LOMBADA

2019

Validation of novel biomarkers for colorectal cancer detection and production of novel antibodies against E-selectin ligands

Roberta Zoppi

Roberta Zoppi

MSc

Validation of novel biomarkers for colorectal cancer detection and production of novel antibodies against E-selectin ligands

Dissertação para obtenção do Grau de Doutor em

Biologia- Especialidade em Biologia Celular

Orientador: Prof. Doutora Paula Alexandra Quintela Videira, Professora Auxiliar da Faculdade de Ciências e Tecnologia da Universidade NOVA de Lisboa

Co-orientador: Inv. Doutor Carlos Manuel Mendes Novo, Investigador Auxiliar do Instituto de Higiene e Medicina Tropical da Universidade NOVA de Lisboa

Júri:

Presidente: Prof. Doutor Paulo Manuel Assis Loureiro Limão Vieira

Arguentes;
Prof. Doutor João Manuel Braz Gonçalves
Prof. Doutor Hugo Ricardo Loureiro Soares

Vogais:
Prof. Doutor Paulo Sérgio de Matos Figueira da Costa
Prof. Doutora Paula Alexandra Quintela Videira
Prof. Doutora Margarida Casal Ribeiro Castro Caldas Braga

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

November, 2019

Validation of novel biomarkers for colorectal cancer detection and production of novel antibodies against E-selectin ligands

Copyright ©Roberta Zoppi, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

The Faculty of Sciences and Technology and the NOVA University of Lisbon have the right, forever and without geographical limits, to file and publish this dissertation through printed copies reproduced in paper or by digital means, or by any other mean known or that is invented, and to disclose it through scientific repositories and to allow its copying and distribution for non-commercial educational or research purposes, provided that the author and editor are credited.

Ad maiora semper

Acknowledgments

The PhD journey is a challenging professional and personal growth.
None of this would be possible without the people that surround and support me.

Firstly, I would like to thank my supervisor Dr. Paula Videira for believing in my potential and for giving me this great opportunity. I will always be thankful for the great care in supporting my decisions.

A special thanks goes to my co-supervisor Carlos Novos, and to my collaborators, Dr. Alexandre Ferreira, Dr. Lucio Santos and Dr Sandra Van Vliet for the unconditional trust.

To Dr Daniel Spencer and Helen Williamson, for being the best supporter and for being fundamental in defining my future path.

A lab is not a lab without the beautiful people sharing joys and disappointments of everyday experiments, thus thanks go to Mylene, Zelia, Danielle, Diana, Gonçalo, Rita and Erika.

To my "MSc" super friends and supporters Constança (stanza), Tiago (T) and Carlota (lota). The best master student a worker could ask for and the best souls a person could look for.

To my PhD besties Lily, Fanny and Tiago (Tj) who shared with me every hope, every laugh and every tear. Probably I would have achieved half of what I have accomplished without you around. I am very lucky to be your friend

To that beautiful group that goes under the name of GLYCOCAN. I could never imagine a more perfect team that this one. You are all special people and I am glad the destiny brought us together. A particular thanks must go to my partner in crime Giulia and my five-star-rating friend Thanos. I am very thankful to have you in my life.

To the family I have chosen and that is distributed around the world. Joji, Anisi, Ghedy, Tian, Pini, Sandra, Alessandra, Anindita, Sarah, Isa, Jul, Jason. You make me a better person and I miss you everyday.

To the family I have chosen in Portugal: Joan, Arianna, Rita, Guily, Bruno and Jack. Thank you for all the good times and for making my life in Portugal unforgettable. To my sister Nicolyini, without who my life in Lisbon would have been another story. You made the difference.

Finally, the best thanks must go to my family, who I did not chose but I am so lucky. to have.
To my brother and best friend Fulvio, who compensate my irruent personality and who always have been my number one supporter. We brought brotherhood to the next level.
To my mom and dad without who I will not be the person who I am today. You taught me how to be strong and independent but also to be kind and generous with the others. Your unconditional support is what motivates me everyday to be the best version of myself. I love you very much.

Resumo

Em 2018, o cancro colorectal (CRC) continua a ser o segundo tipo de cancro mais mortal, com 881,00 mortes dos 1.8 milhões de novos casos. A deteção de estádios avançados tem mais probabilidade de desenvolver recorrências, mesmo após o tratamento, o que leva à necessidade de criar um novo sistema para a deteção de estádios iniciais. Assim, compreender a biologia e a descoberta de bio marcadores são etapas importantes na investigação do cancro.

Vários estudos sobre glicosilação associada ao cancro revelaram que a glicosilação aberrante é uma característica universal em várias fases da transformação maligna e progressão tumoral. A glicosilação aberrante está associada com tempos de sobrevivência inferiores, progressão tumoral e metástase, nomeadamente, a sobre expressão dos antigénios *Thomsen-nouvelle* (Tn)-, T-, e sialil-T (sT). Os principais portadores dos antigénios sT e sTn identificados são a mucina MUC1 e a glicoproteina CD44v6. Por outro lado, os antigénios de Lewis e os seus derivados sialilados (Lex/sLex and Lea/sLeA) são os epitopos associados ao cancro mais proeminentes, em ambos as glicoproteínas e glicolípidos, já que a sua sobre expressão está relacionada com transformações malignas no CRC e pode levar a uma maior adesão e motilidade das células tumorais, resultando em metástase.

O projeto tem como objetivos descobrir novos bio marcadores com potencial para a deteção precoce de CRC e avaliar o potencial terapêuticos de anticorpos contra antigénios associados ao CRC, nomeadamente contra antigénios de Lewis, Tn, T- e sT. Pela mesma razão, glicoproteínas, como a MUC1, CAE e CD44 vão ser estudadas para o desenvolvimento e produção de novos anticorpos monoclonais.

Após a validação de bio marcadores, a tecnologia de hibridoma foi escolhida para produzir novos anticorpos contra sLe e/ou CD44, imunizando murganhos diretamente com proteínas de linhas celulares cancerígenas. O hibridoma contra CD44 mostrou marcação contra CD44 e $sLe^{x/a}$ e foi realizada uma triagem adicional. A segunda linha de hibridoma contra antigénios sLe também mostrou positividade contra vários lizados celulares totais de células de CRC. Dois clones foram selecionados para caracterização adicional. Os clones irão ser validados para a deteção precoce ou, potencialmente, tratamento de CRC.

Palavras-chave: Glicosilação, Cancro Colorectal, sialil Lewis, CD44, E-selectina, biomarcador tumoral, deteção precoce, anticorpos monoclonais, tecnologia de hibridoma, investigação translacional.

Abstract

In 2018, colorectal cancer (CRC) remains the second deadliest kind of cancer with 881,00 deaths of the 1.8 million new cases. Late stages detection is more likely to develop recurrences, even after treatment, leading to the necessity to create a new system to early stages detection. Thus, understanding the biology of the cancer and biomarker discovery are important steps in cancer research. Several studies on cancer-associated glycosylation revealed that aberrant glycosylation is a universal feature in various steps of malignant transformation and tumour progression. Aberrant glycosylation is associated with poor survival, cancer progression, and metastasis, such as overexpression of Thomsen-nouvelle (Tn)-, T-, and sialyl-T (sT) antigens. Main carriers of sT- and sTn-antigens were identified as the mucin MUC1 and CD44v6. On the other hand, the Lewis antigens and their sialylated derives (Lex/sLex and Lea/sLeA) are the most prominent cancer-associated epitopes on both glycoproteins and glycolipids, since their overexpression is related to CRC malignant transformations and may lead to increased tumour cell adhesion and motility, thereby resulting in metastasis.

The project aims to discover new potential biomarkers for CRC early detection and to evaluate the therapeutic potential of antibodies against CRC-associated antigens, namely against Lewis antigens and Thomsen-nouvelle (Tn)-, T-, and sialyl-T (sT) antigens. For the same reason glycoproteins, such as MUC1, CAE and CD44 will be studied for the development of novel monoclonal antibody production.

After biomarker validation, hybridoma technology has been chosen to produce novel antibodies against, sLe and/or CD44, immunizing the mice directly with cancer cell lines proteins. The hybridoma against CD44 show staining against CD44 and $sLe^{x/a}$ and further screening must be performed. The second hybridoma line against sLe antigens also shows positivity against different total cell lysate of CRC cells. Two clones have been selected for further characterization. The clones will be validated either for CRC early diagnosis or CRC treatment potential.

Keywords: Glycosylation, Colorectal cancer, sialyl Lewis, CD44, E-selectin, cancer biomarker, early diagnosis, monoclonal antibodies, hybridoma technology, translational research

Table of Contents

List of Tables

Abbreviations

A

AA	*Adenomatous polyps*
5-FU	*5-fluorouracile*
Ab	*Antibody*
ADCC	*Antibody-dependent cell-mediated cytotoxicity*
AE1/AE3	*Anti-cytokeratin*
Akt/PKB	*Protein kinase B*
APC	*Adenomatous polyposis coli*
APS	*Ammonium persulfate*
ATP	*Adenosine triphosphate*

B

Bcl-2	*B cell lymphoma 2*
BEAMing	*Beads, Emulsion, Amplification and Magnetics*
BFB	*Chromosomal breakage-fusion-bridge*
BSA	*Bovine serum albumin*
BSM	*Bovine submaxillary mucin*

C

C1GALT1	*B1,3-galactosyltransferase*
CA	*Carbohydrate antigen*
CAFs	*Cancer-associated fibroblasts*
CAM	*Cell-cell adhesion proteins*
CAR	*Chimeric antigen receptor*
CD	*Cluster of differentiation*
CDC	*Complement dependent cytotoxicity*
CDK	*Cyclin-dependent kinase*
cDNA	*Complementary Deoxyribonucleic acid*
CDRs	*Complementarity-determining regions*
CEA	*Carcinoembryonic antigen*
cfDNA	*Cell free DNA*
CIMP	*Cpg island methylator phenotype*
circRNAs	*Circular RNA*
CMS	*Consensus molecular subtypes*
COX2	*Cyclooxygenase 2*
CRC	*Colorectal cancer*
CSCs	*Cancer stem cells cancer*
CT	*Computed tomography*
CTCs	*Circulating tumour cells*
CTL	*Cytotoxic T lymphocytes*

D

DAB	*3,3'-diaminobenzidine*
DCBE	*Double contrast barium enema*
DDR	*DNA damage response*
DMSO	*Dimethyl sulfoxide*

E

E-Ig	*E-selectin-Ig*
EBV	*Epstein-barr virus*
ECM	*Extracellular matrix*
EGF	*Epidermal growth factor*
EGF-R	*Epidermal growth factor receptor*
ELISA	*Enzyme-linked immunosorbent assay*
EMT	*Epithelial to mesenchymal transition*
EpCAM	*Epithelial cell adhesion marker*
Eph	*Ephrins*
ER	*Endoplasmic reticulum*

F

FAK	*Focal adhesion kinase*
FAP	*Familial adenomatous polyposis*
FBS	*Fetal bovine serum*
FC	*Flow cytometry*
FFPE	*Formalin-fixed, paraffin embedded*
FGF	*Fibroblast growth factor*
FIT	*Immunochemical faecal occult blood test*
FL-DNA	*Fluorescent long DNA*
FOBT	*Faecal occult blood test*

FRET *Förster resonance energy transfer*

FSC *Forward scatter*

Fuc-TVIII *Fucosyltransferases 8*

FUT *Fucosyltransferases*

G

G6PD *Glucose-6-phosphate dehydrogenase*

GADPH *Glyceraldehyde 3-phosphate dehydrogenase*

GalNAc *N-acetylgalactosamine*

GALNT1 *N-acetylgalactosaminyltransferases 1*

GalT *Galactosyltransferase*

GBPs *Glycan binding proteins*

GF *Growth signals*

GlcNAc *N-acetylglucosamine*

GlcNAcTs *Acetylglucosaminyltransferases*

GLUT1 *Glucose transporters-1 and 3*

GnTs *N-acetylglucosaminyltransferase*

GS *Growth signals*

H

HAMA *Human antibody anti-mouse antibodies*

HAT *Hypoxanthine-aminopterin-thymidine*

HBP *Hexosamine biosynthetic pathway*

HDI *Human development index*

HER2 *Human epidermal growth factor receptor 2*

HGPRT *Hypoxanthine-guanine-phosphoribosyl transferase*

HIF-1α *Hypoxia-inducible factor 1- α*

HNPCC *Hereditary nonpolyposis*

HS *Heperan sulfate*

I

Ig *Immunoglobulin*

IGF-1 *Insulin-like growth factors*

IGF-1R *Insulin-like growth factor 1 receptor*

IgG *Immunoglobulin G*

IHC *Immunohistochemistry*

IP *Immunoprecipitation*

L

LacNAc *N-acetyllactosamine*

LC-ESI-MS/MS *Liquid chromatography – electrospray ionization - tandem mass spectrometry*

LKB1 *Liver kinase B1*

LTR *Truncated long-terminal residues*

M

M2-PK *Pyruvate kinase isoenzyme 2*

mAb *Monoclonal antibody*

MAPK *Mitogen-activated protein kinase*

mCRC *CRC with metastasis*

MDSCs *Myeloid-derived suppressor cells*

MET *Epithelial transition*

MFI *Median fluorescence intensity*

miRNA *Micro RNA*

MMPs *Matrix metalloproteinases*

mRNA *Messenger RNA*

MS *Mass spectrometry*

MSI-H *Microsatellite instability*

MSS *Microsatellite stability*

MTT *3-(4,5-Dimethylthiazol-2-yl-2,5-Diphenyltetrazolium Bromide)*

MUC *Mucins*

N

NCD *Noncommunicable disease*

NeuAc *N-acetylneuraminic acids*

NeuGc *N-glycolylneuraminic acid*

NK *Natural killer*

O

OGA *O-glucanase*

OS *Overall survival*

PBS *Phosphate buffered saline*

PCR *Polymerase chain reaction*

P

PD-1 *Programmed death-1 receptor*

PDGF *Platelet-derived growth factor*

PEG *Polyethylene glycol*

PFA *Paraformaldehyde*

PFK1 *The phosphofructokinase 1*

PFS *Progression-free survival*

PI3 *Phosphoinositide 3-kinase*

PIP3 *Phosphatidylinositol (3,4,5) trisphosphate*

PLA *In situ proximity ligation assay*

PPP *Pentose phosphate pathway*

pRb *Retinoblastoma protein*

PVDF *Polyvinylidene fluoride*

R

R0 *Tumour mass resection*

RLB *Reporter lysis buffer*

RNA *Ribonucleic acid*

ROS *Reactive oxygen species*

RT *Room temperature*

S

scFv *Single-chain variable fragment*

sDNA *DNA in stools*

SDS *Sodium dodecyl sulfate*

SDS-PAGE *Sodium dodecyl sulfate – polyacrylamide gel electrophoreses*

SLe$^{x/a}$ *Sialyl-Lewis$^{x/a}$*

SSC *Side scatter*

ST *And sialyltransferase*

ST3Gal-Ts *A2,3-sialyltransferases*

ST6Gal-I *A-galactoside α-2,6-sialyltransferase I*

STRs *Polymorphic short tandem repeat loci*

T

T- *Thomsen–Friedreich antigen*

TACAs *Tumour-associated carbohydrate antigens*

TBST *Tris-buffered saline- tween*

TE *Trypsin-EDTA*

TEMED *Tetramethylethylenediamine*

TERT *Telomerase reverse transcriptase*

TFH *T follicular helper*

TGF a *Tumour growth factor a*

TGFß *Transforming growth factor ß*

TIMPs *Tissue inhibitor metalloproteinases*

TME *Tumour microenvironment*

Tn- *Thomsen-nuovelle antigen*

Tregs *Regulatory T cells*

TSA *Total sialic acid*

TSP-1 *Trombospondin-1*

V

VEGF *Vascular endothelial growth factor*

VSV-G *Virus G envelope glycoprotein gene*

W

WB *Western blot*

WHO *World health organization*

α

α-KG *A-keto-glutarate*

Funding Institution

Marie Curie ITN-ETN, 676421-Glycocan

1

INTRODUCTION

1

1.1. Cancer

Cancer, also defined as noncommunicable disease, is the leading cause of the death worldwide. Thoroughly, it is the 1st or 2nd leading cause of death worldwide before the age of 70, in 91 of 172 countries[1]. In the western countries, cancer premature mortality remains the highest, regardless the greatest advances in laboratory and clinical studies (Figure 1.1). The need for new strategic approaches, for the creation of personalized therapy, is driving the formation of a multidisciplinary environment and specialist cancer centres, with the mission to improve diagnostic, prognostic and therapeutic tools.

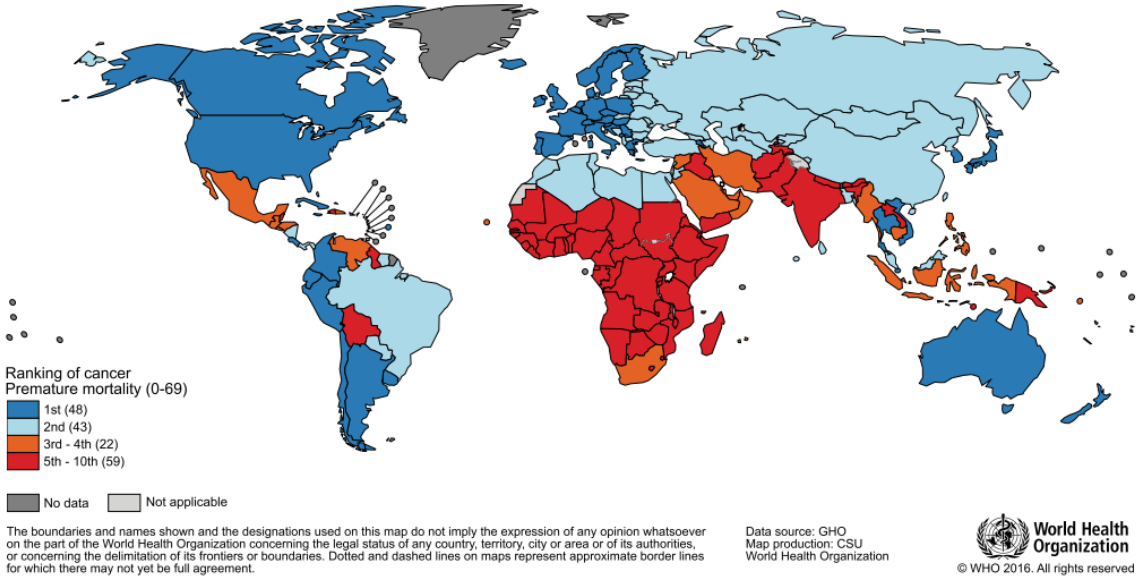


Figure 1.1 Global map ranking of cancer premature mortality worldwide up to 70 years, in 2015. The number of countries is included in the legend. Project GLOBOCAN 2018, adapted from World Health Organization (WHO)

1.1.1 Epidemiology and statistics

Understanding cancer trends and statistics has become fundamental to comprehend pattern and, eventually, causes in cancer increase. Epidemiology programs are crucial for monitoring clinical data, projects and activities. Yet, collection of data worldwide is still requiring standardization of the process. At the date, each country provides their own data to the world health organization (WHO) and no unification of data was provided worldwide up to 2012[2].

For this reason, great efforts have been put in the collecting and re-organizing of statistical data, concerning cancer.

Worth of consideration are thus the projects, GLOBOCAN and CONCORD-2. GLOBOCAN is an interactive web platform presenting incidence of mortality and prevalence of 36 types of cancers in 185 countries in the world; CONCORD-2 aims to register all cancer survival analysis of population-based registry data[1,3].

GLOBOCAN estimations declare that cancer incidence and mortality are rapidly growing worldwide with 18.1 new million cases with 9.6 million cancer deaths, in 2018, compared to the 14.1 cases of 2012, with 8.2 million deaths[1].

Major causes of cancer have been associated with aging, growth of population and, especially, socioeconomic development, defined using the human development index (HDI)[1]. HDI is a significant variable that considers the importance of national policy, in economic growth, in influencing development outcomes. Low-HDI are associated to developing countries and high-HDI to developed ones. Analysis of the variable HDI on cancer data, confirmed that cancer incidence in high-HDI countries is 2-or 3- folds superior compared to low-HDI, but differences in mortality remain smaller, due to higher mortality chances in low HDI countries. Change in trends shows that low-HDI countries present increasing substitution of what-so-defined "poverty-related"/"infectious-related cancer" with cancer types ascribed to the *westernization* kind of life[4,5]. As a result, lung cancer remains the most diagnosed (11.6%) and the deadliest (18.4%) kind of cancer in both sexes worldwide, followed by breast (11.6%) and colorectal (10.2%), which however jumps on the second place as mortality cause with 9.2%[1].

These data suggest not only an improve in diagnostic and treatment in low-HDI, but also a homogenization of lifestyle worldwide.

1.2. Heterogeneity

Cancer is a dynamic disease and various changes can happen during tumour development. As a consequence, the tumour mass is often the result of a diverse collection of cells harbouring distinct molecular signatures with differential levels of sensitivity to treatment[6]. Thus, heterogeneity plays a major role in making cancer a major deadly disease.

In the last 20 years, many studies have been focused in investigating the genesis of the heterogeneity and whether, it is the cause of cancer formation rather than a by-product of tumour progression. Hence, understanding the processes involved in transforming a normal cell into a malignant one could be the key for the creation of personalized diagnostic, prognostic and therapeutic tools. Here will be highlighted the tumour microenvironment and the hallmarks of cancer.

1.2.1 The tumour microenvironment

Cancer cells, through upregulation of oncogenes and downregulation of tumour suppressor, are the responsible entities of tumour initiation and progression. Different cells types are involved in cancer formation, transformation and malignancy and it is important to comprehend the contribution of each one of them in malignancy transformation. However, the great heterogeneity among same cancer types hardly explain so many sub-clonal populations.

