre Ferreira (jose.a.ferreira@ipoporto.min-saude.pt)

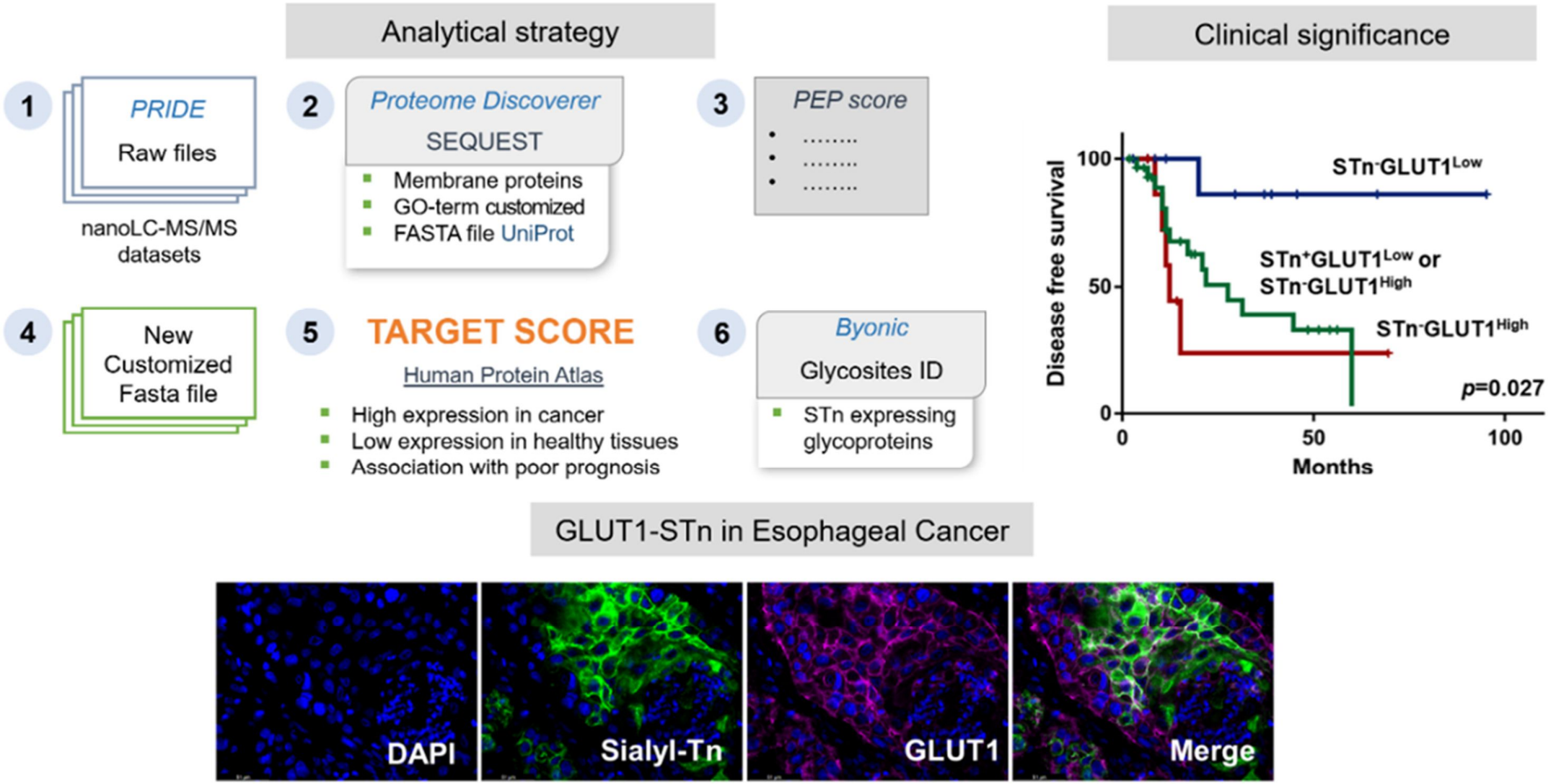
Supplementary data to this article can be found online at https://www.mdpi.com/1422-0067/22/4/1664/s1

Abstract

Esophageal cancer (EC) is a life-threatening disease, demanding the discovery of new biomarkers and molecular targets for precision oncology. Aberrantly glycosylated proteins hold tremendous potential towards this objective. In the current study, a series of esophageal squamous cell carcinomas (ESCC) and EC-derived circulating tumor cells (CTCs) were screened by immunoassays for the sialyl-Tn (STn) antigen, a glycan rarely expressed in healthy tissues and widely observed in aggressive gastrointestinal cancers. An ESCC cell model was glycoengineered to express STn and characterized in relation to cell proliferation and invasion in vitro. STn was found to be widely present in ESCC (70% of tumors) and in CTCs in 20% of patients, being associated with general recurrence and reduced survival. Furthermore, STn expression in ESCC cells increased invasion in vitro, while reducing cancer cells proliferation. In parallel, an ESCC mass spectrometry-based proteomics dataset, obtained from the PRIDE database, was comprehensively interrogated for abnormally glycosylated proteins. Data integration with the Target Score, an algorithm developed in-house, pinpointed the glucose transporter type 1 (GLUT1) as a biomarker of poor prognosis. GLUT1-STn glycoproteoforms were latter identified in tumor tissues in patients facing worst prognosis. Furthermore, healthy human tissues analysis suggested that STn glycosylation provided cancer specificity to GLUT1. In conclusion, STn is a biomarker of worst prognosis in EC and GLUT1-STn glycoforms may be used to increase its specificity on the stratification and targeting of aggressive ESCC forms.

Keywords: cancer biomarkers; circulating tumors cells; esophageal cancer; glycomics; glycoproteomics;
bioinformatics

Graphical Abstract



1. Introduction

Esophageal cancer (EC) poses as a major health risk due to increasing incidence, late diagnosis and poor prognosis [1]. It ranks seventh in terms of incidence and sixth in terms of cancer mortality worldwide, with over 500,000 new cases a year and 5-year overall survival rate of approximately 20% [2]. Esophageal squamous cell carcinoma (ESCC) arising from the non-keratinized stratified squamous epithelium is the most prevalent histological subtype worldwide (60–70%) and occurs more often in the upper and mid esophagus [3]. Although surgery, chemo- and radiotherapy are widely used to manage ESCC patients, responses to treatment remain poor and encompass relevant health side effects [4], urging the introduction of novel and more effective therapeutic modalities to avoid tumor relapse and rapid disease dissemination. However, high inter- and intratumoral molecular heterogeneity poses a significant hurdle for the identification of cancer specific molecular targets, decisively delaying accurate patient stratification and effective targeted therapeutics.

In recent years, many studies have comprehensively interrogated the proteome of EC in search for biomarkers to aid disease management at different levels, including diagnosis, patient stratification, prognosis and responses to treatment [5–7]. Despite promising, conventional proteomics approaches on EC are yet to deliver targetable molecular signatures due to insufficient cancer specificity. Addressing alterations in the glycosylation of membrane proteins holds a great potential towards this objective [8]. In fact, over the past years, we and other research groups have demonstrated that changes in the structure and distribution of glycans in proteins at the surface of cancer cells may provide the necessary bispecificity towards unique cancer molecular signatures [8,9]. Namely, we have identified MUC16, CD44 and nucleolin glycoforms holding potential for more accurate patient stratification and treatment follow-ups in different types of tumors [9–11]. We have also used patient's autoantibodies to demonstrate that changes in EC glycosylation could generate cancer glyconeoantigens, reinforcing the enormous potential of glycoproteomics for biomarker discovery [12]. However, pinpointing clinically relevant molecular signatures from big datasets remains a daunting enterprise, requiring dedicate bioinformatics tools. Recently, we have addressed this aspect by proposing a simple algorithm, termed Target Score, that exploits clinicopathological data deposited in the Human Protein Atlas to extract cancer biomarkers from glycoproteomics experiments [9]. Our approach was specifically designed to rank proteins according to its prognosis value and targetability, favouring proteins located at the cell surface, showing high abundance in tumors in relation to healthy

human organs. Furthermore, it significantly penalized proteins expressed in reproductive, immune and nervous systems, allowing to choose molecules posing minimum probability of significant off-target effects. Target Score-assisted glycoproteomics in gastric cancer led to the identification of nucleolin glycofoms at the cell surface associated with worst prognosis and metastases development, which contrasted with its typical nuclear location in healthy tissues. The cancer-specific nature of this finding elegantly portrayed the potential of Target Score algorithm for identification of unexpected glycosignatures and set a biomarker discovery roadmap that can now be translated and adapted to other human tumors, including EC [9].

Glycobiomarkers discovery often requires the adoption of targeted glycoproteomics approaches focusing on specific alterations in protein glycosylation associated with cancer [9,10,13]. In this context, digestive tract tumors have been long known to express membrane proteins carrying abnormal glycosylation patterns as post-translational modifications [8,9,14]. Namely, these tumors often overexpress sialyl-Tn (STn), which results from a premature stop in O-glycans elongation by sialylation of the Tn antigen [8]. The STn antigen is highly expressed by more aggressive and advanced gastric and colorectal tumors, being generally associated with poor prognosis [8]. Moreover, it plays a relevant functional role in cancer, namely chemoresistance [15], cell invasion [16,17] and immune escape [18,19], decisively contributing to disease progression. The STn antigen is also rarely observed in most healthy tissues, except for low expression in cells specialized in mucin secretion facing the lumen of the gastrointestinal and respiratory tracts [20]. As such, targeting STn and STn-expressing glycoproteins constitutes a relevant approach towards identifying cancer-specific glycobiomarkers in EC. However, little is known about STn expression patterns and its clinical relevance in these tumors. Therefore, the present work devotes to providing a clinical and functional context for STn in ESCC, which constitute the bulk of EC. Furthermore, it comprehensively interrogates a glycoproteomics dataset for STn-expressing glycoproteins. Potentially targetable biomarkers arising from this study were subsequently validated in EC and healthy human tissues foreseeing the molecular rationale for precision oncology.

2. Results and discussion

The present study builds on the hypothesis that the STn antigen may be explored towards the objective of more precise patient stratification and targeting ESCC, given its functional role in the disease and frequent association with poor prognosis in many solid tumors [8,17,18,21]. Therefore, we first focused in setting the clinical context for STn expression using ESCC tissues of different clinicopathological natures and exploiting its functional role

through a glycoengineered cell model. The second part of the study is concerned with the identification of cancer differently expressed proteins which bear the STn antigen, using database information and subsequent validation in tumor tissues.

2.1. STn Antigen in ESCC Primary Tumours, Metastases and CTCs

Little is known about the expression of STn antigen in ESCC, which constitutes the main histopathological group within EC. Therefore, we have evaluated its expression in a retrospective series of 48 cancer tissues reflecting different stages and grades of the disease (Table 1). Approximately 70% of tumor tissues were positive for the antigen, with a marked membrane expression mainly in tumor cells. Notably, cancer cells expressing this antigen were found in scattered clusters that, for most cases, did not exceed 20% of the tumor area (Figure 1A). We could also find STn antigens in adjacent lymph node metastases, suggesting a role in disease dissemination via lymphatic vessels, which warrants future confirmation. The presence of STn was further confirmed by the lack of immunoreactivity for B72.3 plus CC49 antibodies in tissue sections incubated with neuraminidase prior to analysis. Finally, we have used innovative microfluidics devices to capture CTCs from the blood of ESCC patients and assess STn expression in situ. Briefly, peripheral blood was prospectively collected from 10 patients without prior treatment history (Table 2) and processed using the microfluidic chips. Cancer cells with larger dimensions and less deformable were entrapped in the filter area with minimal contamination from smaller and deformable blood cells [22]. Based on our previous observations, the absence of CD45 and presence of pan-CK and/or STn were considered for definitive tumor cell identification [21,22]. Even though early stage ESCC, which constitutes the bulk of the prospective series, are known to shed modest numbers of cancer cells into hematogenous circulation [23], we were able to detect STn positive CTCs (DAPI+, STn+, pan-CK-, CD45-) in 20% of the patients (Figure 1B). Interestingly, STn expressing CTCs did not present pan-CKs expression typical of cells of epithelial nature, suggesting a mesenchymal-like phenotype which also warrants confirmation. Although the low number of cases did not allow the establishment of associations with clinicopathological features, the discovery of STn in ESCC CTCs supports the need for more in-depth investigation on this subject. In fact, the detection of STn antigens in lymph node metastases and CTCs supports a role in disease dissemination, which may be of great value for improved liquid biopsies, as previously suggested for other solid tumors [21,22]. Notably, the STn antigen was also detected in the histologically normal esophageal mucosa adjacent to tumor tissue, mainly in the differentiating cells of the mucosa facing the esophagus lumen (Figure 1A). Notwithstanding, mucosal cells showed considerably lower STn expression than tumor cells

and of predominantly cytoplasmatic nature, demanding a comprehensive characterization of the STn-glycoproteome for improved cancer specificity.

Table 1. Clinicopathological data of patients included in the STn analysis (n=48)

	n (%)
Stage	
I	13 (27)
II	11 (23)
III	23 (48)
IV	1 (2)
Tumour (pT)	
T1	9 (19)
T2	13 (27)
T3	25 (52)
T4	1 (2)
Lymph node metastasis (pN)	
N0	21 (44)
N1	11 (23)
N2	10 (21)
N3	6 (12)
Distant Metastasis (M)	
M0	47 (98)
M1	1 (2)
Distant Recurrence (DR)	
No	31 (65)
Yes	17 (35)
Histological Classification	
Squamous cell carcinoma	48 (100)
Adenocarcinoma	0 (0)
Borrmann Classification	
I	7 (15)
II	21 (44)
III	7 (15)
IV	11 (22)
Missing information	2 (4)
Keratinization Degree	
Non-keratinized	30 (63)
Moderately keratinized	2 (4)

Keratinized	16 (33)
Differentiation Degree	
Well-differentiated	7 (15)
Moderately differentiated	26 (54)
Poorly differentiated	10 (21)
Missing information	5 (10)
Extra-tumoral growth	
Present	12 (25)
Absent	36 (75)
Lymphovascular Permeation	
Present	21 (44)
Absent	27 (56)
Neural Permeation	
Present	13 (27)
Absent	35 (73)

Finally, we evaluated possible links between STn expression in tumors and clinically relevant clinicopathological variables as stage, TNM staging, keratinization degree, differentiation degree, extra tumoral growth, lympho-vascular permeation and perineural permeation. No evident associations were established (Table 3), reflecting the prevalent nature of this antigen across the disease. However, a statistically significant association was observed regarding recurrence (Figure 1C) and decreased disease-free survival (Figure 1D), supporting a role in tumor aggressiveness.

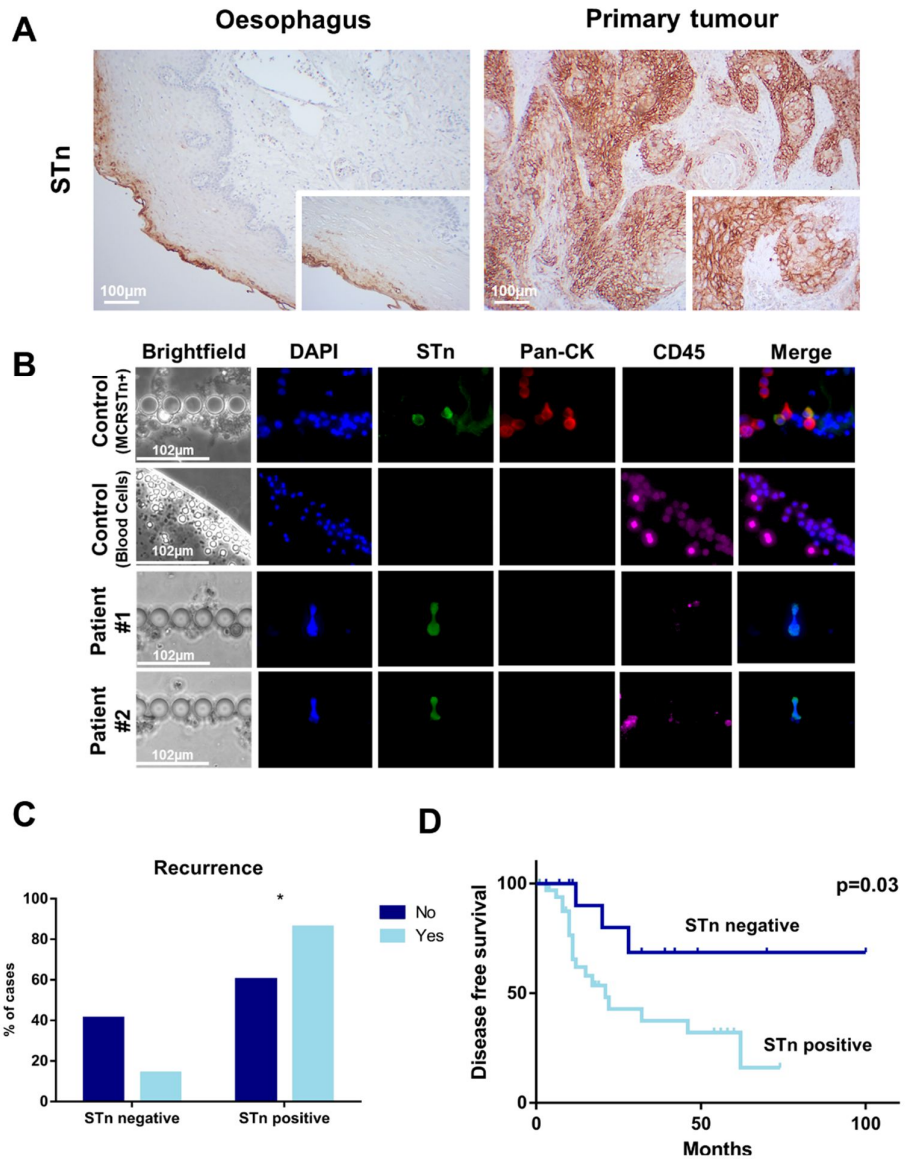


Figure 1. The STn antigen is overexpressed in esophageal squamous cell carcinomas (ESCC) and CTCs, being associated with recurrence and decreased disease-free survival. (A) The STn antigen is overexpressed in approximately 70% of ESCC and adjacent lymph node metastasis and shows little expression in the healthy esophagus. Most ESCC and lymph node metastasis present focal STn expressions showing high intensity staining at the cell membrane. On the other hand, the healthy esophagus shows a diffuse low to moderate cytoplasmatic staining in differentiating cells of the mucosa facing the lumen of the organ. (B) CTCs recovered from the blood of ESCC patients express the STn antigen. CTCs were isolated from whole blood by size-based microfluidics devices

presenting an internal architecture that enables the entrapment of larger cancer cells with minimal contamination from blood cells. In addition to STn (green), cell nuclei were stained with 4′,6-diamidino-2-phenylindole, dihydrochloride (DAPI) (blue), pan-CKs (red) was used as an epithelial marker, generally expressed by some subpopulations of cancer cells and CD45 (purple) was used to identify blood cells. ESCC-derived CTCs presented a DAPI+/pan-CKs-/STn+/CD45- phenotype. MCRSTn+ bladder cancer cells were used as positive control for cancer cells expressing the STn antigen (DAPI+/pan-CKs+/STn+/CD45-) and blood cells from healthy donors were used to demonstrate the absence of STn in CD45+ cells (DAPI+/pan-CKs−/STn−/CD45−). (C) STn expression associates with recurrence (* $p < 0.05$; Chi-square) and (D) disease-free survival ($p = 0.03$, Chi-square).

Table 2. Clinicopathological data of patients included in circulating tumor cells (CTCs) analysis (n = 10).

	n (%)
Stage	
I	0 (0)
II	7 (70)
III	3 (30)
IV	0 (0)
Tumour (pT)	
T1	1 (10)
T2	2 (20)
T3	6 (60)
T4	1 (10)
Lymph node metastasis (pN)	
N0	7 (70)
N1	3 (30)
Distant metastasis (M)	
M0	10 (100)
M1	0 (0)
Histological Classification	
Squamous cell carcinoma	10 (100)
Adenocarcinoma	0 (0)
Lymphovascular Permeation	
Present	3 (30)
Absent	3 (30)

Missing information	4 (40)
Neural Permeation	
Present	1 (10)
Absent	5 (50)
Missing information	4 (40)

2.2. Generation of a STn ECSS Cell Line and Functional Implications

Analysis of clinical samples strongly supported a link between the STn antigen and ESCC progression and dissemination, as previously observed for other solid tumors [17,24]. To gain more insights on this matter, we elected the Kyse-30 cell line as in vitro model to explore the functional impact of STn in ESCC. We started by characterizing the cellular O-glycome, using a benzyl-GalNAc residue, which mimics the Tn antigen, the first sugar in O-glycans biosynthesis [25]. The Tn mimetic was added to the culture medium and used by cancer cells glycosylation machinery as a scaffold for further elongation. After processing across the secretory pathways, more extended benzylated glycans reflecting the cells O-glycome were secreted and easily recovered from the culture media by C18 reverse phase affinity, permethylated and analyzed by nanoLC-ESI-MS/MS. According to Figure 2A, Kyse-30 mainly expressed extended core 2 O-glycans (m/z 1021.53, 1195.62, 1382.71, 1470.76) as well as T (m/z 572.31) and sialyl-T antigens (m/z 933.48). To a lesser extent, this cell line also expressed core 3 O-glycan (m/z 613.33). Moreover, mass spectrometry did not shown signs of the STn antigen (Figure 2A). Collectively, we have demonstrated that Kyse-30 cells presented a glycome marked mainly by extended core 2 O-glycans, therefore not reflecting the presence of STn exhibited by some subpopulations of cancer cells in human ESCC tumors. Notably, the absence of short-chain O-glycans such as the STn antigen in Kyse-30 is consistent with many other reports for human cancer cell lines grown in vitro, denoting a strong dependence on yet incompletely determined microenvironmental features [16,26]. We believe that the identification of the microenvironmental elements underlying O-glycans antagonization in vivo will be critical to provide a clear context for clinical interventions and should be addressed in a close future [27].

Table 3. STn expression in esophageal cancer (EC) according to clinicopathological variables.

STn expression	Positive cases/Total (%)	*p* value
Stage		
I	7/13 (54)	0.414
II	8/11 (72)	
III	18/23 (78)	
IV	1/1 (100)	
Tumour (pT)		
T1	4/9 (44)	0.261
T2	10/13 (77)	
T3	19/25 (76)	
T4	1/2 (50)	
Lymph node metastasis (pN)		
N0	13/21 (62)	0.639
N1	8/11 (73)	
N2	8/10 (80)	
N3	5/6 (83)	
Distant metastasis (M)		
M0	33/47 (70)	0.708
M1	1/1 (100)	
Distant recurrence (DR)		
No	20/31 (65)	0.167
Yes	14/17 (82)	
Borrmann Classification		
I	4/7 (57)	0.583
II	15/21 (71)	
III	6/7 (86)	
IV	9/11 (82)	
Missing information	0/2 (0)	
Keratinization Degree		
Non-keratinized	22/30 (73)	0.762
Moderately keratinized	1/2 (50)	
Keratinized	11/16 (69)	
Differentiation Degree		
Well-differentiated	4/7 (57)	0.577
Moderately differentiated	19/26 (73)	

Poorly differentiated	8/10 (80)	
Missing information	3/5 (60)	
Extra-tumoral growth		
Present	10/12 (83)	0.237
Absent	24/36 (67)	
Lymphatic Permeation		
Present	16/21 (76)	0.347
Absent	18/27 (67)	
Vascular Permeation		
Present	14/21 (67)	0.403
Absent	20/27 (74)	
Neural Permeation		
Present	8/13 (62)	0.301
Absent	26/35 (74)	

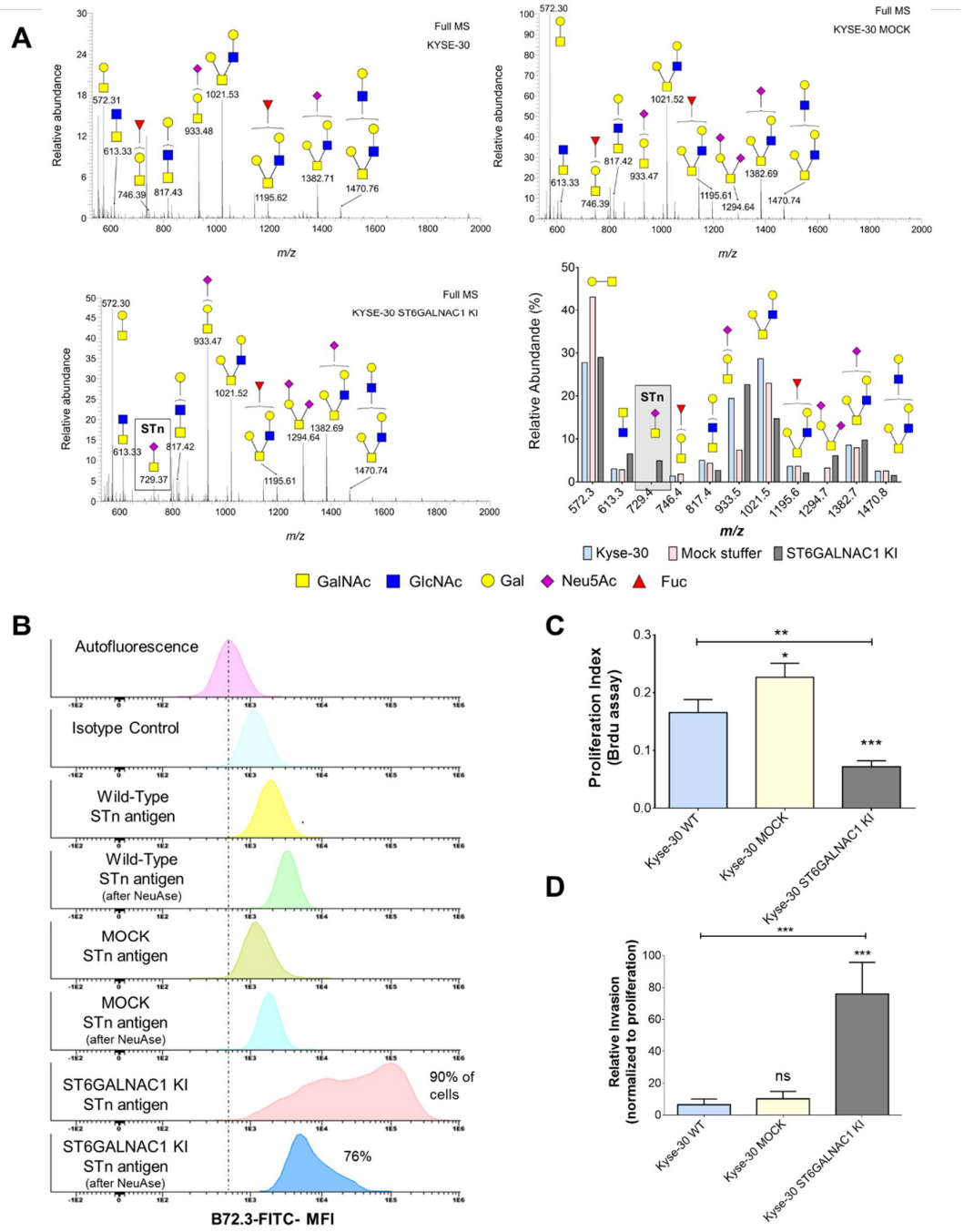


Figure 2. The stable transfection of ST6GALNAC1 leads to STn expression in Kyse-30 cells without inducing profound glycome remodeling and induces a massive decrease in proliferation and a significant increase in invasion. (A) Kyse-30 cells express elongated O-glycans and do not express STn, whereas Kyse-30 ST6GalNAc1 KI express STn with minor alterations in the glycome. Glycomics analysis on Kyse-30 revealed a glycome dominated by several core 2 O-glycans (m/z 817.43; 1021.53; 1195.62; 1382.71;

1470.76) as well as T (m/z 572.31) and sialylated T antigens (m/z 933.48). Core 3 was also detected. (m/z 613.33). Mock cells presented a similar profile, even though showing the T antigen (m/z 572.3) as major ion and presenting low amounts of di-sialyl-T (m/z 1294.7). Neither wild type nor mock cells expressed the STn antigen (m/z 729.4). On the other hand, ST6GALNAC1 transfected cells presented a glycome very similar to wild type cells but were the only cells expressing STn (m/z 729.4). (B) Flow cytometry analysis confirms the low STn levels in Kyse-30 and the massive increase in this antigen after stable transfection with ST6GalNAc1. The stable transfection with ST6GALNAC1 leads to a significant increase in STn in almost 90% of the cells. The signal significantly decreases with the removal of the sialic acid after sialidase (NeuAse) digestion, confirming the presence of STn. Notably, the extension of NeuAse digestions is never complete, explaining the presence of lower amounts of STn in the cells after treatment. (C) The overexpression of STn resulting from ST6GalNAc I overexpression leads to a massive decrease in proliferation. (D) The overexpression of STn resulting from ST6GalNAc I overexpression leads to a massive increase in invasion. ns: not significative; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; Mann-Whitney for three independent replicates in proliferation assays and five independent replicates in invasion assays.