The personalized treatment and drug targeting are based into the recognition of crucial pathways and molecule recognition and it would not be possible without in-depth knowledge of the tumour microenvironment (TME).

Subsequently, the biology of a tumour can only be understood by studying the individual specialized cell types within it and their different behaviour[7].

I. **Endothelial cells** are the most studied class of cells in tumour environment: near the stroma is where the most heterogeneity is presented, and they are the cells responsible of tumour-associated vasculature. Many signalling pathways have been discovered responsible of their proliferation and differentiation among which Notch[8], Neuropilin[9], Robo[10], Ephrin A/B[11], Vascular endothelial growth factor (VEGF)[12], angiopoietin[13] and Fibroblast growth factor (FGF)[14] signals. Most of the immunotherapies nowadays are based on anti -VEGF therapy.

II. **Pericytes** are mesenchymal type cells wrapped around the endothelial tubing of blood vessels and responsible in maintenance of a quiescent endothelium. They collaborate with endothelial cells to produce the vascular endothelial membrane, important for holding the blood pressure.

Interestingly, it has been shown that the therapeutic inhibition of platelet-derived growth factor (PDGF) receptor expressed by tumour pericytes, affects pericytes and decreases tumour vascularization, while healthy pericytes are not affected by the therapy[15].

III. **Cancer Stem Cells (CSCs)** present in most of the tumours, may origin in different phases of cell development. They might be the cell-of-origin or a cell in a proliferative state, that undergo in tumour progression. In any case, multiple distinct classes of increasingly neoplastic stem cells form during cancer progression with their self-renewal and tumour-initiating capabilities[7]. Their phenotypic plasticity can be the cause of such cellular variety in TME. As example, CSC can initiate a epithelial to mesenchymal transition (EMT), and be differentiated into mesenchymal, fibroblast-like cancer cells, behaving as cancer-associated fibroblasts (CAFs)[16,17]. CSCs are also chemotherapy resistant and they might be the reason of tumour recurrence after treatment. They may undergo a dormancy program, where tumour cells persist and suddenly erupts after years[16].

IV. **Stem and progenitor cells of the tumour stroma** can be recruited by the surrounding healthy tissue, but the bone marrow has also been identified as a key source. Some of the recruited cells from the bone marrow can persist in a partially-differentiate status exhibiting different function form the fully differentiate ones[18].

V. **Cancer-associated fibroblast** creates the support for the epithelial tissue. Reprogrammed variants of normal tissue-derived fibroblastic cells can enhance cancer cell proliferation, angiogenesis, and invasion and metastasis[19].

VI. **Immune inflammatory cells** are the most diverse type of cells in TME. They have shown a dual role of tumour antagonizing and tumour promoter, depending on which type of immune cells infiltrates the cancer mass. Cytotoxic T lymphocytes (CTLs) and natural killer (NK) cells are associated with tumour antagonizing abilities while macrophage subtypes, mast cells, and neutrophils, as well as T and B lymphocytes are proven to be tumour promoters[20,21]. They can release into the TME pro-tumour factors, such as epidermal growth factor (EGF), VEGF, FGF2, matrix-degrading enzymes and chemokines[22].

1.2.2 Cancer evolution and Hallmarks of cancer

Several theories have been proposed to explain cancer complexity and heterogeneity.

In model of *clonal evolution*, sub sequential acquisition of genetic mutations and functions alterations, leads to a selection of sub-clones capable of co-existing in an altered tumour micro-environment[23]. However, as depicted by Fouad *et al*. (2017), the clonal evolution theory is not able to explain why the alterations are undertaken in the first place[24].

In the model of *stem cells*, proposed by Tomasetti and Vogelstein (2015), is stated that only a minor number of cells, with self-renewable capacity, is capable of initiating and progressing tumour status[25]. These tumour cells have specific surface proteins (CD133, CD166, or CD44) and after chemotherapy have been shown to be more resistant, explaining relapse following treatments[26].

A third theory by Sottoriva *et al*. (2015), *the big bang model*, sustain that the fundamental mutations, happens early in cancer development, thus explaining why some tumour might overgrow without presenting metastasis and other small tumours present metastasis at early phases[27].

These three theories can be considered complementary one to each other and there is no doubt that the comprehension of tumour cell origin and heterogeneity is the key for cancer diagnosis and treatment.

Subsequently, among these theories Hanahan and Weinberg (2000) followed by update in 2011, proposed a new model, called the hallmarks of cancer [7,28].

In their claim, cancer cells must go through 10 different transformation to form a malignant environment: they sustain that the passage from a normal to a malignant cell is the result of accumulation of defective traits, in a continuous evolution of the TME (Figure 1.2).

In this sense, their theory has common points with the previously proposed *clonal evolution*.

The hallmarks of cancers have represented the first big step towards cancer understanding and personalized therapy. Presently, being aware of how genetic and epigenetic alterations are associated with specific hallmarks of cancer, authors are completing the proposed model by correlating behaviour of different molecules, useful for diagnosis or treatment (i.e. miRNA[29], targeted drug delivery[30], the CRISPR generation[31] and glycosylation[32,33]).

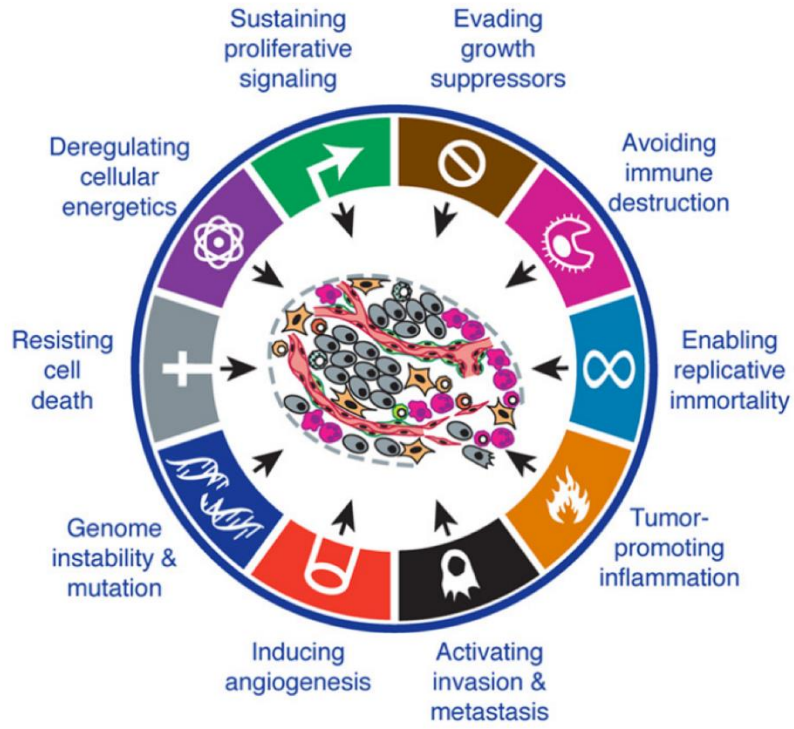


Figure 1.2 The hallmarks of cancer.i) Sustain proliferative signals, ii) evading growth suppressor, iii) activating invasion and metastasis, iv) enabling replicative mortality) inducing angiogenesis and vi) resisting cell death represent the first six hallmarks theorized in 2000. In the revision of 2011, vii) deregulating cellular energetics and viii) avoiding immune destruction where added as emerging hallmarks. ix)Tumour promoting inflammation, x) genome instability and mutation are considered as tumour enabling characteristics. Image adapted from Hanahan and Weinberg (2011).

1.2.2.1 Self-Sufficiency in Growth Signals

Normal cells maintain a quiescent mode until mitogenic growth signals (GS) move them into a proliferative status. Extracellular signals, intracellular transducer and intracellular pathways are the three main factors involved into cell proliferation mechanism. Alteration in any point can induce an uncontrolled growth. It has been shown that while normal cell cannot proliferate in absence of GS, tumour cells have adapted to mimic GS signals in an autocrine way: in glioblastomas and sarcomas, cancer cells produce PDGF and tumour growth factor α (TGFα), two important differentiation and growth signals[34]. Cancer cell can also overexpress GF receptors often carrying tyrosine kinase property in the cytoplasmic domains, i.e. epidermal growth factor receptor (EGF-R-erbB) or human epidermal growth factor receptor 2 (HER2-neu) in stomach tumours. In other occasion, cancer cells can activated a ligand-independent signalling by alteration of the structure of the receptor, such as the EGFR receptor cut in the cytoplasmic domain is always active[35].

At the cytoplasmic level, tumour cells have been displayed an alteration in downstream pathways with SOS-Ras-Raf-MAPK (mitogen-activated protein kinase) cascade, responsible of activation of DNA cell proliferation signals. The alteration of the Ras protein into a constitutive activator of mitogenic and proliferative signals makes it a very powerful oncogene[36]. Ras can also interact directly with the survival promoter PI3 (Phosphoinositide 3-kinase) enabling constitutive growth signals within the cells, including its key protein kinase B (Akt/PKB) signal transducer. When PTEN phosphatase, a catalyser of PI3 kinase product, phosphatidylinositol (3,4,5) trisphosphate (PIP3), is downregulated, generally by promoter hyper methylation, PI3K signalling is amplified inducing tumorigenesis[37]. The same happens when mTOR kinase, a coordinator of PI3K signalling, is inhibited in cancer cells[38].

TME and extracellular matrix (ECM) play an important role in activation of tumour signals. Successful binding of the integrins to the ECM promotes resistance to apoptosis and entrance to active cell cycle, while its unpairing can induce apoptosis and cell cycle arrest. Many cancer cells over express integrins on their cell surface[39].

A successful tumour is the one that cooperate with the microenvironment cells by inducing them to release growth factors[28]. Increase of oncogenic proliferative signals induces cell proliferation and tumour growth. However, it has been demonstrated that high levels of oncoprotein as RAS, RAF and MYC induce cellular senescence and /or apoptosis, as a defensive mechanism. Mice overexpressing these proteins, present cells with the typical senescent morphology: enlarged cytoplasm, the absence of proliferation markers [40].

1.2.2.2 Insensitivity to antigrowth signals

Cell machinery when in healthy status, lives in a perfect equilibrium where growth signals are equally important as the antigrowth signals. The antigrowth signal can act in two main point of a cell cycle: the G_0, where the cell is forced out into a quiescent status, or the G_1, where the cell in a post-mitotic state can proliferate or remain quiescent. Most of the antiproliferative signals are funnelled through the retinoblastoma protein (pRb) that in an hypo-phosphorylated status (Rb) can sequester the E2F transcriptional factors, fundamental to transit the cell from G1 to S phase[28]. Associated to pRb is the transforming growth factor ß (TGFß), which controls the phosphorylation and, thus, the activation of pRb. In some cell types, TGFß represses pro-proliferative genes as *c-myc* gene and, concomitantly, activates the cyclin-dependent kinase (CDK) inhibitors, $p15^{INK4B}$ and p21 proteins: CDK is responsible to phosphorylate pRb[41].

Alteration in any point of this cascade is responsible of inhibition of antigrowth signal and thus promoter of proliferation. TGFß can also activate a cell program named the EMT which confer the cell an high grade of malignancy[42]. Some cancer cell downregulate TGFß receptors by mutation or alteration of the structure[43].Therefore, the role of the TGFß in cancer is still under investigation

Protein in charge of signal transduction can be altered or downregulated in tumour cells, as happens with the protein Smda4 and the $p15^{INK4B}$ [44]. Other proteins like the CDK4, can also change trough mutation and by not-recognition of the inhibitor $p15^{INK4B}$ can be constitutively activated, resulting in pRb constant hyperphosphorylation[45].

Proliferation by escaping of antigrowth signals can be induced into the cell before terminal differentiation. In this event, the protein Myc, which generally is bound to the protein Mad to induce differentiation, is paired to the Max protein, which complex is a proliferation inducer[46].

At the same level the inactivation of the Adenomatous Polyposis Coli (APC) /ß-catenin pathway is inactivated, blocking the entrance of the enterocyte to a differentiate post mitotic state[47].

Another aspect that leads to insensivity to antigrowth signals by cancer cells is the cell-to-cell contact inhibition. The tumour suppressor, Merlin, triggered by NF2 gene, is the responsible for coupling the E-cadherin to the tyrosine kinase receptors EGF. While straightening the bond, merlin is inhibiting the mitogenic signals. Thus the loss of NF2 gene in merlin functions leads to proliferation[48].

1.2.2.3 Evading apoptosis

With apoptosis, cells control internal homeostasis and, in case of serious dysregulation and damage, promote cell death. The role of the apoptosis in cancer is still under discussion since in some cases has been proved to be induced by elevated levels of oncogenes signals, while, in cases of hyper-malignancy and resistance to therapy, it is attenuated[7]. It is induced by different signal ,which activate two main receptors: the inducers FAS and cytochrome C (present in the mitochondria), which activate respectively the ultimate effector of the apoptosis, the caspase-8 and 9. Cytochrome C is activated by B cell lymphoma-2 (Bcl-2), a family of protein whose members are either pro (Bax, Bak, Bid, Bim) or antiapoptotic (Bcl-2, Bcl-XL, Bcl-W)[49].

Bax and Bak share with the Bcl2-like proteins, the same BH3 motif for protein-protein interaction, leading to abnormal protein coupling and anti-/ and pro-apoptotic signals responsible of tumour development[49]. In a normal functioning cell, when p53 tumour

suppressor protein recognize a damage in the DNA, activates BH3-only protein such as Noxa, Puma and Bax that induce apoptosis. On the contrary, in cancer cells p53 is downregulated[50]. External survival factors such insulin-like growth factors (IGF-1) can force Bcl-2 upregulation and FAS death signal disruption. Other cancer cells have been shown to produce FAS ligand decoys in order to avoid downstream signalling[51]. Moreover, the PI3 kinase-AKT-PKB pathway in charge of antiapoptotic signals have been shown to be involved in the evasion[52].

Autophagy represents a cell-homeostasis controls mechanism, in which cellular organelles are broken down allowing the recycle of the metabolites. The degradation happens by the fusion of the autophagosome envelope vesicles with the lysosomes. It is a controlled pathway and PI3, AKT and mTOR kinase, stimulated by survival signals, are the responsible protein for blocking the autophagy program. Beclin-1 is also a very important protein, member of the BH3-pdomain proteins, and in stress-induced situations, such as reactive oxygen species (ROS), is uncouple from Bcl-2 and after liberation induces autophagy[53]. During radiotherapy, nutrient starvation and the use of cytotoxic drugs, induces high level of autophagy that paradoxically results in cancer-cells cytoprotection, by impaired death[54].

Necrosis have been investigated as a cancer deaths program: cells in necrosis increase cytoplasmic volume until membrane integrity loss. The debris liberated induce proinflammatory signals that recruit inflammatory cells specialized in removing the extracellular necrosis debris. Yet, in many tumour masses, inflammatory cells are activated by necrosis, promoting tumour growth angiogenesis and proliferation[55].

1.2.2.4 Limitless replicative potential

Limitless replicative potential or cell immortalization has been evaluated has fundamental step in cancer malignancy.

Cells are capable of a limited number of replicants, called doubles. A normal human cell can replicate from 60 to 70 times until it stops growing and begins a process called senescence. Curiously, when pRb and p53 tumour suppressor protein are silenced, fibroblast in culture avoid senescence and continue multiplying[28]. Yet, telomers, which are the repetition of short 6bp at the end of each chromosome, are the key player in this mechanism. In fact, it has been proven that for each replication, each cell lose a part of the 3' end of the DNA of telomers until the damage affects chromosomes function and bring the cell to death[28].

Cancer cells, however, have developed different system to promote immortalization, by upregulating the enzyme telomerase that adds hexanucleotide at the end of the DNA. An alternative mechanism called alternative lengthening of telomeres, helps maintaining the telomerases ends by recombination-based inter-chromosomal exchange of information, named chromosomal breakage-fusion-bridge[56,57].

Recently, it has been associated a new telomere-independent functions and in particular its protein subunit telomerase reverse transcriptase (TERT). TERT can amplify the signalling cascade of the Wnt pathway, involved in cell migration and cancer progression, by serving as a cofactor of the β-catenin/LEF transcription factor complex, hence inducing proliferation and apoptosis resistance [58].

1.2.2.5 Sustained angiogenesis

Creation of new blood vessels contributes to maintain auto-sufficiency and to promote tumour growth. The development of the vasculature involves the birth of new endothelial cells and their assembly into tubes (vasculogenesis) in addition to the sprouting (angiogenesis) of new vessels from existing ones[7]. The vascularization process is controlled by balanced signals that can promote or silence its initiation[59]. One of the mechanisms is controlled by soluble factors and tyrosine kinase receptors on endothelial surface, where the initiating signals can be VEGF or FGF1/2[60,61]. Endogenous inhibitor factors are present also in the angiogenesis process and they have been recognized in the trombospondin-1 (TSP-1), fragments of plasmin (angiostatin) and collagen 18 (endostatin)[62]. In most of tumour VEGF and FGF are up regulated, while the inhibitor factors such as trombospondin-1 and interferon ß are downregulated. P53 and Ras play main roles also in the vascularization processes: it has been documented that the downregulation of tumour suppressor p53 cause the drop of production of TSP-1 while Ras oncogene causes upregulation of VEGF[63]. Integrins, cell-matrix and cell–cell association also play critical roles. Interference in signals of integrins differently expressed on vascular and capillary surface can promote angiogenesis. Bone marrow-derived cells also contributes to tumorigenesis: cells of the innate immune system (macrophages, neutrophils, mast cells and myeloid progenitors) help in the angiogenetic switch of quiescent cells. It has been shown that the innate cells can contribute into the protection of the vascular tissue from drugs targeting in endothelial tissue [64].

1.2.2.6 Tissue Invasion and Metastasis

Tissue invasion and metastasis is the important final step common to all malignant tumours. The ability of malignant cells to detach form the primary tumour and invade new healthy sites in the body is the cause of the 90% of cancer death.

The process invasion-metastasis cascade, that brings to the colonization of tumour cell into new tissues from the original tumour, begins with local invasion of tumour cells followed by intravasation of the cancer cells into the nearby lymphatic vessel. It ends with the extravasation/escape of the same cells from the vessel Lumina into the parenchyma of distant tissue and formation of micro-metastasis, which can grow into metastatic tumour. Three receptors are involved in this process: cadherin, integrins, lectins and cell-cell adhesion proteins (CAM).

E-cadherin has the function of antigrowth signal by cell-cell interaction and activating the ß-catenin transcription factor. Loss of E-cadherin and ß-catenin is thus a metastasis promoter. N-cadherin which is activated in migrating neurons during embryogenesis, is upregulated in many invasive carcinomas[65].Integrins, in charge of binding ECM, can have different function and in a metastasis carcinoma cells shifts the expression of integrins in favour of epithelial or stromal cells, rather than ECM. Upregulation and activation of protease in the ECM environment can cause detach from integrins and activate invasion by cancer cells in the nearby stroma, blood vessel and epithelial cell layers[66,67].

Tumour cell surface lectins might be involved in cell-cell adhesion, cell attachment to substratum and metastasis. Selectins are a family of three "C-type" lectins that bind to sialofucosylated proteins or lipids in a calcium-dependent manner - E-selectin, P-selectin and L-selectin[68]. Several studies have pointed to a key role played by E-selectin-mediated endothelial adhesion in metastasis[69].In fact, the binding efficiency of colon cancer cell lines to E-selectin is directly proportional to their respective metastatic potential[70].

The CAM family have been shown to be involved in a malignant transformation: N-CAM in cancer cells can switch from a more adherent to a less adhesive form in colorectal cancer[71].

An innovative concept raised up in the "hallmarks of cancer: the next generation" is the epithelial to mesenchymal transition (EMT). It refers to the capacity of epithelial tumour cells to acquire ability to invade, to resist apoptosis and to disseminate. The EMT transition has been found in most of the carcinomas and it can be activated permanently or in a transient way. The main transition factors responsible for the transitions are Snail, Slug, Twist and Zeb1/2: they

cause the loss of adherent's junctions with associated conversion from an epithelial to a fibroblastic morphology, induce resistance to apoptosis and induce the expression of degradation enzymes of the matrix[72,73]. Yet, it remains unclear the rules and the regulation of their expression. It appears that the transition is fundamental for the cell to proceed to metastasis and evidence have been shown that tumour cells and cells of neoplastic stroma interaction is crucial for the capability of invasive growth and metastasis[74]. More and more important is becoming the role of the immune system on the tumour growth: macrophages can supply matrix-degrading enzymes and stimulate EGF and be reciprocally stimulated by cancer cells with colony stimulating factor 1 (CSF-1), promoting an invasive behaviour[20].

The EMT transition can be reverted into mesenchymal to epithelial transition (MET) during initial invasion and metastasis dissemination. In fact, the adaptation of the transitioning cells to foreign tissue microenvironments results into successful colonization growth into macro-tumours. However, some metastasis can remain dormant and they grow only after total resection of the primary tumour, suggesting a major role of the primary tumour in micro-metastasis regulation[75]. Other reasons for metastasis dormancy have been directed to anti-growth signal from the surrounding healthy cells[76], tumour suppression by immune system[77] and high autophagy provoked by nutrient starvation[78]. It remains uncertain when and where the tumour acquires capacity of colonization of tissue. It may depend to selective pressure, where the most adapt cells survive or tissue specific colonization. The *anatomical-mechanical mechanisms* explain the capacity of micro-metastasis of vascular and lymphatic drainage arresting non-specifically in the first organ encountered. On contrast, the theory of the *seed and the soil* sustain, that tumour cells have a preferable subsets of organs and their microenvironment[24].

1.2.2.7 Genome instability and mutation

Genome instability has been well documented throughout the years of cancer research and study of mutation in the DNA has been used as promising biomarker for cancer detection. Most of cancer present different patterns of mutation, yet defects in the mechanisms of DNA maintenance and repair, as well as DNA methylation and histone modification have been resulted as the common feature. In general, the DNA maintenance genes, also named care-taker of the genome, can act in three different points: repairing directly DNA damage, activating the repair machinery or inactivating mutagens before their effects on the DNA[79]. Loss of their function correspond to an increase in cancer manifestation and malignancy.

1.2.2.8 Immune escaping and tumour promoting inflammation

Immunoediting is a dynamic process that describes the relation between the tumour cells and the immune system, and it comprises the elimination, equilibrium and escape phases [80]. In the elimination phase tumour cells are mostly eliminated by the innate and adaptive immune responses to tumour cells. On the equilibrium, tumour cells that escaped the first phase do not present immunogenic antigens and various mutations increase overall resistance to immune attack. In the escaping phase tumour cells continue to grow and expand in an uncontrolled manner and may eventually lead to malignancies[80].

The role of the immune system in tumour progression is still under discussion. The double role of the immune system towards the tumour microenvironment originated two new hallmarks of cancer: the immune escaping mechanism of cancer cells to evade immune activation and the tumour promoting inflammation by the same immune cells.

It has been shown in fact that the infiltration of immune cells in tumour environment has the purpose of tumour eradication. For example, colon cancer patients with a high level of infiltration of CTLs and NKs has been shown a better tumour prognosis and outcome[81].

In vivo experiments in immunocompromised mice, lacking development or function of $CD8^+$ CTLs, $CD4^+$ Th1 helper T cells, or NK, had an increase tumour incidence[77]. At the same time, tumour microenvironment acquire mechanism to escape the immune system such as the expression of the programmed death ligand-1 (PD-L1) which inhibit the activation of T lymphocytes by binding the programmed death-1 receptor (PD-1)[82]. Stromal cells and tumour derived exosome can secrete immune suppressor factors by upregulation of immune suppressor molecules[24]. Other cancer cells are able to secrete TGFß, a potent immunosuppressor, which can paralyse CTLs and NK infiltration, or they can recruit immunosuppressive immune cells, such as regulatory T cells (Tregs) and myeloid-derived suppressor cells[83,84]. Highly immunogenic cells are constantly eliminated by the immune system (immunoediting) while weakly immunogenic cancer cells, barely recognized, are free to grow and expand into solid tumours, following a Darwin clone selection mechanism.

Unexpectedly, it has been demonstrated the role of inflammation in most of tumour progression by supplying active biomolecule in cancer microenvironment: growth factors that sustain proliferative signalling, survival factors that limit cell death, proangiogenic factors, extracellular matrix-modifying enzymes that facilitate angiogenesis, invasion, and metastasis,

and inductive signals that lead to activation of EMT[85]. Moreover, immune cells can release reactive oxygen species(ROS) well known for their mutagenic potential[55].

1.2.2.9 Reprogramming energy metabolism

Glucose and glutamine are the main nutrients and pool of carbons, fundamental for energy production inside the cell[24].The production of energy inside the cells is conducted by firstly the glycolysis (or scission of the glucose molecule) in the cytoplasm, with the production of two molecules of pyruvate. If the reaction is conducted in aerobic condition, the pyruvate is converted in the mitochondria into carbon dioxide and water; if the reaction is happening in anaerobic conditions the pyruvate is converted into lactate or acetic acid, through fermentation[86]. The last hallmarks of cancer, also known as the "Warburg effect" is identified in the choice of cancer cells to conduct a glycolysis process also in anaerobic conditions, with high rates of glucose consumption and lactic acid production, leading to cellular acidosis[87,88]. To compensate the lower efficiency in adenosine triphosphate (ATP) production, the complex organic molecule that provide energy to drive many processes in living cells, the cancer cells upregulates the glucose transporters, i.e. GLUT1, increasing the glucose intake. This has been shown to be associated with activated oncogenes, i.e. RAS, MYC, and mutant tumour suppressors, i.e. p53[89]. In hypoxia condition cells upregulate the glucose transporters and increase the production of glycolytic pathways enzyme and this can be one of the reasons why cancer cells chose this mechanism of energy production.

Cancer cell have also shown to increase the lipid production with upregulation of major components of fatty acyl chain synthesis, probably to favour membrane formation in a high proliferative environment. In this case, glutamine is captured by mitochondria and converted into glutamate which can be used directly or converted into α-keto-glutarate, a glucose alternative to produce citrate or share in fatty acid synthesis under unfavourable conditions[90].

1.3 Colorectal cancer

In 2018, colorectal cancer (CRC) remains the second deadliest kind of cancer with 881,00 deaths of the 1.8 million new cases worldwide. The burden of colorectal cancer is supposed to increase up to 2.2 new million cases with 1.1 million death by 2030[91]. The world health organization (WHO) has predicted an increase in diagnosed cases of 77% and increase of death by 80% by 2030, making CRC one of the leading burdens after cardiac diseases and lung cancer.

Although the mortality rates are equivalent in low and high HDI countries, its incidence is 2-/3- time superior in high HDI states, suggesting high correlation between lifestyle and cancer arisen. As support of this hypothesis, it has been shown an improvement of survival rates when best practice of dietary patter and life style are adopted in the developed countries[92]. Obesity, exceed in red meat consumption, smoking and low physical exercise have been correlated with high colorectal cancer incidence[92–95]. Colorectal cancer develops from the transformation of a normal colonic mucosa to an invasive cancer cell. This transformation can take from 10 to 15 years and for this reason for non-hereditary kind of colorectal cancer the diagnostic routine for the so-called average risk population is every 10 years from 40-50 years old and every 5 years after 50. Age remains the major responsible of CRC carcinogenesis but in general the 70%-80% of CRC formation are sporadic, while the rest 20%-30% have hereditary basis, of which familiar adenomatous polyposis (< 1%), non-polyposis hereditary CRC also known as Lynch syndrome (1%-5%) and MYH-gene associate polyposis (<1%) are the most characterized[96].

Three are the main pathways involved in CRC formation: the chromosomal instability which consists in the accumulation of mutation in oncogene e suppressor genes, that leads to neoplastic transformation[97]. The second involves defects in the DNA reparation mechanisms and the consequent accumulation of mutations, which occur in particular in the DNA microsatellites[98]. The third pathways involves epigenetics and hyper methylation of DNA nucleotides in the promoters gene, also known as CpG island methylator phenotype (CIMP)[99].

1.3.1 Diagnosis and biomarkers

It is well documented that a good screening program combined with sensitive diagnostic method is the key for a reduction in CRC mortality and increase of overall survival (OS). Detecting CRC at the first stages correspond to an increase of 5-year survival up to the 90%,

significantly higher compared to the 15% when detected in the advanced stages. For this reason, many detection methods have been studied and tested in clinics for their specificity and sensitivity[100]. The main screening methods in use can be divided into two groups: analysis of biological samples, by liquid or tissue biopsy, where the doctors are looking for indicative CRC biomarkers and endoscopic techniques, where the identification of advanced adenomatous polyps (AA) pre-neoplastic lesion formation before cancer transformation is crucial for the exit of CRC treatment[96]. Main biological samples test in use are performed in faecal samples (faecal occult blood test, FIT: faecal (FOBT), faecal immunochemical test, faecal (FIT), faecal DNA and RNA, proteins), in blood/plasma (DNA, RNA, protein) and urine (Figure 1.3).

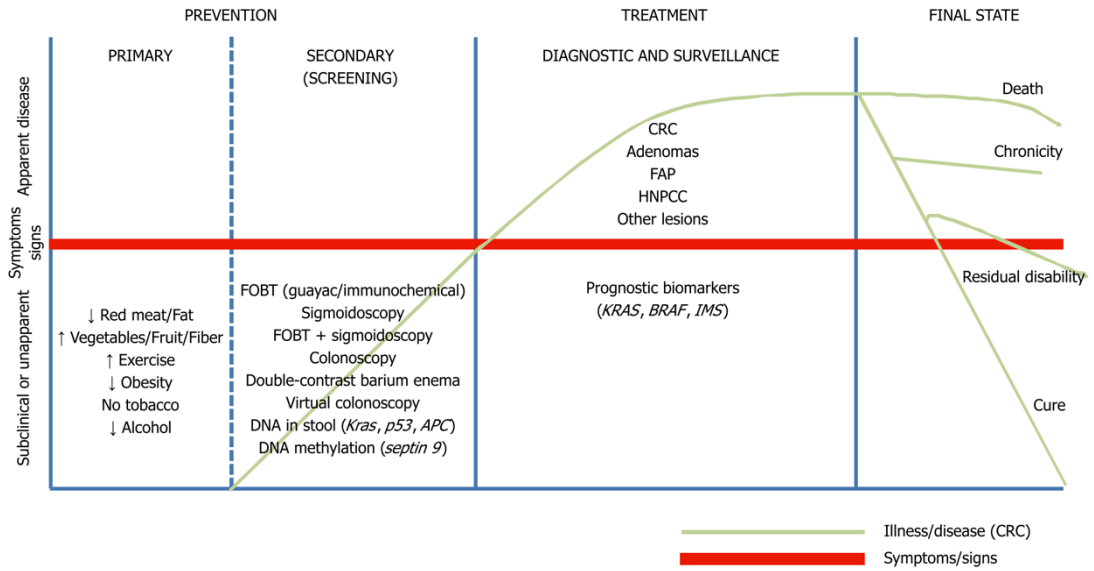


Figure 1.3. Schematic representation of all the alternatives test available for colorectal cancer prevetion, screening, diagnosis, prognosis and treatment. The red line separate symptomatic from asymptomatic patients; in green is represented the stages of CRC. FOBT: fecal occult blood test, FAP: familial adenomatous polyposis, HNPCC: hereditary nonpolyposis CRC. Image adapted from Binefa et al. (2015)

1.3.1.1 Fecal occult blood test and fecal immunochemical test

The faecal blood occult test is a non-invasive test and up-to the date is the most in use. The guanic (g)FOBT, based on guaiac resin, is based on the identification of the enzymatic activity of the peroxidase in the haemoglobin heme subunit. In fact, polyps, AA and neoplastic formation, leads to lesion in the internal mucosa and the identification of residual blood in the faeces may be an indicator of abnormal growths[101]. However, diet, such as red meat, could influence the test outcome and is important to follow dietary restrictions days before the test[102].

The test has been shown to help in mortality reduction up to 33% in cohorts of patients tested every year for 13 years. The identification of pre-neoplastic lesion can in fact be treat immediately by surgical removal, helping fighting cancer formation by early stages detection. gFOBT sensitivity can go from 6.2% up to 83.3% while the specificity remains more constant at 80%[103].

Immunochemical faecal occult blood test, in contrast with gFOBT, not only is able to give semi-quantitative results but is also haemoglobin specific and does not need any dietary restriction before the use. The FIT is based on the use of monoclonal or polyclonal specific antibodies against human haemoglobin and it is performed with immunoassays. Although it has a better sensitivity that could reach the 98%, it could have very little sensitive down to 5.4%[104]. In general, the specificity remains high with a range of 77-99%[104].

1.3.1.2 DNA and RNA

CRC can have high mitotic activity and low adhesion index, that allows the identification of DNA analysis of the gene alterations involved in CRC, such as the KRAS gene and hyper methylation. The analysis of DNA in stools (sDNA) has not shown high sensitivity since the amount of human DNA in the sample is around 0.01%[105]. Moreover, most of the sample contains natural inhibitors of the polymerase chain reaction (PCR), technique in use for DNA amplification and study[106]. On the other hand, the methylation status of many genes displayed a remarkable diagnostic accuracy for CRC and AA. A new recent technique is the fluorescent long DNA (FL-DNA) which allows the identification of longer segments the DNA (>150-200 bp). These segments are generated by cancer cells avoiding the normal apoptotic pathway and generating longer fragments of DNA; the technique have showed a sensitivity of 80%[107].

DNA CRC markers have been searched also in the plasma/blood samples. In order to remove enzymes present in the blood, that could inhibit the PCR, the sample must be pre-treated[108,109]. In general, two components are studied in DNA analysis, genes mutation and/or hyper methylation. Many mutations have been recognized in CRC formation and among these, the APC and KRAS genes being the most present and the earliest formation in the tumorigeneses. 90% of sporadic colon cancer occurs after APC mutations, and they are detected through DNA in plasma up to the 73% specificity[110,111].

Other genes, such as *Septin 9*, are hyper methylated during colon cancer formation and the methylation analysis detects CRC up to 96% with a specificity range that goes from 86% up to 100%[112,113].

RNA is also becoming more and more important in CRC diagnosis, with circular RNA (circRNAs) messenger RNA (mRNA) and micro RNA (miRNA). A study identified 257 new circRNA species of which 67 circRNAs were missing in colorectal cancer (CRC) patients compared with those in healthy controls. Moreover, upregulation of serum circ-KLDHC10 was able to distinguish CRC patients from healthy subjects, suggesting that could be promising non-invasive biomarker for the early detection and screening of CRC [114].

miRNA can be downregulated or upregulated in the CRC. A study of 88 CRC identified five most expressed miRNAs (miRNA-10a-5p, -21-5p, -22-3p, -143-3p, and -192-5p), among a pool of 523 miRNAs, with good potential as diagnostic tools[115]. Further studies, identified upregulation of miRNA92a and miRNA29a and downregulation of miRNA 16-1 associated with CRC profile[17,116]. In particular, miRNA 16-1 plays a role in the epithelial-mesenchymal transition, contributing to the metastasis capacity of the CRC cells[17].

Combination of miRNA and mRNA gene expression could correctly distinguish high microsatellite instability (MSI-H) versus microsatellite stability (MSS) in colon cancer samples. The work suggests that the this technique may represent a general approach for improving bio-molecular classification of human cancer[117].

1.3.1.3 Proteins

Proteins are considered an important biomarker, due to their stability both in faeces and blood sample.

In the stool of patients are detected protein such as the pyruvate kinase isoenzyme 2 (M2-PK), an isoform overexpressed during cell proliferation, the S100 calcium protein and the metallopeptidase inhibitor 1: they have shown to have sensitivity up to 80% with a specificity up to 95%[118,119].

In the plasma, carcinoembryonic antigen (CEA) have been studied as a diagnostic biomarker for colorectal cancer, but it has not shown enough sensitives. Also, the carbohydrate antigen CA19.9 have been investigated with CA 50, CA 72.4 and CO29.11 with not acceptable diagnostic performances[120].

The study of autoantibodies is very equally relevant in CRC diagnosis (epithelial cell adhesion molecule, p53, p62, CEA, HER-2/neu, Ras, topoisomerase Ⅱ-alpha, histone deacetylase 3 and

5, ubiquitin L3, tyrosinase, tropomyosin, cyclin B1), that can be detected by immunoassays[121,122].

1.3.1.4 Cell free DNA and circulating tumour cells

Innovative forms of cancer detection involve also the circulating tumour cells (CTCs) and cell free DNA. The latter is released by cancer cells during the cell turnover, can be isolated from the plasma or other biological fluids. They allow detection of cancer mutations through non-invasive techniques. Emulsion PCR techniques, such as the Droplet Digital PCR and the Beads, Emulsion, Amplification and Magnetics (BEAMing) are able to identify genomic alterations ranging from 0.01 to 0.001%[123].

The study of CTCs has been challenging due to difficulty in detection. In metastasis formation, Wnt signalling and EMT confer chemotherapeutic resistance to CTCs. Until now, CTCs are isolated through positive selection for an epithelial cell surface marker, epithelial cell adhesion marker (EpCAM), overexpressed in cancer cells. However, when the cells undergo to epithelial to mesenchymal (ETM) the sensitivity of the detection, decreases. New studies have revealed the potential of the carbohydrate sTn, for positive selection and isolation of CTCs[123–126].

1.3.2 Second phase of diagnosis

The use of instruments to detect abnormal cellular mass is still crucial for a correct CRC diagnosis. In fact, after the identification of altered levels of CRC biomarkers the clinician performs colonoscopy (or sigmoidoscopy, if performed on the distal-left colon) and proceeded to tissue collection for the final tissue diagnosis, performed by the pathologist[100].

The machines used in CRC diagnostic are in order of importance and specificity: colonoscopy, sigmoidoscopy, computed tomography (CT) colonography and Double Contrast Barium Enema (DCBE)[127].

Colonoscopy is the endoscopic machine performed throughout all the intestine length. It is an invasive technique that require sedation, a specific diet for bowl preparation and might lead to secondary complications, such as perforation or bacterial infection. However, up to the date, is the most effective in detecting polyps >6 mm with specificity up to 98% and sensitivity up to 99%.

It has been shown that a screening every 10-year performed by colonoscopy can reduce mortality and incidence up to 65% and 67%[96]. Likewise, the sigmoidoscopy is becoming of important use in CRC diagnostic: it is a colonoscopy performed only on the distal part of colon, known also as left colon[96].

Colonoscopy and sigmoidoscopy have been exhibited not only to be invasive techniques but they can also cause severe infection by multi-drug resistant Escherichia coli and Klebsiella species, with rates of 7-day post endoscopic infections varied widely, ranging from 0 to 115 per 1000 procedures for screening colonoscopy, 0 to 132 for non-screening colonoscopy[128].

For this reason, other techniques have been experimented for CRC such as CT colonography and DCBE. CT colonography, remains quite invasive technique and its specificity to detect neoplasm is 82% but for polyps >10mm which is still inferior to the colonoscopy while the DCBE has shown even lower results with 48% specificity in polyps>10 mm; both the techniques remains expensive and they are used as complementary tests[129,130].

1.3.3 Colorectal cancer and stages classifiers

Throughout the sample histologic analysis, it is defined the cellular origin of the tumour, the stage and the grade. The pathologist's diagnosis by tissue biopsy is crucial the decision for a correct treatment. CRC is a very heterogeneous disease and can develop form different cellular subsets, however more than 99.5% are carcinomas of which 92% adenocarcinomas originating from the epithelial stratum of colon mucosa[131]. Subsequently, the pathologist defines the cellular grade of the cancer, which consist in the level of differentiation of the tumour cells from its origin, and the stage, which define the level of invasiveness of cancer. Original staging system is composed by stage I, where the polyps appear, stage II/III where the tumour is spreading and stage IV, where the tumour has started metastasis. Another important classificatory, named TNM, helps to for a better description of the cancer, with T describing the mass of the tumour, N the invasion to the lymph nodes and M eventual metastasis present. The TNM system contributes also to the definition of other parameters: the grade (G), the elevation of serum tumour markers (S), the completeness of a resection of the tumour (R), invasion at lymphatic vessel (L) and invasion into the veins (V)[132]. All the factors helps for a better identification of the cancer, yet other aspects of host-tumour interaction are crucial for the tumour microenvironment progression such as extracellular matrix, supporting stromal cells, and immune cells interactions[131,133].

Broussard *et al.* (2011) suggested to include in the TNM scoring an evaluation of the immune cells in the tumour mass. In fact, they have correlated an higher infiltration of cytotoxic $CD8^+$ T cell in the tumours with higher disease-free and overall survival rates[133]. New diagnosed lesions during the course of the follow-up of patients is often necessary in order to confirm the recurrence of the disease[125,132].

1.3.4 New algorithms for staging and diagnosis

Because of the great heterogeneity in colorectal cancer study results, researchers have proposed new algorithms to find correlation among the different markers in cancer development and progression, useful for the choice of proper treatments. The markers would include gene mutation, gene expression, up-/downregulation of proteins, immune infiltration, hypermethylation, Chromosomal instability and drug resistance. This is how Guinney *et al*. (2015) regrouped the data by forming four consensus molecular subtypes (CMS) with distinguishing features. As represented in Figure 1.4, CRC has been divided in:

CSM 1 (14%), *MSI immune*, characterized by an overexpression of genes involved with immune infiltrate combined with a strong activation of immune evasion pathways associated with higher MSI.

CSM2 (37%), *the canonical*, shows an elevated SCNA with an upregulation of WNT and MYC protooncogenes activation.

CSM3 (13%), *the metabolic*, with mutations in KRAS, presents an heterogeneous MSI associated with high metabolic activity.

CSM4 (23%), *the mesenchymal*, with the worst outcome, presents activation of TGF-β, angiogenesis and an upregulation of factors involved in the EMT, necessary for migration and infiltration[134]. Nevertheless, a 5^{th} subgroup containing the 13% of unclassified cases of CRC was later identified, suggesting the necessity of ulterior refinement of the CRC-CSMs, including new variables such as post-transcriptional and epigenetic factors[134].

CMS1 MSI immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
14%	37%	13%	23%
MSI, CIMP high, hypermutation	SCNA high	Mixed MSI status, SCNA low, CIMP low	SCNA high
BRAF mutations		*KRAS* mutations	
Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF-β activation, angiogenesis
Worse survival after relapse			Worse relapse-free and overall survival

Figure 1.4. Molecular consensus subtypes regrouped data of algorithm analysis for CRC classification. For each group is listed the percentage of frequency, genetic and epigenetic features that characterizes each subgroup. Image adapted from Guinney et al. (2015).

Trinh *et al*. (2016) validated the prognostic value of CRC subtyping by identifying a robust IHM assay for different markers, that will help to identify molecular CRC subtypes and uncover subtype-specific therapeutic benefit. A microarray stained for five markers (CDX2, FRMD6, HTR2B, ZEB1 and KER) was used. CSM1 was selected by the MSI status and the CSM2/3 by CDX2, a marker highly expressed in epithelial-like tumour. CSM4 were selected with HTR2B and FRMD6 expressed in mesenchymal like cells. To confirm the ETM transition, the marker ZEB1 and an anti-cytokeratin (AE1/AE3) was used to normalize for epithelial content. The experiment permitted to confirm the stratification of CRC groups and verify the therapeutic benefit for each subtype[135].

In order to find out the most valuable biomarkers and pathways for diagnosis and prognosis in colorectal cancer (CRC) Zangh *et al*. (2018) have created a CRC biomarkers database, based on all the publications found on CRC. The CRC biomarkers were closely related to the CRC initiation and progression. As a results, correlation of multiple protein biomarker of KRAS-PTEN-STAT3-CD44-ZEB1-ZEB2-S1PR1 reveals to have most significant role in CRC diagnosis, such as molecular functions and cellular composition of the protein biomarkers are involved in CRC diagnosis, therapy and prognosis[136].