We then induced the stable overexpression of ST6GalNAc I, responsible by the O-6 sialylation of the Tn antigen originating STn epitopes, originating the Kyse-30 ST6GALNAC1 KI cell line. As shown in Figure 2A, Kyse-30 ST6GALNAC1 KI cells gained the capacity to express the STn, whereas the antigen was not observed in mock cells. This was obtained without inducing major glycome remodeling in comparison to wild type and mock cells (Figure 2A). Flow cytometry analysis confirmed that Kyse-30 cells did not express the STn antigen, whereas Kyse-30 ST6GALNAC1 KI exhibited high levels of this antigen in approximately 90% of cells (Figure 2B). Subsequent functional studies showed that STn overexpression was accompanied by a striking decrease in cell proliferation (55 and 70% in relation to mock and wild type cells; Figure 2C), suggesting that STn expression may contribute to a quasi-quiescent state, which is frequently observed in more aggressive cancer cell phenotypes. Concomitantly, STn-expressing cells significantly increased their invasive capacity in vitro compared to both mock and wild type control cells (Figure 2D), as previously observed for other glycoengineered cancer cell models of different origins (breast, gastric, bladder) [24,28,29]. Collectively, these findings reinforce the close link between STn expression and ESCC aggressiveness, as previously observed for other types of solid tumors [8,17,24]. Namely, it is well described that the substitution of elongated glycans by shorter sialylated antigens, such as STn, promotes a profound remodeling of the

plasma membrane glycoproteome, triggering important oncogenic signaling pathways involved in cell proliferation [30–32]. Moreover, it changes cell-cell and cell-matrix adhesion interfaces, facilitating cell motility [16,27]. Nevertheless, the exact molecular mechanism governing the functional alterations resulting from STn overexpression in Kyse-30 cells remains unknown and should be subject to future investigation, envisaging a more detailed understanding on the biological role of STn in esophageal cancer.

2.3. Identification of STn Modified Proteins

Immunohistochemistry analysis showed that both ESCC and histologically normal esophageal tissues express STn antigens. Even though with distinct intensities and tissue specific patterns, these observations challenge the cancer specificity of the STn antigen alone. This led us to query if investigating the STn-associated glycoproteome can contribute to increase cancer specificity. To address this objective, we opted for a bioinformatics approach comprehending four main steps, summarized in Figure 3A. Namely, (i) identification of ESCC proteomics raw data deposited in relevant primary databases, which could serve as starting point for our research; (ii) identification of membrane proteins that, due to their cell surface nature, may be easily targetable by antibodies and other ligands; (iii) sorting of potentially targetable glycoproteins based on their overexpression in cancer in relation to healthy tissues, with the assistance of the target score algorithm; (iv) identification of STn-glycosites in top ranked proteins holding potential for future design of targeted therapeutics.

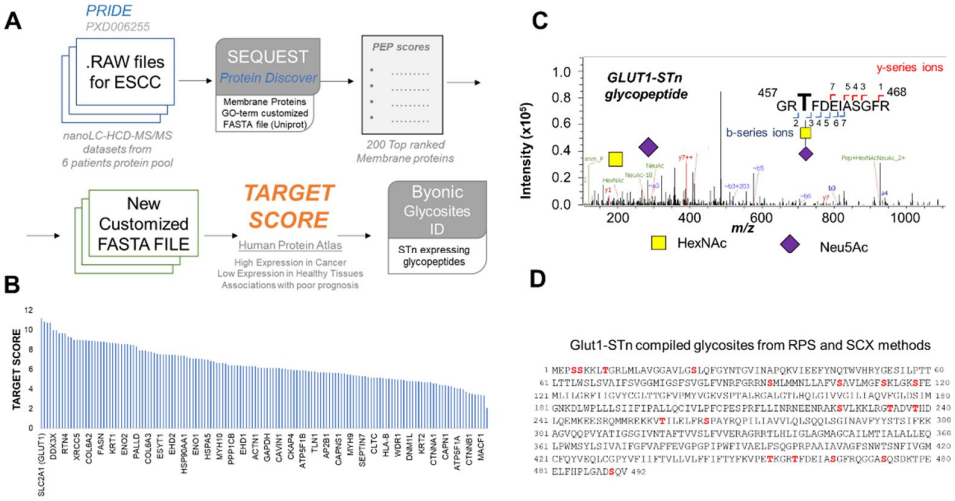


Figure 3. Target score assisted revisitation of ESCC proteomics data identifies SLC2A1 (GLUT1) as a potentially targetable biomarker carrying the STn antigen. A) Bioinformatics workflow for identification of targetable biomarkers in ESCC. Raw nanoLC-MS/MS ESCC proteomics data associated with project PXD006255 were first downloaded from the PRIDE database, and screened for membrane glycoproteins using SEQUEST, a tandem mass spectrometry database search program for Protein Discoverer. The original dataset comprehended nanoLC-MS/MS runs generated from protein digests with Lys-C-trypsin. The proteins were isolated from a pool of 6 ESCC. Two-types of pre-separation methods were employed prior to nanoLC-MS/MS analysis: strong cation exchange and reverse phase. As such, LC-MS/MS data from the two methods were processed separately. Identifications were made against a GO term customized database for plasma membrane proteins and ordered in relation to their PSM. Top ranked 200 proteins for each method were then matched to find common signatures that were subsequently analyzed by the Target Score algorithm. Target score provided an easy strategy for protein classification according to potentially targetability based on pre-existing information about cellular location, expression in tumours and human healthy tissues as well as associations with poor prognosis. The algorithm uses as primary source of information the Human Protein Atlas, but the same workflow maybe applied to similar databases. Finally, we returned to the initial proteomics raw data to access if top scored glycoproteins were modified with the STn antigen, using the Byonic software for more precise glycopeptides identification and glycosites annotation. Collectively, this strategy was designed to identify with high confidence abnormally glycosylated proteins with the STn antigen, showing higher prevalence at the cell membrane of ESCC in comparison to healthy tissues, enabling easy targeting with antibodies and other ligands with minimal potential off-target effects. B) Target Score analysis identified SLC2A1 (GLUT1) as a potential targetable molecule in ESCC. C) Example of a glycopeptide HCD-MS/MS spectrum confirming the existence of GLUT1-Stn proteoglycoforms in ESCC. The MS/MS spectrum highlights characteristic b- and y-series ions derived from fragmentations in the peptide backbone, enabling glycosites assignment, as well as several GalNAc and Neu5Ac oxonium ions from STn. D) STn expressing glycosites in GLUT1. STn glycosites in GLUT1 are highlighted in red. Details on the identified glycopeptides are presented as Supplementary Information.

After an initial screening of the PRIDE database and despite the significant number of proteomics studies on EC, we have identified only one ESCC proteomics dataset freely available to the scientific community at the time of this study. It concerned a pool of tumor proteins from 6 patients, however, without a clear identification of the tumors

clinicopathological natures. Despite the preliminary nature of the study, we decided to use it as starting point to portrait the potential of our approach for glycobiomarkers discovery. As a general approach, we re-analyzed data against a database composed of proteins with Gene Ontology (GO) terms for the plasma membrane (Figure 3A), leading to the identification of 773 and 795 plasma membrane proteins for RPS and SCX pre-purification methods, respectively (Tables S1 and S2-Supplementary Information). To reduce the dataset dimension to very high confidence identifications for downstream analysis, we have focused on the top 250 glycoproteins from each dataset (according to the posterior error probability-PEP score). Notably, we have identified 213 glycoproteins that were common to the RPS and SCX methods (Table S3-Supplementary Information). This list was after investigated with the Target Score algorithm, a tool designed by our group to extract clinically relevant information in a comprehensive manner from large protein lists retrieved from proteomics studies. This algorithm uses protein expression data from the Human Protein Atlas to pinpoint potentially targetable glycobiomarkers by attributing higher scores to proteins expressed in higher abundance at the cell surface in tumors, while penalizing high expression in healthy tissues. Finally, it also potentiates associations with worst cancer prognosis, which help set the pace for addressing more aggressive tumors. However, there is still limited data on ESCC in comparison to other solid tumors. To address this limitation, we used head-and-neck squamous cell carcinomas as models based on similar histology, etiology and molecular features [33–35]. As shown in Figure 3B, SLC2A1, also known as glucose transporter type 1 (GLUT1), was retrieved with the highest target score (Table S4-Supplementary Information). GLUT1 has been previously found to be overexpressed in ESCC in response to metabolic adaptations to rapid tumor growth and hypoxia [36,37]. Moreover, GLUT1 expression has been associated with hematogenous recurrence [38] and considered a surrogate marker of decreased response to cisplatin [39] and worst prognosis following potentially curative surgical resection [40]. Collectively, these observations support a key role for this protein in esophageal cancer, supporting its high target score. However, GLUT1 as all identified cell membrane proteins presents some degree of expression in healthy tissues, limiting their utilization for precise cancer targeting (Table S4-Supporting Information). Therefore, we evaluated the possibility of obtaining a higher degree of cancer specificity when considering the GLUT1-STn glycoproteoforms. We started by re-addressing the mass spectrometry datasets with Byonic, a dedicated protein assignment software which facilitates the assignment of post-translational modifications in glycopeptides [41]. Only glycopeptide spectra presenting typical GalNAc and sialic acid oxonium ions derived from STn were considered for glycosites assignment, as exemplified by the HCD-MS/MS spectra on Figure 3C. Even though the original proteomics experiment

was not designed for glycopeptides identification, we were able to confirm 18 STn-glycosites GLUT1, highlighted in Figure 3D and Supplementary Figures S1 and S2. These, however, should now be carefully validated in individual tumor samples. Despite the exploratory nature of this approach, we were able to identify the GLUT1-STn glycoproteoforms as a potential ESCC specific molecular signature.

2.4. GLUT1 in ESCC

The suggestion of GLUT1-STn as a specific marker for ESCC prompted a reinvestigation of our retrospective tumor series for this glycoprotein. We found GLUT1 widely expressed across tumor tissues at the cell surface of cancer cells, even though with variable intensity and extension between individual cases (Table 4; Figure 4A). GLUT1 could also be observed in all metastasized lymph nodes. The staining extension exceeded 40% of the tumor area for all studied cases, without a defined histological pattern. However, staining intensities varied from weak to very strong. Tumor-associated immune infiltrates also presented strong GLUT1 membrane expression. In addition, the adjacent but histologically healthy esophageal mucosa exhibited a strong membrane staining restricted to cells of the basal layer of the squamous epithelium (Figure 4A). Although lacking tumor specificity, high GLUT1 levels, characterized by both high extension and high staining intensity, showed a trend association with the presence of metastases and associated with distant recurrence ($p = 0.026$, Table 4), reinforcing observations from other authors linking this protein to metastasis [42–44]. Building on previous insights from glycoproteomics analysis, we were then led to investigate the presence of GLUT1-STn in individual tumors.

Table 4. **GLUT1 expression in EC according to clinicopathological variables.**

GLUT1 expression	Low expression (% total)	High expression (% total)	*p* value
Stage			
I	8 (62)	5 (38)	0.081
II	10 (91)	1 (9)	
III	19 (83)	4 (17)	
IV	0 (0)	1 (100)	
Tumour (pT)			
T1	8 (89)	1 (11)	0.406
T2	8 (62)	5 (38)	
T3	20 (80)	5 (20)	
T4	1 (100)	0 (0)	
Lymph node metastasis (pN)			

N0	15 (71)	6 (29)	
N1	10 (91)	1 (9)	0.575
N2	7 (79)	3 (30)	
N3	5 (83)	1 (17)	
Distant metastasis (M)			
M0	37 (79)	10 (21)	0.064
M1	0 (0)	1 (100)	
Distant recurrence (DR)			
No	27 (87)	4 (13)	**0.026**
Yes	10 (59)	7 (41)	
Keratinization Degree			
Non-keratinized	23 (77)	7 (23)	
Moderately keratinized	2 (100)	0 (0)	0.727
Keratinized	12 (75)	4 (25)	
Differentiation Degree			
Well-differentiated	3 (43)	4 (57)	
Moderately differentiated	20 (77)	6 (23)	
Poorly differentiated	9 (90)	1 (10)	0.081
Missing information	5 (100)	0 (0)	
Extra-tumoral growth			
Present	9 (75)	3 (25)	0.563
Absent	28 (78)	8 (22)	
Lymphatic Permeation			
Present	15 (71)	6 (29)	0.316
Absent	22 (81)	5 (19)	
Vascular Permeation			
Present	15 (71)	6 (29)	0.316
Absent	22 (81)	5 (19)	
Neural Permeation			
Present	10 (77)	3 (23)	0.632
Absent	27 (77)	8 (23)	

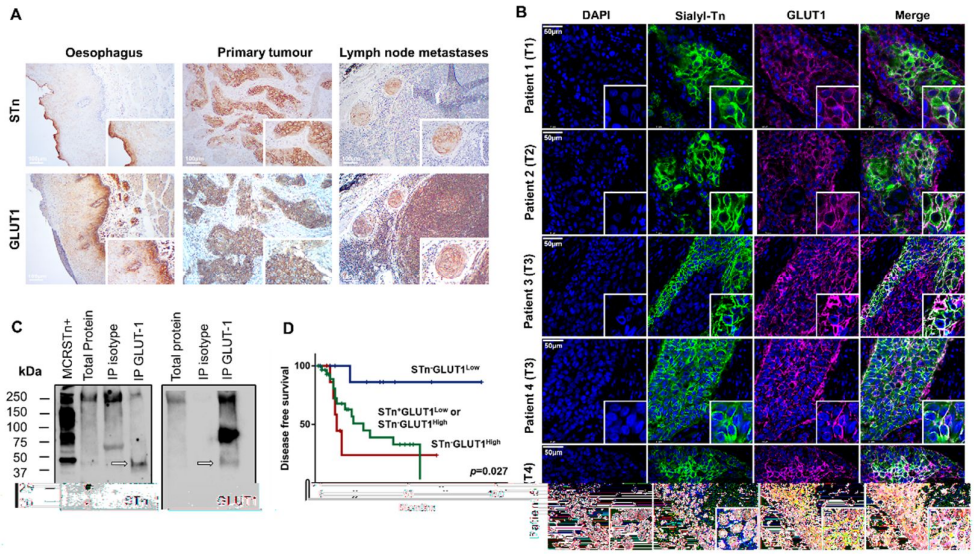


Figure 4. GLUT1-STn is widely expressed in ESCC and correspondent metastases and associates with decreased disease-free survival. (**A**) GLUT1 and STn are co-localized in the same tumor areas in primary tumors, lymph node metastases but not in the healthy esophagus. The immunohistochemistry analysis panels show the co-expression of GLUT1 and STn in the same areas in tumors and lymph node metastases strongly suggesting GLUT1-STn glycoforms. Immunohistochemistry of the healthy esophagus suggests that these epitopes are present in different areas. (**B**) GLUT1 and STn are expressed in the same cancer cells in all studied tumors. Double immunofluorescence for GLUT1 and STn show co-localization in the same cells in all studied tumors, across different stages and histopathological natures of the disease. Nuclei stained in blue; STn in green; GLUT1 in purple; Co-expression (white). (**C**) Western blot confirms the existence of GLUT1-STn glycoproteoforms. GLUT1 was immunoprecipitated from ESCC and blotted for both GLUT1 and STn. Isotype immunoprecipitated proteins were used as negative controls. MCRSTn+ bladder cancer cell line was used as positive control for STn expression. All blots showed bands at 250 kDa and above most likely resulting from unspecific co-precipitation of mucins carrying STn. On the other hand, GLUT1 blots evidence a major band just above 75kDa not present in STn blots. A less intense band at approximately 50 kDa was observed in both GLUT1 and STn blots reinforcing the existence of GLUT1-STn. (**D**) GLUT1-STn overexpressing tumors present reduced decreased survival when compared to tumors showing no STn expression or high STn expression and decreased levels of GLUT1.

A comparative analysis of STn and GLUT1 on consecutive immunohistochemically labeled tissue sections revealed the co-expression of GLUT1 and STn in the same tumor areas as well as in the metastases (Figure 4A). Immunofluorescence double staining for these

antigens in the same tumor section confirmed that both antigens were in the same cancer cells, irrespectively of the stage of the tumor (Figure 4B), strongly supporting the existence of GLUT1-STn, which was later confirmed by western blot (Figure 4C).We started by screening whole protein extracts isolated from FFPE tissues showing high GLUT1 and STn by western blot with no success (Figure 4C), suggesting low abundance of membrane proteins [9]. To overcome this limitation, GLUT1 was immunoprecipitated from tumors showing high GLUT1 and STn co-expression in the same area and immunoblotted for both GLUT1 and STn. We started by noting that all blots presented a strong band above 250 kDa that we believe to be mucins known to carry the STn antigen (Figure 4C). These glycoproteins are abundantly present along the gastrointestinal tract and overexpressed in ESCC and may often co-precipitate in immunoaffinity assays. Besides these bands, GLUT1 immunoprecipitates showed a major band above 75 kDa, mostly likely corresponding to highly glycosylated forms of the glycoprotein and a faint band at approximately 50 kDa that was also observed in the anti-STn blots and most likely belongs to GLUT1-STn glycoproteforms (Figure 4C). Notably, none of these bands were present in isotype negative controls, reinforcing previous assignments. Moreover, the lower abundance of GLUT1-STn band in relation to the most expressed form of GLUT1 was consistent with the focal expression presented by STn in comparison to the diffuse nature of GLUT1 in ESCC. Notably, according to Uniprot (https://www.uniprot.org/ (accessed on 1 January 2021)), GLUT10s canonical form has approximately 54 kDa, which may be significantly increased by post-translational modifications explaining the band at 75 kDa. In fact, bioinformatics predictions showed for this protein two potential N-glycosites (according to NetNGlyc; http://www.cbs.dtu.dk/services/NetNGlyc/ (accessed on 1 January 2021) [45]) and up to 14 predicted O-glycosites, depending on available polypeptide N-acetylgalactosaminyltransferases (according to ISOGlyP; https://isoglyp.utep.edu/ (accessed on 1 January 2021) [46]). Replacement of extended by shorter glycoforms such as the STn antigen may explain the existence of a lower molecular weight GLUT1 band. The fact that this band was found slightly below the estimated for the canonical form of the protein may derived from alterations in electrophoretic mobility, which are common upon changes in glycosylation, namely sialic acids content [47,48]. Furthermore, it should be noted that samples were obtained from FFPE tissues and it is well known that pre-analytical factors (both the fixation and the solubilization protocols) may introduce protein modifications, such as amino acid side chain PTMs or protein backbone hydrolysis, which could result in the alteration of the protein molecular weight and electrophoretic migration [49–52]. These modifications, together with difficulties in predicting the MW shift induced by glycosylation make it difficult to interpret WB singly on the MW predicted from protein amino

acid sequence. Finally, we cannot exclude the expression of shorter GLUT1 proteoforms, as frequently observed for other proteins in cancer [53] and have been predicted by Uniprot. Faced with these preliminary observations, we are engaging in more comprehensive glycoproteogenomics studies to fully characterize the nature of GLUT1 glycoforms in ESSC. Nevertheless, collectively, western blot supports the existence of GLUT1-STn glycoproteoforms in ESCC suggested by immunoassay, as previously hinted by our *in silico* approach. Interestingly, STn and GLUT1 were not co-expressed in the same areas of the healthy esophagus (Figure 4A).

Having validated the presence of GLUT1-STn in ESCC, we then devoted to seeking potential associations with clinical variables. We found that patients with STn negative tumors and expressing low levels of GLUT1 exhibited longer disease-free survival compared to patients with tumors presenting other glycophenotypes ($p = 0.027$, log-rank; Figure 4D). In addition, STn positive tumors with high GLUT1 favored worse patient survival, reinforcing the close link between the GLUT1-STn molecular signatures and cancer aggressiveness. According to the Cox Regression model, the GLUT1-STn overexpression phenotype in ESCC constitutes an independent prognosis factor regarding disease-free survival.

In summary, we have presented strong evidence that the GLUT1-STn glycophenotype is characteristic of more aggressive ESCC, associating with decreased survival, which may be helpful for patient stratification. Moreover, it could not be observed in healthy esophageal tissue, holding high cancer specificity.

2.5. STn and GLUT1 in Health Tissues

To address the cancer-specific dimension of GLUT1-STn, we started by comparing GLUT1 and ST6GalNAc I expressions in a wide variety of healthy human tissues, exploring the Human Protein Atlas (Figure 5A). ST6GalNAc I, the glycosyltransferase responsible by the O-6 sialylation of the Tn antigen, was used as surrogate for STn biosynthesis potential, overcoming the lack of systematized information on O-glycans expression in human tissues. However, we note that ST6GalNAc I detection cannot be regarded as definitive proof for presence of STn, given the complex nature of events governing O-glycosylation. According to Figure 5A, GLUT1 is present in few healthy tissues compared to ST6GalNAc I, namely the placenta, kidney, skin, tonsil and the cerebral cortex. Interestingly, the Human Protein Atlas did not report GLUT1 in the esophagus, which contrasts with our experimental data demonstrating low to moderate expressions in the basal layer of the squamous epithelium. These discrepant observations may result from differences in the specificity of the used antibodies and support more detailed reinvestigation of GLUT1 in healthy tissues by

quantitative high-throughput technologies such as mass spectrometry. On the other hand, ST6GalNAc I was mostly found across the gastrointestinal and respiratory tracts, female and male organs and the skin, in agreement with previous reports for STn [20]. Interestingly, the organs that presented GLUT1 were also positive for ST6GalNAc I. However, a closer look showed significant differentiation at the cell level in the placenta and the cerebral cortex (Figure 5B), as previously observed by us for the esophagus (Figure 4A). On the other hand, GLUT1 and ST6GalNAc I were observed in the same type of cells in the kidney and skin. The subsequent screening of healthy kidney tissue sections by immunohistochemistry for GLUT1 and STn supported the presence of STn in tubular cells of the kidney but not GLUT1 (Figure 5C) as suggested by the Human Protein Atlas, which possibly translates differences in antibody affinities. On the other hand, we could find GLUT1 in the healthy skin but not STn, vowing for the absence of GLUT1-STn glycoforms (Figure 5C). We then expanded this study to a wider array of healthy tissues consisting of thyroid, liver, gallbladder, testis, lung, stomach, pancreas and colon samples. STn presented a restricted distribution in these tissues, mostly circumscribed to the cytoplasm of the parietal and goblet cells of the respiratory and gastrointestinal tracts as well as in sertoli cells of the testicles, thus in agreement with the presence of ST6GalNAc I (Figure 5A). Nevertheless, the intensity of expression was much lower in comparison to esophageal tumors. On the other hand, using this approach GLUT1 expression in healthy tissues was only detectable in the cell membrane of immune cells in stomach, liver, colon, lung and pancreatic tissues that did not express STn but not in the cells of the targeted organs, again in agreement with Figure 5A. To fully access the possibility of GLUT1 and STn expressions in the same types of cells, we also performed double staining immunofluorescence for GLUT1 and STn in the same sections (data not shown because it retrieved negative). Again, we could not find any evidence of GLUT1-STn glycoproteoforms, reinforcing its highly specific cancer-associated nature. Nevertheless, at this stage, the existence of GLUT1-STn glycoproteoforms in the kidney remains a possibility, which warrants confirmation foreseeing targeted therapeutics. Moreover, detailed glycoproteomics studies by mass spectrometry are also required to unequivocally validate these preliminary observations.

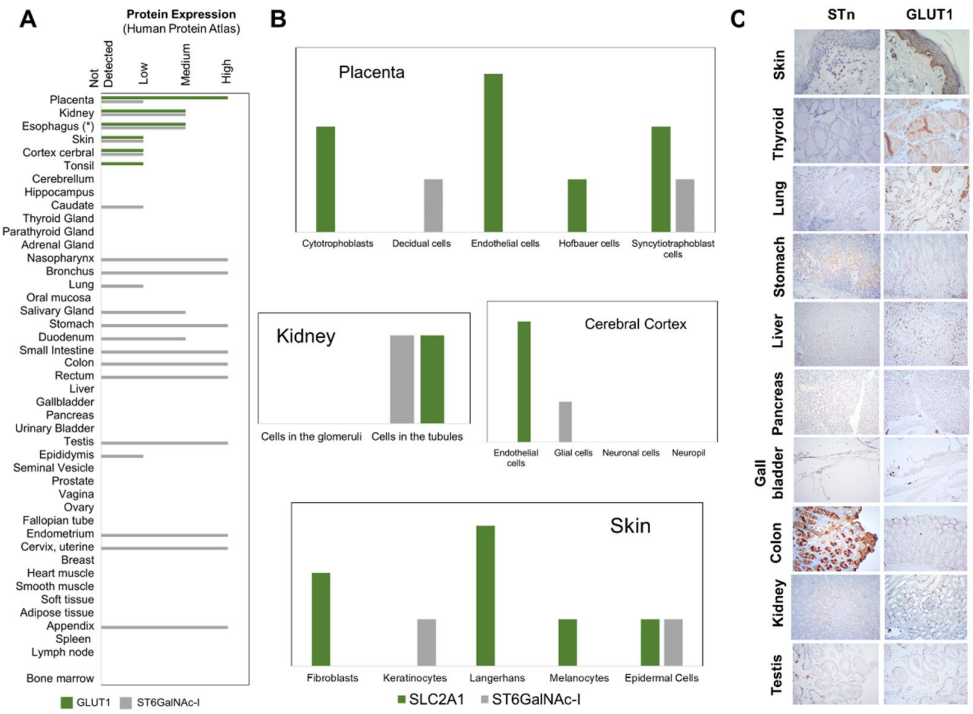


Figure 5. STn and GLUT1 expressions are rarely overlapping in healthy tissues. (**A**) GLUT1 and ST6GalNAc I (a key glycosyltransferase for STn biosynthesis) expressions in healthy human tissues. According to the Human Protein Atlas, GLUT1 and ST6GalNAc I are co-expressed in the placenta, kidney, skin and the cerebral cortex, suggesting potential to express GLUT1-STn. GLUT1 for the esophagus concerned experimental data from the present report, according to Figure 4A. (**B**) GLUT1 and ST6GalNAc I expressions in different cell types of the placenta, kidney, cerebral cortex and skin. Syncytiotraphoblast cells of the placenta, cells in the kidney tubules and epidermal cells of the skin co-express GLUT1 and ST6GalNAc I, supporting capacity to present GLUT1-STn. (**C**) GLUT1 and STn expressions in the skin, thyroid, lung, stomach, liver, pancreas, gallbladder, colon, kidney, testis. GLUT1 was detected at the cell membrane in the immune cells' component of the stomach, liver, colon, lung and pancreas. STn was found mostly in the cytoplasm and, to less extent, the membrane of parietal and goblet cells of the respiratory and digestive tracts. No evidence supporting co-expression in the same areas were found.

3. Material and methods

3.1. Patient Samples and Ethics Statement

Sialyl-Tn and Glucose transporter type 1 (GLUT1) analysis was performed retrospectively in a series of 48 formalin-fixed paraffin-embedded (FFPE) ESCC tissues obtained from Portuguese Institute of Oncology of Porto (IPO-Porto) biobank. Esophageal squamous cell carcinomas were surgically removed from 38 men and 10 women, ranging from 44 to 83 years of age (median 63 11 years), admitted and treated at IPO-Porto between 2010 and 2015 (Table 1). Overall survival (OS) was defined as the period between surgical treatment and patients' death by cancer. Recurrence was defined as the appearance of disease, locally or distant, after the first treatment. Clinicopathological information was obtained from patients' clinical records, upon IPO institutional Ethics Committee approval (202/017 approved on 20 July 2017, addendum 202a/017 approved on 9 November 2017) and after patient's informed consent. A broad library of healthy tissues (skin, thyroid, lung, stomach, liver, pancreas, gallbladder, colon, kidney, testis) was also screened for these markers.

3.2. Isolation and Characterization of Circulating Tumor Cells

The isolation and characterization of the STn antigen in EC-derived circulating tumor cells (CTCs) were performed prospectively in 10 male ESCC patients (age of diagnosis 63 11 years). This corresponds to a group of patients admitted at IPO-Porto and enrolled in the study after informed consent and before initiation of any therapeutic procedure. The clinicopathological nature of the tumors is summarized in Table 2. CTCs were enriched from peripheral blood using a size-based microfluidics device developed at International Iberian Nanotechnology Laboratory (INL)-Braga and described in detail elsewhere [22,54]. Captured cells were stained in situ for CD45 (R&D Systems, Minneapolis, MN, USA), pan-Cytokeratin (pan-CK; Novus Biologicals, Centennial, CO, USA) and STn (B72.3+CC49; Abcam, Cambridge, UK) by immunofluorescence using a Leica DMI6000 FFW microscope (Leica Microsystems, Wetzlar, Germany) equipped with the Las X software (Leica Microsystems, Wetzlar, Germany). This enabled the estimation of STn positive cancer cells (DAPI+/CD45• /pan-CK+ or pan-CK• /STn+) as well as the assessment of hematopoietic contamination (DAPI+/CD45+/pan-CK• /STn•). Spiking experiments were performed using the MCRSTn+ bladder cancer cell line as a positive control of epithelial cancer cells expressing the STn antigen and CKs but lacking CD45 expression [24]. Blood cells from healthy donors were used as negative controls, containing cells positive for CD45 and negative for STn (DAPI+/CD45+/pan-CK• /STn•). In parallel, we incubated microchips

containing isolated CTCs with neuraminidase (NeuAse) from *Clostridium perfringens* (Sigma-Aldrich, St. Louis, MO, USA) prior incubation with primary antibodies, as a negative control for STn. This was done based on a protocol optimized by us [22].