1.3.4 Colorectal cancer and treatment

The choice of the treatment is based on cancer location, staging and identification of its biomarkers. When possible, colon cancer is always treated surgically, with R0 tumour mass resection (R_0), which correspond the entire remission and removal, improving long-term survival from 20% up the 50% of the patients with metastasis[100]. R1 to microscopic residual tumour, R2 to macroscopic residual are also operated whenever possible[137]. As a prognostic biomarker, level of CEA is checked for cancer total remission[138,139]. In rectum is more difficult to obtain a cut with adequate distal and circumferential margins (at least 1 cm), due to the absence of serosa and proximity to other pelvic organs. For this reason, pre- or/post-chemotherapy is used as adjuvant treatment in order to increase the 10 years' overall survival (OS) and progression-free survival (PFS). Conventional chemotherapeutic includes 5-FU, capecitabine (one fluorouracil oral prodrug), Leucovorin (5-FU and folinic acid), FOLFOX (5-FU, folinic acid and oxaliplatin-inhibits DNA repair-), FOLFIRI (5-FU, folinic acid and irinotecan-topoisomerase inhibitor-) and FOLFOXIRI (5-FU, folinic acid and oxaliplatin and irinotecan)[100,140,141].

1.4. Glycosylation

Glycosylation is one of the most important post-translational modification on proteins and lipids and it influences most of the biological processes, such as protein folding, cell adhesion, molecular trafficking/clearance, receptor activation, signal transduction and endocytosis[146]. The study of glycans that compose a certain cell or tissue, or glycome, has become more influencing with the discovery of the glycocalyx in the mid of the 20th, conferring a new role to the carbohydrates, known to be involved in energy generation and metabolism[147]. The glycocalyx is an assemble of glycoproteins, glycolipids and receptors-binding-glycan and it covers the surface of mammalian cells, as well as most of the extracellular molecules. Among the functions are included mechano-transduction, homeostasis, signalling and blood cell–vessel wall interactions[148]. The word glycan comprehend oligosaccharide, polysaccharide, and carbohydrates; the structural diversity of glycans depends on several factors, namely differences in monosaccharide composition, glycosidic linkage, branching, the presence of non-carbohydrate substituted components (phosphorylation, sulfation, acetylation, etc.) and linkage to their aglycones (peptide, lipid, etc.)[149].

Nine monosaccharides are used in the enzymatic process of glycosylation in mammals which are recovered from conserved biosynthetic pathways that provides the monosaccharides[150]. It has been difficult to predict glycan's sequences since are not directly translated from the DNA. In other words, glycan diversification exclusively depends on regulation, translation and activity of glycoenzymes. These enzymes, responsible of glycosylation processes, are divided into glycotransferases which synthetize sugar chains and glycosidases, which hydrolyses sugar bonds in the glycans structures. Both enzymes contribute to the final sugar structure and the paradigm one sugar one enzyme is most of the times respected; yet, glycosyltransferases can also act as isozymes, transferring different types of sugar to the chain. Different enzyme can also compete for the same acceptor substrate and add diverse sugars with different linkages.

Glycans can be found attached to proteins and lipids. The glycosylation process occurs in the lumen of the endoplasmic reticulum (ER) and in the Golgi apparatus in co-translational or post-translational processes. Some bulky sugar is formed in the cytoplasmic and flipped inside of the ER, but most of the growing chain happens directly inside the ER or Golgi. The glycosylation process follows precise rules, where not all the monosaccharides can bind all lipid and protein (Figure 1.5) and the expression of many glycans has specific functions: thus, glycans are as universal in nature as nucleic acids, proteins, lipids and metabolites[151].

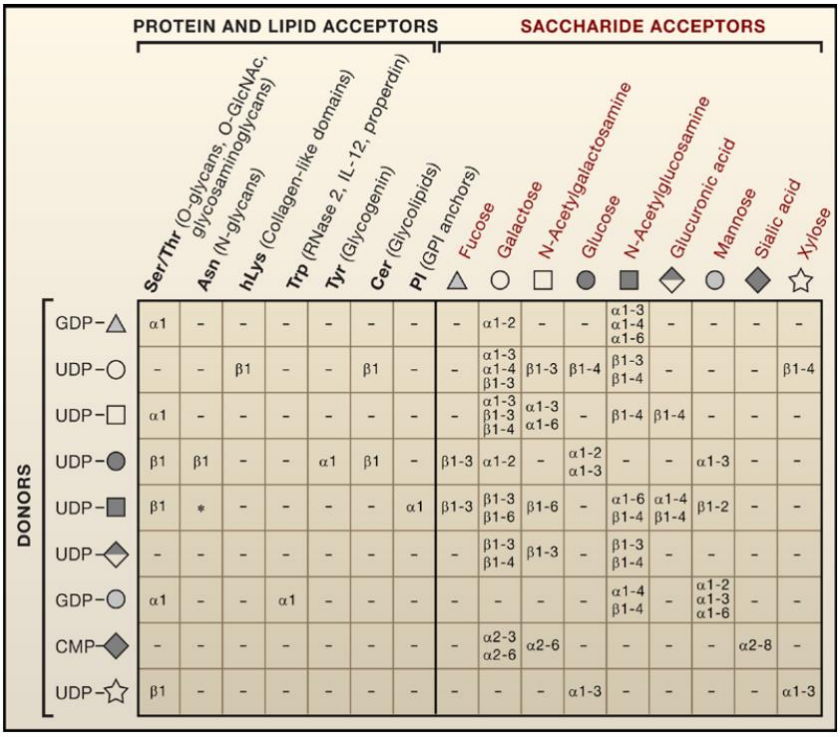


Figure 1.5 Representation of the possible linkages between sugar donors (on the left) and their respective protein and lipid acceptors (up-left) or/and their saccharide acceptors (up-right). α and ß correspond to the type of glycosidic bondage of the sugars depending on the anomeric position of the stereocenter linked though position 1-2. Numbers from 2 to 6 identify position on the acceptor saccharide. Image adapted from Ohtsubo et Marth (2006).

Glycans can be attached to lipids in order to generate glycolipids, such as glycosphingolipids (GSLs) (Figure 1.6)[152]. GSLs is the major product of lipids glycosylation and include the sialic acid-bearing gangliosides. 50–70% of the serum proteins are also glycosylated[153]. Protein glycosylation includes N-glycans, O-glycans and glycosaminoglycans, or proteoglycan (Figure 1.6)[152]. O-glycan and N-glycan is the major class of protein glycosylation and their aberration has vastly documented been involved with malignancy in tumour progression. For this reason, N- and O- glycosylation will be the major focus of this thesis.

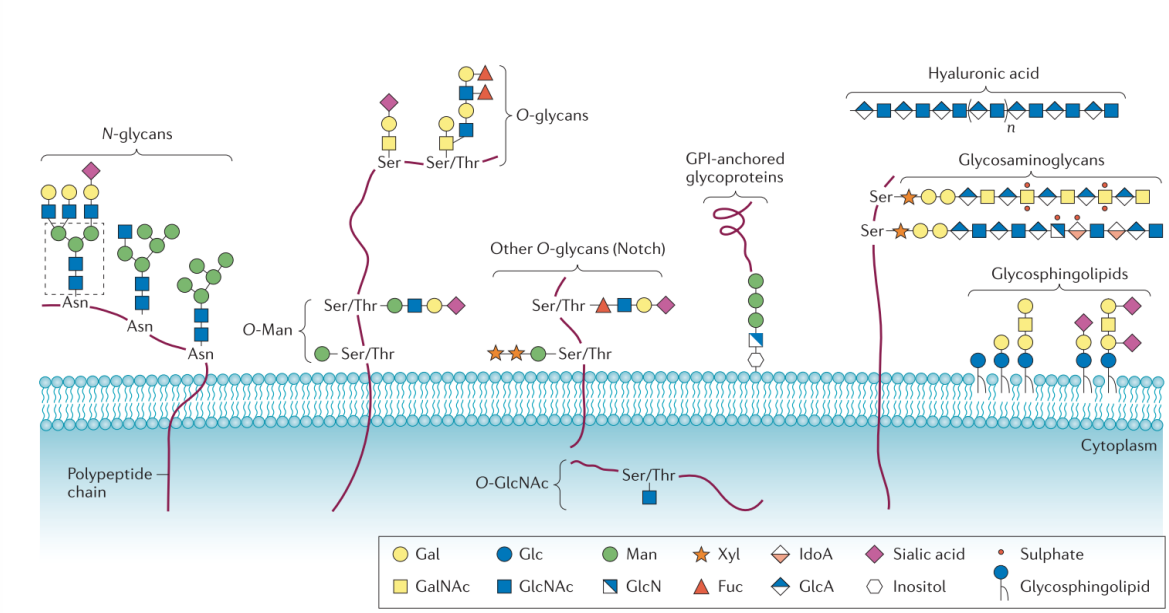


Figure 1.6 Common classed of glycoconjugates in mammalian cells. N-glycans, O-glycans, GPI-anchored glycoproteins and glycosaminoglycans are the major class of protein linked to glycans; differs for preferences in aminoacidic linkage and structures. Hyaluronic acid belongs to the glycosaminoglycans family but is found free in the ECM. Glycosphingolipids (GSLs) are the major class of glycans bound to lipids. on the bottom-right is the coloured legend of monosaccharide components. image adapted from Pinho et Reis (2015)

1.4.1 N-glycans

N-glycans binds to the asparagine of protein by recognizing the Asn-X-Ser/Thr motif. It is estimated that two thirds of this consensus sequences present on the protein databased is N-glycosylated. Some exception has been proved to be also in the Asn-X-Cys motif, given the cysteine present in a reduced form. All the N-glycans share a common penta-saccharides core region and can be divided in three main classes: i) high mannose type, in which only mannose residues are attached to the core, ii) the complex-type, in which "antennae" initiated by N-acetylglucosaminyltransferases (GlcNAcTs) are attached to the core and iii) the hybrid-type, in which only mannose residues are attached to the Manα1–6 arm of the core and one or two antennae are on the Manα1–3 arm[154] (Figure 1.7).

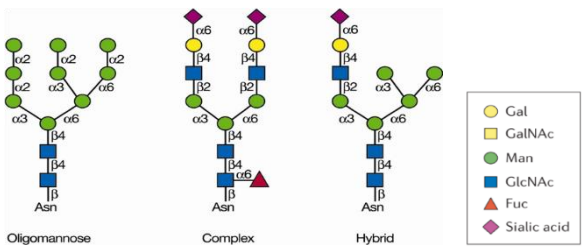


Figure 1.7 Three main types of N-Glycans. Oligomannose, complex and hybrid, share the same pentasaccharides core but differs for the elongation. Image adapted from Essential of glycobiology (2009).

No evidence proved that N-glycans can occur on cytoplasmic or nuclear proteins nor on the cytoplasmic portions of membrane proteins. The transfer happens in the luminal side of the ER membrane while the protein is translated on ribosomes and translocating in the ER membrane[150]. In the trans-Golgi occurs the final sugar addition: i) sugar additions to the core, ii) elongation of branching N-acetylglucosamine residues by sugar additions, and iii) "capping" or "decoration" of elongated branches. The role of different glycoenzymes is critical for the final differentiation of N-glycans chains (Figure 1.8). This converts the limited repertoire of hybrid and branched N-glycans into an extensive array of mature, complex N-glycans[154].

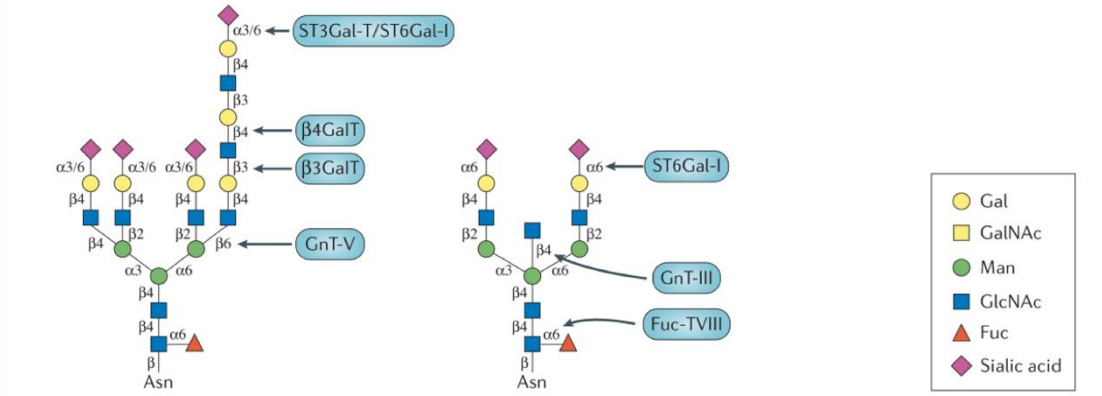


Figure 1.8 Example of different N-glycans with specific glycoenzymes, responsible for each linkage (in blue). Sialyltransferases, such as α-galactoside α-2,6-sialyltransferase I (ST6Gal-I), α2,3-sialyltransferases (ST3Gal-Ts); fucosyltransferases 8 (Fuc-TVIII); N-acetylglucosamine (GlcNAc) transferases (GnTs); galactosyltransferase (GalT). Image adapted from Pinho et Reis (2015)

1.4.2 O-GalNAc glycans

O- Glycans are attached to the hydroxyl oxygen of serine and threonine residues on mammal proteins. O-glycans are mostly covalently α-linked via an N-acetylgalactosamine (GalNAc) moiety to the -OH of serine or threonine by an O-glycosidic bond[155]. Other types of O-glycans include those attached via O-mannose, and the β-N-acetylglucosamine, which tend to be more present in the intra-cellular protein[156]. Mucin are the glycoproteins carrying the greatest number of O-GalNAc glycans, and therefore, this type of modification is also called mucin type O-glycan. The GalNAc can be extended with galactose, N-acetylglucosamine, fucose, or sialic acid, but not mannose, glucose, or xylose residues.

There are eight GalNAc core structures, of which the first four (core1 to 4, the most common) and the mucin O-Glycan can be branched or modified with O-acetylation of sialic acid and O-sulfation of galactose and N-acetylglucosamine[150]. Most of mucin O-glycan are antigenic (Table 1.1 O-GalNAc glycans different cores and antigenic epitopes of mucins. image adapted from Essential of glycobiology (2015)).

Table 1.1 O-GalNAc glycans different cores and antigenic epitopes of mucins. image adapted from Essential of glycobiology (2015)

Core	
Tn antigen	GalNAcαSer/Thr
Sialyl-Tn antigen	Siaα2-6GalNAcαSer/Thr
Core 1 or T antigen	Galβ1-3GalNAcαSer/Thr
Core 2	GlcNAcβ1-6(Galβ1-3)GalNAcαSer/Thr
Core 3	GlcNAcβ1-3GalNAcαSer/Thr
Core 4	GlcNAcβ1-6(GlcNAcβ1-3)GalNAcαSer/Thr
Epitope	
Blood group H	Fucα1-2Gal-
Blood group A	GalNAcα1-3(Fucα1-2)Gal-
Blood group B	Galα1-3(Fucα1-2)Gal-
Linear blood group B	Galα1-3Gal-
Blood group i	Galβ1-4GlcNAcβ1-3Gal-
Blood group I	Galβ1-4GlcNAcβ1-6(Galβ1-4GlcNAcβ1-3)Gal-
Blood group Sd(a), Cad	GalNAcβ1-4(Siaα2-3)Gal-
Blood group Lewis a	Galβ1-3(Fucα1-4)GlcNAc-
Blood group Lewis x	Galβ1-4(Fucα1-3)GlcNAc-
Blood group sialyl-Lewis x	Siaα2-3Galβ1-4(Fucα1-3)GlcNAc-
Blood group Lewis y	Fucα1-2Galβ1-4(Fucα1-3)GlcNAc-

Most of the glycosylation occur in the secretory pathway (ER and Golgi), and therefore almost all membrane and secreted proteins are glycosylated. However, it has been found many nuclear and cytoplasmic protein can be O-glycosylated[155]. O-glycosylation may serve for the viability of many mammalian cell types perhaps by acting as a nutrient sensor, preventing protein phosphorylation, or regulating protein turnover[155].

The simplest O-glycan is called Tn and it consist of the single GalNAc bond to the serine or threonine, and when is sialylated is called sialyl Tn (sTn). The core 1, name T antigen, is composed by Galβ1-3GalNAc- and it forms the core of many longer and complex O-glycan structures. Equally the core 1 can be sialylated and it take the name of sialyl T (sT). When the core one is attached to a branching N-acetylglucosamine is called core 2 and it is found on most of mucin and glycoproteins; linear core 3 and branched core 4 are found only is specific mucin and tissue (Figure 1.9). Core 5-8 are most rare and are reported to be bound to specific mucins; in particular Bovine submaxillary mucin (BSM) is found containing core 7 structure. One of the important complex O-glycans carries antigens, such as the ABO and Lewis blood group determinants, on core 1-4 and 6[155].

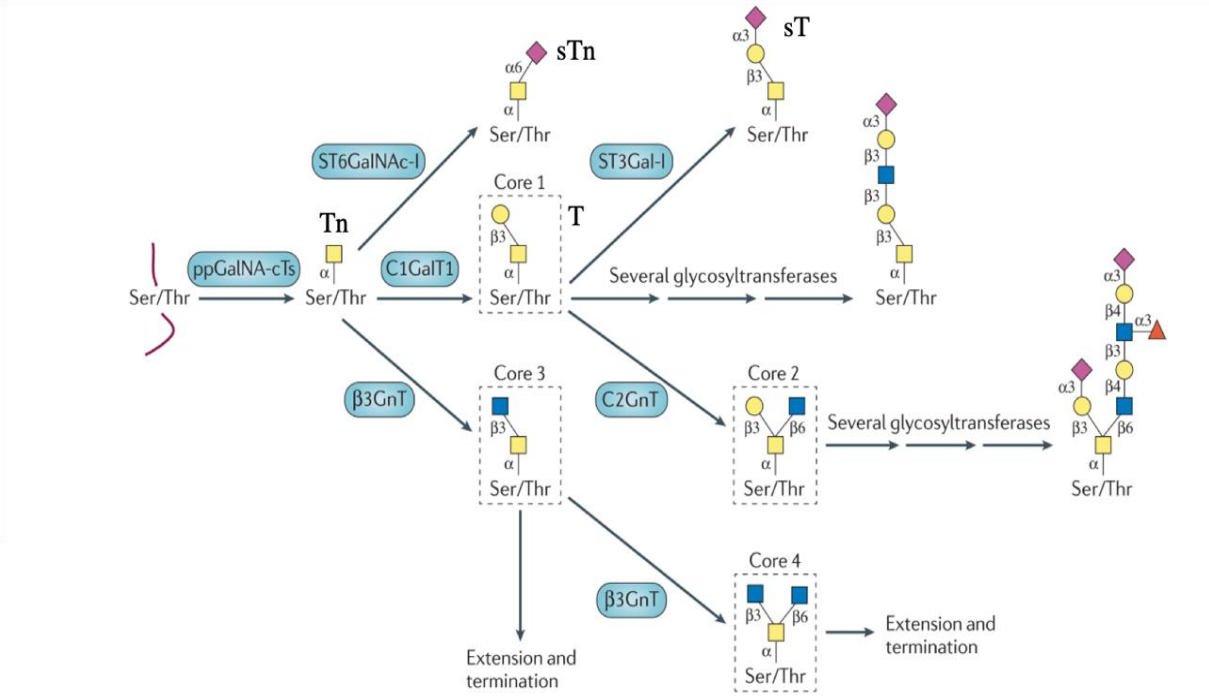


Figure 1.9 O-glycans core structure assembling pathway. Different glycoenzymes (in blue) are involved in the catalysis. Sialyltransferases, such as α-GalNAc ST6Gal-I (ST6GalNAc-I),α2,3-sialyltransferases (ST3Gal-Ts); fucosyltransferases 8 (Fuc-TVIII); N-acetylglucosamine (GlcNAc) transferases (GnTs); galactosyltransferase (GalT). Image adapted from Pinho et Reis (2015)

The terminal structures of O-GalNAc glycans may contain fucose, galactose, N-acetylglucosamine, sialic acid in α-linkages, N-acetylgalactosamine in both α- and β-linkages, and sulphate. Many of this sugar are antigenic and they are recognized by the lectins, receptor recognizing glycans. In particular, the sialylated and sulphated Lewis antigens are ligands for selectins: increase of sialylation and binding of aberrant Lewis antigens has been proved to play a major role in cancer invasion and metastasis.

1.4.3 Function of glycosylation

Glycosylation has important roles that spans from affecting protein structure and function to cell signalling, and migration[157].

Glycosylation is very important in regulating cell-cell communication by promoting or inhibiting molecular binding. For example, the adhesion molecule, NCAM, is known to be negatively charged by a poly-sialic structure in order to inhibit homotypic NCAM protein-protein binding[151].

Glycosylation function is regulated by glycans' receptor, named lectins. Lectins can recognize different sequences-motifs and are defined as C-type lectins, S-type lectins, P-type lectins, and the Siglecs[158,159]. Lectins have generally low affinity but high avidity, due to its multivalence allowing different sugar to bind the same lectin. Some lectins also have specificity for specific protein, suggesting a participation in protein recognition during lectin selective binding[151].

In general, glycans are very important in controlling different cellular mechanisms such as cell adhesion, self/non-self-recognition, receptor activation, molecular trafficking and endocytosis. In cell-cell adhesion one of the most key lectin is the selectin. Selectins bind the Sialyl-Lewis$^{x/a}$ ($SLe^{x/a}$) antigens on the endothelium of the vasculature and on the leucocytes, contributing to the immune system trafficking responses[160]. Selectin are single chain transmembrane protein and they share the calcium dependent C-lectin/N-terminal and the epidermal growth factor (EGF)-like domains. Their subtypes are divided into E-selectin (endothelial), P-selectin (platelet) and L-selectin (lymphocytes), where the first is expressed by immune cells and endothelial cells, the second by platelets and endothelial cells and the third mostly by leukocytes[68].

Glycans are beneficial in self/non-self-recognition as it happens in the activation of the toll-like receptor on immune system cells by binding of specific pathogens glycan.

An important example of molecular trafficking is the glycoprotein turnover promoted by the Asialoglycoprotein lectin which bind and internalize glycoproteins lacking or bearing insufficient sialic acid linkages. Asialoglycoprotein receptors can be expressed in endothelial cells, dendritic cells and macrophages[161]. Sometimes glycosylation can contribute to the intracellular regulation. For instance, O-fucosylation of notch receptor is in charge to control its trafficking, ligand-binding and activation. Similarly, the receptors MHC, EGF, TGF-β (on epithelial cells), the IgM B cell antigen receptor (BCR) and the glucose transporter 2 glycoprotein (on pancreatic β cells) are shown to be highly regulated by glycosylation. In fact,

glycans expressed on these receptors have demonstrated to contribute to the endocytosis and consequent internalization[162,163].

Glycosylation appears, thus, to be involved not only in all the main cellular functions, but also in receptors regulation and cells-cells communication. Hence, any disruption in the mechanisms can cause serious alterations in cellular homeostasis. This abnormal behaviour is often seen and studied in tumour microenvironments.

1.5. Hallmarks of cancer glycosylation

Alteration in glycosylation pathways, in glycoenzymes expression or in glycan sequences is involved in many diseases. Aberrant glycosylation, associated with regulation of cell adhesion, migration and proliferation, has been connected with the acquisition of cancer capabilities and suggested to be considered as new hallmark of cancer[32,33]. Efforts have been spent in finding cancer-characteristic glycans that could be used in early diagnosis and drug targeting.

Alteration in glycotransferases and glycosidases can origin at different levels: epigenetic, transcription, post-transcription, altered glycosidase activity, altered expression of glycoconjugate acceptor, altered sugar nucleotide transporter activity and improper function of the Golgi structure[164].

Overall, sialylation and fucosylation are the major typical terminal modification of proteins and they are involved with cancer formation and progression.

Sialylation is in charge of many cellular functions such as cell-cell communication and protein targeting; modification in total sialic acid and sialoprotein has been reported contributing to tumour progression[152]. Two enzymes, sialidase and sialyltransferase (ST), are responsible for controlling the sialic acid cellular contents. As the amount of sialic acid in tumour depends on many different STs, they have been considered promising potential biomarker for cancer early diagnosis, treatment and tumour prognosis[165].

Fucosylation is a modification which involve the addition of the fucose sugar unit by the fucosyltransferases (FUT) enzymes. FUT enzymes, expressed in many tissues, incorporates fucose to reducing end of complex glycans in the linkage specific manner. The fucose homeostasis is controlled by a lysosomal enzyme called α-L-fucosidase, which catalyse the hydrolytic cleavage of the terminal fucose. Fucoprotein are glycoproteins which contain fucose and have been reported useful in cancer prognosis and diagnosis. Alteration in FUT genes or/and proteins expression has been suggested involved in cancer formation[164].

Furthermore, different types of FUTs and STs are involved in the formation of Lewis antigens, which over or altered expression is involved in cancer malignant phenotypes and recognised as tumour antigens (Table 1.2).

Table 1.2 Association of different fucosyltranfersases (FUTs) and sialyltranferase (STs) in glycan tumour antigens formation. Table adapted from Vajaria and Patel (2016)

α-1,2 Fucosyl transferase 1(*FUT1)*	Le^b, Le^y
α-1,2 Fucosyl transferase 2 (*FUT2)*	Le^b, Le^y
α-1,3/4 Fucosyl transferase 3 (*FUT3)*	Le^a, Le^b, sLe^a
α-1,3Fucosyl transferase 4 (*FUT4)*	Le^x, Le^y, sLe^x
α-1,3Fucosyl transferase 5 (*FUT5)*	Le^x, Le^y, sLe^x
α-1,3Fucosyl transferase 6 (*FUT6)*	Le^x, Le^y, sLe^x
α-1,3Fucosyl transferase 7 (*FUT7)*	Le^x, Le^y, sLe^x
α-1,6Fucosyl transferase 8 (*FUT8)*	Trimannosyl core fucose
α-1,3Fucosyl transferase 9 (*FUT9)*	Le^x, Le^y
Beta galactoside α-2,3 Sialyl transferase I (*ST3Gal1)*	ST
Beta galactoside α-2,3 sialyl transferase 3 (*ST3Gal3)*	SLe^a
Beta galactoside α-2,3 sialyl transferase 3 (*ST3Gal3)* Beta galactoside α-2,3 sialyl transferase 4 (*ST3Gal4)*	SLe^x
Beta galactoside α-2,3 sialyl transferase 6 (*ST3Gal6)*	
ST6-*N*-acetyl galactosaminide alpha-2,6-sialyltransferase I *(ST6GalNAc1)* ST6-*N*-acetyl galactosaminide alpha-2,6-sialyltransferase 2 (*ST6GalNAc2)*	S6 T

Incomplete O-glycosylation pathways is linked with the formation of T- (Thomsen–Friedreich antigen or unmodified core 1 structure), Tn- (Thomsen-nuovelle antigen) and $SLe^{x/a}$ antigens which has been linked with tumorigenesis and bad prognosis[166]. The decrease in E-cadherin and increase in $SLe^{x/a}$ expression, with a correspondent overexpression in E-selectin, favourites the movement of cancer cells to distant organs, creating the basis for metastatic tumour formation[167]. Loss of E-cadherin increases, also, the Matrix metalloproteinases (MMPs) expression, enzymes involved in the degradation of the extracellular matrix[65]. In parallel, different class of tyrosine kinase, including EGFR, Her2-neu, insulin-like growth factor 1 receptor (IGF-1R) and c-Met are able to induce the EMT by inhibiting E-cadherin adhesion[168,169].

Glycosylation and, in particular, sialylation has been demonstrated to be involved in apoptosis evasion by the hyper-sialylation of the Fas receptor, which inhibits the recognition of Fas ligand and thus apoptosis activation[170]. In addition, glycoproteins such as mucins (MUC) produced by epithelial cancer cells or the MUC-like-glycoproteins like carcinoembryonic antigen (CEA)125, CEA19–9, CEA15–3, CEA72–4, are involved with tumour progression and are considered very important in study of early detection. For instance, MUC1 overexpression is involved with E-cadherin function disruption in adherents junctions in cancer cell [171].

Following the classification of the different hallmarks of cancer, different authors have proposed glycosylation as a missing hallmarks of cancer, integrating the role of aberrant glycans in each step of tumorigenesis (Figure 1.10)[32,164,172].

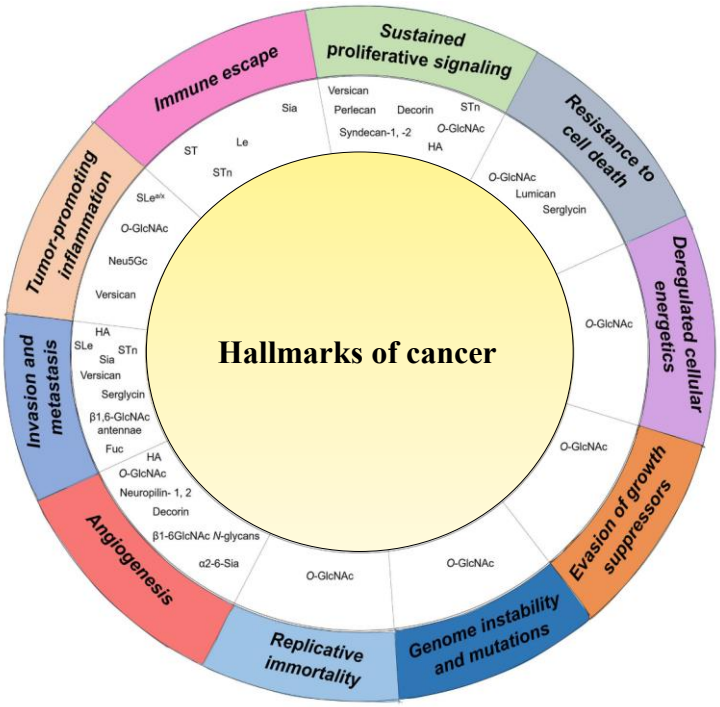


Figure 1.10 Association of expression of aberrant glycans for each cancer hallmark. General increase in O-GlcNAc is mostly common to all steps. Increase of Le, sLe, sTn is mainly involved with invasion and metastasis, tumour-promoting inflammation and immune escape. Image adapter from

1.5.1 Sustaining proliferative signalling

O-glycosylation of proteins can regulate cell cycle signals, including the transcription factor forkhead protein M1(FoxM1), cyclin D1 and cMyc[173,174]. For instance, incorrect glycosylation can contribute to oncogenesis by stabilizing the cMyc protein and competing with phosphorylation. Moreover, in ovarian cancer cells, the knock-down of the enzyme *Cosmc*, the molecular chaperon fundamental for a correct O-glycosylation, results in an increase of T- and Tn- antigen and their sialylated counterparts (sT- and sTn-), which are tumour-associated carbohydrates, not usually present in peripheral tissues and blood cells[175].

Degrees of branching in N-glycosylation are also involved in modulating proliferative signals, including EGFR, FGFR, PDGF, MET and IGFR[32,176].

Further role of glycosylation has been associated within the ECM: it facilitates the integrin dependent growth factor signalling to promote cell growth and survival and modify the function and signalling of the multifunctional cell surface molecule CD44[177,178].

1.5.2 Evading Growth suppressors

O-GlcNAcylation regulates the two main suppressors of proliferation, p53 and RB. In particular, when p53 is O-GlcNAcylated on ser 149, ubiquitin-dependent proteasome degradation is blocked, maintaining the p53 function activated. On the other hand, when O-GlcNAcase (OGA) is overexpressed, it is induced the activation by phosphorylation of MDM2, resulting in p53 degradation[179].
Similarly, RB is positively controlled by glycosylation and, in particular by O-GlcNAc in phase G1, by preventing phosphorylation and promoting binding of E2F-1 transcription factor. During final G1 cell cycle, phosphorylation prevails on glycosylation leading to the release of E2F and the entering in the S phase of cell cycle and thus promoting cell proliferation[180].

1.5.3 Deregulating cellular energetics

As previously described, one of the main hallmarks of cancer cells is the passage from the oxidative phosphorylation towards the aerobic glycolysis, also known as 'Warburg effect'[88]. In this hypoxic context, Hypoxia-inducible factor 1- α (HIF-1α) upregulates the expression of glucose transporters-1 and 3 (GLUT1, GLUT3), allowing the intracellular uptake and phosphorylation of glucose. This passage can influence most of metabolic pathways, such as the hexosamine biosynthetic pathway (HBP). The final product of the HPB pathway is the UPD-GlcNAc responsible for O-GlcNAcylation, critical metabolite in O- and N-glycosylation[181]. It has been shown that O-GlcNAcylation is involved in cancer-associated metabolic reprogramming: Glucose-6-phosphate dehydrogenase (G6PD), an enzyme of the pentose phosphate pathway (PPP) is O-GlcNAcylated in response to hypoxia and is involved with cell cancer proliferation[182]. O-GlcNAcylation at the serine 149 of the phosphofructokinase 1 (PFK1) has proven linkage with cell proliferation[183]. Moreover, O-GlcNac can regulate transcription factors activity and c-Myc stability, confirming the main role that Hyper-O-GlcNAcylation contributes to oncogenicity, frequently accompanied by cell dedifferentiation and acquisition of mesenchymal characteristics[89,174].

1.5.4 Resisting cell death

Glycosylation plays a major role in controlling the function of death receptors as Fas (CD95) and TNFR1; cancer cells use the glycosylation machinery to regulate apoptosis by disrupting ligand-receptor interactions, modulating the formation of signalling complexes and influencing ligand secretion from effector cells. For instance, the lectin galectin-3 binds the Fas ligand repressing the apoptotic signal and increase tumour survival[184]. Likewise, the addition of sialic acids by ST6Gal-I in an α2-6 linkage to the N-glycans of Fas provides protection against Fas-mediated apoptosis in colon carcinoma cells[185].

Other glycosyltransferases, as N-acetylgalactosaminyltransferases 1 (GALNT1) has been reported responsible for glycosylation at integrin α3β1 which modify its conformation and induces focal adhesion kinase (FAK) activation in bladder cancer cells[186]. Inhibition or de-phosphorylation of FAK is indeed an apoptotic inducer and has been shown to inhibit also the PI3K/AKT signalling, suggesting a role in the downstream apoptotic regulation[187].

O-GlcNAcylation has proved a main role in resisting cell death: both NF-κB p65 subunit and its upstream kinases IKKα/IKKβ are O-GlcNAcylated. Reduction in O-GlcNAcylation reduces NF-κB activity and target gene expression[188].

1.5.5 Enabling replicative immortality

The ability of cancer cell to reactivate the human telomerase reverse transcriptase (hTERT) and activation of telomerase, is the main mechanism to overcome senescence. There are still no direct proves of the linkage between glycosylation and telomerase activation, yet the role of c-MYC, direct activator of TERT, is discussed. In fact, c-MYC is a glycosylated protein and it has been shown to be stabilized by O-GlcNAc and with it, high level of O-GlcNAcylation has been shown in cancer cells[174,189].

1.5.6 Angiogenesis

Glycans are involved in many steps of angiogenesis (sprouting) and vasculogenesis (new tubes) and a distinct set of glycosylation genes have been associated into the cascade.

Important key inducer of angiogenesis is the VEGF and its correspondent receptor VEGFR: glycosylation of both is associated in angiogenesis. Heperan sulphate (HS) proteoglycans can

modulate angiogenesis by effecting the bioavailability and interaction of VEGFs with VEGFRs or interacting with endostatin, an anti-angiogenic factor[190,191]. The upregulation of O-GlcNAcylation on VEGF and aberrant glycosylation on VEGFR modulates the interaction with galectins[192,193]. For instance, galectin-1 (Gal-1) binds β1-6GlcNAc branched N-glycans present on VEGFR2 in endothelial cells surface to activate a VEGF-like signalling[194]. The interruption of the β1-6GlcNAc branching led to converted refractory tumours into anti-VEGF-sensitive, which suggest the main role of glycosylation in tumour therapy when treated with anti-VEGF treatments[194]. It has been also proved to impact angiogenesis through EGF receptor signalling and influencing the expression of angiogenic cytokines, in ovarian cancer[195].
O-glucose, O-GlcNAc, and O-GalNAc have shown to regulate angiogenesis by affecting Notch signalling, responsible for stabilization of arterial endothelial fate and angiogenesis [196,197].

1.5.7 Activating invasion and metastasis

The ability of glycans to activate invasion and influence metastatic conditions in cancer cells has been well documented. The role shown by glycosylation seems dual: from a side reduces adhesion of cancer cells, thus inducing CTCs formation, but from the other side promotes adhesion of cancer cells, inducing nesting of metastatic cells[33,124,160]. High levels of sialylated glycans present on tumour cells surface increase local negative charges that physically disrupt cell-cell adhesion and promote detachment from the tumour mass through electrostatic repulsion[198]. High level of sialic acid is indeed correlated with malignancy and bad prognosis. Likewise, the presence of truncated O-glycan, as sTn, due to inhibition of the COSMC chaperone or increase in ST6GalNAc1enzyme, has been associated with cancer cell growth and invasion[175,199]. More overexpression of glycosyltransferases, such as GnT-V, induce β1,6GlcNAc branching of N-glycans on E-cadherin, leading to non-function adherent junction and misleading cell-cell adhesion. GnT-III catalyses the addition of bisecting GlcNAc to complex N-glycans and is thought to influence interactions with galectins. Downregulation of GnT-III has been associated with increase of cell migration and metastasis[200].
SLewis antigens are instead associated with the tumour cells capability of promoting adhesion. The $sLe^{x/a}$ antigen is overexpressed in most of tumour and promotes adhesion of tumour cells by interaction with selectins on the endothelial surface[201]. Addition of a sulphate group at the sixth position of GlcNAc generates 6-sulfo-sLe^x, a ligand for L-selectin and E-selectin in bladder cancer[202]. In agreement, cell lines overexpressing Le^x have revealed an increase of FUT6 and FUT7 fucosyltransferases, two enzymes involved in the catalysis of sLe^x [203].

Sialyltransferase ST6GalNAc2 has a role in metastasis in breast cancer: loss of the enzyme corresponds to alteration of O-glycans, leading to an increase in metastasis[204].

Galetin-1 binding to CD44 or CD326 can promote attachment to the ECM and endothelial cells[205].

1.5.8 Genome Instability and mutations

As previously mentioned, p53 is glycosylated by O-GlcNAc and O-phosphate modifications which increase its stability. On the other hand, p53 plays the major role as DNA sentinel in maintaining its stability[206]. The protein kinase ataxia-telangiectasia mutated (ATM), a key regulator of DNA damage response (DDR), is also glycosylated: O-GlcNAc and phosphorylation play major roles in regulation of the DNA damage pathway[207].

1.5.9 Tumour promoting inflammation and tumour escape

The dual role of the immune system in cancer development is supported also from glycan change in the tumour system. The selectin proteins (E-, P- and L-) are important for the entry of circulating lymphocytes into peripheral lymph nodes and leukocyte emigration into inflamed tissues through binding of sialylated and fucosylated glycans[68].

Glycosylation is able to influence NF-kb and cyclooxygenase 2 (COX2), important proinflammatory molecules[188,208]. Other tumours have shown stroma containing several pro-inflammatory cytokines, as IL-1β and IL-6, which regulate the expression of biosynthetic glycosyltransferases to increase the expression sialylated antigens, as $SLe^{x/a}$ [209].

On the other hand, glycans can interfere in immune cell recognition, leading to the emergency of cancer cells resistant. In general, hyper-sialylation phenotype has been associated with the mechanisms of immune evasion.

Lectins have a main role in tumour escaping surveillance: galectins and siglecs (sialic acid-binding immunoglobulin-type lectins) are thought to be involved in immune evasion. In particular, siglecs can induce an antigen-specific tolerogenic programming, enhancing Treg cells and reducing the generation and propagation of inflammatory T cells, as happens in the binding of the macrophage siglecs-15 with sTn- inducing the TGF-ß secretion[210]. sTn can be also act on the membrane of MUC1 and CD44 by promoting DC maturation inducing tolerogenic mechanism[211]; MUC-sT on breast cancer cells, recognize siglecs-9 on

macrophages initiate inhibitory immune signalling through the activation of the MAPK/ERK pathway[212]. N-glycosylation of the immune checkpoint molecule PD-L1 can reduce proteasomal degradation which increase its immunosuppressive activity over T- cells[213].
Moreover, also the fucosylated group on Lewis antigens, bound to the C-type of lectins on the macrophages, induces a tolerogenic response by activation of cytokines IL-10 and IL-27, with consequent activation of TH2, T follicular helper (TFH) or Treg cells[214].

1.6. Colorectal cancer and glycosylation

Aberrant glycosylation on colorectal cancer progression have been significantly reported in the last 15 years. Incomplete and aberrant synthesis of glycans have been revealed in all the stages of CRC and two mechanisms can explain functional consequences: i) a direct alteration of function and structures of protein and lipids or ii) a shift in glycan binding proteins (GBPs) through galactose-binding proteins (galectins), sialic acid-binding immunoglobulin-type lectins (siglecs), and selectins[153]. Nevertheless, glycosylation revealed to be a challenging topic to investigate, being its regulation highly dependent on the context and on epigenetic signals. Consequently, many techniques have been applied in glycosylation studies in order to find patterns in carcinogenesis events.
CRC cell lines, tissues and biological sample from patients are analysed by gene expression (gene mutation, miRNA and mRNA expression), binding assay (antibodies/lectins used in enzyme-linked immunosorbent assay [ELISA], Flow cytometry, microarrays, immunofluorescent microscopy, immunohistochemistry and western blot), glycoenzyme modification (with synthetic enzymes, or knock-in knock-down and knock-out experiments) and mass spectrometry techniques, which allow the identification and characterization of glycan structures. The use of different samples and of multiple techniques is useful to overcome eventual bias that a single analysis may bring in the interpretation of glycan structures and behaviours. General alterations in N-, O- and GSL- glycans, fucosylation, sialylation, sulfation and (sialyl) Lewis antigens has been reported in CRC (Figure 1.11).

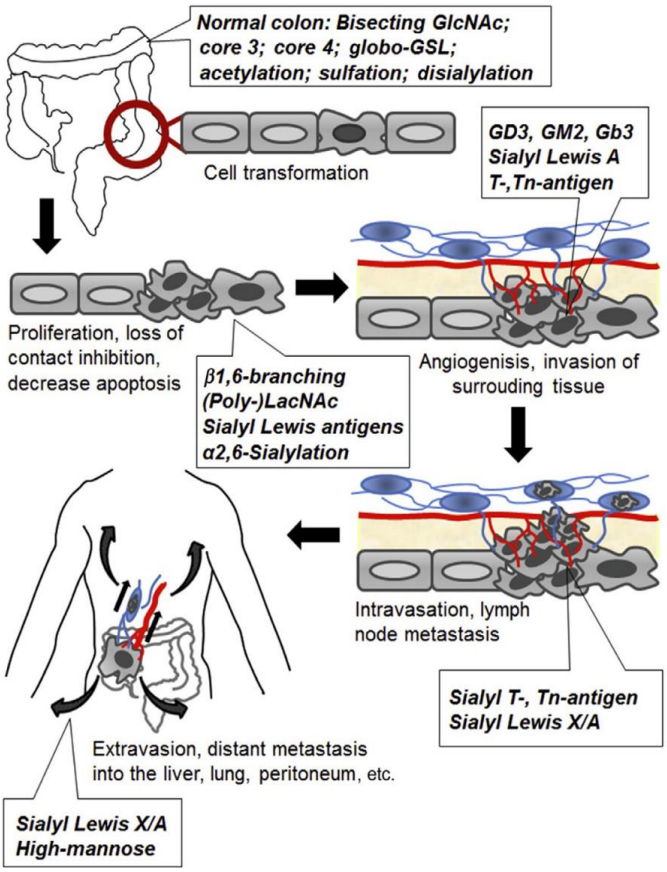


Figure 1.11 Alteration of glycan structure in different phases of CRC formation and progression. A general increase in SLe antigens, sialylation and fucosylation is common throughout the different steps. Image adapted from Holst et al. (2015)

In CRC, N-glycans alterations show a general intensification of high-mannose type structures as well as higher branching and core-fucosylation. The increase of high mannose type has been correlated in CRC tissue and cell lines with metastasis[215]. A hypothesis suggest that this increase is due to a precursor accumulation to incomplete maturation during N-glycan biosynthesis[216].