3.3. Cell Lines and Culture Conditions

The Human ESCC cell line Kyse-30 was kindly provided by Prof. Fátima Baltazar (Life and Health Sciences Research Institute from the University of Minho, Braga, Portugal). Upon arrival, cell line authentication by autosomal short tandem repeat (STR) DNA Profiling was performed [55] and the DSMZ STR profile database was used to query STR profiles, confirming its origin. Cells were maintained with 10% heat-inactivated fetal bovine serum (FBS) (Thermo Fisher Scientific, Waltham, MA, USA) and 1% penicillinstreptomycin (10,000 Units/mL penicillin; 10,000 mg/mL streptomycin; Thermo Fischer Scientific,Waltham, MA, USA) supplemented RPMI 1640+GlutaMAXTM-I medium (Thermo Fisher Scientific,Waltham,MA, USA). The cell line was cultured at 37 C in a 5% CO2 humidified atmosphere using a BINDER C-150 incubator (BINDER GmbH, Tuttlingen, Germany).

3.4. Generation of an STn ESCC Cell Line

Kyse-30 cells were plated onto 24-well plates to be 70% confluent at the time of transfection. Human ST6GALNAC1 (hST6GALNAC1, NM_018414.5) knock-in (KI) was achieved by conventional mammalian gene expression vector transfection using jetPRIME® transfection reagent (Polyplus-transfection, Illkirch, France), according to the manufacturer's instructions. In parallel, a mock system containing a 300 bp stuffer ORF in an empty vector was established. The KI system was purified through puromycin selection of positively transfected cells. Cells were subsequently sorted in a BD FACSAria™III sorter for selecting high level STn expressing cells. Cell sorting was preceded by cellular detachment using Versene solution (Thermo Fisher Scientific, Waltham, MA, USA) and staining with mouse anti-TAG72 antibody [B72.3] (ab691; Abcam, Cambridge, UK) at 2 g/106 cells dilution in PBS 2% FBS for 1 h at room temperature. Polyclonal rabbit anti-mouse immunoglobulins/ FITC (F0313; Agilent, Santa Clara, CA, USA) was used as secondary antibody for STn detection at a 1:100 dilution in phosphate buffered saline (PBS) 2% FBS for 15 min at room temperature.

3.5. Proliferation Assays

Cell proliferation was evaluated using a colorimetric immunoassay that measures the incorporation of 5-bromo-20-deoxyuridine (BrdU) during DNA synthesis. Briefly 5 103 cells/well were cultured into 96 well plates and the Cell Proliferation ELISA, BrdU

(colorimetric) Kit (Roche Diagnostics GmbH, Mannheim, Germany) was used for the quantification of cell proliferation, according to the manufacturer's instructions. Measurements of samples absorbances were taken at 450 nm using the iMARKTM microplate reader (Bio-Rad, Hercules, CA, USA). Cell death negative controls composed of 1% Triton-X in complete cell culture medium were used in all experiments. All experiments were performed in triplicate and three replicates were conducted for each independent experiment. The results are presented as the average and standard deviation of these assays.

3.6. Invasion Assays

Corning BioCoat Matrigel Invasion Chambers (Corning, Corning, NY, USA) were used to assess the invasive properties of Kyse-30, Kyse-30 Mock and Kyse-30 ST6GALNAC1 KI cells in vitro. Briefly, 5 104 cells/mL were plated onto invasion inserts conditioned according to the manufacturer's instructions. After 24 h assay, the membranes were washed and mounted with VECTASHIELD® mounting medium with DAPI and cells were counted in a Leica DM2000 microscope (Leica Microsystems,Wetzlar, Germany). Three independent assays were performed and cells were seeded in quintuplicates for each cell line.

3.7. Flow Cytometry

Cells were detached using Versene solution (Thermo Fisher Scientific,Waltham, MA, USA), fixed with 2% paraformaldehyde (PFA; Sigma-Aldrich, St. Louis, MO, USA) and stained with mouse anti-TAG72 antibody [B72.3] (ab691; Abcam, Cambridge, UK) according with conditions previously described in Generation of an STn ESCC cell line topic. Mouse IgG1 [MOPC-21] isotype control (ab18443; Abcam, Cambridge, UK) was included as a negative control. In parallel, 106 cells were digested with 70 mU neuraminidase from Clostridium perfringens (Sigma-Aldrich, St. Louis, MO, USA) in appropriate sodium acetate buffer at 37 C overnight under mild agitation prior to STn staining, as a proper negative control. Data analysis was performed through CXP Software and results represent the standard deviation of three independent experiments performed in a FC500 Beckman Coulter flow cytometer (Beckman Coulter Life Sciences, Indianapolis, IN, USA).

3.8. Glycomics

ESCC cell models O-glycome were characterized using the Cellular O-glycome Reporter/Amplification method [25] and analyzed by nanoLC-ESI-MS/MS, as previously described by us [9]. Glycan structures were assigned based on m/z identification, retention

time, characteristic product ion spectra and previous knowledge about O-glycosylation pathways [9]. Glycans were expressed in terms of relative abundance in comparison to the sum of all individual contributions to the glycome.

3.9. Immunohistochemistry

Three micrometers FFPE esophageal tumor sections were deparaffinized, rehydrated and subjected to antigen retrieval for 20 min with boiling citrate buffer for STn (Vector Laboratories, Burlingame, CA, USA) or 1 mM EDTA pH 9 for GLUT1 analysis. Endogenous peroxidases were inactivated by 3% hydrogen peroxide (Merck KGaA, Darmstadt, Germany) incubation for 5 min, followed by blockage of unspecific links for 5 min with Protein Block (Leica Biosystems,Wetzlar, Germany). Finally, tissue sections were separately incubated with 0.5 g/mL monoclonal antibody (moAb) anti-STn (B72.3+CC49; Abcam, Cambridge, UK) and moAb anti-GLUT1 (EPR3915; Abcam, Cambridge, UK) at the dilution of 1:750. STn and GLUT1 were posteriorly detected using the Novolink Max Polymer DS Kit (Leica Biosystems, Wetzlar, Germany), according to the manufacturer's instructions. Positive and negative controls were tested in parallel, including enzymatic controls with 0.2 U/mL -neuraminidase from Clostridium perfringens (Sigma-Aldrich, St. Louis, MO, USA) overnight at 37 C. Additionally, a broad panel of healthy tissues (skin, thyroid, lung, stomach, liver, pancreas, gallbladder, colon, kidney, testis) was also screened for both glycan and protein to assess the expression pattern in normal conditions. Cancer tissues were considered positive for STn whenever staining was observed, whereas the expression of GLUT1 was determined by multiplying the extension (0–100) against the intensity of expression (1: weak; 2: medium; 3: strong). Tumors showing GLUT1 levels above 200 were considered to overexpress this antigen. Healthy tissues were considered positive for the two antigens whenever staining was observed.

3.10. Double Staining Immunofluorescence Microscopy

STn and GLUT1 were evaluated in the same tumor sections by double staining immunofluorescence microscopy. Briefly, FFPE esophageal tumor sections were deparaffinized, rehydrated and incubated for 20 min with 1 mM EDTA pH 9 to recover STn and GLUT1 epitopes. Unspecific background was blocked with Protein Block (Leica Biosystems, Wetzlar, Germany) for 5 min and both markers (STn and GLUT1) were incubated with primary monoclonal antibodies, as previously described for immunohistochemistry staining. Posteriorly, moAb anti-STn and anti-GLUT1 were detected using an Alexa Fluor 488 goat anti-mouse IgG (Thermo Fisher Scientific,Waltham, MA, USA) and an Alexa Fluor 594 goat anti-rabbit IgG (H+L), respectively, at the dilution of 1:100

for 30 min at room temperature. Nuclear counterstaining was obtained using a 40,6-diamidino-2-phenylindole, dihydrochloride (DAPI; Thermo Fisher Scientific,Waltham, MA, USA) solution for 10 min. Fluorescence images were acquired on a Leica DMI6000 FFW microscope (Leica Microsystems, Wetzlar, Germany) using the Las X software (Leica Microsystems, Wetzlar, Germany).

3.11. Immunoprecipitation and Western Blot

Proteins were extracted from 10 m FFPE ESCC tissues using Qproteome FFPE tissue kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Prior to quantification of total protein extracts using a DC Protein Assay kit (Bio-Rad, Hercules, CA, USA), the elution buffer was exchanged to a Lysis buffer (50 mM Tris, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS). For each immunoprecipitation (IP) reaction, PierceTM Protein G Agarose (Thermo Fisher Scientific,Waltham, MA, USA) was blocked with 1% Bovine Serum Albumin (BSA; Sigma-Aldrich, ST. Louis, MO, USA) for 1 h at 4 C. Then, 100 g whole cell protein extracts were pre-incubated with agarose beads blocked with BSA for 2 h at 4 C to remove proteins showing unspecific binding to the columns. The protein extracts were then recovered and subsequently incubated with 3 g of moAb anti-GLUT1 (EPR3915; Abcam, Cambridge, UK) at 4 C for 2 h. BSA blocked agarose beads were then added to the medium and allowed to incubate at 4 C overnight under gentle stirring. After five washing steps with the buffer provided by the vendor, the IP products were eluted from beads using an acidic IP elution buffer (0.1 M citrate pH 3). IPs controls were conducted in parallel using a rabbit IgG isotype control (02-6102; Thermo Fisher Scientific, Waltham, MA, USA). Finally, the IP products were neutralized, run on 4–20% gradient SDS-PAGE gels (Bio-Rad, Hercules, CA, USA), transferred into nitrocellulose membranes (GE Healthcare Life Sciences, Chicago, IL, USA) and screened for STn and GLUT1 antigens byWestern Blot. Briefly, membranes were blocked with 1% Carbo-Free Blocking Solution (Vector Laboratories, Burlingame, CA, USA) and probed with each moAb, mentioned formerly, for 1 h at room temperature. MoAb anti-STn detection was performed using a peroxidase affiniPure goat anti-mouse IgG (H+L) polyclonal antibody, while moAb anti-GLUT1 was conjugated with a goat anti-rabbit IgG (H+L) HPR antibody for 30 min at room temperature. Finally, the Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences, Chicago, IL, USA) was used as developing reagent. Chemiluminescence detection was performed in a ChemiDoc XRS+ with Image LabTM Software (Bio-Rad, Hercules, CA, USA).

3.12. Bioinformatics for Biomarker Discovery

STn bearing glycoprotein discovery was performed exploiting the only available proteomics dataset for ESCC (project PXD006255; https://www.ebi.ac.uk/pride/archive/projects/PXD006255 (accessed on 1 January 2021)) deposited in the EMBL-EBI proteomics identification database (PRIDE; https://www.ebi.ac.uk/pride/ (accessed on 1 January 2021)) until April 2020. The dataset reports data previously published by Puttamallesh et al. [56] and consists of the nanoLC-MS analysis of a combined protein pool extracted from 10 m FFPE tissue sections from 6 ESCC patients by high resolution tandem mass spectrometry. The dataset comprises nanoLC-HCD-MS/MS data generated on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer for the analysis of tryptic digests after previous fractionation by strong cation exchange (SCX) or reverse phase sulfonate (RPS) chromatography. Data were re-analyzed using the SequestHT search engine with the Percolator algorithm for validation of protein identifications (Proteome Discoverer 2.5; Thermo Fisher Scientific, Waltham, MA, USA). A double search strategy was devised first aiming at identifying relevant cell membrane proteins and secondly trying to map the proteins O-glycosites. Initially, data were searched against the human plasma membrane proteome obtained from the SwissProt database in March 2020. Trypsin was selected as the enzyme and up to 2 missed cleavages were allowed. Precursor ion mass tolerance was set at 10 ppm and 0.02 Da for product ions. Carbamidomethylcysteine was selected as a fixed modification, while oxidation of methionine (+15.9949) was defined as variable modification. High confidence identifications were ordered in relation to their protein scores and the top 250 membrane proteins from the SCX and RPS pre-fractionation were combined for common molecular signatures. The resulting list was re-organized using a Target Score, as previously described in detail by us [9]. The Target Score algorithm was designed to attribute higher scores to proteins overexpressed in cancer, located at the cell membrane and associated with poor prognosis, while penalizing proteins observed at the same cellular location in healthy tissues. As primary source of information for differently expressed proteins, we used the Human Protein Atlas database (https://www.proteinatlas.org/ (accessed on 1 January 2021)) consulted in March 2020 [57]. More details on the algorithm are presented as Supporting Information. Since the Human Protein Atlas at the time of consultation presented no information concerning ESCC, the present study adopted protein expression levels in headand- neck squamous cell carcinoma as a model, based on exposure to the same carcinogens, similar histopathology and molecular traits [33–35]. For the second search, the top 25 Target Score ranked proteins were constituted as a customized database and interrogated for STn-glycosites using the Byonic software v3.10-52 (Protein Metrics, Cupertino, CA, USA) [41]. Briefly, precursor ion mass tolerance was set at 10 ppm and 0.01

Da for product ions. Carbamidomethylcysteine was selected as a fixed modification, while oxidation of methionine (+15.9949 Da) and the presence of the STn antigen (+494.1748 Da) were considered as variable modifications.

3.13. Statistical Analysis

Statistical data analysis was performed with IBMStatistical Package for Social Sciences—SPSS for Windows (version 20.0, IBM, Armonk, NY, USA). Non-parametric Mann-Whitney test was used for proliferation and invasion assays. Chi-square analysis was used to compare categorical variables. Kaplan-Meier curves were used to analyze the influence of the biomarkers expression in the context of time-to-event (death). Multiple Cox regression analysis was used to assess the effect of the studied biomarkers on the time to recurrence and to adjust for potentially confounding effects. Comparison of estimates was done using log-rank test. A 95% significance threshold for the null hypothesis was considered.

4. Conclusions

ESCC remains a challenging disease due to the lack of cancer-specific cell surface biomarkers to aid therapeutic decision and precisely target cancer cells. Herein, we have demonstrated that more aggressive tumors facing worst prognosis overexpress the STn antigen, potentially mirroring profound de-regulation in protein O-glycosylation pathways yet to be fully uncovered for these tumors. High STn expression was associated with increased probability of developing distant metastases, which was supported by the identification of STn in CTCs and reinforced by studies in vitro linking STn to increased cellular invasion. Notably, this is also the first report of STn in esophageal CTCs, which follows similar observations for bladder, colorectal and gastric tumors [21,22]. Future studies must assess its potential clinical value in the context of liquid biopsies, namely for early detection of disseminated disease, prognosis, treatment follow-up and other molecular-based clinical decisions. Nevertheless, we must emphasize the exploratory nature of this study, involving a low number of tumor tissues and blood samples, warranting confirmation in larger patient cohorts. Notably, we have also observed low amounts of STn in histologically normal esophagus and healthy cells of the gastrointestinal and respiratory tracts. These findings prompted an investigation of the ESCC STn-glycoproteome envisaging specific glycosignatures that could overcome limitations associated with possible off-target effects. As proof-of-concept we have applied the Target Score algorithm to pinpoint potentially clinically relevant glycoproteins identified from large proteomics datasets. Faced with absence of information on protein expression for EC, we used as

primary source of information squamous cells head-and-neck tumors, which share similar etiology, clinicopathological and molecular natures. Even though of preliminary nature and strongly depending on the quality of the initial database, this approach led to the identification of GLUT1 as a cancer-associated differentially expressed protein. Building on these findings, we have demonstrated that GLUT1 was modified with STn antigens. We also showed that GLUT1-STn overexpression may hold potential for ESCC patient stratification, enabling the identification of subgroups facing worst prognosis, as previously suggested when the two antigens were evaluated separately. These findings now warrant definitive confirmation in a broader patient series including different clinicopathological natures and detailed clinical histories, which is currently ongoing. Furthermore, we are addressing the functional implications of GLUT1-STn molecular signature for ESCC to help shaping the necessary rationale for biomarker interventions. Foreseeing this objective, we have also conducted an exploratory screening of relevant healthy tissues of distinct systems (reproductive, respiratory, digestive, immune), which suggested that GLUT1-STn glycoforms may be mainly present in cancer. Nevertheless, our studies were based on immunoassays that strongly depend on the specificity of the antibodies used for detection. As such, a more comprehensive interrogation of the human proteome using high throughput mass spectrometry is warranted to support this claim. Finally, our investigation supports previous reports from our group highlighting the clinical relevance of specific protein glycosignatures, namely for improved patient stratification and increased cancer specificity. Yet, we reinforce that this was a proof-of-concept study exploiting a pre-existing proteomics data generated from a pool of patients without a defined clinical characterization. The original study was not initially designed to accommodate the subtilities of glycoproteomics analysis, including prior enrichment of the samples for glycoproteins and mass spectrometry methodologies to support precise glycosites assignment and higher protein coverages. Nevertheless, it was sufficient to provide molecular signatures of interest, setting a rationale that may be generalized to other tumors. Moreover, it strongly encourages a revisitation of ESCC with dedicated glycoproteomics protocols towards true precision medicine settings based on glycobiomarkers.

References

1. Smyth, E.C.; Lagergren, J.; Fitzgerald, R.C.; Lordick, F.; Shah, M.A.; Lagergren, P.; Cunningham, D. Oesophageal cancer. Nat. Rev. Dis. Primers 2017, 3, 17048.
2. Then, E.O.; Lopez, M.; Saleem, S.; Gayam, V.; Sunkara, T.; Culliford, A.; Gaduputi, V. Esophageal Cancer: An Updated Surveillance Epidemiology and End Results Database Analysis. World J. Oncol. 2020, 11, 55–64.
3. Jain, S.; Dhingra, S. Pathology of esophageal cancer and Barrett's esophagus. Ann. Cardiothorac. Surg. 2017, 6, 99–109.
4. Liao, Z.; Cox, J.D.; Komaki, R. Radiochemotherapy of esophageal cancer. J. Thorac. Oncol. 2007, 2, 553–568.
5. Uemura, N.; Kondo, T. Current advances in esophageal cancer proteomics. Biochim. Biophys. Acta 2015, 1854, 687–695.
6. Cai, X.-W.; Yu, W.-W.; Yu, W.; Zhang, Q.; Feng, W.; Liu, M.-N.; Sun, M.-H.; Xiang, J.-Q.; Zhang, Y.-W.; Fu, X.-L. Tissue-based quantitative proteomics to screen and identify the potential biomarkers for early recurrence/metastasis of esophageal squamous cell carcinoma. Cancer Med. 2018, 7, 2504–2517.
7. Uemura, N.; Kondo, T. Current status of proteomics of esophageal carcinoma. Expert Rev. Proteom. 2016, 13, 1029–1040.
8. Fernandes, E.; Sores, J.; Cotton, S.; Peixoto, A.; Ferreira, D.; Freitas, R.; Reis, C.A.; Santos, L.L.; Ferreira, J.A. Esophageal, gastric and colorectal cancers: Looking beyond classical serological biomarkers towards glycoproteomics-assisted precision oncology. Theranostics 2020, 10, 4903–4928. [CrossRef]
9. Fernandes, E.; Freitas, R.; Ferreira, D.; Soares, J.; Azevedo, R.; Gaiteiro, C.; Peixoto, A.; Oliveira, S.; Cotton, S.; Relvas-Santos, M.; et al. Nucleolin-Sle A Glycoforms as E-Selectin Ligands and Potentially Targetable Biomarkers at the Cell Surface of Gastric Cancer Cells. Cancers 2020, 12, 861.
10. Cotton, S.; Azevedo, R.; Gaiteiro, C.; Ferreira, D.; Lima, L.; Peixoto, A.; Fernandes, E.; Neves, M.; Neves, D.; Amaro, T.; et al. Targeted O-glycoproteomics explored increased sialylation and identified MUC16 as a poor prognosis biomarker in advancedstage bladder tumours. Mol. Oncol. 2017, 11, 895–912.
11. Campos, D.; Freitas, D.; Gomes, J.; Magalhães, A.; Steentoft, C.; Gomes, C.; Vester-Christensen, M.B.; Ferreira, J.A.; Afonso, L.P.; Santos, L.L.; et al. Probing the O-glycoproteome of gastric cancer cell lines for biomarker discovery. Mol. Cell Proteom. 2015, 14, 1616–1629.

12. Fernandes, E.; Peixoto, A.; Neves, M.; Afonso, L.P.; Santos, L.L.; Ferreira, J.A. Humoral response against sialyl-Le(a) glycosylated protein species in esophageal cancer: Insights for immunoproteomic studies. Electrophoresis 2015, 36, 2902–2907. [CrossRef]
13. Carrascal, M.A.; Silva, M.; Ferreira, J.A.; Azevedo, R.; Ferreira, D.; Silva, A.M.N.; Ligeiro, D.; Santos, L.L.; Sackstein, R.; Videira, P.A. A functional glycoproteomics approach identifies CD13 as a novel E-selectin ligand in breast cancer. Biochim. Biophys. Acta Gen. Subj. 2018, 1862, 2069–2080.
14. Ferreira, J.A.; Magalhães, A.; Gomes, J.; Peixoto, A.; Gaiteiro, C.; Fernandes, E.; Santos, L.L.; Reis, C.A. Protein glycosylation in gastric and colorectal cancers: Toward cancer detection and targeted therapeutics. Cancer Lett. 2017, 387, 32–45.
15. Santos, S.N.; Junqueira, M.S.; Francisco, G.; Vilanova, M.; Magalhaes, A.; Dias Baruffi, M.; Chammas, R.; Harris, A.L.; Reis, C.A.; Bernardes, E.S. O-glycan sialylation alters galectin-3 subcellular localization and decreases chemotherapy sensitivity in gastric cancer. Oncotarget 2016, 7, 83570–83587.
16. Peixoto, A.; Fernandes, E.; Gaiteiro, C.; Lima, L.; Azevedo, R.; Soares, J.; Cotton, S.; Parreira, B.; Neves, M.; Amaro, T.; et al. Hypoxia enhances the malignant nature of bladder cancer cells and concomitantly antagonizes protein O-glycosylation extension. Oncotarget 2016, 7, 63138–63157.
17. Freitas, D.; Campos, D.; Gomes, J.; Pinto, F.; Macedo, J.A.; Matos, R.; Mereiter, S.; Pinto, M.T.; Polónia, A.; Gartner, F.; et al. Oglycans truncation modulates gastric cancer cell signaling and transcription leading to a more aggressive phenotype. EBioMedicine 2019, 40, 349–362.
18. Carrascal, M.A.; Severino, P.F.; Guadalupe Cabral, M.; Silva, M.; Ferreira, J.A.; Calais, F.; Quinto, H.; Pen, C.; Ligeiro, D.; Santos, L.L.; et al. Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. Mol. Oncol. 2014, 8, 753–765.
19. Van deWall, S.; Santegoets, K.C.M.; van Houtum, E.J.H.; Bull, C.; Adema, G.J. Sialoglycans and Siglecs Can Shape the Tumor Immune Microenvironment. Trends Immunol. 2020, 41, 274–285.
20. Julien, S.; Videira, P.A.; Delannoy, P. Sialyl-tn in cancer: (how) did we miss the target? Biomolecules 2012, 2, 435–466.
21. Lima, L.; Neves, M.; Oliveira, M.I.; Dieguez, L.; Freitas, R.; Azevedo, R.; Gaiteiro, C.; Soares, J.; Ferreira, D.; Peixoto, A.; et al. Sialyl-Tn identifies muscle-invasive bladder cancer basal and luminal subtypes facing decreased survival, being expressed by circulating tumor cells and metastases. Urol. Oncol. 2017, 35, 675.e1–675.e8.

22. Neves, M.; Azevedo, R.; Lima, L.; Oliveira, M.I.; Peixoto, A.; Ferreira, D.; Soares, J.; Fernandes, E.; Gaiteiro, C.; Palmeira, C.; et al. Exploring sialyl-Tn expression in microfluidic-isolated circulating tumour cells: A novel biomarker and an analytical tool for precision oncology applications. New Biotechnol. 2019, 49, 77–87.
23. Li, Y.;Wu, G.; Yang,W.;Wang, X.; Duan, L.; Niu, L.; Zhang, Y.; Liu, J.; Hong, L.; Fan, D. Prognostic value of circulating tumor cells detected with the CellSearch system in esophageal cancer patients: A systematic review and meta-analysis. BMC Cancer 2020, 20, 581.
24. Ferreira, J.A.; Videira, P.A.; Lima, L.; Pereira, S.; Silva, M.; Carrascal, M.; Severino, P.F.; Fernandes, E.; Almeida, A.; Costa, C.; et al. Overexpression of tumour-asociated carbohydrate antigen sialyl-Tn in advanced bladder tumours. Mol. Oncol. 2013, 7, 719–731.
25. Kudelka, M.R.; Antonopoulos, A.; Wang, Y.; Duong, D.M.; Song, X.; Seyfried, N.T.; Dell, A.; Haslam, S.M.; Cummings, R.D.; Ju, T.
Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. Nat. Methods 2016, 13, 81–86.
26. Bernardo, C.; Costa, C.; Amaro, T.; Gonçalves, M.; Lopes, P.; Freitas, R.; Gärtner, F.; Amado, F.; Ferreira, J.A.; Santos, L. Patientderived sialyl-Tn-positive invasive bladder cancer xenografts in nude mice: An exploratory model study. Anticancer Res. 2014, 34, 735–744.
27. Peixoto, A.; Relvas-Santos, M.; Azevedo, R.; Santos, L.L.; Ferreira, J.A. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front. Oncol. 2019, 9, 380.
28. Julien, S.; Krzewinski-Recchi, M.A.; Harduin-Lepers, A.; Gouyer, V.; Huet, G.; Le Bourhis, X.; Delannoy, P. Expression of sialyl-Tn antigen in breast cancer cells transfected with the human CMP-Neu5Ac: GalNAc alpha2,6-sialyltransferase (ST6GalNac I) cDNA. Glycoconj. J. 2001, 18, 883–893.
29. Pinho, S.; Marcos, N.T.; Ferreira, B.; Carvalho, A.S.; Oliveira, M.J.; Santos-Silva, F.; Harduin-Lepers, A.; Reis, C.A. Biological significance of cancer-associated sialyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. Cancer Lett. 2007, 249, 157–170.
30. Clément, M.; Rocher, J.; Loirand, G.; Le Pendu, J. Expression of sialyl-Tn epitopes on beta1 integrin alters epithelial cell phenotype, proliferation and haptotaxis. J. Cell Sci. 2004, 117 Pt 21, 5059–5069.
31. Wang,W.-Y.; Cao, Y.-X.; Zhou, X.;Wei, B.; Zhan, L.; Sun, S.-Y. Stimulative role of ST6GALNAC1 in proliferation, migration and invasion of ovarian cancer stem cells via the Akt signaling pathway. Cancer Cell Int. 2019, 19, 86.

32. Loureiro, L.R.; Sousa, D.P.; Ferreira, D.; Chai,W.; Lima, L.; Pereira, C.; Lopes, C.B.; Correia, V.G.; Silva, L.M.; Li, C.; et al. Novel monoclonal antibody L2A5 specifically targeting sialyl-Tn and short glycans terminated by alpha-2–6 sialic acids. Sci. Rep. 2018, 8, 12196.

33. Sunpaweravong, S.; Bunbanjerdsuk, S.; Pongrujikorn, T.; Naktang, C.; Sunpaweravong, P.; Nitiruangjaras, A.; Dechaphankul, T.; Jinawath, N. Clonal relationship of synchronous head and neck cancer and esophageal cancer assessed by single nucleotide polymorphism-based loss of heterozygosity analysis. BMC Cancer 2019, 19, 1174.

34. Yan, W.; Wistuba, I.I.; Emmert-Buck, M.R.; Erickson, H.S. Squamous Cell Carcinoma—Similarities and Differences among Anatomical Sites. Am. J. Cancer Res. 2011, 1, 275–300.

35. Kim, J.S.; Kim, B.W. Esophageal cancer and head and neck cancer: The earlier, the better. Gut Liver 2015, 9, 131–132.

36. Chiba, I.; Ogawa, K.; Morioka, T.; Shimoji, H.; Sunagawa, N.; Iraha, S.; Nishimaki, T.; Yoshimi, N.; Murayama, S. Clinical significance of GLUT-1 expression in patients with esophageal cancer treated with concurrent chemoradiotherapy. Oncol. Lett. 2011, 2, 21–28.

37. Lin, Z.; Weinberg, J.M.; Malhotra, R.; Merritt, S.E.; Holzman, L.B.; Brosius, F.C. 3rd, GLUT-1 reduces hypoxia-induced apoptosis and JNK pathway activation. Am. J. Physiol. Endocrinol. Metab. 2000, 278, E958–E966.

38. Sawayama, H.; Ishimoto, T.; Watanabe, M.; Yoshida, N.; Baba, Y.; Sugihara, H.; Izumi, D.; Kurashige, J.; Baba, H. High expression of glucose transporter 1 on primary lesions of esophageal squamous cell carcinoma is associated with hematogenous recurrence. Ann. Surg. Oncol. 2014, 21, 1756–1762.39. Sawayama, H.; Ogata, Y.; Ishimoto, T.; Mima, K.; Hiyoshi, Y.; Iwatsuki, M.; Baba, Y.; Miyamoto, Y.; Yoshida, N.; Baba, H. Glucose transporter 1 regulates the proliferation and cisplatin sensitivity of esophageal cancer. Cancer Sci. 2019, 110, 1705–1714.

40. Blayney, J.K.; Cairns, L.; Li, G.; McCabe, N.; Stevenson, L.; Peters, C.J.; Reid, N.B.; Spence, V.J.; Chisambo, C.; McManus, D.; et al. Glucose transporter 1 expression as a marker of prognosis in oesophageal adenocarcinoma. Oncotarget 2018, 9, 18518–18528.

41. Bern, M.; Kil, Y.J.; Becker, C. Byonic: Advanced peptide and protein identification software. Curr. Protoc. Bioinform. 2012, 40, 13.20.1–13.20.14.

42. Younes, M.; Lechago, L.V.; Lechago, J. Overexpression of the human erythrocyte glucose transporter occurs as a late event in human colorectal carcinogenesis and is

associated with an increased incidence of lymph node metastases. Clin. Cancer Res. 1996, 2, 1151–1154.
43. DeWit, M.; Jimenez, C.R.; Carvalho, B.; Belien, J.A.; Delis-van Diemen, P.M.; Mongera, S.; Piersma, S.R.; Vikas, M.; Navani, S.; Pontén, F.; et al. Cell surface proteomics identifies glucose transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. Gut 2012, 61, 855–864.
44. Goos, J.A.; de Cuba, E.M.; Coupé, V.M.; Diosdado, B.; Delis-Van Diemen, P.M.; Karga, C.; Beliën, J.A.; Menke-Van der Houven van Oordt, C.W.; Geldof, A.A.; Meijer, G.A.; et al. Glucose Transporter 1 (SLC2A1) and Vascular Endothelial Growth Factor A (VEGFA) Predict Survival After Resection of Colorectal Cancer Liver Metastasis. Ann. Surg. 2016, 263, 138–145.
45. Gupta, R.; Jung, E.; Brunak, S. Prediction of N-glycosylation sites in human proteins. 2004, In preparation. 2004. In preparation.
46. Mohl, J.E.; Gerken, T.A.; Leung, M.Y. ISOGlyP: De novo prediction of isoform specific mucin-type O-glycosylation. Glycobiology 2020, cwaa067.
47. Gahmberg, C.G.; Andersson, L.C. Role of sialic acid in the mobility of membrane proteins containing O-linked oligosaccharides on polyacrylamide gel electrophoresis in sodium dodecyl sulfate. Eur. J. Biochem. 1982, 122, 581–586.
48. Liu, C.; Erickson, A.K.; Henning, D.R. Distribution and carbohydrate structures of high molecular weight glycoproteins, MUC1 and MUCX, in bovine milk. J. Dairy Sci. 2005, 88, 4288–4294.
49. Scicchitano, M.S.; Dalmas, D.A.; Boyce, R.W.; Thomas, H.C.; Frazier, K.S. Protein extraction of formalin-fixed, paraffin-embedded tissue enables robust proteomic profiles by mass spectrometry. J. Histochem. Cytochem. 2009, 57, 849–860.
50. Thompson, S.M.; Craven, R.A.; Nirmalan, N.J.; Harnden, P.; Selby, P.J.; Banks, R.E. Impact of pre-analytical factors on the proteomic analysis of formalin-fixed paraffin-embedded tissue. Proteom. Clin. Appl. 2013, 7, 241–251.
51. Shi, S.R.; Taylor, C.R.; Fowler, C.B.; Mason, J.T. Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies. Proteom. Clin. Appl. 2013, 7, 264–272.
52. Zhang, Y.; Muller, M.; Xu, B.; Yoshida, Y.; Horlacher, O.; Nikitin, F.; Garessus, S.; Magdeldin, S.; Kinoshita, N.; Fujinaka, H.; et al. Unrestricted modification search reveals lysine methylation as major modification induced by tissue formalin fixation and paraffin embedding. Proteomics 2015, 15, 2568–2579.

53. Azevedo, R.; Gaiteiro, C.; Peixoto, A.; Relvas-Santos, M.; Lima, L.; Santos, L.L.; Ferreira, J.A. CD44 glycoprotein in cancer: A molecular conundrum hampering clinical applications. Clin. Proteom. 2018, 15, 22.

54. Ribeiro-Samy, S.; Oliveira, M.I.; Pereira-Veiga, T.; Muinelo-Romay, L.; Carvalho, S.; Gaspar, J.; Freitas, P.P.; López-López, R.; Costa, C.; Diéguez, L. Fast and efficient microfluidic cell filter for isolation of circulating tumor cells from unprocessed whole blood of colorectal cancer patients. Sci. Rep. 2019, 9, 8032.

55. Dirks, W.G.; MacLeod, R.A.; Nakamura, Y.; Kohara, A.; Reid, Y.; Milch, H.; Drexler, H.G.; Mizusawa, H. Cell line crosscontamination
initiative: An interactive reference database of STR profiles covering common cancer cell lines. Int. J. Cancer 2010, 126, 303–304.

56. Puttamallesh, V.N.; Patel, K.; Gowda, H. Rapid Processing of Archival Tissue Samples for Proteomic Analysis Using Pressure- Cyclin Technology. J. Proteins Proteom. 2017, 8, 127.

57. Uhlen, M.; Fagerberg, L.; Hallstrom, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. Science 2015, 347, 1260419.

Supporting Information

Sofia Cotton[1,2,4,5], Dylan Ferreira[1,2,4,5], Janine Soares[1,2,3], Andreia Peixoto[1,4,5], Marta Relvas-Santos[1,2, 4,5,6], Rita Azevedo[7], Lorena Dieguez[8], Carlos Palmeira[1,9,10], Luís Lima[1], André M.N. Silva[6], Lúcio Lara Santos[1,2,10,11,12], José Alexandre Ferreira[1,2,12]

[1]Experimental Pathology and Therapeutics Group, IPO Porto Research Center (CI-IPOP), Portuguese Oncology Institute (IPO Porto), 4200-072 Porto, Portugal; [2]Department of Pathology and Molecular Immunology, Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, 4050-013 Porto, Portugal; [3]QOPNA/LAQV, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal; [4]Institute for Research and Innovation in Health (i3S), University of Porto, 4200-135 Porto, Portugal; [5]Institute for Biomedical Engineering (INEB), Porto, Portugal, 4200-135 Porto, Portugal; [6]REQUIMTE-LAQV, Department of Chemistry and Biochemistry, Faculty of Sciences of the University of Porto, 4169-007 Porto, Portugal; [7]Institute of Biomedicine, University of Turku, FI-20014 Turun yliopisto, Finland; [8]International Iberian Nanotechnology Laboratory, Department of Life Sciences, 4715-330 Braga, Portugal; [9]Departament of Immunology, Portuguese Institute of Oncology of Porto, 4200-072 Porto, Portugal; [10]Health School of Fernando Pessoa University, 4249-004 Porto, Portugal; [11]Department of Surgical Oncology, Portuguese Institute of Oncology, 4200-072 Porto, Portugal; [12]Porto Comprehensive Cancer Center (P.ccc), 4200-072 Porto, Portugal

Corresponding author:

José Alexandre Ferreira (jose.a.ferreira@ipoporto.min-saude.pt)

Experimental Pathology and Therapeutics Group, Research Centre, Portuguese Oncology Institute of Porto, R. Dr. António Bernardino de Almeida 62, 4200-072 Porto, Portugal; Tel. +351 225084000 (ext. 5111).

Running head: Targetable glycoproteins in esophageal cancer

Keywords: cancer biomarkers; circulating tumours cells; esophageal cancer; glycomics; glycoproteomics; bioinformatics

Supporting Material and Methods

Proteomics data generates a high number of protein identifications, warranting tools that allow to easily pinpoint biomarkers of potential clinical interest. The Target Score algorithm was developed to provide an easy and comprehensive navigation throughout pre-existing molecular databases, enabling to extract information that maybe used to rank identified biomarkers in relation to prognosis and potential targetability. It was designed to attribute higher scores to glycoproteins overexpressed in cancer, located at the cell membrane, and associated with poor prognosis, while penalizing glycoprotein remaining at the same subcellular location than previously observed in healthy tissues. As primary source of information for this study, we used the Protein Atlas database (https://www.proteinatlas.org/) consulted in March 2020. Nevertheless, it sets a roadmap that maybe used to navigate through other databases containing information on protein expression in cancer and healthy tissues, their subcellular location and cancer prognosis. Accordingly, the score system results from the sum of the following nine variables: i) location in healthy cells (membrane: 0 points; other subcellular locations: 1 point); ii) location in cancer cells (membrane: 1 point; cytoplasm and/or other subcellular locations: 0 points); iii) expression in healthy esophageal epithelium at the plasma membrane (negative: 3 points; low: 2 points; moderate: 1 point; high: 0 points); iv) expression in EC (negative: 0 points; low: 1 point; moderate: 2 points; high: 3 points); v) prognosis value in cancer (high expression is not prognostic and/or associates with favorable prognosis: 0 points; high expression associates with poor prognosis: 1 point); vi) expression in lymphoid tissues at the plasma membrane (not expressed: 1 point; expressed: 0 points); vii) expression in gametes (not expressed: 1 point; expressed: 0 points); viii) index of expression in healthy tissues at the plasma membrane (varies from 3 points (no expression) to 0 (high expression)).

Figure S1. GLUT1-STn glycopeptides (green) and glycosites (red) identified in ESCC using the RPX method. Assignments were made using the Byonic software.

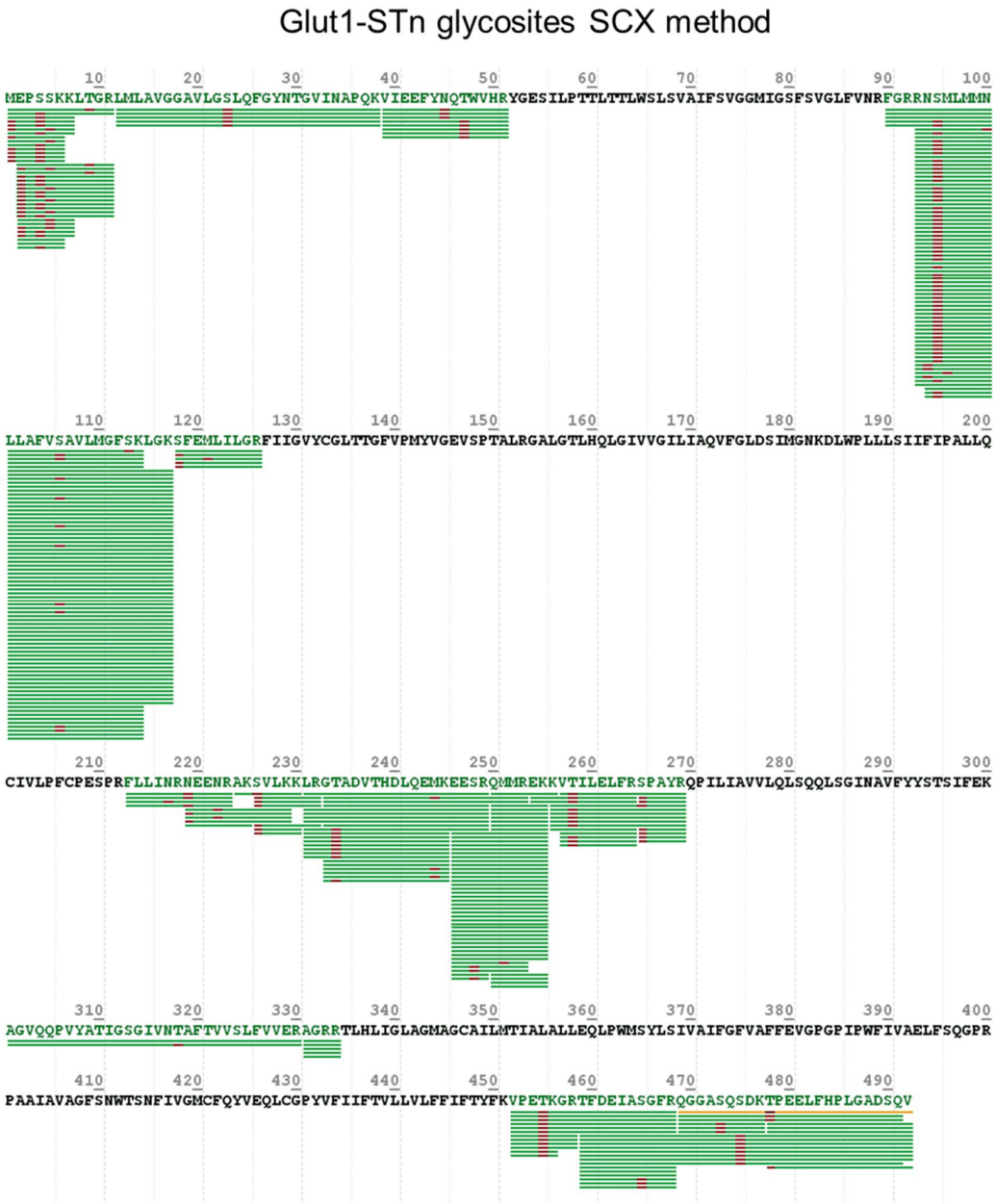


Figure S2. GLUT1-STn glycopeptides (green) and glycosites (red) identified in ESCC using the RPX method. Assignments were made using the Byonic software.

CHAPTER II

GLYCOBIOMARKERS DISCOVERY IN COLORECTAL CANCER

E-selectin Affinity Glycoproteomics Reveals Neuroendocrine Proteins and the Secretin Receptor as a Poor Prognosis Signature in Colorectal Cancer

Sofia Cotton[1,2], Dylan Ferreira [1,2,3], Marta Relvas-Santos [1,2,3,4], Andreia Brandão [1], Luís Pedro Afonso [1,5], Andreia Miranda [1,2,3], Eduardo Ferreira 1, Beatriz Santos [1,2], Martina Gonçalves [1,2], Paula Lopes [5], Lúcio Lara Santos [1,6,7], André M N Silva [2,4], José Alexandre Ferreira [1,2]

1.Research Center of IPO-Porto (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO-Porto) / Porto Comprehensive Cancer Center (P.ccc) Raquel Seruca, Porto, Portugal;
2.Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal;
3.i3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal;
4.LAQV-REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Porto, Portugal;
5.Pathology Department, Portuguese Oncology Institute of Porto, Porto, Portugal;
6.Health School of University Fernando Pessoa, Porto, Portugal;
7.Department of Surgical Oncology, Portuguese Oncology Institute of Porto, Porto, Portugal.

Corresponding author: José Alexandre Ferreira
Experimental Pathology and Therapeutics Group, Research Centre, Portuguese Oncology Institute of Porto, R. Dr. António Bernardino de Almeida 4200-072 Porto, Portugal; Tel. +351 225084000 (ext. 5111).
Email: jose.a.ferreira@ipoporto.min-saude.pt

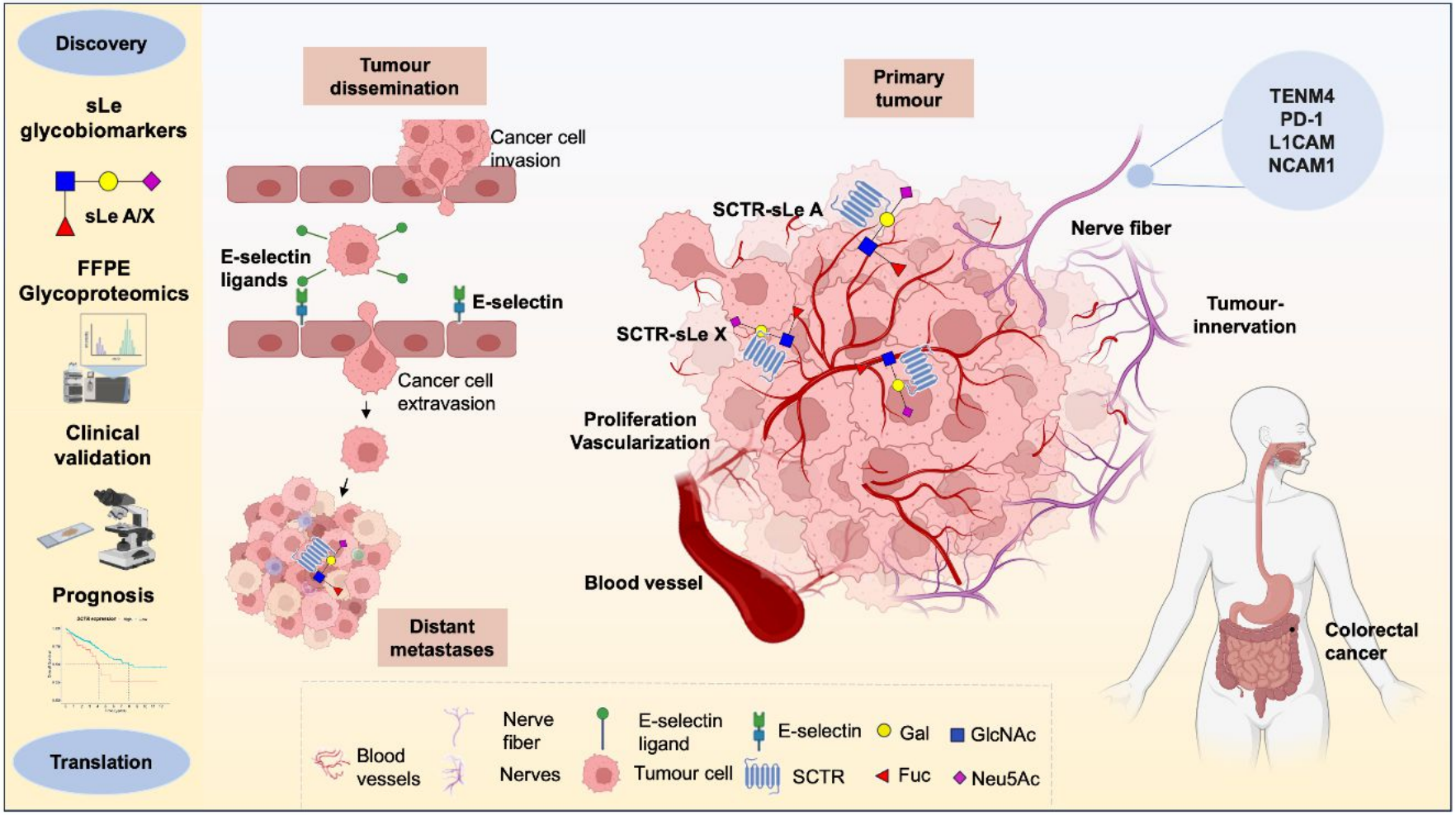
This article has been accepted for publication in Molecular Oncology.

Abstract

Colorectal cancer (CRC) cells express glycoproteins with sialylated Lewis antigens (sLe), crucial for metastasis via E-selectin binding. However, these glycoepitopes lack cancer specificity, and E-selectin-targeted glycoproteins are unknown. Here we established a framework for identifying metastasis-linked glycoproteoforms for precision oncology. We found that over 70% of colorectal tumours overexpressed sLeA/X, yet without associations with metastasis or survival. This prompted deeper exploration on the sialoglycoproteome of metastatic tumours. Nearly 600 glycoproteins were identified using E-selectin affinity enrichment and mass spectrometry, significantly broadening our understanding of potential metastasis-related glycoproteins. This glycoproteome was linked to cell adhesion, active transport of molecules and ions across the plasma membrane, oncogenic pathways, angiogenesis, and, unexpectedly, neuroendocrine functions. Employing an in-house algorithm for targetability prioritization, the less studied secretin receptor (SCTR) emerged as a top-ranked glycoprotein. The screening of tumours confirmed SCTR's association with poor prognosis and metastasis. *N*-glycosylation carrying sLe antigens added cancer specificity to SCTR. Prognostic links were reinforced by TCGA-based investigations. In summary, SCTR, a relatively unknown CRC glycoprotein, holds potential as a biomarker of poor prognosis and as a E-selectin ligand, suggesting an unforeseen role in metastasis warranting confirmation. Future investigations should also focus on this glycoproteoforms' biological foreseeing clinical applications.

Keywords: colorectal cancer; cancer glycoproteome; secretin receptor; E-selectin; metastasis

Graphical Abstract



Introduction

Colorectal cancer (CRC) remains a significant health burden, ranking high in both incidence and mortality rates (1). The challenges associated with CRC are amplified by late-stage diagnoses and the limited efficacy of existing therapeutics against advanced disease (2). To address these limitations, there is a crucial need to identify molecular signatures associated with metastasis for precise cancer targeting.

Recent progress in glycomics has significantly advanced our understanding about the alterations in the glycocalyx accompanying CRC development and progression (3-5). Utilizing high-throughput mass spectrometry, researchers have comprehensively characterized primary tumours, metastatic sites, and healthy colon mucosa from diverse patient populations (5, 6). These investigations have unveiled key alterations, particularly the expression of (sialyl-)LewisA/X antigens (sLeA/X) that cap extended *N*- as well as *O*-glycosidic chains on glycoproteins (5). Traditionally linked to poor prognosis, these antigens play a pivotal role in mediating metastasis by facilitating cancer cell adhesion to endothelial cells, intravasation into the bloodstream, and homing to distant sites (3). This makes sLe antigens attractive targets for controlling metastasis. Glycomics analyses have identified a substantial presence of glycoproteins in CRC cases featuring core 2 *O*-glycans with terminal sLe antigens, a characteristic not found in normal colon tissue (5). This pattern is notably pronounced in mucinous adenocarcinomas (T2 and T3), commonly classified as MSI tumours (7). However, it's worth noting that sLe antigens may also be present in core 3 structures in normal colon mucosa, albeit with a downregulation in cancer (5). In summary, these findings provide a compelling molecular basis for targeting sLe glycoepitopes in CRC. However, they also emphasize the imperative for a comprehensive characterization of the sLe-glycoproteome linked to metastasis, envisioning the essential cancer-specificity required for the development of targeted therapeutics.

Building upon these findings, the present study focuses on creating a framework to examine the CRC glycoproteome, aiming to identify molecular features linked to metastasis. This involves a combination of glycomics and E-selectin affinity glycoproteomics to identify sLe-glycoproteoforms in metastatic tumours. Additionally, we employed an in-house developed algorithm that draws from a well-established protein expression data repository to pinpoint unforeseen and potentially clinically relevant and cancer-specific glycosignatures (8-10). These glycosignatures have been further validated in clinical samples and their cancer specificity confirmed by screening of healthy tissues. We anticipated that this may constitute a relevant milestone towards a comprehensive understanding of metastases signatures, paving the way for future advancements in patient care.

Material and Methods

Patient Sample Set and Healthy Human Tissues

SLeA, sLeX, and SCTR were retrospectively analysed in 35 formalin-fixed paraffin-embedded (FFPE) advanced CRC tissues obtained from the IPO-Porto biobank, reflecting the incidence of the disease. The patient cohort included 15 female and 20 male patients, aged between 28 and 76 years (median age: 61 years), who underwent surgical resection of colorectal carcinomas at IPO-Porto from 2005 to 2012. The expression of these markers was also evaluated in 6 lymph node, 7 hepatic, and 3 peritoneal metastases, which are available for some of these primary tumours. All clinicopathological information used to assess the clinical relevance of SCTR, sLeA, and sLeX expressions is summarized in **Table S1**. Clinicopathological data were extracted from the patients' medical records. This study received approval from the IPO institutional Ethics Committee (Approval No. CES 64/2017, dated 16 March 2017) and was conducted following the acquisition of informed consent from all participating patients. Overall survival (OS) was calculated from the date of surgical intervention to the date of death attributable to cancer. Recurrence was defined as the manifestation of the disease, either locally or at a distant site, following the initial treatment. A series of healthy tissues, including appendix, colon, lung, liver, pancreas, stomach, skin, testis, and thyroid were also screened for these sialoglycans and the SCTR to assess cancer specificity.

Immunohistochemistry

Three-micrometre sections from FFPE colorectal tumour samples were screened for sLeA, sLeX, and SCTR expression. Briefly, tissue sections were deparaffinized, hydrated, and subjected to antigen retrieval using a low pH citrate buffer (*Vector Laboratories*). Subsequently, the sections were treated with a 3% hydrogen peroxide solution (Leica) to inhibit endogenous peroxidase activity. To minimize non-specific binding, the sections were exposed to Protein Block (Leica) before incubation with monoclonal antibody anti-CA19.9 (1:100; ab116024; Abcam), monoclonal antibody anti-sLeX (1:100; 368102; BioLegend), and polyclonal antibody anti-Secretin Receptor (1:100; ab224236; Abcam) for the determination of sLeA, sLeX, and SCTR, respectively. Antibody binding to the tissue sections was assessed using the Novolink Max Polymer DS Kit (Leica), following the manufacturer's protocol. A double-blinded analysis was carried out by independent observers and subsequently validated by an experienced pathologist. Immunohistochemical analysis included the assessment of expression patterns, encompassing extension (percentage of positive tumour cells) and intensity (degree of chromogenic deposit), as well as the subcellular location of the staining. Quality assurance

measures involved parallel testing of both positive and negative controls. Enzymatic controls utilizing α-neuraminidase from *Clostridium perfringens* (Sigma-Aldrich) were employed to ensure the specificity and accuracy of the immunohistochemical results for the two sialoglycans (sLeA and sLeX). CRC tissues used in glycoproteomics studies were further screened for the vascular marker CD34 (1:100; M7165; Dako) and two neuroendocrine markers (Synaptophysin and Chromogranin A – 1:100, M7315, Dako; 1:1000, M0869, Dako, respectively).

Glycomics

N- and *O*-glycomics were performed on metastatic colorectal tumours to disclose the nature of the glycans terminated with sLe epitopes, as previously described (11). Briefly, FFPE tissues were deparaffinized, rehydrated, and subjected to heat-induced antigen retrieval using a citrate-based solution (Vector Laboratories). Then, proteins were denatured and reduced by incubation with 150 µl denaturation mix (145 µl 8 M GuHCl and 5 µl 200 mM DTT) at 60 ºC for 30 min. *N*-glycan release was achieved after digestion with PNGase F (1 U/10 µg protein at 37 ºC overnight; Promega). Released *N*-glycans were hydrolysed with 25 µl of 100 mM ammonium acetate at pH 5 for 1 h at room temperature (RT), removing the glycosylamine form of the non-reducing terminus, and subsequently reduced with 20 µl of 1 M $NaBH_4$ in 50 mM KOH for 3 h at 50 ºC. The reaction was quenched by adding glacial acetic acid. *O*-glycans were released by reductive *β*-elimination (1 M $NaBH_4$ in 50 mM KOH overnight at 50 ºC). Finally, reduced *N*- and *O*-glycans samples were desalted using a cation exchange resin (AG 50W-X8; Bio-Rad). All glycans were permethylated and analysed by reverse phase nanoLC-ESI-MS/MS as previously described by us (9), using a 3000 Ultimate nano-LC coupled to a QExactive mass spectrometer (Thermo Fisher Scientific). Glycan structures were identified considering previous knowledge on glycosylation, chromatography retention times, *m/z* identification, and corresponding product ion spectra. Manual annotation was supported by GlycoWorkbench tool (12).

Glycoproteomics and data curation

Proteins were extracted from three metastatic CRC FFPE tissues using the QProteome FFPE tissue kit (QIAGEN) according to the manufacturer's instructions and pooled to a total of 700 mg of protein. Samples were then enriched for glycoproteins carrying sLe antigens by E-selectin affinity chromatography, as previously described by our group (13), using a recombinant mouse E-selectin Fc Ig chimera protein (575-ES-100; R&D Systems, Minneapolis, MN, USA). Negative controls included immunoprecipitations with an IgG1 isotype control (LTI-02-6102; Thermo Fisher

Scientific) and with the E-selectin chimera in the absence of Ca^{2+}. After electrophoresis, glycoproteins were reduced and digested with chymotrypsin, following the protocol described in Fernandes *et al.* (9). The mass spectrometry analysis was performed by nanoLC-MS/MS using an Ultimate 3000 RSLCnano system coupled to a QExactive mass spectrometer (Thermo Fisher Scientific). Eluent A was aqueous formic acid (0.2%), while eluent B was formic acid (0.2%) in acetonitrile. Briefly, samples were injected into a trapping column (C18 PepMap 100, 5 µm particle size) and washed with an isocratic flux of 98% eluent A and 2% eluent B at a flow rate of 10 µL/min. After 3 min, the flux was redirected to the analytical column (EASY-Spray C18 PepMap, 100 Å, 150 mm x 75 µm I.D and 3 µm particle size) at a flow rate of 0.25 µL/min. The column temperature was set at 35 °C. Glycopeptide separation occurred using a linear gradient of 10–40% eluent B over 50 min. Column wash and re-equilibration were warranted before the following injection. The mass spectrometer was operated in the positive ion mode, with an *m/z* range from 300 to 2000, with a spray voltage of 1.9 kV and a transfer capillary temperature of 275 °C. Q-Exactive HF settings were full scan resolution 120000, fragment scan resolution 15000, fragment scan fixed first mass at 110 *m/z*, normalized collision energy 30%, isolation window 4.0 *m/z*. Mass spectrometry data was processed using the SequestHT search engine and the Percolator algorithm (Proteome Discoverer 3.0, Thermo Fisher Scientific) to validate protein identifications. Data were searched against the human membrane proteome from the SwissProt database. Chymotrypsin was selected as the digestion enzyme, considering up to 2 missed cleavage sites, a precursor ion mass tolerance of 10 ppm, and a product ion tolerance of 0.02 Da. Fixed and variable modifications included carbamidomethylcysteine (+57.021 Da) and oxidation of methionine (+15.995 Da), respectively. The following *N*-glycans were also considered as variable modifications, building on glycome analysis: H5N4F1S1 (+2059.735 Da; asparagine); H5N4F1S2 (+2350.830 Da; asparagine). Alternatively, the following *O*-glycans were considered: Hex(2)HexNAc(2)NeuAc(1)dHex(1) (+1167.418 Da), Hex(2)HexNAc(2)NeuAc(2)dHex(1) (+1458.513 Da), and Hex(3)HexNAc(3)NeuAc(2)dHex(2) / +1969.703 Da) on serine and threonine. The final glycopeptide list included protein species solely detected in E-selectin pulldowns, identified with high confidence as well as with low confidence but presenting manually validated glycopeptides. Glycoproteins molecular and biological functions and subcellular localization were assessed by GO term analysis using ClueGO version 2.5.10 for Cytoscape version 3.10.1 (14-16).