Core-fucosylated high- mannose glycans as well as truncated versions with or without fucose, also called pauci-mannosidic structure, were identified in tumours. This increase of core-fucosylation in cell lines was correlated with the upregulation of fucosyltransferase FUT8, responsible for the addition of α1,6-core-fucose on the innermost GlcNAc of the N-glycan core. Increasing of FUT8 expression have been associated with malignancy in CRC[217].

Upregulation of N-acetylglucosaminyltransferase V (GnT-V) is responsible for the extended β1,6-linked GlcNAc branching. The increase of the GnT-V has shown loss of contact

inhibition, increased cell motility, and morphological transformation, and in correlation, increase in higher branching have been associated with increase in invasion and metastasis[218]. The same branching can be extended by (poly-)N-acetyllactosamine (LacNAc), whose higher expression is associated with cancer progression, metastasis, and poor survival[219].

O-glycosylation is heavily present on the MUC membrane protein surfaces and their alteration have been studied. In general, an increase in core 1 structures which are generally truncated, sialylated and fucosylated facilitation tumour progression and metastasis. In colon cancer tissue the upregulation of core 1 β1,3-galactosyltransferase (C1GALT1, T-synthase) correspond to an increase in Thomsen-Friedenreich (T)-antigen leading to poor survival and metastasis[220]. Equally, the down regulation of C1GALT1 brings to a diminution in Thomsen-nouvelle (Tn)-, T-, and sialyl-T (sT) antigens and might be used in target therapies[153].

Overexpression of T-, O-acetylated sTn- as well as Tn-antigens have been associated also with tumorigenesis and progression, in the early events of malignant transformation. Main carriers of sT- and sTn-antigens were identified as the mucin MUC1 and CD44v6[221,222]. Sialylation of the T- antigen is performed by α2,6-sialyltransferase gene (ST6GalNAc2), while an increase in sTn- corresponded to an upregulation of the sialyltransferase ST6GalNAc1. Interestingly, in colon cancer carcinomas the enzyme core 2 β1,6-N- acetylglucosaminyltransferase (C2GnT), responsible for conversion of the T- and Tn-antigen to core 2 structures, is upregulated allowing the biosynthesis of sialyl Lewis$^{x/a}$, proved cancer antigens[223]. Correspondently, the enzyme β1,4-galactosyltransferase, responsible for type 2 chains (precursor for SLe^{x}) is upregulated and the β1,3-galactosyltransferases forming type 1 chains are downregulated[224]. Overall, an increased in density of O-glycan has been associated with tumour malignancy and progression.

Holst *et al* (2013) have investigated modifications in GSL-glycans, reporting a general pattern in colorectal cancer formation: (i) increased fucosylation, (ii) decreased acetylation, (iii) decreased sulfation, and (iv) reduced expression of globo-type glycans, as well as (v) disialyl gangliosides[225].

A general increase of fucosylation on glycans attached on lipids and proteins have been reported on N-, O- and GSL-glycans, by the action of various fucosyltransferases. In particular, FUT3, 4, 5, 6, 7, and 9 on antennary N-glycans are elevated in CRC and their upregulation brings elevated expression of cancer-associated blood group Lewis antigens[153]. It has been hypothesised that fucosylation plays a major role in early stages of the cancer, while a process of de-fucosylation is promoted in more advanced stages.

Sialylation plays important duties in healthy cells and sialic acid decorates the end position of glycans of proteins and lipids. The most common is the negatively charged N-acetylneuraminic acids (NeuAc), while adult humans lack the expression of another variant of sialic acids, N-glycolylneuraminic acid (NeuGc). However, adult tumour environment not only shows a general increase in sialylation but also appearing of NeuGc expression[226].

Dysregulation of silalyation is due to sialyltransferase and/or glycosidases or enhanced possible sialylation sites. In detail, increase in α2,6-linked sialic acids on N-glycans and α2,3-linked sialic acid have been associated with poor prognosis and metastatic profiles. In accordance, the α2,6-sialyltransferase ST6Gal1 is upregulated in colon tumours, as well as increased expression of α2,3-sialyltransferases with cancer cell migration and metastasis[227]. Further studies have to be performed in order to understand the role and the up regulation of the two type of sialic acid linkage, but hypothesis attribute main role in early stages cancer development to the α2,6-linked sialic acids, and critical role in metastasis with regard of epithelial–mesenchymal transition of the α2,3-linked sialic acid[153].

Decreases in sulfation have been observed, however contradictory results in cancer progression show the urge to further investigate the subject. It has been suggested that the negative charge of sulfation plays a critical role in protease resistance and interaction with functional molecules. On the contrary, the blood related group (sialyl) Lewis antigens X and A have been deeply investigating and are responsible of increased malignancy, poor prognosis and increased metastasis for creation of new ligands. The Le^a on the contrary diminish in cancer progression. Supposedly, incomplete synthesis of 6-sulfo sLe^x and disialyl Le^a cause the accumulation of $sLe^{x/a}$. Equally, neosynthesis by changes in glycosyltransferases have been ascribed to accumulation of the antigen[228]. In hypoxia conditions, overexpression of FUT7 leads to increase of (s)Le^x and of ST3Gal1 to $Le^{a/x}$. FUT3 and FUT6 responsible for sLe^a and sLe^x, respectively, are used as tumour markers, being overexpressed in metastatic conditions[229]. In general, FUT3, 4, 5, 6, and 9 as well as α2,3-sialyltransferases are involved in neosynthesis of Lewis antigens[229]. However, FUT 1 and 2 play a major role in converting $Le^{x/a}$ into antigens Lewis Y (Le^y) and Lewis B (Le^b) by addition of a fucose to the galactose of this motif[230]. Study reveals increase expression of these epitopes and their respective enzymes in colon tumour and carcinomas. An increase of FUT4 has been also observed and possibly related to an enhanced expression of LeY[231].

In conclusion, incomplete synthesis or neosynthesis is the common ground for aberrant glycosylation formation. Alteration in glycosyltransferases or sialidases brings to

tumorigenesis and formation of new ligand binding recognitions, most of the times responsible for malignant transformation and metastasis. Although many studies must be performed in order to fully understand the glycosylation role in cancer formation, it is undeniable its potential for targeted therapies. Among the all aberrant glycan just described, O-glycans and Lewis antigens, shows the most important role: not only they are, in most of the cases, cancer specific but also expressed from the early to the late stages of cancer. This is an important factor, that need to be exploited in biomarker discovery and targeted therapies. Also, Sialylation being responsible of malignant transformation and present in the final part of glycans structure, represents a good objective for cancer treatment.

1.7. Glycoproteins and biomarker discovery

Treatment in Colorectal cancer has higher chances to be effective during early stages of the disease, with chances up to 90% during the 5-year rate survival[232]. Consequently, most of the studies are driven towards the discovery of new and effective early detection biomarkers.

A good molecular biomarker must be sensitive and specific to the disease: detection of malignancy at first stages of the cancer and use of non-invasive tools are the central focus in biomarker discovery nowadays. As a matter of fact, it has been noted an increase in testing population when screening exams are quick and less invasive. Biomarker can be of different origin, including altered DNA, RNA, proteins, glycosylation and immune cells.

Up to the date, colonoscopy followed by tissue biopsy is the main technique adopted for CRC diagnosis and prognosis. However, use of human specimen and testing of liquid biopsy is also increasing. Each type of biomarker can be detected with different techniques and the use of diverse types of biomarker helps for a correct diagnosis and prognosis. Proteins are considering excellent biomarkers not only because important for cellular mechanisms (i.e. cell growth, cell signalling, protein metabolic process and cell motility) but also for their stability.

Moreover, some upregulated protein when glycosylated can also carry aberrant glycosylation even when secreted, revealing promising potential as liquid biomarker.

As mentioned, glycan profiles of proteins can be affected by of glycosyltransferases and glycosidases and by the availability of environmental factors (e.g., glucose, growth factors) and sugar nucleotides leading to alterations of glycan structures[153]. Some glycoproteins are already in use in clinics, such as CEA and CA19.9 as good prognostic factor.[233] Levels of CEA are checked after total resection of tumour and decreasing of its level in the blood correspond

to proper remission of the tumour[100]. CEA and CA19.9 have been in use also for diagnostic proposes, however, they lack specificity precluding them to be used as early detection tools. Increase in core-fucosylation in serum protein in pancreatic cancer have been shown and fucosylated haptoglobin has been also proposed as new biomarker[234]. The measure of aberrant glycans variation has diagnostic tool have been proposed. Increased levels of sLe^x and sLe^a in serum of CRC patients were highly associated with distant metastasis[201]. Likewise, Tn antigen and its sialylated form, sTn, have shown great specificity in cancer detection. Wei *et al* (2016.) reported 6 O-glycoproteins carrying Tn (Keratin 8, Keratin 18, Decorin, Sorbin, SORBS1 and CD44 antigen) to correlate with the initiation and progression of CRC[235].

Nowadays, many glycoproteins have been already investigated for CRC and in the next paragraphs will be highlighted the most discussed.

1.7.1 TIMP1

Tissue inhibitor metalloproteinases (TIMPs) is a glycoprotein that contributes to the inhibition of MMPs, which have been implicated with tumour invasion and metastasis by degradation of the ECM.

However, in colon cancer TIMP-1 is upregulated compared to their normal counterparts, eventually increasing the growth of cancers. It induces cell growth and inhibit apoptosis. Moreover, TIMP-1 promotes colon cancer progression and accumulation of CAFs (Cancer Associated Fibroblasts) in colon cancer [236]. TIMP-1 has been identified has a new early-stage biomarker with blood plasma sample having significantly elevated TIMP-1 levels compared with healthy individuals[237]. In CRC, TIMP-1 is aberrantly glycosylated with enhanced β1,6-branching in CRC[238]. The aberrant glycosylation decreases the inhibition power of the MMP2 and 9, causing increase in cell motility and metastatic phenotype.

The pro-tumoral activity, the upregulation and the presence of characteristic aberrant glycosylation make TIMP-1 a novel potential target for cancer therapy.

1.7.2 MUC1

Mucin is a class of high molecular weight glycoproteins expressed on epithelial tissue. In general, mucins are highly covered in glycans. Mucin can be classified as secreted gel forming mucins (MUC2, MUC5AC, and MUC6), transmembrane mucins (MUC1) and other types.

MUC1 is found on the apical part of most epithelial surface, yet during tumorigenesis, expression of specific mucins may be reduced or the organ specificity lost and new mucins may be expressed aberrantly[239]. MUC1 one can contribute in tumour growth by activation of ß-catenin trough binding of siglecs-9 and aberrant O-glycosylation expressed on MUC1 surface[240]. In metastasis, the expression of T-antigen on MUC1 serves as a binding of galectin-3, which promotes cell surface polarization of MUC1 and consequent exposure of E-selectin, increasing adhesion of cancer cells[241]. MUC1 was found playing a role also in immune modulation. The Tn-antigen on its surface interacts with macrophage galactose lectin (MGL) and instruct the DC to activate a Type 2 T helper cells (TH2)-mediated responses, which is not implied in tumour eradication[242]. Likewise, MUC1 can interact with NK cells and inactivate NK cell-mediated cytotoxicity[243]. A metanalysis from Zeng et al. (2015) not only confirmed a great correlation between CRC and MUC1 expression, but also suggested it as a good potential diagnostic tool in predicting status of metastasis. However, the use of different monoclonal antibodies to test the MUC1 can bias the specificity in MUC1 recognition, with ranging from 20% up to 92.7%, for this remains very important the creation of tools for specific and sensible recognition of the marker[244]. Vaccine against MUC1 tumour-associated antigens (MUC1-TAA) have been designed. Kimura et al. (2013) have reported high immunogenicity in ~ 50% of patient with colon adenocarcinoma tested. High levels of anti-MUC1 immunoglobulin G (IgG) and long-lasting immune memory was not associated with significant toxicity, suggesting an important role of this vaccine for cancer prevention[245].

1.7.3 Carcinoembryonic antigen

The carcinoembryonic antigen (CEA) is a glycoprotein of molecular weight ≤200 kDa and belongs to the immunoglobulin's superfamily [246]. CEA is well known to be a serum protein biomarker and has been acting as cell adhesion mediator for cancer cells. Many drugs are in use or in trail with target CEA, of which most have action on controlling tumour development in colon cancer patients. CEACAM5 exhibiting an aberrant glycosylation which is recognized by C-type lectin receptors expressed on dendritic cells (DCs) and modulates the innate and adaptive tumour immune response[247]. Expression of Lewis antigens on CEA and CEACAM1 proteins impairs DC maturation and increases the secretion of immuno-suppressive cytokine interleukin-10 (IL-10)[248]. Thus, CEA produced by tumour cells is not recognized by the immune system, for this reason have been developed new strategies to enhance immune reactions against CEA. Vaccines against CEA have been designed and under clinical trial,

whose porpoise is the activation of the immune system against cells overexpressing CEA. Recent studies showed that the CEA expression is directly linked and increases with cancer stage [249]. Yet, it has been calculated a low sensitivity and specificity in screening cancer early stages I-II [250]. However, it has been identified that its receptor, CEAR, mediates CEA pro-metastatic activity to other organs, thus it would probably play an important role in drug development for cancer treatment[251]. CEA increase in metastatic cancer that develop liver and spontaneous pancreatic and lung metastasis. Zhao *et al.* (2018), using lectin staining revealed higher expression levels of fucose, mannose and T antigen on CEA in the tumour tissues relative to the tumour-adjacent normal tissues; sufficient to distinguish CRC tumour tissues from healthy tissues with high sensitivity and specificity[252]. These results suggest that the use of combinatorial lectins and detection of aberrant glycosylation on CEA could be used as potential effective and non-invasive biomarker.

1.7.4 CA 19-9

Carbohydrate antigen (CA) 19.9, correspond to the carbohydrate sLewis^a and is used in clinic as prognosis biomarker with the CEA. CA 19.9 name has been defined by a monoclonal antibody produced by hybridoma of mice immunized with CRC cell line, SW 1116. Sensitivity and specificity to detect CRC are still under discussion since is less effective than CEA. However, when used in combination with CEA it increases up 71.7%[139]. These suggest the sensitivity of this combination increases significantly [139], suggesting that even though it cannot be considered a good single biomarker but could be a great booster in diagnosis and prognosis when combined. Other carbohydrate antigens, such as CA 50, CA 195, CA 242, CA M26, CA M25, CA M43 and CA 72.4, have been tested but excluded due to their reduced sensitivity, stage dependency and specificity[253]. Although testing of carbohydrate antigens reveals a dwell in specificity and sensitivity, the combined use with glycoprotein discloses an interesting raise in sensitivity, which could be use in early detection.

1.7.5 CD44

CD44 is a ubiquitous multifunctional glycoprotein. Main duties involve cell-cell and cell-matrix adhesion and interaction with growth factor and extracellular ligands for signal transduction and cytoskeleton rearrangements[254]. CD44 is also involved in lymphocytes

homing haematopoiesis, cell migration and adhesion, tumour invasion and metastasis. It is a single chain molecule composed by a distal extracellular domain, a membrane-proximal region, a transmembrane-spanning domain, and a cytoplasmic tail. Its molecular weight may vary from 85-230kDa as a consequence of the many isoforms originated by alternative splicing and the heavy glycosylation[255]. CD44 gene is located in the short arm of the 11th chromosome and consist of 19 exons of which 1–16 encode the extracellular domain, exon 17 encodes the highly conserved transmembrane domain, and exons 18 and 19 encode the intracellular domain. Azevedo *et al.* (2018) performed an in-silico study of the possible mRNA transcript and results possible 21 combinations of which just 8 are proved experimentally[256]. Analytical difficulties are aggravated by the glycosylation, and possible glycosylation sites present on the splicing sites: glycans often present a non-templated and context-dependent nature, increasing molecular diversity and functionality[256]. The known 8 isoforms are gp116 and gp85, glycoproteins named after their molecular weight, CD44H and E observed on hematopoietic and epithelial cells respectively, HCELL a CD44 hematopoietic cell E-/L-selectin ligand mostly implicated with tumour progression and metastasis, CD44v3, CD44v6, CD44v9, named after their monoclonal antibodies recognition, involved with cancer, including chemoresistance and prognosis. Being the study of this protein very complex due to the different isoforms, roles and possible posttranslational modifications, Azevedo *et.al* (2018) pointed out the necessity for a better characterization of the glycosylation of this protein, having specific cancer-associated isoforms, which have a great potential as tumour biomarker[256]. The role of CD44 as a diagnostic and prognostic tool is still under discussion since results remains controversial.

Recently, Wang *et al.* (2019) have performed a metanalysis of 48 studies with total CD44 isoforms overexpression: as results the expression of CD44 was showing direct correlation with worse OS of CRC patients and was associated with lymph node metastasis, distant metastasis, poor differentiation, and CD44v6 with tumour size[257]. However, other studies high levels of CD44 mRNA was associated with benign outcome survival rate of gastric tumour[258]. In intestinal CSCs, the variants 4-10 are found the most abundant, however other isoforms are mostly abundant. In breast cancer, the CD44v4 is responsible to mediate cell adhesion to endothelial monolayer[259]. Soluble isoform of CD44 is also available. Proteolytic cleavage release CD44 from the membrane by detaching the cytoplasmatic tail. In has been found serum concentration at 2 µg/ml in mouse and in tumour-bearing strains increase up to approximately 10 µg/ml. Immunodeficient mice have low concentrations of 0.5 µg/m indicating that malignant activities can induce release from cell surface CD44[222,260].

1.8. Antibodies

The immune system plays an everyday central role in maintaining cellular homeostasis by fighting non-self-agents. The immune response is defined innate when the immune cells do not rearrange their receptors and therefore are not affected by previous interaction with infectious agents. In contrast, the acquired immune response is when the immune cells suffer genetic rearrangement on their receptor genes and its effectiveness can be greatly improved after contact. A central role in the acquired immune system is played by the antibodies, or immunoglobulins (Igs). They are glycoproteins produced specifically by B lymphocytes, which expose on the cell surface an antibody with a unique antigen binding site, that is used as receptor or secreted. In fact, antibodies have the ability to recognize with high sensitivity and specificity a particular sequence (epitope) of an antigen[262].

The remarkable multifunctionality of the antibodies, is given by its structure. The "Y" shape is formed by a constant part, Fc, which correspond to the lower part, and a variable part composed by two antigen-binding fragments, Fab, at the top. The Fab portion is composed by a heavy and a light chain both in the constant and variable part (CH, CL/VH, VL) (Figure 1.12a). The antigen specificity is created through continuous genetic recombination of the hypervariable complementarity-determining regions (CDRs) present both in the light and heavy chain[262]. The Fc constant part is instead responsible of the effector function of the antibody, activating effector molecules and stimulating the cells of the immune system. There are five classes of immunoglobulins named IgG, IgM, IgA, IgD and IgE which maintain the four-chain antibody structure but differ for the constant region of the heavy chain (Figure 1.12b) and function[263]. For instance, IgM, as monomeric form is a cell surface receptor on B lymphocytes, and as a pentameric form is the first class of antibodies secreted by B lymphocytes once activated, i.e. plasma cells. Upon specific stimuli, the plasma cells can suffer class switching and then express other class of antibodies, but with the same antigen specificity and differentiate into memory B cells. After the primary response, and due to point mutations, the plasma cells may express antibodies with higher affinity. The IgG can activate several mechanism of action as antibody dependent cell cytotoxicity (ADCC), complement dependent cytotoxicity (CDC), phagocytosis, or blockage of the action of specific molecules(Figure 1.12b).[264].

In detail the ADCC is activated by the engaging of Fc receptor on Natural Killer (NK) cells, leading to their activation and exocytosis of the cytolytic granule complex perforin/granzyme and thus activation of apoptosis. The CDC is a cytolytic cascade mediated by a series of

complement proteins (C1–C9) that are abundantly present in the serum: C1q binds to the Fc region of IgM or IgG bound to antigens, which triggers a cascade of signal ending with the lysis of the targeted cell. Other soluble antibodies can function as decoy for signal homeostasis or cross-linking of receptors that are connected to mediators of cell apoptosis to induce cell death[262]. Antibodies are present in the serum in a polyclonal form and different clones in fact can recognize different epitopes on the same antigen. However, each B cell clone can produce only one antibody. The antibody and the antibodies produced by one B cell clone is then called monoclonal. Nowadays, monoclonal antibodies are one of the most important tools used in research for characterization and in therapy as leaders of the development of personalized treatment.