Target Prioritization

The in-house Target Score algorithm, building on the Human Protein Atlas database (proteinatlas.org) (17), was employed to prioritize glycoproteins for precise cancer targeting, as previously described by us (8).

***In situ* proximity ligation assay (PLA)**

SCTR-sLeA/X glycoproteoforms were indirectly assessed in CRC tissues by PLA employing Duolink *in situ* Detection Reagents Red (Sigma-Aldrich). After FFPE tissue deparaffinization and antigen retrieval with acid and heat treatment, the tissue sections were incubated overnight at 4 °C with primary antibodies: anti-SCTR (ab224236; Abcam) and anti-CA19-9 (ab116024; Abcam) or anti-sLeX (368102; BioLegend) in a humidity-controlled chamber. The tissues were subsequently incubated with anti-rabbit MINUS and anti-mouse PLUS PLA probes at 37 °C for 1 h. Subsequent stages included a 30 min ligation process at 37 °C and a 120 min amplification phase 37 ºC to induce the formation of fluorescent rolling dots. Tissue sections were then stained with 4',6-diamidino-2-phenylindole (DAPI) for 10 min at RT and analysed by fluorescence microscopy. The PLA results underwent evaluation by two independent observers and were further validated by a seasoned pathologist. Fluorescence images were acquired on a Leica DMI6000 FFW microscope using Las X software (Leica).

Immunoprecipitation (IP) and Western Blot

The SCTR was immunoprecipitated from CRC protein extracts on lysis buffer (50mM Tris, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) using the PierceTM Protein G Agarose kit (Thermo Scientific). Briefly, Protein G agarose beads were pre-blocked with 1% Bovine Serum Albumin (BSA; Sigma-Aldrich) for 1 h at 4 °C and protein extracts were precleared using the BSA-blocked agarose beads for 2 h at 4 °C to reduce unspecific binding. Then, the supernatant was incubated with the SCTR monoclonal antibody (5mg; sc-166112; Santa Cruz) at 4 °C for 2 h, followed by overnight incubation with freshly BSA-blocked agarose beads at 4 °C. After washing, the protein-antibody complexes were eluted with loading buffer at 95 ºC. The IP products were resolved on a 4–20% gradient SDS-PAGE gel (BioRad) followed by protein blotting as described by Fernandes et al. (9), using anti-SCTR (1:100, 1 h at RT; ab224236, Abcam) and anti-CA19.9 (1:100, 1h at RT; ab116024; Abcam). Chemiluminescence was captured and analysed

using a ChemiDoc XRS+ system with Image LabTM Software (Bio-Rad).

Flow Cytometry

Circulating leukocytes were screened for sialylated Lewis antigens and SCTR by flow cytometry. Peripheral blood samples were obtained from healthy donors at Immunohemotherapy Department of Centro Hospitalar São João (CHSJ), Porto, Portugal. Procedures were approved by the Hospital ethical committee (Protocol reference 260/11). Whole blood, after red blood cell lysis with RBC lysis buffer 1x (420301, BioLegend), was incubated with primary antibodies: rabbit anti-sLeA (CA19-9, NBP2-53220, Novus Biologicals), mouse anti-sLeX (CD15s, 551344 BD), and anti-secretin receptor (sc-166112, Santa Cruz Biotechnology). Subsequently, leucocytes were incubated 30 min with anti-rabbit Alexa 647 and anti-mouse Alexa 488. Further incubation with anti-CD45-PE/Cy7 (368532, BioLegend) was performed. After additional washes, cells were acquired on a NAVIOS Flow Cytometer (Beckman Coulter). The FACS data was analysed using the Kaluza C software (version 1.2) to discriminate cell populations, determine the mean fluorescence intensity (MFI) and the percentage of positive cells.

TCGA dataset analysis

Gene expression data of 623 colorectal (COADREAD dataset) tumour tissues and 51 normal adjacent tissues from the Genomic Data Common (GDC) database, were downloaded from the UCSC Xena database (http://xena.ucsc.edu/). Additionally, clinicopathological characteristics, including age, gender, tumour stage, and histological type, as well as survival information, were also obtained. After excluding patients with missing clinical data, a final dataset comprising 598 tumour samples and 51 histologically normal adjacent tissue samples was included in the study. The clinical data for COADREAD patients are presented in **Table S2**. The prognosis value of glycosyltransferases involved in sialyl-Lewis epitopes, SCTR and secretin (SCT) in CRC was performed by univariate, and subsequent multivariate Cox regression. Kaplan–Meier analysis and log-rank test were used to compare OS and progression-free survival (PFS) curves, using the “survival” and “survminer” R packages, respectively. Optimal cut-off value for each gene expression was determined using the surv_cutpoint function, for both OS and PFS outcomes. Shapiro–Wilk normality test was used to determine variable normality, and Wilcoxon rank test was performed to assess the difference between tumour stages and sample tissue types. Correlation analysis was performed using Spearman’s correlation and visualized by correlograms using the “corrplot” R package. P-value < 0.05 was considered as a

statistically significant difference. All statistical analysis were performed using R software (version 4.2.1).

Statistical analysis

Mann-Whitney tests were performed to compare the expressions of sLeA, sLeX, and SCTR in non-metastatic vs metastatic tumours from IPO-Porto patients, after Shapiro–Wilk normality test. A significance threshold of 95% for the null hypothesis was applied.

Results

This work builds on a well-established notion that sLe antigens are key mediators of metastasis by showing high affinity to E-selectin. It highlights the abundant expression of these glycans in CRC and proposes a roadmap for identifying cancer-specific glycoproteoforms, with implications for targeted therapeutics in the future. Moreover, the study aims to establish a foundation for identifying potentially clinically relevant molecular signatures for precision medicine, utilizing glycosylation as a starting point.

sLe Antigens in CRC and Healthy Tissues

We first devoted to characterizing the expression pattern of sLeA and sLeX in advanced CRC, building on a patient series of 35 tumours comprehending more advanced stages of the disease (stage III/IV). Over 75% of the tumours expressed both sLeA and sLeX antigens, in general in approximately 30% of the tumour area, in the same regions and without a defined expression pattern (**Figure 1A**). Notably, the percentage of sLe glycoepitope-expressing cases was higher in T4 compared to T3 tumours (**Figure. S1A**). In addition to screening primary tumours, we also examined lymph nodes and distant metastases in 10 cases where metastatic lesions were available. We found a slightly higher number of tumours expressing sLeX (85% sLeX *vs* 76% sLeA) and that this antigen was present in 67% of lymph nodes (4/6) and 80% of distant metastases (8/10), irrespective of their origin (peritoneal or hepatic; **Figure S1B**). Furthermore, sLeA showed expression in over 80% of both lymph node and distant metastases. Also, 80% of the metastases mimicked the expression of sLe antigens found in the primary tumour, reinforcing the close link between these sialoglycans and metastasis. However, no associations were found between the expression of these antigens, alone or in combination, with stage, prognosis, or presence of metastases amongst advanced stage patients. These findings may be directly linked to the nature of the patient sample set used for this study, which is mostly

representative of advanced stage disease. We further devoted to analysing healthy tissues with the aim of understanding the cancer specificity of these glycoepitopes (colon, skin, lung, stomach, pancreas, liver, appendix, thyroid, and testis; **Figure. 1B**). We found sLeA in colon goblet cells as well as the pancreas, goblet cells of the appendix, and the *stratum lucidum* of the skin. On the other hand, sLeX could not be detected in the colon but was present in the pulmonary alveoli and mucus secreting cells in the appendix. In white blood cells, sLeX was notably overexpressed, particularly in granulocytes (75% granulocytes; 16% lymphocytes; 7% monocytes), showing a prevalence of 70%. In contrast, sLeA was expressed at a lower rate of 10%, predominantly in monocytes (63%) and lymphocytes (28%) when compared to its expression in granulocytes (9%; **Figure 1C; Figure S2**). Furthermore, only a few cells co-expressed both glycoepitopes (<10%), mostly on monocytes (>60%). 80%). In summary, these findings reinforce the notion that sLe antigens exhibit low cancer specificity, particularly sLeX, which shows significant expression in leukocytes.

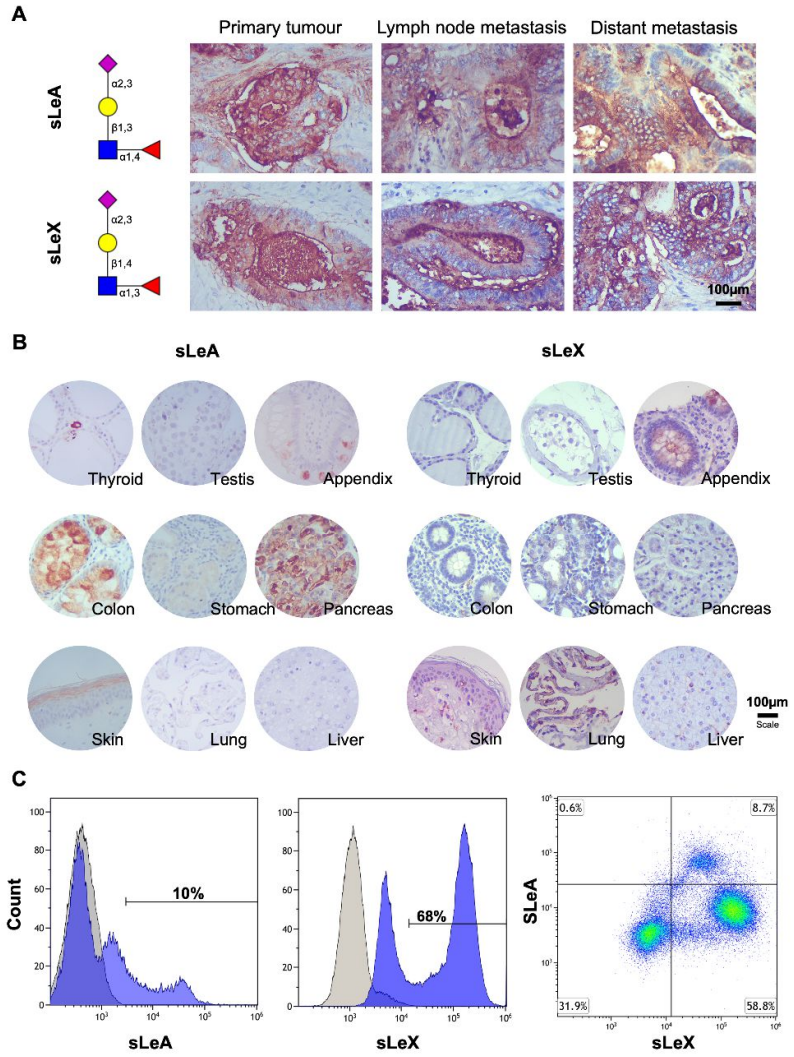


Figure 1. Sialylated Lewis antigens exhibit high expression levels in primary tumors, as well as in lymph node and distant metastases in contrast to healthy organs and leukocytes. A) SLeA and sLeX present a diffuse expression pattern in CRC primary tumours and metastases. Over 75% of CRCs express sialylated Lewis antigens, regardless of their stage. Lymph node and distant metastases of tumours positive for sLe antigens also express these glycoepitopes, mirroring the pattern observed in the primary tumour. **B) Sialylated Lewis antigens present a restricted expression pattern in histologically normal human tissues.** SLeA is mostly expressed in the colon and pancreas, whereas sLeX is mostly found in the lungs. Residuals expressions of these antigens could be found in the appendix and skin. **C) SLeX is significantly more expressed in circulating leukocytes in comparison to sLeA.** Approximately 10% of circulating leukocytes express sLeA, whereas 68% express sLeX. Notably, sLeX is mostly

expressed by lymphocytes (X%) and granulocytes (59%), while approximately 9% of lymphocytes plus monocytes express both antigens.

sLe Glycogenes in CRC

To gain more insights into the glycophenotype of CRC, we conducted a comprehensive analysis leveraging data from approximately 600 CRC cases, spanning all stages of the disease, along with 51 adjacent healthy mucosa samples retrieved from the TCGA database. Our investigation focused on key glycogenes, specifically fucosyltransferases (*FUT3*, *FUT4*, *FUT5*, *FUT6*, *FUT7, FUT9*) and sialyltransferases (*ST3GAL3*, *ST3GAL4*, and *ST3GAL6*), known to play a direct role in the biosynthesis of sLe antigens (**Figure 2A**). We found that *FUT4* and *FUT7* transcripts were elevated in CRC compared to healthy tissues, whereas the other glycosyltransferases were downregulated (**Figure S3**). Notably, *FUT4* and *FUT7* are preferentially responsible for *O*-3 fucosylation of type 2 chains leading to the formation of sLeX (**Figure 2A**). On the other hand, *FUT3* linked to *O*-4 fucosylation was downregulated in CRC, suggesting a predominance of type 2 over type 1 fucosylated structures, thus in accordance with immunohistochemistry analyses (**Figure 1A**). *FUT5* was residually expressed in few healthy and CRC cases and therefore disregarded in future evaluations. Focusing on glycosyltransferases expression in tumour tissues, we observed two distinct expression clusters, segregating *FUT3*, *FUT4*, *FUT6*, and *ST3GAL4* transcripts, from *FUT7*, *FUT9*, *ST3GAL3*, and *ST3GAL6* (**Figure 2B**). Correlation analysis further reinforced these observations (**Figure 2B-C**). Accordingly, one group showed strong positive correlations between *FUT3*, *FUT6*, and *ST3GAL4*, along with negative correlations with *FUT9*, *ST3GAL3* and *ST3GAL6* (**Figure 2C)**. The other group presented the opposite phenotype, with *FUT9* and *ST3GAL3* being significantly correlated positively with *ST3GAL6* and negatively with *FUT4*. In addition, *ST3GAL3* also correlated positively with *FUT7* and negatively with *FUT3*. Collectively, these findings reinforce the existence of two distinct groups of glycosyltransferases (*FUT3*, *FUT4*, *FUT6*, *ST3GAL4 vs FUT7*, *FUT9*, *ST3GAL3*, *ST3GAL6*) potentially responsibly by sLe antigens biosynthesis in CRC, suggesting patterns of co-expression or regulatory relationships among them. Finally, we analysed their expression according to the severity of disease by comparing early (stage I/II) with late stage (III/IV) tumours. Despite differences between cancer and healthy tissues, most glycosyltransferases' expression remained unchanged with disease stage (data not shown). The exceptions were *FUT9* and *ST3GAL3* that were statistically overexpressed in advanced (III/IV) compared to early-stage tumours (I/II; **Figure 2D**), whereas *FUT4* and *FUT6* were downregulated (**Figure 2D**). In agreement with these observations, we found

ST3GAL3 and *ST3GAL6* overexpression to be associated with worse prognosis, translated by decreased overall and disease-free survivals (**Figure 2E**). Furthermore, tumours exhibiting low *ST3GAL3* expression had a hazard ratio (HR) of 0.7 for overall survival (p=0.05) and 0.5 for disease-free survival (p=0.02; **Figure 2F**). Similarly, for low *ST3GAL6* expression, the HR was 0.5 for overall survival (p=0.002) and 0.7 for disease-free survival (p=0.06; **Figure 2F**). Conversely, tumours overexpressing *FUT3*, *FUT4*, and *FUT6* showed better prognosis (**Figure S4**). Tumours showing a downregulation of these glycosyltransferases exhibited significantly higher probabilities of cancer related death and disease progression (*FUT3*low: HR-1.8, p=0.04 for overall survival/ HR-1.7, p=0.04 for progression free survival; *FUT4*low: HR-2.4, p=0.02/ HR-1.8, p=0.002; *FUT6*low: HR=1.6, p=0.006/ HR-1.6, p=0.01*; FUT9*low: HR=0.7, p=0.09/ HR-0.7, p=0.19 (**Figure S4**).

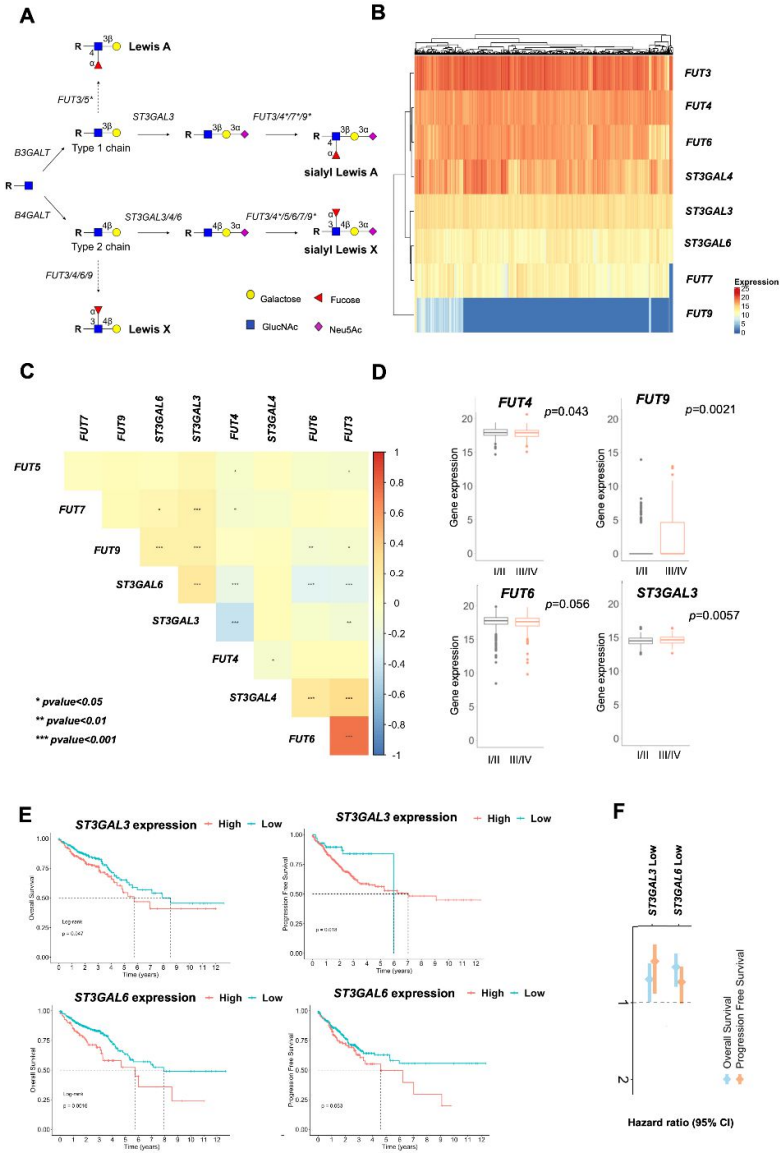


Figure 2. CRC exhibits two distinct sLe-related glycogenes signatures linked to prognosis. A) Schematic representation of the biosynthesis of sLeA and X glycoepitopes. This schematic representation highlights the main biosynthesis routes for sLeA and sLeX terminal glycoepitopes, which are present in more elongated *N*- and *O*-glycosidic chains in proteins as well as lipids. It particularly emphasizes the fucosyltransferases (*FUT3*, *4*, *5*, *6*, *7*, *9*) and sialyltransferases (*ST3GAL3*, *4*, *6*) involved in the modifications of Type 1 and Type 2 glycosidic chains that form the backbones of sLeA and sLeX. FUTs marked as "*" fucosylate *O*-3 GlcNAc residues on Type 2 glycosidic chains, even though O-4 fucosylation cannot be overulled. FUTs marked with "*" preferentially fucosylate *O*-3 GlcNAc residues on Type 2 glycosidic chains, even though *O*-4 fucosylation cannot be excluded. **B) Clustering analysis highlights the existence of two groups of FUTs and ST3GALTs driving sLe biosynthesis in CRC.** CRC tumours segregate into two distinct groups based on the expression of glycogenes involved in sLe biosynthesis (*FUT3*, *FUT4*, *FUT6*, *ST3GAL4 vs FUT7*, *FUT9*, *ST3GAL3*, *ST3GAL6*). Notably, the expressions *FUT3*, *FUT4*, *FUT6*, *ST3GAL4* are markedly more elevated in CRC compared to *FUT7*, *FUT9*, *ST3GAL3*, *ST3GAL6*. **C) A correlation analysis supports the existence of two glycogenes groups in CRC.** The correlogram shows significant positive correlations between *FUT3*, *FUT4*, *FUT6*, *ST3GAL4* and negative correlations with *FUT7*, *FUT9*, *ST3GAL3*, *ST3GAL6*, characterizing the second group of sLe-related glycogenes. On the other hand, *FUT7* and *FUT9* correlated positively with *ST3GAL3* and *ST3GAL4* and negatively with the other group of glycogenes, reinforcing the existence of two distinct clusters of potentially co-expressed glycogenes. **D) *FUT3* and *FUT4* were significantly decreased in more aggressive tumours (stages III/IV *vs* I/II), whereas *FUT9* and *ST3GAL3* transcripts were significantly increased.** These observations link the group of tumours enriched for *FUT9* and *ST3GAL3* to cancer aggressiveness. **E and F) Overexpression of *ST3GAL3* and *ST3GAL6* required for sLe antigens biosynthesis is associated with decreased survival in CRC.** *ST3GAL3* and *ST3GAL6* overexpressing tumours presented significantly decreased overall and disease-free survivals. Furthermore, tumours showing low transcripts for these glycogenes presented lower risk of death by cancer (*ST3GAL3*: HR=0.7, p=0.05; *ST3GAL6:* HR=0.5, p=0.002) and disease progression (*ST3GAL3*: HR=0.5, p=0.02; *ST3GAL6:* HR=0.7, p=0.06).

In summary, we found two distinct glycosyltransferases expression patterns linked to different prognoses. Namely, we found indications that elevated *ST3GAL3* and *ST3GAL6* transcripts may be used as surrogates of poor prognosis, underscoring a connection

between hypersialylation driven by these glycosyltransferases and CRC aggressiveness. Their strong correlation with *FUT7* and *FUT9* supports the preferential expression of sLeX antigens in more aggressive tumours, aligning with our observations in patient samples. On the other hand, we found the high *FUT3*, *FUT4*, and *FUT6* transcripts were surrogates favourable prognosis. Our findings further highlight the importance of gaining a deeper understanding of the molecular pathways underlying sLe biosynthesis and accurate mapping of these terminal sialoglycans through glycomics, particularly in the context of precision oncology.

E-selectin affinity glycoproteome

To overcome the limitations arising from the non-cancer specificity of sLe antigens, we undertook a thorough characterization of the glycoproteome to pave the way for precision oncology (**according to the scheme in Figure 3A**). Our approach involved using proteins extracted from FFPE tissues as the starting material. The goal was to highlight the underexplored yet promising potential of this material for conducting comprehensive glycome and glycoproteome studies. A pool of metastatic tumours overexpressing sLeA/X in more than 50% of the tumour area was selected for a thorough investigation of the sLe-glycoproteome. This involved utilizing E-selectin affinity chromatography followed by bottom-up identification through mass spectrometry (**Figure 3A**). Glycomics analysis, involving the *in situ N*- and *O*-deglycosylation of sLe hotspots in tumour tissues, was conducted in parallel to facilitate the annotation and validation of glycoproteins. We found high levels of sLe antigens and little alterations in the *N*- and *O*-glycome across the metastatic CRC samples used for this study (**Figure S5; Table S3, S4**). These glycoepitopes were mainly observed on extended core 2 *O*-glycans and on complex *N*-glycans (**Figure S5A, -B**), also in accordance with previous reports on the CRC glycome (5). The presence of sLe antigens on *N*- and *O*-glycans was further confirmed by MS/MS analysis (**Figure S5C, -D**). This data was later used for unequivocal identification of sLe glycoproteoforms by mass spectrometry.

We have identified 554 glycoproteins carrying glycopeptides indicative of the presence of sLe antigens **(Figure 3B; Table S5, S6**). Notably, the number of *O*-glycoproteins (n=349; **Table S5**) potentially carrying sLe glycoepitopes was higher than that of *N*-glycoproteins (n=231; **Table S6**). Notably, it was only possible to ascertain the simultaneous occurrence of both types of glycosylation in 26 proteins. These included mucins (MUC-15; MUC-16; **Figure 3B**) and proteins associated with cell adhesion, such as

FAT2 (commonly found in the nervous system) (18) and CD6, which may serve as a co-stimulatory molecule for T-cell activation and proliferation (19). Additionally, adhesion-related molecules like E-selectin and L-selectin, typically associated with lymphocytes, hinted at the possibility that these receptors may also carry sLe antigens. Nevertheless, bleeding from the enrichment strategy cannot be overruled at this stage. The comprehensive integration of this data using GO term analysis revealed markedly different biological functions (**Figure3C -D**). Identified *O*-glycoproteins predominantly play a crucial role in facilitating the transport of inorganic molecules and ions across the cell membrane, accounting for over 70% of their involvement (**Figure 3C**). They also contribute, albeit to a lesser extent, to cellular functions such as adhesion, migration, chemical homeostasis, and participation in the G-protein coupled receptors (GPCRs) signalling pathway (**Figure 3C**). In contrast, the *N*-glycoproteins carrying sLe epitopes reveal distinct molecular networks (**Figure 3C**). Beyond mediating GPCRs and ion channel activities, as well as cellular adhesion, these proteins are associated with vasculature development and positive regulation of the Mitogen-Activated Protein Kinase (MAPK) cascade, a pathway linked to promoting proliferation, resistance to apoptosis and poor prognosis in CRC (20, 21). Notably, the most significant and markedly different biological function is related to axon guidance (>30% of the glycoproteins; **Figure 3D**). This involves the processes through which developing nerve cells navigate and extend their axons to establish connections within the nervous system. Interestingly, a more detailed revisitation of the identified *N*- and *O*-glycoproteins, supported by the Human Protein Atlas database, identified several protein species typical of neural cells, namely CNTN6, NRCAM, NFASC, TENM4, and NCAM1. The presence of glycans carrying sLe antigens on these glycoproteins was further confirmed by manual annotation of MS/MS spectra (**Figure S6**). These observations strongly suggest that these sialoglycoproteins may originate from the neuroendocrine system, constituting novel findings that warrant more in-depth confirmation in future studies. Interestingly, in explorative settings, we observed that all three tumours pooled for this study presented significant vascularization (high CD34; **Figure S7**), and one also displayed high levels of the neuroendocrine marker synaptophysin (**Figure S7**). Such observations align with the connection between the identified glycoproteins, angiogenesis, and neural features. Another noteworthy discovery was the identification of the immune checkpoint PD-1, now being described as an E-selectin binder for the first time. However, the implications of these findings for protein functionality as well as its clinical significance remain unaddressed and merit future investigations. We then compared our findings with the limited information on E-selectin ligands from CRC cell lines (5, 22)(**Figure S8)** While the number of glycoproteins identified by our study significantly broadens the understanding of the E-

selectin affinity CRC glycoproteome compared to current literature, we were able to identify some common grounds with a recent investigation on a sLeX-expressing cell line (22) (**Figure. S8**). We particularly highlight nervous system associated glycoprotein L1CAM, which suggests that cells of epithelial origin may gain capacity to express neuronal markers carrying sLe antigens.

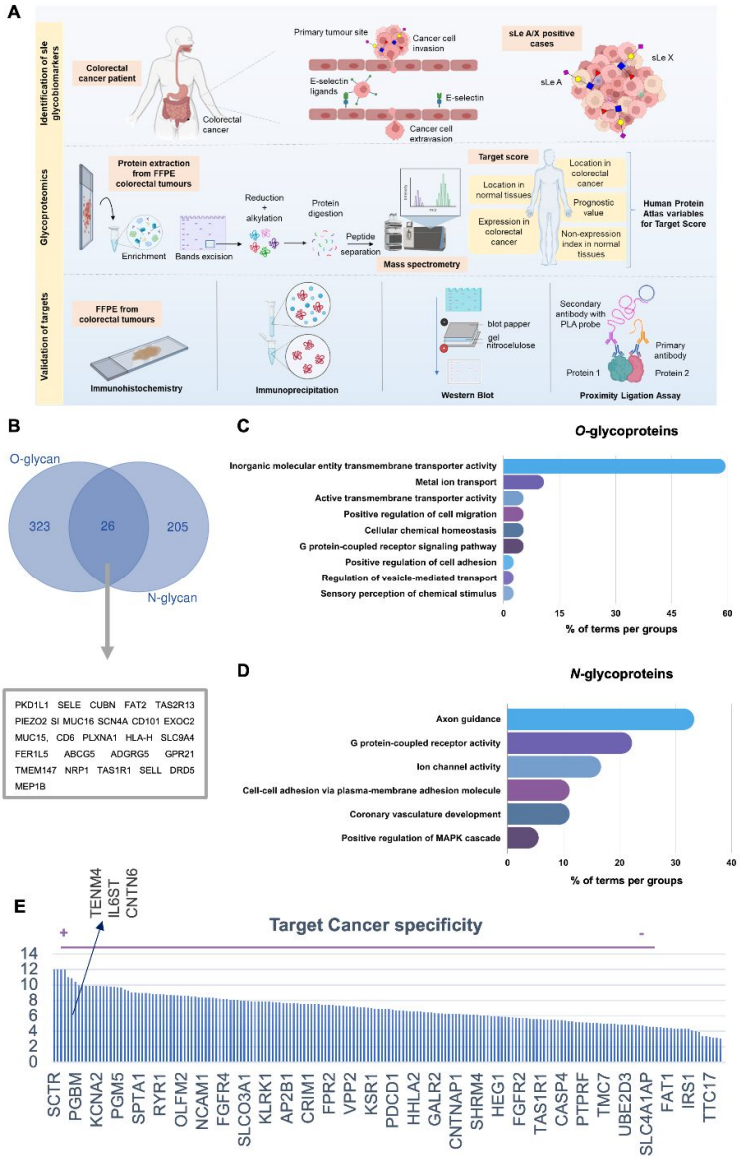


Figure 3. E-selectin affinity glycoproteomics identifies relevant functional molecular signatures and targetable glycoproteins potentially carrying sialylated Lewis antigens. A) Schematic representation of the analytical workflow explored for identification of E-selectin ligands in CRC. The top panel illustrates how sialylated Lewis antigens may facilitate the intravasation and extravasation of cancer cells through interactions with E-selectin on endothelial cells. The middle panel outlines the workflow for

the isolation and identification of glycoproteins, along with the curation of glycoproteomics data to identify potentially targetable molecular signatures. In brief, protein extracts from FFPE metastatic CRC tissues were pooled, enriched for sialylated Lewis antigens expressing glycoproteins through affinity for E-selectin and separated by SDS-PAGE. The bands were excised, and the proteins were reduced, alkylated, and digested with trypsin before being analysed by nanoLC-ESI-HCD-MS/MS. Identified glycoproteins were ranked using an in-house algorithm (target score), which utilizes information from the Human Protein Atlas to sort proteins based on their targetability, by summing cancer specificity and association with poor prognosis, while penalizing expression in healthy tissues. Relevant glycoproteins were then orthogonally validated in cancer tissues for sLe antigen expression using different immunoassays such as immunohistochemistry, PLA, and western blot**. B) More *O*-glycoproteins potentially carrying sialylated Lewis antigens were identified compared to *N*-glycoproteins.** Notably, only 26 glycoproteins were found carrying both modifications. **C) The most notable distinction in terms of biological functions between the *O*- and *N*-glycoproteomes was the association of *N*-glycoproteins with neuronal traits.** GO term analysis showed that *O*-glycoproteins were mainly involved in plasma membrane molecular transport and cellular homeostasis regulation, as well as in cell adhesion, migration, and G protein-coupled receptor signalling pathways. Furthermore, proteins linked to sensory functions suggest an association with neuroendocrine functions, mirroring observations for *N*-glycoproteins. *N*-glycoproteins are primarily associated with axon guidance, and to a lesser extent, with G protein-coupled receptor and ion channel activities, as well as cell adhesion, similarly to *O*-glycans. Notably, glycoproteins involved in vasculature development and positive regulation of the MAPK cascade were also observed. **D) SCTR emerged as one of the top ranked targetable glycoprotein in CRC.** Most abundant glycoproteins were ranked using our in-house target score algorithm, which builds on the Human Protein Atlas database to ascertain cancer specificity and associations with poor prognosis. Notably, this approach identified SCTR, along with TENM4, IL6ST, and CNTN6, as top-ranked glycoproteins.

Target Selection

Finally, we have conducted a thorough exploitation of the identified glycoproteins to uncover potential targetable signatures, considering cancer specificity and prognostic links. This comprehensive analysis utilized an in-house algorithm, previously successful in identifying cancer-specific glycoproteoforms in esophageal (8), gastric (9), and bladder cancer (10). The algorithm systematically utilizes the protein databank to prioritize

associations with cancer, poor prognosis, preferential localization at the cell membrane for enhanced targetability, and low expression in relevant healthy organs to minimize potential off-target effects. Conversely, the algorithm-imposed penalties for the presence of glycoproteins in healthy tissues. Accordingly, the most abundant glycoproteins were then analysed by this approach. Notably, amongst the top ranked glycoproteins carrying sLe antigens was TENM4 and the SCTR (**Figure 3E; Table S7**). Interestingly, both glycoproteins have been observed in nerve cells (23, 24).

SCTR and SCTR sialoproteoforms in CRC

In this study, we focused on assessing the cancer specificity and the clinical relevance of the SCTR in CRC, aiming to gain more information about this protein, which has been poorly studied in this context. This membrane glycoprotein was expressed in all primary advanced cases and was significantly increased in tumours with distant metastasis (**Figure 4A, -B, -C**). Notably, the SCTR was diffusely expressed across the tumour tissue (>70% of the tumour area). It was also detected in 50% of lymph node metastases (3/6) and 30% of distant metastasis (3/10; **Figure 4B**, **Figure S9)**. Strikingly, all distant positive metastases were observed in the liver, suggesting some degree of tropism. Complementary transcript analysis from TCGA confirmed the observations made in our patient series, linking the *SCTR* receptor with CRC aggressiveness. It revealed that *SCTR* is overexpressed in more advanced tumours (stage III/IV) compared to less aggressive lesions (stage I/II; **Figure 4D**). Also, *SCTR* overexpression was associated with decreased overall ($SCTR^{low}$ tumours: HR=0.5; p= 0.004) and progression-free ($SCTR^{low}$ tumours: HR= 0.7; p= 0.05; **Figure 4E-H**) survivals. Interestingly, TCGA analysis also unveiled a significant positive correlation between *SCTR* and *SCT* for a relevant subset of colorectal tumours **(Figure 6B**). Notably, SCT binding to the SCTR receptor has been found required for initiating physiological responses (25). Furthermore, *SCT* was also significantly overexpressed in stage III/IV CRC and associated to worse prognosis (overall survival for SCT^{low} tumours HR=0.5, p= 0.001; disease free survival HR=0.5, p= 0.001; **Figure 4F-H**). Notably, clustering analysis revealed an association between SCRT and SCT with glycosyltransferases linked to the unfavourable prognosis glycosyltransferase phenotype. (**Figure 6A**). These findings were further reinforced by correlation analysis, where *SCTR* and *SCT* expressions correlated positively with *FUT4*, *5*, and *9* linked to the worst prognosis (**Figure 6A-B**). Conversely, *SCRT* correlates negatively with *FUT3* and *FUT4,* associated with favourable prognosis (**Figure 6A-B**). *SCT* expression correlated positively with *FUT7* and negatively with *FUT3* and *FUT4*. Additionally, it correlated positively with poor

prognosis marker *ST3GAL3* (**Figure 6B**). These observations reinforce a link between SCTR and SCT cancer aggressiveness and sLe antigens driven by a particular subset of glycosyltransferases, which merits future validation.

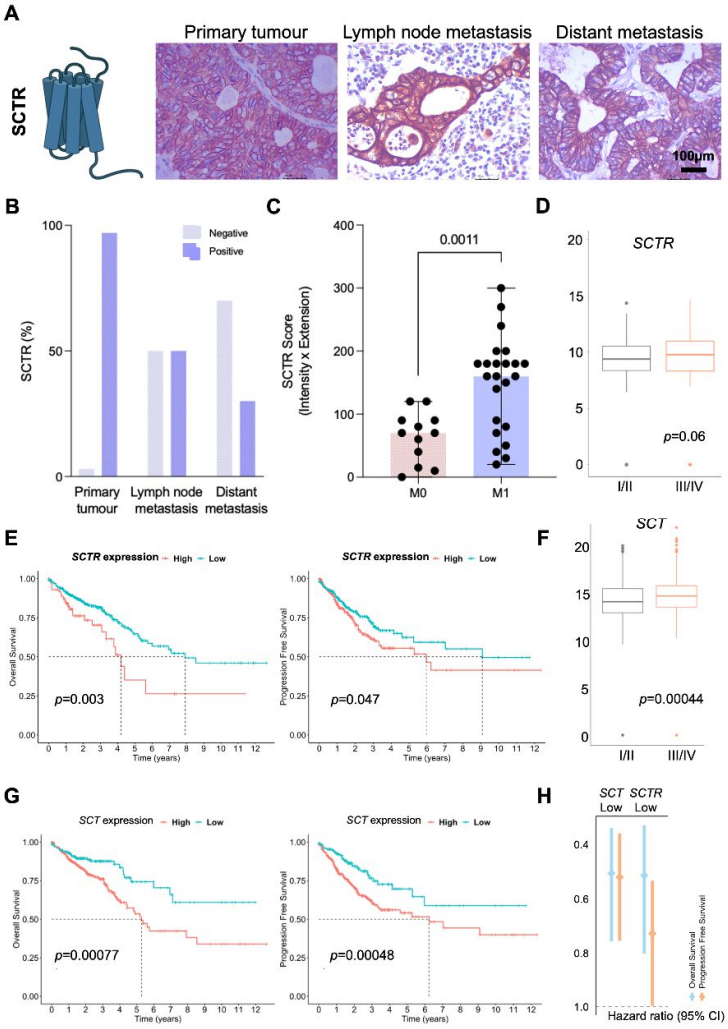


Figure 4. The SCTR is overexpressed in metastatic tumours as well as in the metastases and associates with poor prognosis in CRC. A-B) The SCTR is highly expressed in both aggressive primary tumours as well as lymph node and distant metastases. The SCTR exhibits diffuse expression in most colorectal tumours and may also be present in some lymph nodes (in 50% of cases) and distant metastases (in 25% of cases), predominantly in the liver. Notably, no signs of SCTR expression were observed in peritoneal metastases. **C) The SCTR is significantly more expressed in primary tumours showing metastasis. D) The *SCTR* gene is overexpressed in advanced stage CRC.** TCGA analysis showed that the number of *SCTR* transcripts was higher for advanced tumours (stages III/IV vs I/II). **E) *SCTR* overexpression associates with decreased survival.** Tumours exhibiting *SCTR* overexpression showed significantly lower overall and

progression-free survivals compared to tumours with low *SCTR* expression. **F) *SCT* is also overexpressed in advanced stage tumours.** Advanced stage tumours present significantly higher number of *SCT* transcripts, which encode for a key activating ligand of SCTR. **G) *SCT* overexpression associates with worse survival.** Tumours exhibiting high *SCT* transcript levels present significantly lower overall and disease-free survival rates when compared to *SCT*-low tumours**. H) Low *SCTR* or *SCT* transcript levels are associated with a decreased risk of death.**

We then assessed the glycosylation of SCTR, aiming to determine cancer specificity. Approximately 70% of the cases positive for SCTR, either tumours or metastases, co-expressed sLe antigens in the same tumour area (**Figure 5A**). PLAs targeting sLeA and sLeX provided additional support for the hypothesis that SCTR could function as a carrier for these glycoepitopes (**Figure 5A**). Furthermore, PLA analysis of the same tumour tissue sections, post-digestion with PNGase F to eliminate *N*-glycans, led to the disappearance of the SCTR-sLe signals (data not shown). The hypothesis that sLe antigens could be attached to *N*-glycans was further supported by tandem mass spectrometry, revealing glycopeptides with complex *N*-glycans terminated by sLe antigens. (**Figure 5B**. PLA analysis further showed a higher density of SCTR-sLeA in comparison to the SCTR-sLeX glycoproteoform in CRC (**Figure 5A**). SCTR immunoprecipitation and western blotting reinforced these observations (**Figure 5C**). Additionally, in our Western blot analysis of SCTR-immunoprecipitates, multiple bands were observed, but mass spectrometry confirmed the presence of the SCTR glycoprotein only below 75 kDa and at approximately 50 kDa and 37 kDa. The blots for sLeX displayed a faint band at 75 kDa, suggesting also residual cross-reactivity with bands at 50 kDa and 75 kDa, in line with the low abundance of SCTR-sLeX observed through PLA. Notably, E-selectin did not recognize the 75 kDa band, likely due to low expression levels. Conversely, both E-selectin and the anti-sLeA antibody exhibited noticeable cross-reactivity with bands above 50 kDa and a low molecular weight proteoform at 37 kDa. Notably, non-specific signals were detected in the sLeA and sLeX blots at 25 kDa (corresponding to the Ig light chain) where SCTR was absent. Additionally, non-specific cross-reactivities with the 50 kDa band, potentially containing IgG heavy chains, cannot be ruled out. Nevertheless, our findings strongly support the existence of a short SCTR-sLeA glycoproteoform with significant affinity for E-selectin.

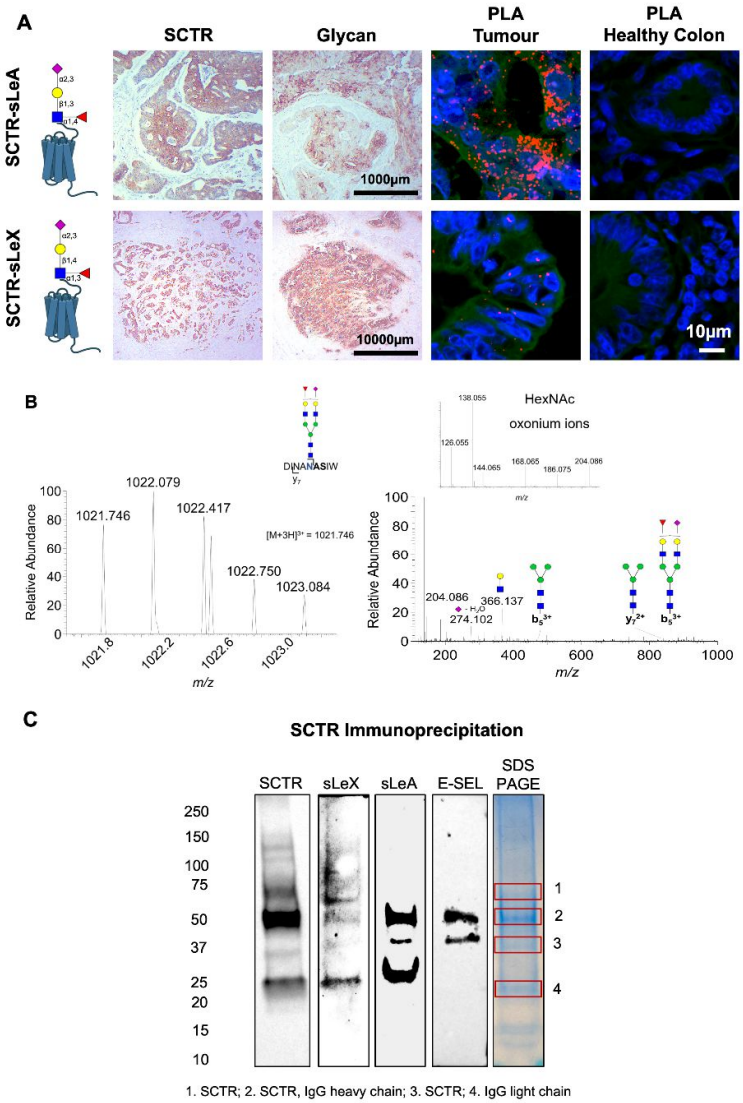


Figure 5. The SCTR is a major carrier for sialylated Lewis antigens in CRC but not on healthy tissues. A) The SCTR predominantly carries sLeA in CRC. Immunohistochemistry analysis revealed a co-expression of both sLeA and sLeX glycoepitopes along with SCTR in the same tumour area, which was not observed in healthy tissues denoting some degree of cancer specificity. PLA analysis further confirmed SCTR as a potential carrier for these glycans, a finding absent in healthy colon samples. Moreover, PLA analysis strongly suggests that SCTR is predominantly modified with glycans carrying sLeA. **B) Western blot analysis of SCTR-immunoprecipitates provides evidence for the existence of SCTR-sLeA glycoproteoforms acting as ligands for E-selectin.** Western blot analysis of SCTR immunoprecipitates revealed multiple bands, with the most intense below 75 kDa, at 50 kDa, and 25 kDa. However, SCTR was detected by mass spectrometry only at approximately 75 kDa (band 1), 50 kDa (band 2), and 37 kDa (band 3). The anti-sLeX antibody displayed less reactivity, with faint bands observed

predominantly between 75 kDa and 50 kDa and at 25 kDa. The sLeA antigen stained strongly, exhibiting clear bands at 50 kDa, 37 kDa, and 25 kDa. Furthermore, E-selectin showed affinity for bands at 50 kDa and 37 kDa. Notably, all blots also displayed signals from the antibody used for immunoprecipitation at 50 kDa (IgG heavy chain) and 25 kDa (IgG light chain). Collectively, these findings strongly suggest the existence of a low molecular weight SCTR-sLeA glycoproteoform with affinity for E-selectin at 37 kDa. The existence of another glycoproteoform at 50 kDa cannot be fully confirmed due to antibody signals. **C) MS and MS/MS spectra for an SCTR *N*-glycopeptide potentially carrying sialylated Lewis antigens.** The MS and MS/MS spectra support the existence of a complex type *N*-glycan potentially carrying sialylated Lewis antigens. Collectively, multiple orthogonal validations support the modification of SCTR with glycans carrying sialylated Lewis antigens, most likely sLeA.

In summary, we have established a connection between SCTR overexpression in CRC, poor prognosis, and the presence of metastases. Additionally, we have demonstrated that, despite the high expression of sLeX in CRC cancer, the receptor is a preferential carrier for sLeA and can be targeted by E-selectin, potentially explaining its presence in metastatic contexts.

SCTR and SCTR sialoproteoforms in Healthy Tissues

Finally, we conducted a more in-depth investigation on the cancer specificity of the SCTR exploiting documented analysis in the human protein atlas. The receptor has been found highly expressed in the gastrointestinal tract and, to a lesser degree, in the respiratory system as well as in the pancreas, kidney, and urinary bladder. Low expressions could also be found in proximal digestive tract, in male, connective and soft tissues (**Figure S10**). Additionally, and for validation purposes, we also analysed the cohort of healthy tissues previously screened for sLe antigens. We confirmed the expression of SCTR across the digestive tract (stomach, colon, appendix; **Figure 6B**) in cells of the glandular epithelium as well as in respiratory tract in cells that compose the bronchus and bronchiole epithelia. The thyroid, testicle, liver, and pancreas were negative (**Figure 6B**), which contrasted with the low expression described by the human protein atlas. However, despite the expression of sLe antigens in some of these tissues (lung, pancreas, colon, appendix, skin; **Figure 1B**), we found no indications of overlapping with SCTR-sLe expressing areas, a finding later confirmed by negative PLA (**Figure 6C**). Furthermore, we identified SCTR in around 50% of leukocytes, predominantly present in granulocytes (approximately 70%), and to a lesser

extent, in 20% of lymphocytes and 10% of monocytes (**Figure 6D; Figure S11**). However, less than 10% of these cells also exhibited positivity for sLeA, with a comparable distribution observed among granulocytes, lymphocytes, and monocytes (**Figure 6D; Figure S11**). In summary, our findings indicate a constrained expression pattern for SCTR-sLeA glycoproteoforms in healthy tissues, in clear contrast with their elevated presence in colorectal primary tumours and metastases.

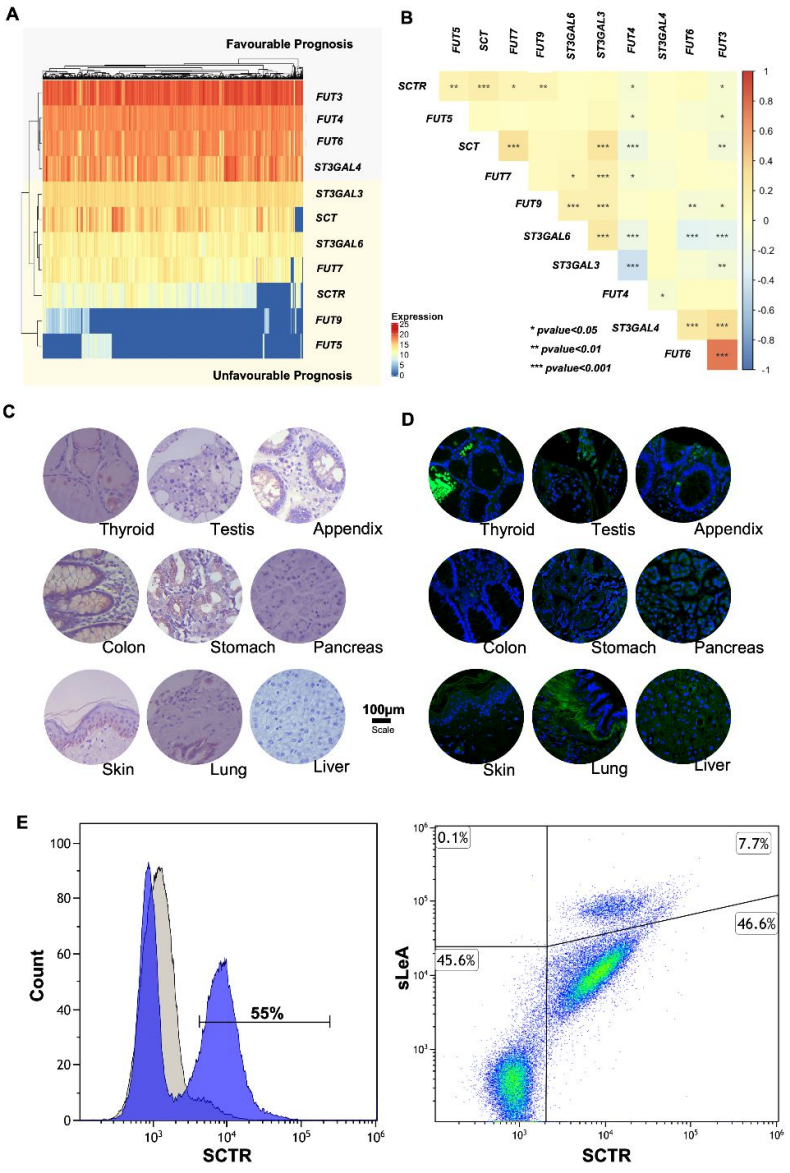


Figure 6. SCTR-sLe glycoproteoforms associate with worst prognosis and present a very restricted pattern in human healthy tissues and cells. A) The SCTR and its activating ligand SCT cluster with sLe-related glycogenes linked to worst prognosis in CRC. *SCTR* and *SCT* expressions clusters with the expression of *FUT5*, *FUT7*, *FUT9*, *ST3GAL3* and *ST3GAL4* linked to worst prognosis in CRC. **B) The SCTR is significantly**

correlated with SCT and the glycogenes linked to worst prognosis. The SCTR receptor positively correlates with SCT and glycogenes linked to a poor prognosis, while both are negatively correlated with the glycogenes associated with a favourable prognosis. **C) SCTR presents relevant expression in several human healthy organs (respiratory system; gastrointestinal tract; pancreas; kidney and urinary bladder).** The analyses of healthy organs revealed moderate SCTR expression in the healthy glandular epithelium of the gastrointestinal tract, specifically in the stomach, colon, and appendix, as well as in the bronchi and bronchioles. **D) SCTR-SLeA glycoproteoforms were not detected in healthy tissues.** PLAs conducted on healthy tissues, including SCTR-positive/SLeA tissues (colon, pancreas, skin) did not reveal signs of proximity, consistent with the observation that the glycoepitope and the SCTR are not expressed in the same region. **E) SCTR-sLeA glycoproteoforms are residually expressed by circulating leukocytes.** SCTR is highly expressed in leukocytes, with approximately 50% of them showing expression, primarily in granulocytes (80%) and some lymphocytic populations. Additionally, a residual population of approximately 8% of monocytes expresses both SCTR and sLeA.

Discussion and Concluding Remarks

This work establishes a roadmap for identifying glycoproteoforms carrying sialylated antigens in CRC, with potential applicability to other tumours. The study was motivated by the crucial role played by these glycans driving metastasis (26, 27). They facilitate cancer cell engagement with endothelial cells, aiding intravasation, enabling travel through the bloodstream, and ultimately directing them to distant locations via interactions with E-selectin (3, 28). In fact, blocking cancer cell binding to E-selectin has been demonstrated to prevent CRC metastasis (29, 30). However, this binding is influenced not only by the structural characteristics of the terminal sLe epitope but also by the protein scaffold carrying these post-translational modifications (31, 32). Still, the identity and functions of glycoproteoforms carrying these glycoepitopes in CRC and other tumours remain mostly elusive, hindering a comprehensive understanding of their biological roles and precision oncology. In fact, most studies concerning CRC have been exclusively conducted on cancer cell lines, leading so far to the identification of a limited number of glycoproteins such as CD44 (33), carcinoembryonic antigen (34), podocalyxin-like protein (35), lysosomal membrane glycoprotein-1 and 2 (28), and several types of integrins (22). The low cancer specificity of sLe antigens also poses a limitation for targeted therapies.

To address these challenges, we applied E-selectin affinity chromatography and tandem mass spectrometry, a methodology successfully employed in breast (13), gastric

(9), and CRC cancer cell models (22) to analyse glycoproteomics in tumour tissues. Notably, FFPE tissues, predominant in cancer biobanks, proved invaluable for glycome and glycoproteome characterization, expanding beyond recent demonstrations on CRC glycome characterization in these tissues (5, 36). Our investigation revealed a broader spectrum of CRC glycoproteins with affinity for E-selectin compared to previous reports in cell models (22, 33). We observed significant differences between the *O*- and *N*-glycoproteomes in terms of identified glycoproteoforms and key biological functions. Crucial glycoproteins were identified for cell adhesion, maintenance of cellular homeostasis, response to microenvironmental stimuli, and activation of relevant oncogenic pathways, particularly MAPK cascade (37, 38). We also unveiled a significant number of glycoproteins associated with neovascularization and introduced novel observations regarding the substantial presence of glycoproteins characteristic of the colorectal enteric nervous system, recognized for autonomously regulating physiological functions (39) but poorly characterized in CRC. Among these glycoproteins is the neuron-associated glycoprotein LICAM1, also described in a recent study on the *FUT6*-transfected sLeX-overexpressing SW620 cell line as a major E-selectin binding molecule (22). LICAM1, characteristic of nervous cells, has been implicated in neuronal migration, neurite fasciculation, and synaptic plasticity, observed in several types of tumours, including CRC (40-42). Its identification and overexpression, promoted by *FUT6*/sLeX in an adenocarcinoma cell line with epithelial morphology (22), pose a provocative hypothesis. It suggests that CRC cancer cells may acquire neurological traits potentiated by sLe, emphasizing the need for further confirmation. Our study demonstrates that glycosylation with sLe antigens is a molecular feature shared by glycoproteins associated with vascular and neuro functions, reinforcing potential yet undetermined links between these two systems. Additionally, we identified PD-1, a well-known immune checkpoint receptor targeted in clinics to enhance anti-cancer immune responses (43, 44). While PD-1 *N*-glycosylation is critical for receptor functionality and recognition by therapeutic antibodies (45, 46), the presence of sLe antigens is being reported for the first time, and its biological role remains unknown. Nevertheless, these findings strongly indicate the presence of T cells in the tumour bulk carrying sLe antigens, known to play a key role in T cell recruitment to inflammation sites by enabling adhesion to E-selectin (47, 48). Overall, we hypothesize that the diversity found in the sLe glycoproteome could potentially arise from shared microenvironmental pressures affecting different types of cells (immune, vascular, neuronal, tumour) within the tumour ecosystem. Specifically, it is well-established that sLe antigens expression may be triggered by pro-inflammatory signals in various conditions, including cancer (49, 50). In chronic gastric inflammation induced by *H. pylori*, the expression of sLe antigens has been linked to the

modulation of glycosyltransferases by pro-inflammatory cytokines, namely TNF-α (51). Additionally, in a primary invasive ductal carcinoma cell line, the inhibition of fucosylation suppressed sLeX antigen and E-selectin ligands expression, resulting in lower migration capacity and reduced IL-8 and TGF-β expressions (37). Furthermore, *FUT6* silencing has been shown to reduce TGF-β-mediated epithelial-mesenchymal transition and inhibit migration in CRC (52). These findings highlight the pivotal role of the microenvironment on the expression of sLe antigens. Future studies should explore the functional impact of these seminal observations and their consequences for organ functionality and disease progression. Emerging single cell proteomics approaches will be unvaluable tools towards this objective (53-55).