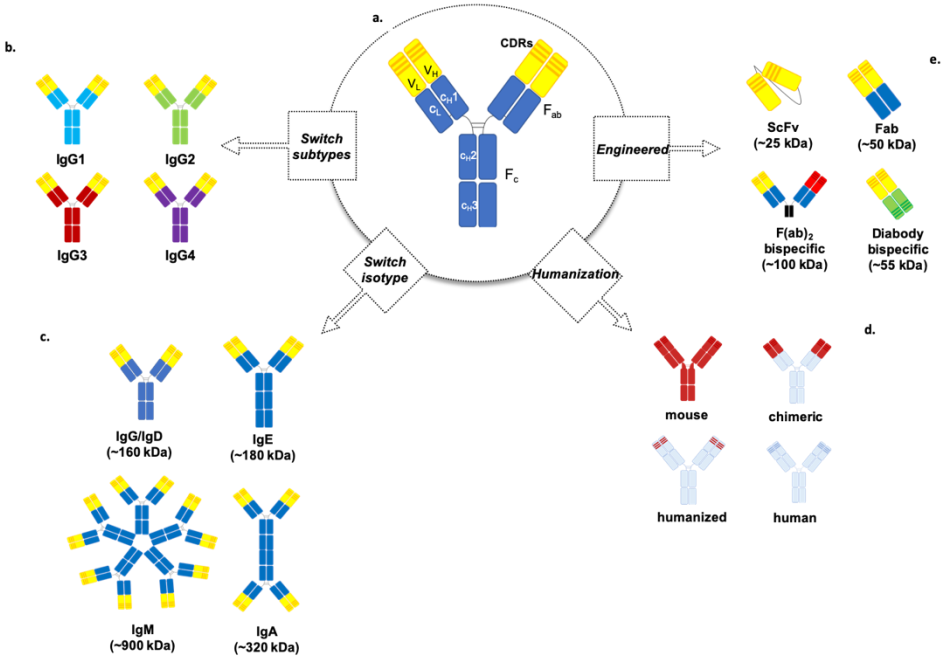


Figure 1.12 Broad overview of antibody structure and diversity. a) structure of an immunoglobulin G (IgG) antibody's molecule is composed by constant (C) and variable (V) domains for each light (L) or heavy (H) chain. The heavy chain comprises three constant domains (CH) and one variable (VH), and the light chain contains one constant (CL) and one variable (VL) domain. Specificity of an antibody is determined by its complementarity-determining regions (CDRs). b) IgG can switch different subtypes depending on the Fc effector function. c) switch in class of antibody improves the effector function of the antibody and increases its ability to eliminate the pathogen that induced the response. engineering of antibodies comprehends the different modification to the Y structure with the intent to reduce immune response and increase efficacy. d) humanization of the antibody sequence and e) creation of shorter fragments have been the most successful modification up to the date.

1.8.1 Production of monoclonal antibodies

The first monoclonal antibody was generated in 1975 and fully licenced in 1986. Since then, the market and the production of monoclonal antibodies has raised, becoming fundamental in research and in clinics[265]. After 40 years, around 47 monoclonal antibodies have been approved

in EU or USA for treatment of many different diseases[266]. Major problems in antibody production is to obtain a highly specific antibody, obtain high yield in production and guarantee the less immunogenicity or secondary effect, as possible, in therapies. For this reason, advancing in genetic sequencing and biotech techniques as led to efficient modification of antibodies constructing for example short fragments antibodies like single chain variable fragment (scFv) with reduced size and immunogenicity[267]. Engineering the Fc region in modulation of its interaction with the neonatal Fc receptor (FcRn) promotes IgG half-life extension (Figure 1.12e)[268]. Likewise, the glycosylation of the antibodies can modify and modulate the Fc effector functions of antibodies[269]. To promote efficacy, monoclonal antibodies can also be modified by conjugation of coupling effectors such as, plant/bacterial toxins, enzymes, radionuclides, cytotoxic drugs etc[265,270].

Three are the main techniques to produce monoclonal antibodies: hybridoma technology (the oldest), immortalized B cells and phage display[265].

1.8.1.1 Hybridoma technique

Hybridoma technique was used to produce the first monoclonal antibody in 1975. The generation of this cells begins with immunization of a particular animal species, with the epitope or antigen target and collection of the B cells from the spleen of the sacrificed animal. The polyclonal mix of B cells is fused by using chemicals, by electrofusion or virus-assisted, with an immortal myeloma cell line lacking the hypoxanthine-guanine-phosphoribosyl transferase (HGPRT) gene[271]. The fused or hybridoma cells are then cultured in hypoxanthine-aminopterin-thymidine (HAT) selection medium in which only the hybridoma cells are able to survive and proliferate since they have inherited immortality conferred by the myeloma cells and capacity to survive in culture medium in the presence of aminopterin conferred by B cells. Aminopterin blocks the principal pathway for the synthesis of nucleotides and unfused myeloma cells cannot produce nucleotides by the salvage pathway as they lack the enzyme HGPRT[272]. The selection is about two weeks long and after the cells, which are selected in a polyclonal environment, are then separated usually by limiting dilution cloning techniques. The cell medium supernatant of single clone cell is then tested, usually by ELISA, to select the specific clone producing the desired antibody. The selected hybridomas can then be stored away in liquid nitrogen[273]. Drawbacks of this technique consist in a non-stable production of clones, but mostly the creation of murine antibodies that are then recognized as non-self by the

immune system when used in therapy (HAMA effect). For this reason, Winter *et al.* (1988) initiated the method of humanization of the Fc and Fv framework regions by replacing the murine sequences with human germline amino acids, revolutionized the field of antibody therapeutics (Figure 1.12d)[274]. Nonetheless, difference in species in the sequence remains very important in research, which take advantage of the diversity in complex immunoassays studies.

1.8.1.2 Immortalized B cells

B cells from peripheral blood of a donor can also be immortalized with the help of the Epstein-barr virus (EBV), which belongs to the herpes family and has the ability to efficiently immortalize *in vitro* nearly all human B cells. The B cells with the desired antibody, are selected and expanded before immortalization due to the low frequency of the cells. After immortalization although they are able to produce continuously the antibody, the EBV-transformed cells cannot grow indefinitely since they do not have the phenotype of malignant cells[275]. These cells are also difficult to grow and do not produce the antibody with a high yield. However, when the B cell is extracted from the patient it allows the production of antibodies with unique specificity against neoantigens. In order to increase its production and stability, techniques of cell fusion have been applied also to these cells[276].

1.8.1.3 Phage display

Phage display is a technique invented in 1985 to overcome the problems of immunogenicity of full antibodies produced in animals. In fact, with the phage display is possible by sequencing the variable fragments to produce short fragments antibody (scFv) which consist only of the heavy and light chains variable parts of the antibody (Fig12e)[267]. The first step in phage display is the production of phage libraries, created by isolating B-lymphocytes from the blood of humans and then isolating the mRNA and converting it into cDNA using PCR to amplify all the VH and VL segments. Once the segments are cloned into a vector next to the protein pIII of the phage genome, the phage are expanded by infection of *E. coli* in order to generate a library containing approximately 10^{10} cells by inoculating the library with an additional helper phage. The linkage to the pIII allows the VH and VL to be exposed on the outer surface of phage which is then tested against different antigens. After several cycles of binding of the phage with the targeted epitope, the selected phage is expanded a second time through infection

of the *E. coli*, the phage expressing the desired scFv isolated and the nucleotide sequences of the scFv VH and VL sequenced, enabling the useful creation of different antibodies[166]. The tricky part in phage display is to obtain a proper library against the desired antigen. In fact, different libraries can be tested more times and the process of selection is quicker, due to the absence of immunization and sacrifice of animals. However, without a good phage library becomes very hard to produce the desired antibody. Phage display is particularly indicated when toxic agents must be used (not suitable for animal immunisation) or when the molecule are very small (not immunogenic). For this reason one of the successful scFv antibody against Tn has been produced by Kubota *et al*.[277]

1.8.2 Monoclonal antibody in cancer

In the last 20 years the use of monoclonal antibody in cancer and solid tumour treatment has revealed to be one the most successful therapy strategy. The selection of the target for monoclonal antibody treatment required a deep knowledge in cancer dynamic and immunoglobulin activity. In early days the use of murine antibodies provoked the development of immune response HAMA (antibody anti-mouse antibodies) which reduced drastically their clinical utility (Figure 1.12d-e)[278]. However, with humanization in transgenic mice of the Fc and Fv sequences, by Winter and the production of short fragments by phage display, have helped reducing drastically the side immune response[274]. The mechanism of killing the cells by antibody, can occur directly by receptor blockade or agonist activity, induction of apoptosis, or delivery of a drug, radiation, or cytotoxic agent; immune-mediated cell killing mechanisms; regulation of T cell function; and specific effects on tumour vasculature and stroma. As previously mentioned, the Fc is fundamental in the activation of the ADCC and CDC systems[262]. Up to 2012, 12 antibodies have been approved by FDA for solid and hematologic cancers[278]. The most successful includes the one belonging to the ErbB family, to which belong VEGF and EFGR. In colorectal cancer, the use of anti-EGFR is effective in wildtype KRAS patients, while is considered not useful with mutated KRAS[143]. For this reason, many clinical trials have been developed around the molecular tumour characterization of the patients and the effectiveness of the monoclonal antibodies in combination with the classic chemo and radio therapies. Furthermore, the use of monoclonal antibodies can help T cell activation. T cells targeting to the tumour is improved by antibody-mediated blockade of T cell inhibitory check points (e.g., CTLA4 and PD1)[279,280]. In general, many tumours associated antigens have been

identified, among which glycoprotein cluster of differentiation (CD) CD20, CD30, CD33, CD44 and CD52, growth factors including CEA and EGR and vascular growth factors including VEGF[143,169,222,281].

Finally, the introduction of the CAR-T cells in tumour treatment has become on central role in antigen discovery and antibody production. In fact, CARs (Chimeric Antigen Receptor) are fusion proteins of a selected single-chain fragment variable from a specific monoclonal antibody and one or more T-cell receptor intracellular signalling domains. The modification can occur by CRISPR-CAS or direct transfer *in vivo* of mRNA by electroporation. CAR-T cells are based on the genetic modification of patient's autologous T-cells to express a CAR specific for a tumour antigen, following by *ex vivo* cell expansion and re-infusion back to the patient. The CAR-T would then be able to redirect and activate the immune system against tumour cells[282,283]. Clinical trials have shown a recovery up to the 92% Acute Lymphocytic Leukaemia.

1.8.3 Glycan-binding monoclonal antibodies and lectins

Monoclonal antibodies and lectins have been used in research as important tools to characterize glycan expression, on protein and cell surfaces. Many of the lectins used are extracted from plants or animals, while antibodies are generated by immunization of the animal with antigen carrying the desired glycan or Glycan-protein conjugates (glycans coupled to carrier proteins such as bovine serum albumin [BSA] or keyhole limpet hemocyanin [KLH])[166].

In glycan identification, antibodies and lectins have different advantages. In fact, while lectins recognize specifically general ligands, such as α2-6-linked sialic acid and α1-2-linked fucose, monoclonal antibodies can recognize specifically other certain antigens, as sLe^a[284]. However, in case of antigens such Le^x and sLe^a is not possible to distinguish N-glycans, O-glycans or glycolipids with neither of them. In other situations, some lectins can recognize just the N-glycan from the O-glycans: the plant lectin Concanavalin A (ConA) does not bind mucin-type O-glycans in animal cells but binds only to some specific classes of N-glycans[284]. Most of the lectins and antibodies are used for glycan purification, either with affinity chromatography or affinity binding and immunoprecipitation (using protein A or G for non-covalent binding with the antibody) or lectin-induced precipitation. The same antibody and lectins are used in western blot experience or in cell characterization experiments, through flow cytometry, immunohistochemistry and ELISA. The use of glycosyltransferases and glycosidases in combination with antibodies and lectins is optimal for glycan characterization and function assays.

The production of monoclonal antibodies against glycans that could be use in therapy is also becoming a central focus in research. As previously described, the glycan tumour profiling shows constant alteration compared to healthy tissue and overexpression of tumour-associated carbohydrate antigens (TACAs), as a result of aberrant glycosylation in tumours, is usually correlated with poor prognosis and survival of cancer patients[285]. TACAs are associated both with inducing cancer metastasis and invasiveness and using TACAs for immunotherapy targeting, could play a major role, not only as diagnostic targets, but in also in preventing tumour metastasis and thus increasing OS. Several criteria have been chosen for selection of anti-tumour mAbs: i) the antibody binds the cell tumour surface ii) with high affinity, iii) the antibody recognize a specific tumour marker that has a limited expression in healthy tissue, iv) the mAb must be potent immune-mediated with non-immune-mediated cytotoxicity effects, v) and the antibody could directly kills tumour cells and/or vi) the mAb internalises into target cells so it can delivery toxic agents[285].

In 2015, Dinutuximab a mAb against GD2, a disialoganglioside overexpressed in neuroblastomas and melanomas, have been approved as a good mAb for cancer treatment. In therapy, combined with granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin-2 (IL-2) and 13-cis-retinoic acid (RA), has been shown in a phase III trial an effective therapeutic agent in children with high risk of neuroblastoma. Its action induces cell lysis via ADCC and CDC[286].

Another monoclonal antibody against Le^y, named BR96, have shown to induce ACDD and CDC directed against tumour cells. When conjugated with BR96 was conjugated to doxorubicin and docetaxel (also known as SGN-15), showed remission power in advanced non-small-cell lung cancer patients[287]. In order to reduce HAMA effects, a humanized mAb (hu3S193) against Le^y have been also created. It contains only 3%-5% mice content and in phase one have shown minimal toxicity in 15 patients with different tumours. In MCF-7 xenografts mice have demonstrate capability to reduce tumour mass compared to a placebo treatment[288].

Monoclonal antibodies against sLe^a have been also produced and tested. KM231 is a murine anti- sLe^a antibody which have been shown to react against human gastrointestinal cancer and detection in patient serum. It have been also conjugated with ricin A chain immunotoxin and tested on xenograft: the KM231-ricin significantly inhibited the growth of established subcutaneous tumours suggesting tumour toxicity efficacy[289]. 5B1 (IgG1) and 7E3 (IgM) are also mAbs anti- sLe^a produced by KLH conjugated vaccine immunization. Both have shown

great CDC activation[290]. NCCT-ST-421 (IgG3) is a murine mAb which recognize Le^a and induces ADCC and CDC to antigen-positive gastric cells. FG88 are internalizing murine antibody, which recognize $Le^{a\text{-}c\text{-}x}$ glycans. A subclone FG88.2 have shown great binding associated with poor outcome in the colorectal sample cohort, suggesting good potential as diagnostic and therapeutic tools in most aggressive colon cancers[285]. Other antibodies against sTn have been produced: B72.3, TKH2, 3F1 and 3P9. Some of these were test in clinics with few successes. However, 3P9 have demonstrated inhibit proliferation and migration of STn-expressing cells as well as on tumour growth by CDC and apoptosis. A new monoclonal antibody against sTn have been produced, L2A5. The results demonstrated the high sensitivity of L2A5 mAb to detect tumour tissue with higher extension, in a sialic acid-dependent binging, compared to the widely used anti-STn antibodies. Loureiro *et al.* (2018) have proved that LA25 have a great potential not only as diagnostic tool but also effective targeting and blocking of cancer-associated antigens involved in mechanisms underlying tumour progression in multiple types of cancer[199].

In conclusion, nonetheless the difficulties, great progression has been made in monoclonal antibody production against cancer-aberrant glycans. The advances in clinical trial suggest a great potential of monoclonal antibodies as tools for diagnosis, prognosis and treatment.

Rationale and aims of the thesis

CRC is considered the second deathliest cancer-type and one of the most common human malignant diseases. Due to its localization in the body, it is still very difficult to be detected in the early stages, when treatment is more effective. CRC detected at late stages is more likely to develop recurrences, even after treatment, leading to the necessity to create a new system to early stages detection. Colonoscopy is the main technique used to obtain tissues, to be used for diagnosis, followed by molecular analysis, supporting both diagnosis and prognosis. However, there is still a necessity for more reliable biomarkers.

The understanding of the pathological molecular pathways has contributed not only to molecular diagnosis and allowed the identification of novel targets for treatment. Nowadays, two targeted antibodies, Cetuximab (Erbitux), an anti-epidermal growth factor receptor monoclonal antibody (anti-EGFR) and Bevacizumab (Avastin), a vascular endothelial growth factor monoclonal antibody (anti-VEGF), are in use as treatment to all the CRC stages and for liver and lung metastasis. Although these treatments have been successfully used combined with chemotherapies, they are not efficient for all patients and have associated toxicity, because the targets are not cancer specific as their main function is the inhibition of cell proliferation processes and angiogenesis.

Nevertheless, understanding the biology of the cancer and biomarker discovery are important steps in cancer research; it is essential to develop new sensitive, low cost, minimally invasive and accurate blood test to meet clinical needs and to improve cancer detection.

Many modifications have been found and investigated during the stages of CRC. Many carriers, such as the mucin MUC1 CD44v6 and CEA, have been already identified carrying T- sT- (sialylated/Thomsen-Friedenreich (T)-antigen) and sTn-antigens (sialylated-Thomsen-nouvelle (Tn)-antigen). This aberrant glycosylation is associated with poor survival, cancer progression, and metastasis. On the other hand, the Lewis antigens and their sialylated derivatives (Le^x/sLe^x and Le^a/sLe^a) are the most prominent cancer-associated epitopes on both glycoproteins and glycolipids, since their overexpression is related to CRC malignant transformations and may lead to increased tumour cell adhesion and motility and, thereby, result in metastasis.

Therefore, the project aimed to two main goals: discovering of new biomarkers for CRC early detection and producing specific antibody against the new identified target.

This thesis has been focussed in validating Lewis antigens as well as others E-selectin ligands expression in each stages of the cancer progression. In fact, the role of E-selectin ligands, such as $sLe^{x/a}$ and glycoprotein CD44, has been reinforced in cancer progression and metastasis by several basic researches, thus studies of E-selectin ligands behaviour could be of great utility in biomarker research. For the same reason glycoproteins, such as CEA and CD44 have been studied for the development of novel monoclonal antibody production. Biomarker validation was performed using crossed experiments on CRC cell lines and on CRC patients' tissue. For the development of the project, three main method of screening have been adopted in order to look for new biomarkers for Colorectal cancer (CRC): Flow cytometry, western blot and immunohistochemistry. In flow cytometry, live cells of CRC have been screened for the principal glycoproteins known to be involved in cancer progression and for the most prominent aberrant glycans present in CRC tissues. From this experiment, four mayor antigens have been selected for western blot-total cell lysate analysis: carcinoembryonic antigen (CEA), CD44, Lewis (Le) and sialyl Lewis (sLe). In parallel, screening of CRC tissues with a panel of the same antibodies used in flow cytometry have been performed. As a results of the screening CD44 and $sLewis^{x/a}$ have been selected has potential antigens for new antibodies productions. From the biomarker validation with immunoassay, CD44 decorated with aberrant $sLe^{x/a}$ resulted to be the most promising biomarkers. Thus, monoclonal antibodies have been produced with the use of hybridoma technology against these targets. Novel antibodies against $sLe^{x/a}$ and CD44 have been produced and validated for antigens recognition. Two clones have been selected for further characterization either for early diagnosis and/or CRC treatment

MATERIAL & METHODS

2.1 Cell lines

2.1.1 Stock

During the development of the project, different colorectal cancer cell (CRC) lines were handled. All the cell lines were genetically profiled in order to guaranty cell line authentication. Profiling of human cells was performed by *Microsynth* company, who analysed specific polymorphic short tandem repeat loci (STRs). STR loci were amplified using the PowerPlex® 16 HS System (Promega). Fragment analysis was done on an ABI3730xl (Life Technologies) and the resulting data were analysed with Gene Marker HID software (Softgenetics). Cell line samples matching at ≥80 % of alleles across the eight reference loci were said to be related.

Follow the list of cell lines used during the PhD project:

1. CCD 841 CoN (ATCC® CRL-1790™) cell line established from normal tissue from 21 weeks' gestation female fetus.
2. LS174T (American tissue and cell collection reference: ATCC®CL-188TM) cell line was established from Dukes' type B adenocarcinoma of colon from a 58-year-old Caucasian female. Grade I- Stage 2
3. HT29 (ATCC®HTB-38TM) derived from an adenocarcinoma of colon from a 44-year-old Caucasian female. Grade I- Stage 3
4. SW948 (ATCC®CCL-237TM) was established from Dukes' type C, adenocarcinoma of colon from an 81-year-old Caucasian female. Grade III- Stage 2
5. SW48 (ATCC®CCL-231TM) derived from Dukes' type C, adenocarcinoma of colon from an 82-year-old Caucasian female. Grade III → IV – Stage 3
6. SW620 (ATCC®CCL-227TM) cell line derived from a Dukes' type C metastatic adenocarcinoma from colon in lymph node from a 51-year-old Caucasian male. Grade IV-Stage 3
7. HEK293 T (ATCC® CRL-1573™) cell line derived from embryonic kidney from epithelial tissue.
8. JY1 [LASJY1] (ATCC® 77441™) non-adherent lymphoblastoid releasing IL-6 in the supernatant. Cells gently handled by Prof. *Sandra van Vliet from VUMC, The Netherlands.*

I. HT29 mock and HT29 FUT6. Transfected with void plasmid or plasmid containing human Fuc-TVI cDNA, with G418 resistance.
II. SW948 mock and SW948 ST6GAL1. Transduced with void plasmid or plasmid containing human ST6Gal1 cDNA, with blastacidin resistance.
III. SW48 mock and SW48 ST6GAL1. Transduced with void plasmid or plasmid containing human ST6Gal1 cDNA, with blastacidin resistance.
IV. SW620 mock and SW620 FUT6. Transfected with void plasmid or plasmid containing human Fuc-TVI cDNA, with G418 resistance.