Finally, we have conducted a thorough analysis of identified glycoproteins to uncover potential targetable signatures using an in-house algorithm successful in various cancers (8-10). The G protein-coupled SCTR emerged as a top-ranked glycoprotein and was selected for in-depth evaluations due to its relatively unexplored nature in CRC. Notably, the SCTR has been previously identified in both glandular cells and subsets of putative secretomotor/vasodilator neurons in the healthy colon (56). This reinforces the hypothesis that many sLe-expressing glycoproteins may originate from the peripheral nervous system. While it remains poorly studied in CRC, the SCTR has been found overexpressed in gastrinomas (57), bronchopulmonary carcinoid tumours (58) cholangiocarcinoma (57), esophageal and pancreatic ductal adenocarcinomas (57) and considered a candidate target for molecular tumour imaging as well as for peptide receptor radioligand therapytherapy (59).(59). Here, we describe that both SCTR and its activating ligand SCT are overexpressed at advanced stages and associated with distant metastases, being predictors of worse prognosis. The correlation between elevated SCTR and SCT levels in CRC may imply yet unknown alterations in fluid balance, colonic motility, and interactions with the enteric nervous system, with implications for disease progression and dissemination. These constitute novel observations raising fundamental questions about SCTR's role in CRC biology. Furthermore, we have verified SCTR as a primary carrier of sLe antigens, proposing that these glycoepitopes might confer cancer specificity to the glycoprotein and enable adhesion to E-selectin. However, further in-depth analysis is essential to validate this hypothesis, evaluate the impact of glycosylation on SCTR functionality, and explore targeted therapeutic opportunities. Interestingly, we also linked the SCTR and SCT to a group of sLe-related glycogenes defining worst prognosis, reinforcing the association with CRC aggressiveness. Furthermore, we have uncovered a novel clinical context and a potential co-regulatory mechanism among sialyl Lewis (sLe)-

related glycogenes in CRC, providing novel insights to better understand their biological functions and role in disease, which remain controversial.

In summary, this study presents a strategy for unravelling the cancer glycoproteome's complexity, identifying clinically relevant signatures, and shedding new lights on the sialoglycoproteome's links to the colonic peripheral nervous system. The recently recognized impact of nerves and neoinnervation on various cancer types (60-62) underscores the imperative need for deeper exploitation. Specifically, while interactions between cancer, vascular and nervous cells have been shown to contribute to cancer progression (63, 64), the precise mechanisms driving CRC and the role played by protein glycosylation in this context remain unexplored. We anticipate that this knowledge may provide a novel research avenue at the intersection of neurology and oncology, paving the way for precision oncology and novel strategies to counteract metastases development. Furthermore, our investigation revealed an overexpression of SCTR carrying sLeA and targeted by E-selectin in metastatic colorectal tumours, as well as a significant number of associated lymph node and distant metastases. Notably, this overexpression contrasts with the restricted expression pattern observed in healthy human tissues and immune cells. While the possibility of cancer specificity remains, further studies leveraging a comprehensive characterization of the human glycoproteome are essential. Nonetheless, these findings provide a foundation for targeted approaches to address aggressive CRC tumours and metastases more effectively.

Acknowledgments

The authors wish to acknowledge the Portuguese Foundation for Science and Technology (FCT) for PhD grants 2020.09394.BD (SC), 2020.09384.BD (DF), SFRH/BD/146500/2019 (MRS), 2022.12980.BD (AM), Principal researcher contract 2022.08311.CEECIND (JAF), and Junior researcher contact 2021.03835.CEECIND (AB). FCT is co-financed by European Social Fund under Human Potential Operation Programme from National Strategic Reference Framework. The authors also acknowledge FCT/MCTES funding within the project RESOLVE (DOI 10.54499/PTDC/MED-OUT/2512/2021) and funding for the IPO research center (PEst-OE/SAU/UI0776/201, CI-IPOP-29-2016-2022), and the LAQV research unit (Project DOI: 10.54499/LA/P/0008/2020; 10.54499/UIDP/50006/2020 and 10.54499/UIDB/50006/2020). This article is also a result of the project NORTE-01-0145-FEDER-000012 and “The Porto Comprehensive Cancer Center Raquel Seruca” with the reference NORTE-01-0145-FEDER-072678 - Consórcio PORTO.CCC—

Porto.Comprehensive Cancer Center Raquel Seruca, supported by NORTE 2020, under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund. Finally, the authors also acknowledge the ICBAS PhD Program in Biomedical Sciences

Author Contributions

Conceptualization: SC,DF, JAF; Methodology: SC, DF, MRS, JAF; Software: AM, AB, AMNS, JAF; Validation: LPA, LLS; Investigation: SC, DF, MRS, AB; AM, EF, BS, MG; Resources: AB, LLS, AMNS, JAF; Data Curation: SC, DF, MRS, AM, LPA, JAF; Writing-Original Draft: SC, DF, JAF; Writing- Review & Editing: SC, DF, LLS, AMNS. Visualization: SC, DF, MRS, JAF, Supervision: JAF; Project Administration: JAF; Funding Acquisition: LLS, JAF.

Competing interests

The authors declare no competing interests.

References

1. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023;72(2):338-44.
2. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? Int J Mol Sci. 2018;19(12).
3. Fernandes E, Sores J, Cotton S, Peixoto A, Ferreira D, Freitas R, et al. Esophageal, gastric and colorectal cancers: Looking beyond classical serological biomarkers towards glycoproteomics-assisted precision oncology. Theranostics. 2020;10(11):4903-28.
4. Moran AB, Elgood-Hunt G, van der Burgt YEM, Wuhrer M, Mesker WE, Tollenaar R, et al. Serum N-Glycosylation RPLC-FD-MS Assay to Assess Colorectal Cancer Surgical Interventions. Biomolecules. 2023;13(6).
5. Madunic K, Mayboroda OA, Zhang T, Weber J, Boons GJ, Morreau H, et al. Specific (sialyl-)Lewis core 2 O-glycans differentiate colorectal cancer from healthy colon epithelium. Theranostics. 2022;12(10):4498-512.
6. Boyaval F, van Zeijl R, Dalebout H, Holst S, van Pelt G, Farina-Sarasqueta A, et al. N-Glycomic Signature of Stage II Colorectal Cancer and Its Association With the Tumor Microenvironment. Mol Cell Proteomics. 2021;20:100057.
7. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol. 2005;23(3):609-18.
8. Cotton S, Ferreira D, Soares J, Peixoto A, Relvas-Santos M, Azevedo R, et al. Target Score-A Proteomics Data Selection Tool Applied to Esophageal Cancer Identifies GLUT1-Sialyl Tn Glycoforms as Biomarkers of Cancer Aggressiveness. Int J Mol Sci. 2021;22(4).
9. Fernandes E, Freitas R, Ferreira D, Soares J, Azevedo R, Gaiteiro C, et al. Nucleolin-Sle A Glycoforms as E-Selectin Ligands and Potentially Targetable Biomarkers at the Cell Surface of Gastric Cancer Cells. Cancers (Basel). 2020;12(4).
10. Peixoto A, Ferreira D, Azevedo R, Freitas R, Fernandes E, Relvas-Santos M, et al. Glycoproteomics identifies HOMER3 as a potentially targetable biomarker triggered by hypoxia and glucose deprivation in bladder cancer. J Exp Clin Cancer Res. 2021;40(1):191.
11. Zhang T, Madunic K, Holst S, Zhang J, Jin C, Ten Dijke P, et al. Development of a 96-well plate sample preparation method for integrated N- and O-glycomics using porous graphitized carbon liquid chromatography-mass spectrometry. Mol Omics. 2020;16(4):355-63.

12. Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. J Proteome Res. 2008;7(4):1650-9.

13. Carrascal MA, Silva M, Ferreira JA, Azevedo R, Ferreira D, Silva AMN, et al. A functional glycoproteomics approach identifies CD13 as a novel E-selectin ligand in breast cancer. Biochim Biophys Acta Gen Subj. 2018;1862(9):2069-80.

14. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

15. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091-3.

16. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics. 2013;29(5):661-3.

17. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

18. Aviles EC, Goodrich LV. Configuring a robust nervous system with Fat cadherins. Semin Cell Dev Biol. 2017;69:91-101.

19. Rambaldi B, Kim HT, Arihara Y, Asano T, Reynolds C, Manter M, et al. Phenotypic and functional characterization of the CD6-ALCAM T-cell co-stimulatory pathway after allogeneic cell transplantation. Haematologica. 2022;107(11):2617-29.

20. Slattery ML, Lundgreen A, Wolff RK. MAP kinase genes and colon and rectal cancer. Carcinogenesis. 2012;33(12):2398-408.

21. Zhou G, Yang J, Song P. Correlation of ERK/MAPK signaling pathway with proliferation and apoptosis of colon cancer cells. Oncol Lett. 2019;17(2):2266-70.

22. Deschepper FM, Zoppi R, Pirro M, Hensbergen PJ, Dall'Olio F, Kotsias M, et al. L1CAM as an E-selectin Ligand in Colon Cancer. Int J Mol Sci. 2020;21(21).

23. Suzuki N, Numakawa T, Chou J, de Vega S, Mizuniwa C, Sekimoto K, et al. Teneurin-4 promotes cellular protrusion formation and neurite outgrowth through focal adhesion kinase signaling. FASEB J. 2014;28(3):1386-97.

24. Bai JJ, Tan CD, Chow BKC. Secretin, at the hub of water-salt homeostasis. Am J Physiol Renal Physiol. 2017;312(5):F852-F60.

25. Afroze S, Meng F, Jensen K, McDaniel K, Rahal K, Onori P, et al. The physiological roles of secretin and its receptor. Ann Transl Med. 2013;1(3):29.

26. Ferreira JA, Magalhaes A, Gomes J, Peixoto A, Gaiteiro C, Fernandes E, et al. Protein glycosylation in gastric and colorectal cancers: Toward cancer detection and targeted therapeutics. Cancer Lett. 2017;387:32-45.

27. Holst S, Wuhrer M, Rombouts Y. Glycosylation characteristics of colorectal cancer. Adv Cancer Res. 2015;126:203-56.

28. Tomlinson J, Wang JL, Barsky SH, Lee MC, Bischoff J, Nguyen M. Human colon cancer cells express multiple glycoprotein ligands for E-selectin. Int J Oncol. 2000;16(2):347-53.

29. Kobayashi K, Matsumoto S, Morishima T, Kawabe T, Okamoto T. Cimetidine inhibits cancer cell adhesion to endothelial cells and prevents metastasis by blocking E-selectin expression. Cancer Res. 2000;60(14):3978-84.

30. Kohler S, Ullrich S, Richter U, Schumacher U. E-/P-selectins and colon carcinoma metastasis: first in vivo evidence for their crucial role in a clinically relevant model of spontaneous metastasis formation in the lung. Br J Cancer. 2010;102(3):602-9.

31. Sackstein R. The lymphocyte homing receptors: gatekeepers of the multistep paradigm. Curr Opin Hematol. 2005;12(6):444-50.

32. Hidalgo A, Peired AJ, Wild M, Vestweber D, Frenette PS. Complete identification of E-selectin ligands on neutrophils reveals distinct functions of PSGL-1, ESL-1, and CD44. Immunity. 2007;26(4):477-89.

33. Burdick MM, Chu JT, Godar S, Sackstein R. HCELL is the major E- and L-selectin ligand expressed on LS174T colon carcinoma cells. J Biol Chem. 2006;281(20):13899-905.

34. Thomas SN, Zhu F, Schnaar RL, Alves CS, Konstantopoulos K. Carcinoembryonic antigen and CD44 variant isoforms cooperate to mediate colon carcinoma cell adhesion to E- and L-selectin in shear flow. J Biol Chem. 2008;283(23):15647-55.

35. Larsson A, Johansson ME, Wangefjord S, Gaber A, Nodin B, Kucharzewska P, et al. Overexpression of podocalyxin-like protein is an independent factor of poor prognosis in colorectal cancer. Br J Cancer. 2011;105(5):666-72.

36. Boyaval F, Dalebout H, Van Zeijl R, Wang W, Farina-Sarasqueta A, Lageveen-Kammeijer GSM, et al. High-Mannose N-Glycans as Malignant Progression Markers in Early-Stage Colorectal Cancer. Cancers (Basel). 2022;14(6).

37. Carrascal MA, Silva M, Ramalho JS, Pen C, Martins M, Pascoal C, et al. Inhibition of fucosylation in human invasive ductal carcinoma reduces E-selectin ligand expression, cell proliferation, and ERK1/2 and p38 MAPK activation. Mol Oncol. 2018;12(5):579-93.

38. Colomb F, Krzewinski-Recchi MA, Steenackers A, Vincent A, Harduin-Lepers A, Delannoy P, et al. TNF up-regulates ST3GAL4 and sialyl-Lewisx expression in lung

epithelial cells through an intronic ATF2-responsive element. Biochem J. 2017;474(1):65-78.

39. Breen KC, Coughlan CM, Hayes FD. The role of glycoproteins in neural development function, and disease. Mol Neurobiol. 1998;16(2):163-220.

40. Schultheis M, Diestel S, Schmitz B. The role of cytoplasmic serine residues of the cell adhesion molecule L1 in neurite outgrowth, endocytosis, and cell migration. Cell Mol Neurobiol. 2007;27(1):11-31.

41. Yamasaki M, Thompson P, Lemmon V. CRASH syndrome: mutations in L1CAM correlate with severity of the disease. Neuropediatrics. 1997;28(3):175-8.

42. Bateman A, Jouet M, MacFarlane J, Du JS, Kenwrick S, Chothia C. Outline structure of the human L1 cell adhesion molecule and the sites where mutations cause neurological disorders. EMBO J. 1996;15(22):6050-9.

43. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. Nature. 2017;541(7637):321-30.

44. He X, Xu C. Immune checkpoint signaling and cancer immunotherapy. Cell Res. 2020;30(8):660-9.

45. Liu K, Tan S, Jin W, Guan J, Wang Q, Sun H, et al. N-glycosylation of PD-1 promotes binding of camrelizumab. EMBO Rep. 2020;21(12):e51444.

46. Lu D, Xu Z, Zhang D, Jiang M, Liu K, He J, et al. PD-1 N58-Glycosylation-Dependent Binding of Monoclonal Antibody Cemiplimab for Immune Checkpoint Therapy. Front Immunol. 2022;13:826045.

47. Silva M, Martin KC, Mondal N, Sackstein R. sLeX Expression Delineates Distinct Functional Subsets of Human Blood Central and Effector Memory T Cells. J Immunol. 2020;205(7):1920-32.

48. Silva M, Videira PA, Sackstein R. E-Selectin Ligands in the Human Mononuclear Phagocyte System: Implications for Infection, Inflammation, and Immunotherapy. Front Immunol. 2017;8:1878.

49. Cohen EN, Fouad TM, Lee BN, Arun BK, Liu D, Tin S, et al. Elevated serum levels of sialyl Lewis X (sLe(X)) and inflammatory mediators in patients with breast cancer. Breast Cancer Res Treat. 2019;176(3):545-56.

50. Trinchera M, Aronica A, Dall'Olio F. Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. Biology (Basel). 2017;6(1).

51. Magalhaes A, Marcos-Pinto R, Nairn AV, Dela Rosa M, Ferreira RM, Junqueira-Neto S, et al. Helicobacter pylori chronic infection and mucosal inflammation switches the human gastric glycosylation pathways. Biochim Biophys Acta. 2015;1852(9):1928-39.

52. Hirakawa M, Takimoto R, Tamura F, Yoshida M, Ono M, Murase K, et al. Fucosylated TGF-beta receptors transduces a signal for epithelial-mesenchymal transition in colorectal cancer cells. Br J Cancer. 2014;110(1):156-63.

53. Geiger T. Tackling tumor complexity with single-cell proteomics. Nat Methods. 2023;20(3):324-6.

54. De Vargas Roditi L, Jacobs A, Rueschoff JH, Bankhead P, Chevrier S, Jackson HW, et al. Single-cell proteomics defines the cellular heterogeneity of localized prostate cancer. Cell Rep Med. 2022;3(4):100604.

55. Li M, Zuo J, Yang K, Wang P, Zhou S. Proteomics mining of cancer hallmarks on a single-cell resolution. Mass Spectrom Rev. 2023.

56. Drokhlyansky E, Smillie CS, Van Wittenberghe N, Ericsson M, Griffin GK, Eraslan G, et al. The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell. 2020;182(6):1606-22 e23.

57. Klussmeier A, Aurich S, Niederstadt L, Wiedenmann B, Grotzinger C. Secretin Receptor as a Target in Gastrointestinal Cancer: Expression Analysis and Ligand Development. Biomedicines. 2022;10(3).

58. Lee M, Waser B, Reubi JC, Pellegata NS. Secretin receptor promotes the proliferation of endocrine tumor cells via the PI3K/AKT pathway. Mol Endocrinol. 2012;26(8):1394-405.

59. Buchsbaum DJ. Experimental tumor targeting with radiolabeled ligands. Cancer. 1997;80(12 Suppl):2371-7.

60. Li X, Peng X, Yang S, Wei S, Fan Q, Liu J, et al. Targeting tumor innervation: premises, promises, and challenges. Cell Death Discov. 2022;8(1):131.

61. Gysler SM, Drapkin R. Tumor innervation: peripheral nerves take control of the tumor microenvironment. J Clin Invest. 2021;131(11).

62. Cervantes-Villagrana RD, Albores-Garcia D, Cervantes-Villagrana AR, Garcia-Acevez SJ. Tumor-induced neurogenesis and immune evasion as targets of innovative anti-cancer therapies. Signal Transduct Target Ther. 2020;5(1):99.

63. Ferreira IG, Carrascal M, Mineiro AG, Bugalho A, Borralho P, Silva Z, et al. Carcinoembryonic antigen is a sialyl Lewis x/a carrier and an E-selectin ligand in non-small cell lung cancer. Int J Oncol. 2019;55(5):1033-48.

64. Silverman DA, Martinez VK, Dougherty PM, Myers JN, Calin GA, Amit M. Cancer-Associated Neurogenesis and Nerve-Cancer Cross-talk. Cancer Res. 2021;81(6):1431-40.

Supporting Figures

Sofia Cotton[1,2], Dylan Ferreira[1,2,3], Marta Relvas-Santos[1,2,3,4], Andreia Brandão[1], Luís Pedro Afonso[1,5], Andreia Miranda[1,2,3], Eduardo Ferreira[1], Beatriz Santos[1,2], Martina Gonçalves[1,2], Lúcio Lara Santos[1,6,7], André M N Silva[2,4], José Alexandre Ferreira[1,2]

[1]Research Center of IPO-Porto (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO-Porto) / Porto Comprehensive Cancer Center (P.ccc) Raquel Seruca, Porto, Portugal; [2]Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal; [3]i3S – Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal; [4]LAQV-REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, Porto, Portugal; [5]Pathology Department, Portuguese Oncology Institute of Porto, Porto, Portugal; [6]Health School of University Fernando Pessoa, Porto, Portugal; [7]Department of Surgical Oncology, Portuguese Oncology Institute of Porto, Porto, Portugal.

Corresponding author

José Alexandre Ferreira
Experimental Pathology and Therapeutics Group,
Research Centre, Portuguese Oncology Institute of Porto,
R. Dr. António Bernardino de Almeida 4200-072 Porto,
Portugal; Tel. +351 225084000 (ext. 5111).
Email: jose.a.ferreira@ipoporto.min-saude.pt

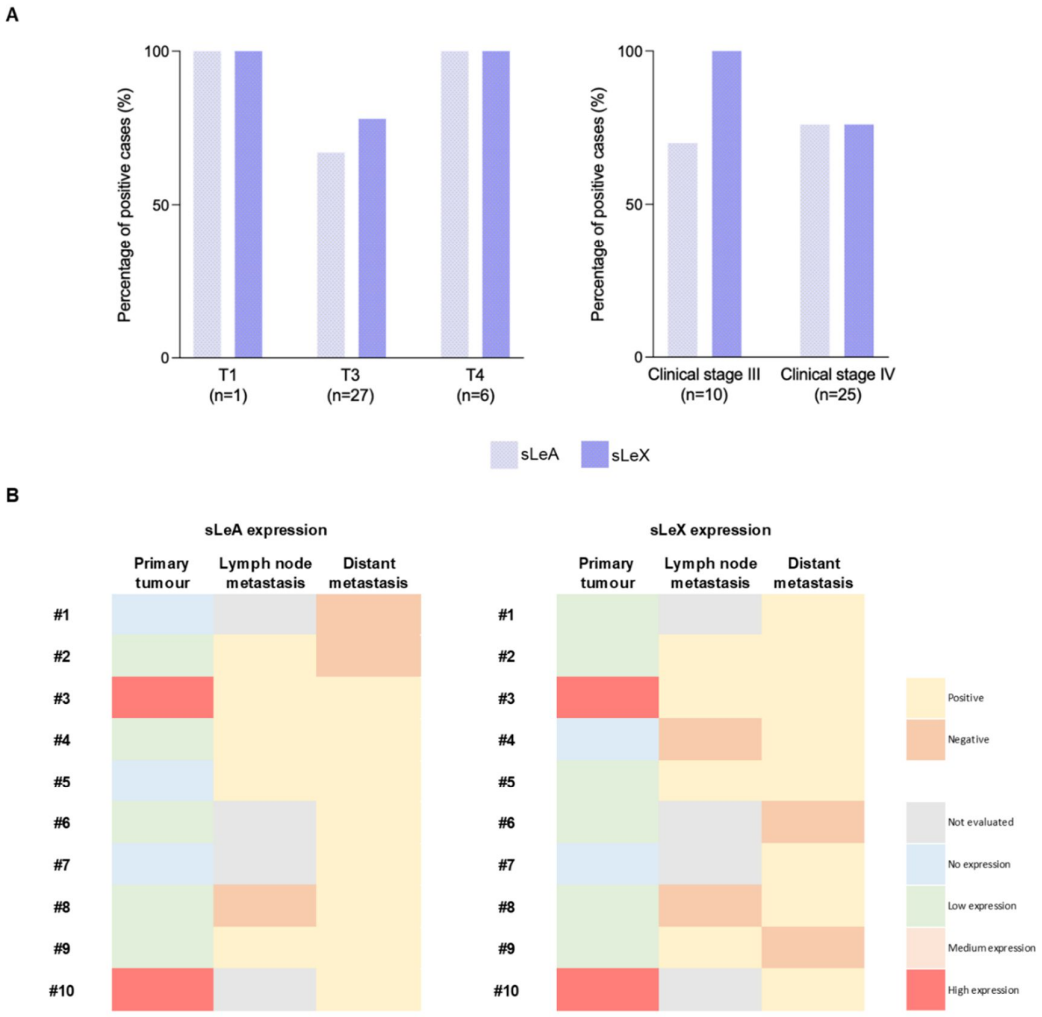


Figure S1. SLeA and sLeX are abundantly expressed in primary tumours and to a high extent in lymph node and distant metastases. A) SLe glycoepitopes expression according to the degree of invasion (pT) and clinical staging. A Chi-square analysis was performed and showed no associations of sLeA and sLeX with T (sLeA: p=0.25; sLeX: p=0.512) and clinical stages (sLeA: p=0.389; sLeX: p=0.109). **B) SLe glycoepitopes expression in primary tumours vs lymph node and distant metastases.**

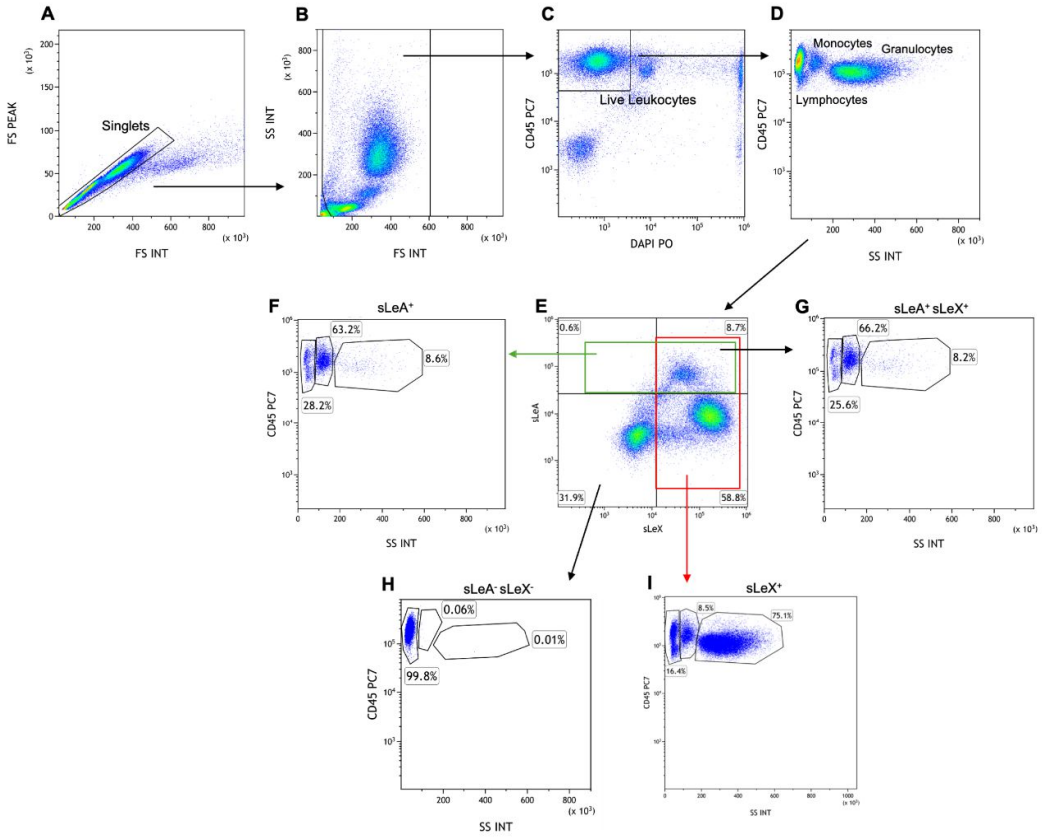


Figure S2. Flow cytometric gate strategy to identify sLeA and sLeX positive leukocytes in peripheral blood of healthy donors. (A) Singlets are selected according to forward scatter-height (FSC-H) versus forward scatter-area (FSC-A). (B) Leukocytes are selected, and erythrocytes and debris were excluded using side scatter-area (SSC-A) versus FSC-A. (C) Live leukocytes are isolated by selecting the CD45 positive and DAPI negative cells. (D) Live leukocytes are subtyped into lymphocytes, monocytes, and granulocytes according to CD45 expression and SSC-A dispersion. (E) sLeA+/sLeX+ leukocytes are isolated using fluorochrome-coupled specific antibodies. (F) Isolating the sLeA positive leukocytes, 30.0% of the cells are lymphocytes, 55.4% are monocytes, and 7.9% are granulocytes. (G) Isolating the double positive leukocytes for sLeA and sLeX, 24.9% of the cells are lymphocytes, 60.0% are monocytes, and 7.9% are granulocytes. (H) Isolating the double negative leukocytes for sLeA and sLeX, all the events are lymphocytes. (I) Isolating the sLeX positive leukocytes, 13.6% of the cells are lymphocytes, 0.3% are monocytes, and 84.5% are granulocytes.

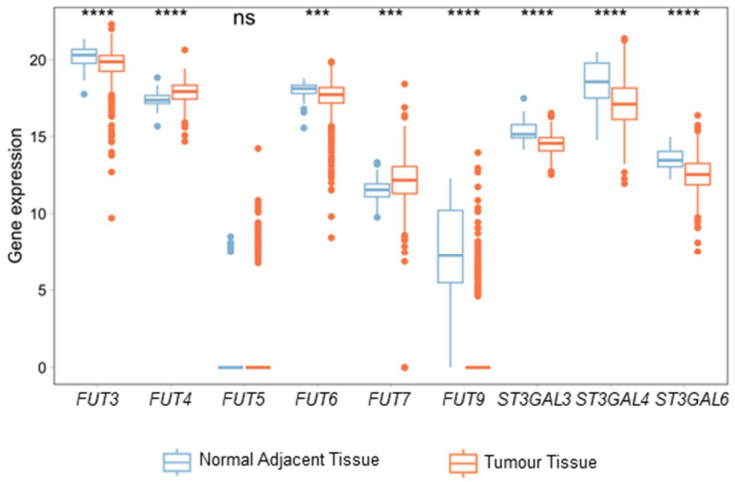


Figure S3. Expression of glycosyltransferases involved in sLe biosynthesis in CRC and normal adjacent tissues. The analysis of transcripts encoding key glycosyltransferases involved in sLe antigen biosynthesis was conducted on 598 tumour samples and 51 normal adjacent tissue samples from the TCGA database. *FUT4* and *FUT7* mostly linked to sLeX biosynthesis, were significantly overexpressed, whereas all the other glycosyltransferases were downregulated. *FUT5* is residually expressed by few cases in healthy tissues and CRC.

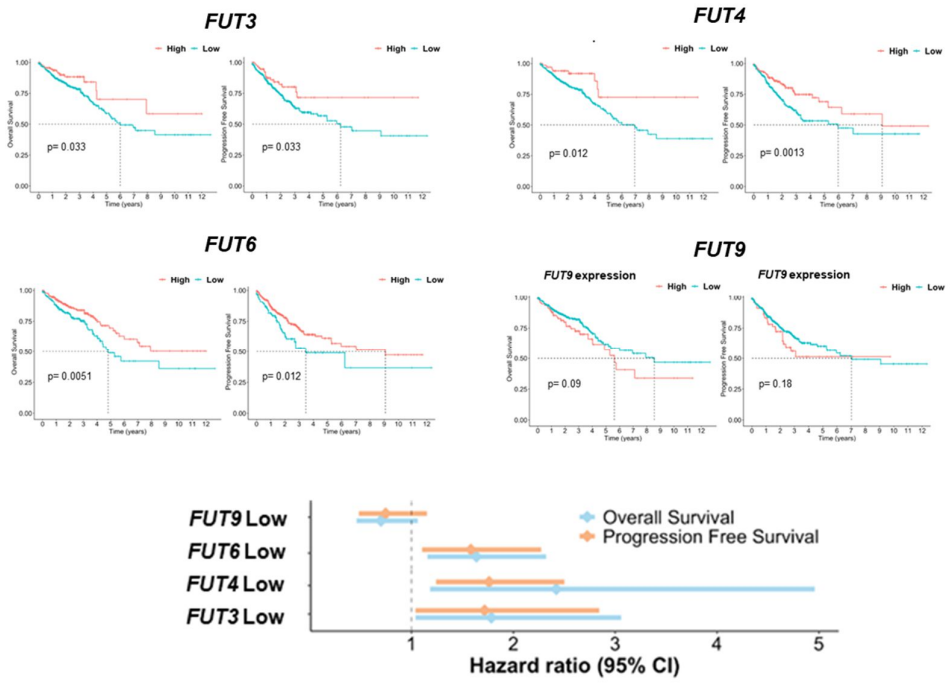


Figure S4. Overexpression of highly correlated *FUT3*, *FUT4* and *FUT6* associate with a favorable prognosis in CRC, whereas *FUT9* presents a trend link to worst prognosis. TCGA analysis revealed that high *FUT3*, *FUT4*, and *FUT6* transcripts associated with favourable overall and disease-free survivals. Also, decreased *FUT3*, *FUT4*, and *FUT6* correlated with higher cancer-related mortality and progression risk (FUT3low: HR-1.8, p=0.04/ HR-1.7, p=0.04; *FUT4*low: HR-2.4, p=0.02/ HR-1.8, p=0.002; FUT6low: HR=1.6, p=0.01/ HR-1.6, p=0.01). Notably, *FUT3* and *FUT6* showed a strong positive correlation. Conversely, *FUT9* overexpression, inversely correlated with *FUT3* and *FUT6*, trended with decreased overall survival. Additionally, tumors with low *FUT9* mRNA levels had lower hazard ratios for overall (HR: 0.7, p=0.09) and disease-free survivals (HR:0.7, p=0.18).

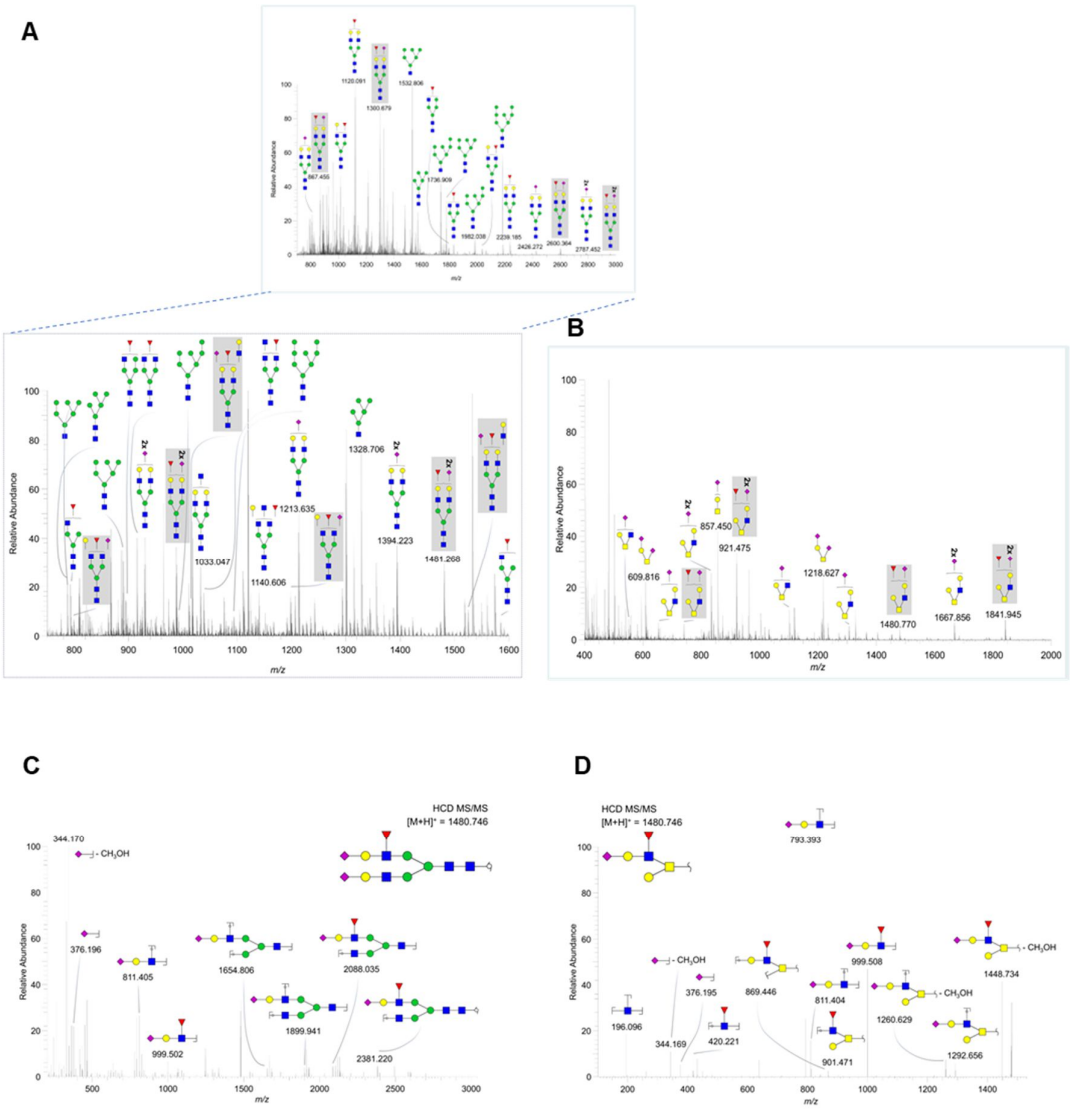


Figure S5. A) Typical MS spectra reflecting the *N*-glycome of metastatic colorectal tumor tissues used in this study. Dashed glycans are consistent with carrying sialylated Lewis antigens as terminal epitopes. **B) Typical MS spectra reflecting the *O*-glycome of metastatic stage IV colorectal tumor tissues used in this study. C) MS/MS of a complex *N*-glycan carrying sialylated Lewis antigens.** The spectrum highlights typical glycosidic linkage fragments consistent with the presence of sialylated Lewis antigens. **D) MS/MS of an extended core 2 *O*-glycan carrying sialylated Lewis antigens.**

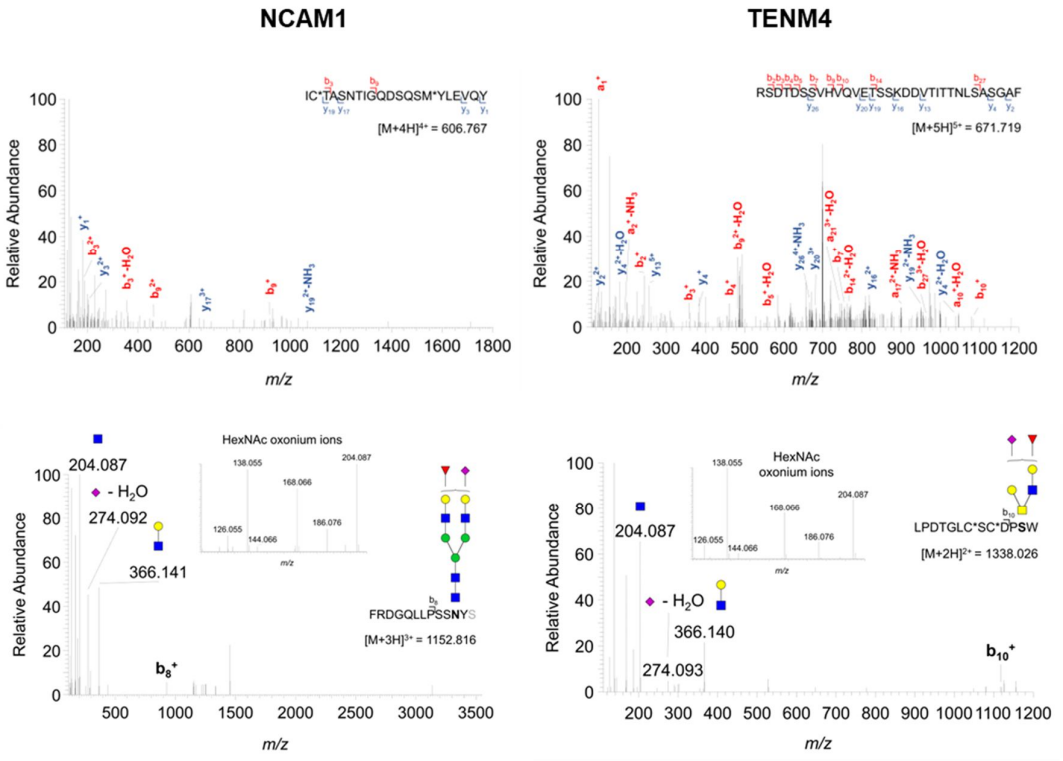


Figure S6. MS/MS spectra for unglycosylated peptides (top panels) and glycopeptides (bottom panels) supporting the expression of neuroendocrine-related glycoproteins NCAM1 and TENM4 potentially carrying sialylated Lewis antigens. * Modification of cysteine with carbamidomethyl

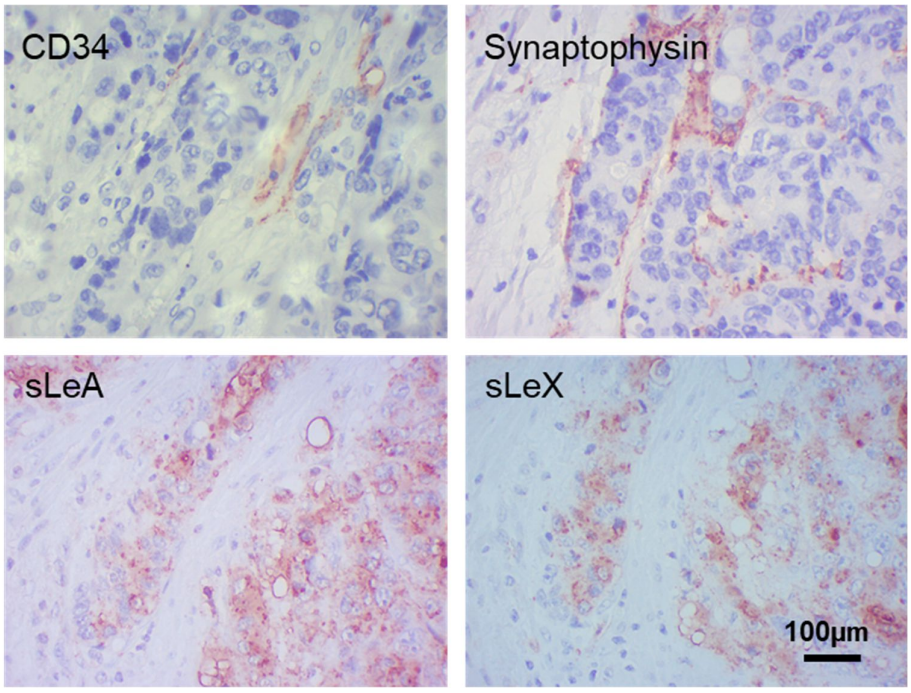


Figure S7. CRC tumours pooled for E-selectin enrichment and glycoproteomic analysis showed evidence of relevant neovascularization (CD34+), neuroendocrine features (Synaptophysin+), and high levels of sialylated Lewis antigens in the same tissue area.

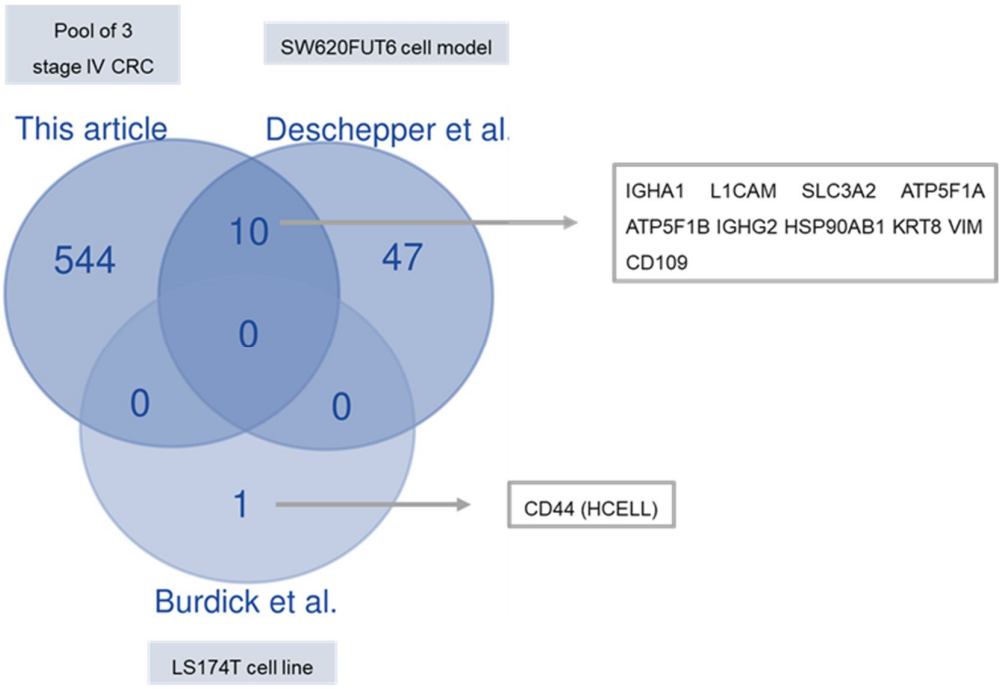


Figure S8. Venn Diagram highlighting E-selectin affinity glycoproteins identified to date in the context of CRC. The Venn Diagram highlights some common proteins between our study and the analysis of a sialyl-Lewis overexpressing cell line, including L1CAM which is a typical nervous system-associated glycoprotein.

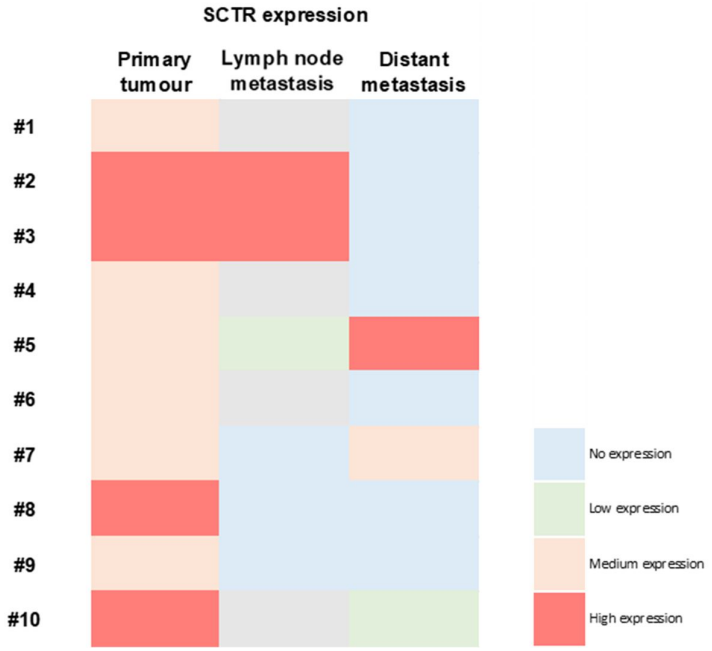


Figure S9. The SCTR receptor is highly expressed in colorectal tumours and may also be found in lymph node and distant metastases.

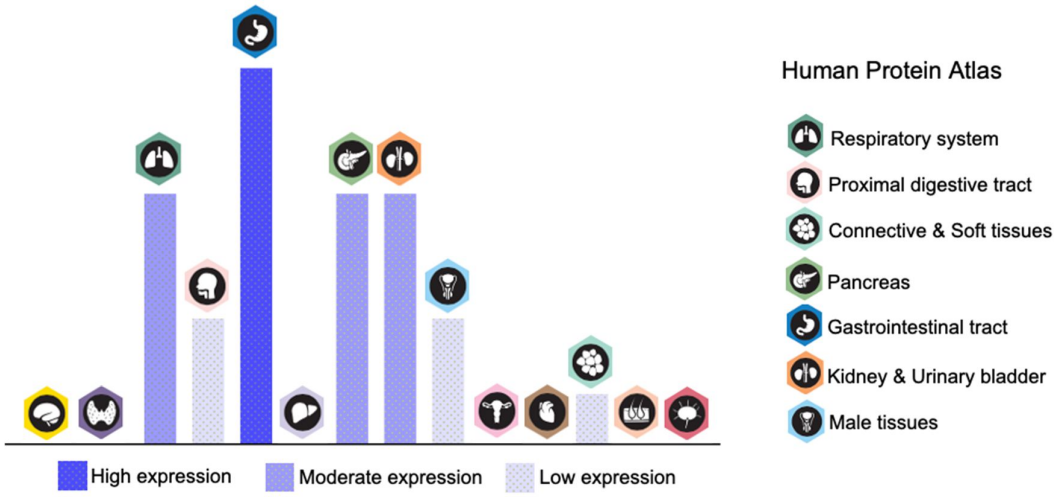


Figure S10. SCTR presents relevant expression in several human healthy organs (respiratory system; gastrointestinal tract; pancreas; kidney and urinary bladder). According to the Human Protein Atlas (A), SCTR exhibits high expression in the gastrointestinal tract, moderate expression in the respiratory system, pancreas, kidney, and urinary bladder, and low to vestigial expression in the proximal digestive tract, male tissues, and connective and soft tissues. Expression levels are defined as high, moderate or low, according to the Human Protein Atlas (www.proteinatlas.org/).

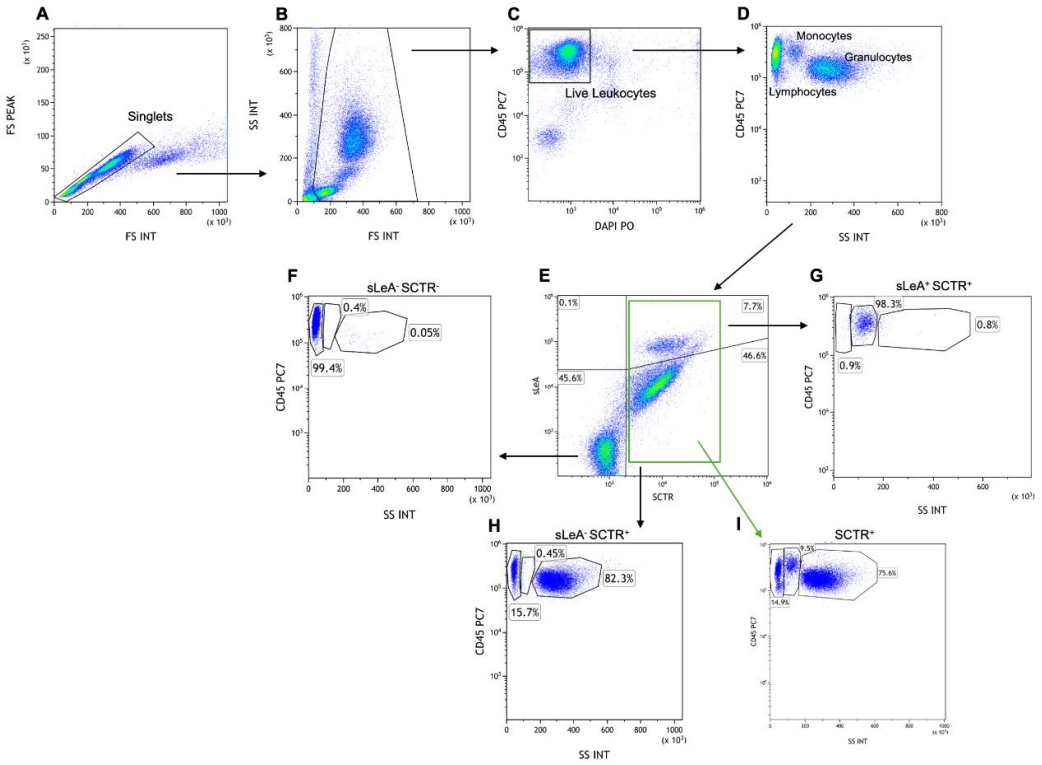


Figure S11. Flow cytometric gate strategy to identify sLeA and SCTR positive leukocytes in peripheral blood of healthy donors. (A) Singlets are selected according to FSC-H versus FSC-A. (B) Leukocytes are selected, and erythrocytes and debris were excluded using SSC-A versus FSC-A. (C) Live leukocytes are isolated by selecting the CD45 positive and DAPI negative cells. (D) Live leukocytes are subtyped into lymphocytes, monocytes, and granulocytes according to CD45 expression and SSC-A dispersion. (E) sLeA+/SCTR+ leukocytes are isolated using fluorochrome-coupled specific antibodies. (F) Isolating the double negative leukocytes for sLeA and SCTR, around 100% of the cells are lymphocytes. (G) Isolating the double positive leukocytes for sLeA and SCTR, 22.5% of the cells are lymphocytes, 67.6% are monocytes, and 3.0% are granulocytes. (H) Isolating the sLeA negative and SCTR positive leukocytes, 15.7% of the cells are lymphocytes, 0.45% are monocytes, and 82.3% are granulocytes. (I) Isolating the SCTR positive leukocytes, 17.2% of the cells are lymphocytes, 8.6% are monocytes, and 72.1% are granulocytes.

CONCLUDING REMARKS & FUTURE PERSPECTIVES

CONCLUDING REMARKS & FUTURE PERSPECTIVES

Precision oncology for esophageal and colorectal cancer hinges on molecular understanding, with systems biology and omics data integration being pivotal. This approach identifies key proteins for intervention, requiring multi-layered biomolecular information, combining genomics, transcriptomics, proteomics, and post-translational modifications (PTM) identification. Proteomics, facilitated by genomics-informed workflows and PTM analysis, is essential for pinpoint neoantigens for clinical implementation. Glycosylation patterns, especially the presence of sialylated glycans, such as STn and sLeA antigens, on cell surface correlate with cancer severity [1]. Recent glycoproteomic studies, specially, in the context of esophageal and colorectal cancers, have unveiled the impact of cancer-associated glycans, like STn and sLeA, which modify protein function and serve as potential biomarkers. These glycans, while not exclusive to malignancies, have a limited presence in healthy tissues, marking them as distinctive cancer signatures. Functional analyses indicate that the presence of STn promotes cancer metastasis by enabling cell movement, tissue invasion, and evasion of the immune system [2]. Indeed, emerging research on circulating tumour cells (CTC) evidences the expression of this short chain O-glycan on these cells that have the capacity to recapitulate the primary lesion on distant lesions. Moreover, studies suggest that STn antigen analysis could enhance liquid biopsy sensitivity with an impact on patient management. The current implemented serological analysis of these antigens inform clinical prognosis and treatment monitoring, however without a clear impact on patient management [1]. The accumulated evidence from both clinical and functional perspectives now strongly supports the development of interventions targeting these glycan alterations in clinical practice.

The current challenge lies in distinguishing cancer-specific glycopeptides from those in benign conditions to enhance the efficacy of glycan focused therapies. Despite the lack of extensive research on tumour-specific STn/sLeA glycoproteomes, the pursuit of such studies is crucial for the advancement of glycoproteomics in clinical applications. Integrating glycoproteomics with proteogenomics could unravel functional glycoepitopes, enhancing our understanding of tumour development and aiding in the refinement of serological tests. Additionally, mapping glycosignatures on therapeutic targets like PD-L1 may reveal resistance mechanisms to current innovative treatments, providing avenues for personalized medicine. Moreover, the glycosylation status of cancer cells plays a role in immune evasion and metastasis, with glycoproteins interacting with siglecs and selectins on immune cells, potentially leading to new immune checkpoint inhibitors [3]. The link

between STn expression and chemoresistance also suggests that glycoproteomic profiling could identify patients who might benefit from alternative therapeutic strategies [4].

Focusing on the glycosylation changes in metastasis and disease progression, especially the roles of STn and sLeA, can reveal critical insights for clinical applications. Integrating glycoproteomics into a multi-omic approach is anticipated to not only refine patient stratification but also to identify novel glycobiomarkers for targeted therapies, ensuring a more personalized and effective cancer treatment paradigm.

Clinical management of EC remains challenging due to the absence of cancer-specific biomarkers, either for prognosis or treatment. On this matter, we conducted an analytical study focusing the identification of relevant glycoproteoforms on EC that could represent a step forward into the personalized medicine. Our study demonstrates that aggressive tumours with poor prognosis overexpress the STn antigen, potentially reflecting dysregulation in O-glycosylation pathways. High STn expression correlates with increased distant metastases, supported by its presence in CTC and association with cellular invasion in vitro. To our knowledge, this is the first report of STn expression in esophageal CTC. Future studies must evaluate its clinical value in liquid biopsies for early detection, prognosis, and treatment decisions. Appling an in-house developed algorithm (Target Score algorithm), allied with a glycoproteomic analysis, we were able to identify clinically relevant glycoproteins. Future research should validate our findings in larger patient populations and elucidate the underlying mechanisms of STn dysregulation. This underscores the need for personalized approaches to improve EC outcomes. In our study, preliminary findings using a database identified GLUT1 as a cancer-associated protein modified with STn antigens. This suggests potential for EC patient stratification based on GLUT1-STn overexpression. Immunoassay-based studies imply dependency on antibody specificity, warranting comprehensive mass spectrometry interrogation of the human proteome for validation. Our investigation supports previous reports on protein glycosignatures' clinical relevance, albeit limited by the original proteomics data's scope. While not initially designed for detailed glycoproteomic analysis, our study sets a precedent for tailored approaches in EC research, advocating for precision medicine guided by glycobiomarkers.

The study in CRC extensively analyses glycoproteins to identify potential targetable signatures using the Target Score algorithm, successful in various cancers [5, 6]. Multi-omics approach has identified the G protein-coupled SCTR as a top-ranked glycoprotein, warranting in-depth evaluations due to its relatively unexplored nature in CRC. Notably, SCTR has been identified in healthy colon cells, suggesting origins from the peripheral nervous system. Although poorly studied in CRC, SCTR is overexpressed in various cancers and associated with worse prognosis, suggesting potential roles in disease progression. Elevated SCTR and SCT levels in CRC may indicate unknown alterations in

fluid balance and colonic motility, raising fundamental questions about their roles in colorectal carcinogenesis. We propose SCTR as a primary carrier of sLe antigens, potentially conferring cancer specificity and enabling adhesion to E-selectin, but further validation is necessary. Additionally, we linked SCTR to a group of sLe-related glycogenes associated with worse prognosis, providing insights into their biological functions in CRC. This study unveils a strategy for unravelling cancer glycoproteome complexity and identifying clinically relevant signatures. The link between the sialoglycoproteome and the colonic peripheral nervous system is highlighted, suggesting a novel avenue for research at the intersection of neurology and oncology. Deeper exploration is needed to understand the precise mechanisms driving CRC and the role of protein glycosylation in cancer progression. Understanding these mechanisms may pave the way for precision oncology and novel strategies to counteract metastasis development. Additionally, our investigation revealed SCTR overexpression in metastatic colorectal tumours, suggesting potential cancer specificity. Further studies leveraging comprehensive glycoproteome characterization are necessary to validate these findings and develop targeted approaches for aggressive CRC tumours and metastases.

Overall, this work provides sufficient grounds on the pattern of glycosylation of digestive tumours, vowing the development of more accurate diagnosis and prognosis tools, such as STn-based liquid biopsy, and putative glycan-based therapeutic solutions. Furthermore, the identification of cancer-specific glycoproteoforms of GLUT1 and SCTR creates the necessary rationale to foster more in-depth studies envisaging glycan-based precision medicine. Future studies should focus on elucidating the biological roles of both glycoproteins in esophageal and colorectal cancer pathogenesis. Additionally, comprehensive investigations are warranted to assess the presence of GLUT1 and SCTR antigens in the sera and CTC of cancer patients. Furthermore, a systematic mapping of GLUT1 glycoforms, alongside SCTR and other cancer-associated glycoproteins (TENM4, PD-L1, e.g.), holds promise for providing crucial structural data essential for the development of highly specific biomarkers and targeted therapeutic strategies.

References

1. Fernandes, E., et al., Esophageal, gastric and colorectal cancers: Looking beyond classical serological biomarkers towards glycoproteomics-assisted precision oncology. Theranostics, 2020. 10(11): p. 4903-4928.

2. Pinho, S., et al., Biological significance of cancer-associated sialyl-Tn antigen: modulation of malignant phenotype in gastric carcinoma cells. Cancer Lett, 2007. 249(2): p. 157-70.

3. Peixoto, A., et al., A roadmap for translational cancer glycoimmunology at single cell resolution. J Exp Clin Cancer Res, 2022. 41(1): p. 143.

4. Peixoto, A., et al., Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front Oncol, 2019. 9: p. 380.

5. Peixoto, A., et al., Glycoproteomics identifies HOMER3 as a potentially targetable biomarker triggered by hypoxia and glucose deprivation in bladder cancer. J Exp Clin Cancer Res, 2021. 40(1): p. 191.

6. Fernandes, E., et al., Nucleolin-SIe A Glycoforms as E-Selectin Ligands and Potentially Targetable Biomarkers at the Cell Surface of Gastric Cancer Cells. Cancers (Basel), 2020. 12(4).