2.1.2 Cell culture of colon cell lines

All cell lines were adherent and were grown in T25 and T75 culture flasks (SARSTEDT), in a 37°C incubator (Panasonic) with a humidified atmosphere and 5% CO_2. SW48, SW948 and SW620 cell lines were cultured in L-15 Leibowitz's medium (Lonza) supplemented with 10% of Fetal Bovine Serum (FBS) (Gibco), 1% 200mM L-glutamine (Gibco) and 1% 10,000 U/mL Penicillin/Streptomycin (Gibco).
LS174T cell lines were cultured in MEM (Biowest) supplemented with 10% FBS (Gibco), 1% 200mM L-glutamine (Gibco), 1% 10,000 U/mL Penicillin/Streptomycin (Gibco), 1% 100X Non-Essential Amino acids (Gibco) and 1% 100mM Sodium Pyruvate (Gibco); HT29 (wild type and transfected), HEK293 T, CCD 841 CoN and LY ebv transformed (source of IL-6 for hybridoma stability) (ATCC® 77441™) cell lines were cultured with DMEM (Gibco) supplemented with 10% FBS (Gibco), 1% 200mM L-glutamine (Gibco) and 1% 10,000 U/mL Penicillin/Streptomycin (Gibco); SW48, SW948 and SW620 cell lines were cultured with L-15 (Lonza) were supplemented with 10% FBS (Gibco), 1% 200mM L-glutamine (Gibco) and 1% 10,000 U/mL Penicillin/Streptomycin (Gibco) in closed tap T-flasks. In fact, cells growing in L-15 (Lonza) medium don't need the use of CO_2.
Regardless specific experiments medium was substituted maximum every four days, respecting the stability time of the components. In normal passages, cells were harvested at 80/90% of confluency, washed with 1X phosphate buffered saline (PBS) and detached with 0.05% Trypsin-EDTA (TE) (Gibco) for cells-depended variable time 1'-5'. Trypsinization of cells must be require the minimum time possible not to intact glycosylation. Centrifugation (Eppendorf) at 200x *g* for 5 min and cultured according to the desired dilution and cells behaviour. All cell lines resuspended in culture medium and 10% (v/v) of dimethyl sulfoxide (DMSO) (Sigma) for storage at -80°C.

2.1.3 Cell growth curve

Cell growth curves were tested for doubling times and cells behaviour. Growth curved were tested for a maximum of 15 days. Cells were maintained in starvation (0% FBS) for 24h and then distributed in T25 flasks (SARSTEDT). 20˙000 cells were seeded in 15 T25 Flasks (SARSTEDT). Medium was changed every 4 days for 15 days. Pictures were taken every 24h hours. Cells were counted with Neubauer Chamber diluted in Trypan blue (Invitrogen).
The curves were then constructed using the Log_{10} value and analysed in Excel (Microsoft).

2.1.4 Influence of confluency in glycosylation expression

In order to test changing in glycosylation depending on cell confluency, cells at time 0 (24h starvation to synchronize cell cycle) were seeded at 50%, 40%, 30%, 20%, 10%. Concentration were calculated by proportion of 100% confluency/cm^2, revealed from cell growth curves (paragraph 2.1.3). Depending on doubling time (24h, 36h and 48h). cells were harvested and tested by flow cytometry (paragraph 2.3.1) for $SLewis^{x/a}$ (HECA-452 clone, Biolegend), $SLewis^{x}$ (CD15s -CSLEX1 clone, BD Pharmigen), $SLewis^{a}$ (CA19.9-203 Abcam), STn (B72.3 from hybridoma supernatant), CD44 (2C5, R&D).

2.1.5 IC50 5-FluoroUracil

SW620 wild type (wt), mock and FUT6 cells ($lxl0^4$) were seeded in 96-well microtitration plates 24 h before exposure to various concentrations of 5-FluoroUracile (5-FU) (0.1 to 100 µM) for 72h. Drug sensitivity was determined by MTT [(3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide)] cytotoxicity assay (Promega) for 4h. After exchange of MTT with DMSO and incubation for 30 min at 37˚C, sample were analysed. Each concentration was performed in four replicate wells. Untreated cells were used as the control. The colorimeter test was measured at Absorbance 497 nm. The average growth inhibition rates compared with the control were calculated from the result with at least three independent experiments. The 5-FU concentrations causing a 50% growth inhibition compared with the controls (IC-50) were

calculated from a semilogarithmic dose-response curve by linear interpolation, using GraphPad Prism 7 Software.

2.1.6 Transduction of LS174T cell line with lentivirus

A three-plasmid system was used that contemplates a lentiviral construct with the cDNA sequence of interest, an expression vector with the vesicular stomatitis virus G envelope glycoprotein gene (VSV-G) and a packing vector. They also have present truncated long-terminal residues (LTR) for self-inactivation. Firefly luciferase cDNA-LUC1(*Photinus pyralis*) is a 1,6kb fragment that codes for 550 amino acids residues and a 61kDa protein. Firefly luciferase (*Photinus pyralis*) genes were before amplified by polymerase chain reaction (PCR) and cloned into different the pLenti6/V5 Directional TOPO cloning vector containing the CMV promoter. Lentiviral particles were generated by co-transfection of HEK 293 T cells with the three plasmids, using pPACK Lentivector Packing Systems (Bioscience).
$5x10^5$ cells per well were seeded with 10^6-10^9 TU/ml of virus to a concentration of 8 µg/ml of polybrene (Sigma-Aldrich) in 2 ml of medium. The plate was shaken and incubated overnight, 37º, 5% CO_2. After 24h the new culture medium was placed with 3,2 µg/ml of Blasticidin (Thermo Fisher) for two weeks. Cells that were already transduced (i.e. SW948/48 ST6Gal1) carrying the Blasticidin antibody selection gene, were selected manually/ To performed it, the cell cultures were kept growing until they could form visible and isolated colonies. Twenty-four random colonies were then picked by small papers imbibed in trypsin that would detach them. The colonies were cultured in 24 well plates (Greiner Bio-One Cellstar) or in no-treated culture dish (Stemcell) for 3D culture.

2.1.6.1 Luciferase assay

The enzymatic reaction requires the substrates luciferin, with co-factors ATP, Mg^{2+} and O_2: the reaction was performed using Luciferase Assay System Kit (Promega).
Firefly luciferase emits a yellow-green light corresponding to a wavelength of 560nm.The luciferase assay offers a nearly instantaneous measure of total luciferase expression in the lysate; therefore, the emitted light is proportional to the luciferase concentration. Although the high efficiency of the assay, the luciferase half-life is less than three hours due to its low intrinsic stability, as a disadvantage.

24 hours before the detection assay, 10^5 cells *per* well were cultured in a 96 well plate (Oregon scientific). After 24h the medium was removed, and the cells were washed with 1X PBS. 1X Reporter Lysis Buffer (RLB) from the kit was added to each well and a single freeze-thaw cycle at -80˚C was performed to ensure the complete lysis. At room temperature, luciferase assay reagent was added, composed by luciferase assay buffer and luciferase assay substrate. Samples were read at ChemiDoc Touch (Biorad) to check for chemiluminescence.

2.2 Antibodies and lectins

The use of antibodies and lectin-chimeras was necessary in all the immunoassay experiments performed: flow cytometry (FC), immunofluorescence, Förster resonance energy transfer (FRET), immunoprecipitation (IP), immunohistochemistry (IHC), enzyme-linked immunosorbent assay (ELISA) and western blot (WB).), western blot (WB). Primary conjugated and not conjugated used are listed in the following Table 2.1. Secondary conjugated antibodies used are specified the appropriate technique section. In the table are listed three main categories of antibody used: specific for CRC classification experiments and glycoprotein and glycans used throughout the entire project development.

Table.2.3 List of primary antibodies used in immunoassays.

	ANTIBODY EPITOPE	**ANTIBODY NAME (lectin name)**	**ANTI-**	**PRODUCED IN**	**CLONE**	**ISOTYPE**	**µg/mL**	***Brand***
CRC CLASSIFICATION	5-hydroxytryptamine receptor 2B	**HTRB2**	human	rabbit	polyclonal		0.1	*Sigma*
	FERM domain containing 6	**FRMD6**	mouse, human	rabbit	polyclonal		0.2	*Sigma*
	Caudal Type Homeobox Transcription Factor 2	**CDX2**	human	rabbit	polyclonal	IgG	1	*Novus biological*
	Zinc finger E-box-binding homeobox 1	**ZEB1**	human	rabbit	polyclonal		0.2	*Sigma*

	Human epidermal keratin	**AE1/AE3**	Bovine, Chicken, Human, Mouse, Non-human primate, Rabbit, Rat	mouse	monoclonal	IgG1	0.2	*Thermo*
GLYCOPROTEINS	CEA (CEACAM5)	**CD66e**	human	mouse	3E10-3	IgG1 kappa	1	*Immunotools*
	CD13	**CD13**	human	mouse	3D8	IgG2a	0.5	*BD Pharmigen*
	CD44	**CD44-APC**	human	mouse	P2A1	IgG2a,k	0.5	*USBiologicals*
	CD44	**CD44**	Mouse, Rat, Human	rabbit	polyclonal	n.a.	0.1	*AbCam*
	CD44s Pan Specific	**CD44**	human	mouse	2C5(monoclonal)	IgG2A	0.5	*R&D*
	SLe x/a // CLA	**HECA**	Human, Mouse	rat	452	IgM, κ	0.5	*Biolegend*
	E-selectin ligands	**E-selectin /CD62E (lectin)**	human	Mouse myeloma cell line	X	chimera IgG	0.1	*R&D*
GLYCANS	SLe a	**CA19.9**	human	mouse	CA19-9-203	IgG, μ	0.2	*Abcam*
	SLe x	**CD15s**	human	mouse	CSLEX1	IgM, κ	0.5	*Bd pharmigen*
	Le x	**CD15**	human	mouse	HI98	IgM, λ	0.5	*Biolegend*
	Lewis y	**A70/C-C8**	human	mouse	7le	IgM	0.2	*Santa Cruz*
	sTn	**B72.3**	human	mouse	monoclonal	IgG1	n.a.	
	sTn	**LA25**	human	mouse	monoclonal	IgM	n.a.	
	sTn	**3F1**	human	mouse	monoclonal	IgG1	n.a.	
	Tn	**5F4**	human	mouse	monoclonal	IgM	n.a.	

2.3 Immunoassays

2.3.1 Flow Cytometry

The expression at cell surface was analysed by flow cytometry. Approximately 3×10^5 cells were harvested per condition and resuspended in PBS buffer. Cells were washed and incubated for 30 min at 4˚C with primary conjugated and not-conjugated antibodies. Subsequent washing steps were performed at 1000/1500 RPM for 2 min, and not-conjugated primary antibodies were detected with FITC conjugated anti-mouse Ig (Dako, F0232; dilution 1μg/mL) or APC anti-rat Ig (BD Pharmingen, 1μg/mL) for 15 min in the dark. After washing at 1000/1500 RPM,

data acquisition was carried out using Attune Acoustic Focusing Cytometer (Applied Biosystems) in which 10 000 events, gated on forward scatter (FSC) vs. side scatter (SSC), were collected for each sample. Data were collected through BL1-A and RL1-A channels. After 1X PBS cell suspension aspiration, the density plot graph FSC-A *vs* SSC-A is generated and then the gate is drawn to cover the cell population within 10 000 events. This gate is used to generate a histogram graph of fluorescence channel *vs* count where the positive cells peak is counted from the end of the unstained peak. At this point the median fluorescence intensity (MFI) or % of positive cells is extrapolated.
If required, the cells were fixed in 4% Paraformaldehyde (PFA) before to be analysed. Data was subsequently analysed using FlowJo vX 0.7 software.

2.3.2 Antibodies blocking migration by wound scratch assay

Cells were cultured in Culture-Insert 2 Well in µ-Dish 35 mm (Ibidi) inside 12-well plate (Invitrogen) until 100% confluence was reached; after silicon membrane removal, cell were maintained outside incubator in L-15 medium (Lonza) at 37˚C in gentle motion. Antibodies diluted at concentration of 0.5µg/mL were added every 3 days [SLewis^{x} (CD15s -CSLEX1 clone, BD Pharmigen), SLewis^{a} (CA19.9-203 Abcam), CD44 (anti-CD44-pan 2C5, R&D)] and relative controls prepared. Pictures were taken every 24h with an inverted microscope equipped with a digital camera (Xerox). For analysis, a fixed area of the wound was defined. Every 24 hours, the area occupied by migratory cells was measured and discounted to the initial area until all the wound was filled by cells.
The wounded area in the photographs was measured using ImageJ software and the percentage

2.3.3 Immunofluorescence

0,4 x 10^{6} cells/mL cells were cultured on top of round coverslips inside 12 well plates (Orange Scientific) and cultivated for 24h. In case of intracellular staining, after 1X PBS wash, they were fixed and permeabilized with 200µl of Fixation/Permeabilization Solution (Kit from BD Biosciences) for 20 minutes at 4ºC. After three washes with 1X PBS for 5 min, cells were blocked with 1% Bovine Serum Albumin (BSA) (Sigma-Aldrich) and washed three times again. In case of extracellular staining, the cells were directly incubated with 1 ml of Fixation/Permeabilization Solution (Kit from BD Biosciences) for 30 minutes. Dilution of

primary antibody was chosen with the same criteria of flow cytometry in Fixation/Permeabilization Solution (Kit from BD Biosciences); incubation for 30 min at room temperature (see paragraph 2.3.1). After x3 5 min wash with 1X PBS, cells were incubated with secondary antibodies with anti-rat IgM-FITC (1:50) (BD Pharmingen), anti-mouse IgG-FITC (1:100) (Dako) or anti-mouse Ig-FITC (1:100). To stain the nucleus, To-Pro3-Alexa Fluor 633 (Invitrogen) or 4',6-diamidino-2-phenylindole (DAPI) (1µg/mL) (ThermoFisher) were used by covering the cells for 15 minutes and then washed 3Xx3 with 1X PBS, 5 min. Finally, the coverslip sealed with mounting medium Mowiol (Dabco) and analysed in the Fluorescence Microscope. The microscope used is Axiolmager D2 – Zeiss, with a 40x objective and the laser filters of 358nm, 488 nm and 633 nm to excite FITC and Alexa Fluor 633, respectively. The images obtained were processed by Fiji software.

Fluorescence intensities from five randomly selected microscopic fields of cells were quantitatively analysed with ImageJ software using the Corrected Total Cell Fluorescence (CTCF) formula:

CTCF = Integrated density – (Area of selected cell x Mean fluorescence of background readings)

2.3.4 Immunohistochemistry

A cohort of 40 clinical cases of colon and rectum cancer (adenocarcinomas and adenomas) were obtained from the Portuguese Institute for Oncology of Porto (IPO-Porto, Portugal), according to the local committee of ethics. All procedures involving tumour tissues have been authorized by the Portuguese Institute of Oncology of Porto ethics committee. Tissue comprehend a variety in stages and cancer location. All the tissue come from 5-fluorouracile (5-FU) treated patients.

Formalin-fixed, paraffin embedded (FFPE) tissues were screened by immunohistochemistry using the biotin/streptavidin system. FFPE tissue sections were deparaffinized with xylene, rehydrated with a graded series of alcohol washes and subjected to heat-induced antigen retrieval using citrate buffer pH=6.0 (Vector Laboratories) for 15 min in the microwave, after pre-heating of the solution at maximum power rating for 5 min. Sections were incubated with 0.3% hydrogen peroxide (Merck KGaA) for 25 min, blocked with UV Block® (Thermo Scientific, TL-060-HD) and incubated overnight at 4ºC in a wet chamber for glycan antigens

[anti-CA19.9 1:100, anti- CD15s 1:100, anti-CD15 1:100, anti-Lewis^{y} 1:75, HECA-452 1:100], for glycoproteins [anti-CD44 polyclonal 1:4000, anti-CEA 1:50] for CSM-CRC markers [anti-HTR2B 1:3000, anti-FRMD6 1:3000, anti-AE1/AE2 1:500, anti-ZEB 1:4000 and anti-CDX2 1:2000].

After washing with PBS-Tween, biotinylated secondary antibody anti-mouse IgG (H+L) was added to tissue sections, before incubation with streptavidin (Thermo Scientific, 434301). Binding was detected by incubation with 3,3'-diaminobenzidine (ImmPACT™ DAB, Vector Laboratories, SK-4105) for 4 min. Nuclei were counterstained with hematoxylin for 1 min. Positive and negative control sections were tested in parallel. The negative control sections were performed devoid of primary antibody. Positive controls consisted of known positive tumour tissues for the antigen in study. Tumours were classified as positive when mAb immunoreactivity was observed by microscopic presence of brown chromogenic product in tumour cells. Staining was assessed double-blindly by two independent observers and validated by an experienced pathologist. Whenever there was a disagreement, the slides were reviewed, and consensus was reached.

2.3.4.1 In situ proximity ligation assay

In situ proximity ligation assay (PLA) was performed in FFPE sections selected with the best score of intensity for anti-CD44 polyclonal, anti-CD15s and anti-CA19.9 for detection of CD44-sLe$^{x/a}$ glycoform using the Duolink *in situ* Detection Reagents (Olink Bioscience, Uppsala, Sweden). FFPE tissues were deparaffinized, rehydrated and submitted to acid and heat-induced antigen retrieval, followed by incubation with 3% hydrogen peroxide and blocking solution in a humidity chamber. The Duolink in situ Probemaker was used according to the manufacturer instructions to conjugate the PLA oligonucleotide arms directly to primary antibodies also used in immunohistochemistry. Conjugated mAbs to CD44 (1:50) and anti-CA19.9 (1:100) or anti-CD15s (1:100) (labelled with PLA probes) were diluted in PBS with 5% BSA and with 1:20 of assay reagent (Olink Bioscience) and slides incubated for 1h at room temperature. Slides were then washed with wash buffer A (0.01 M Tris, 0.15 M NaCl, and 0.05% Tween 20). For the ligation step, the two oligonucleotides (1:5) and the ligase (1:40) both diluted in pure water, was added to slides and incubated for 30 min at 37 °C. After washing, the amplification solution, consisting of nucleotides and fluorescently labelled oligonucleotides (1:5) together with polymerase (1:80) (both diluted in pure water), were added

to the samples. The oligonucleotide arm of one of the PLA probes acted as a primer for a rolling-circle amplification reaction using the ligated circle as a template, generating a repeated sequence product. The fluorescently labelled oligonucleotides were hybridized to the amplification product. The slides were washed with 0.2 M Tris and 0.1 M NaCl. To visualize the nuclei, cells were incubated DAPI (Sigma-Aldrich, 0.4 mg/ml). Samples were examined under a Zeiss Imager.Z1 Axio fluorescence microscope (Zeiss, Welwyn) PLA results were evaluated by two observers and validated by an experienced pathologist, who independently registered cytolocalization of staining. Negative controls included analysis without primary antibody and sialidase treated tissue sections.

2.3.5 RT q-PCR

Total Ribonucleic acid (RNA) was extracted using GenElute^{TM} Mammalian Total RNA Purification kit (Sigma-Aldrich), according to manufacturer's instructions. Optional on-column DNaseI digestion (Qiagen) was also performed. Next, 10 µL of purified RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The conversion mix was prepared by adding, for each PCR tube (VWR), RT Buffer, RT Random Primers, deoxynucleotides (dNTPs), MultiScribe Reverse Transcriptase enzyme and RNase free water (NZYTech), following the manufacturer's instructions. Complementary Deoxyribonucleic acid (cDNA) conversion was performed in a thermocycler (Programmable Thermal Controller PTC-100^{TM} – MJ Research) with the following program: Step I -10 minutes at 25ºC, Step II - 120 minutes at 37ºC and Step III - 5 seconds at 85ºC. Real-time PCR was performed starting with preparing the RT-PCR mix, adding for each RT-PCR tube (Simport), 2 µL of cDNA, 2 µL of diluted probe 1:4, 1 µL of RNase free water and 5 µL of the Master Mix Fast (Applied Biosystems). The probes used was St6GalNAc1(Applied Biosystems). All experiments were performed in duplicates and two endogenous controls were used, β-actin and Glyceraldehyde 3-phosphate dehydrogenase (GADPH). Samples were run in RotorGene 600 (Corbett) with the following program: Step I – 1 cycle, 20 seconds at 95ºC and Step II – 40-50 cycles, 3 seconds at 95ºC and 30 seconds at 60ºC. Finally, results were analysed by calculating the relative mRNA levels using the adapted formula (2.2), which infers the number of messenger RNA (mRNA) molecules of the interest gene per 1000 molecules of the endogenous controls[291]. GraphPad Prism 7 Software was also used.

(2.2) Relative mRNA level = $2^{-\Delta Ct}$ x1000

2.4 Biomarker isolation and study

2.4.1 Protein extraction

Total protein extraction: $5x10^6$ cells/mL were lysed in IP Lysis/Wash buffer (Thermo Scientific) complemented with 1% Protein Inhibitors (Roche) and incubated for 20 min in ice. Incubation was interrupted every 5 min with vortexing for 1 min. Subsequently, cell debris were precipitated by centrifuged at 10000 x *g* for 10 min and protein concentration was quantified using Pierce BCA Protein Assay Kit (Thermo Scientific) following manufacturer's instructions.

Membrane protein extraction: membrane protein extraction was performed with the Mem-PER™ Plus Membrane Protein Extraction Kit (Thermo scientific) following manufacturer's instructions (Protocol 1: Adherent Mammalian Cells). 1% Protein Inhibitors (Roche) was added before the use of the permeabilization buffer. Regardless the instruction, cell concentration may depend on the density of the protein required.

All the protein extracts were stored frozen at -80˚C, in order to reduce variables and to guaranty experiment continuity.

2.4.2 Immunoprecipitation

For protein isolation, 100-200 µg of cell membrane extracted from LS174T and HT29FUT6 with Mem-PER™ Plus Membrane Protein Extraction Kit (Thermo scientific) were precleared with Protein G-agarose (Invitrogen). Precleared samples were incubated with 3µg of anti-CD44 or HECA-452, or E-selectin-Ig (in presence of calcium) for 2h and then incubated with Protein G-agarose beads (pre-blocked with 1% BSA) overnight. Mix with beads was washed three times with lysis buffer and boiled in SDS-PAGE sample loading buffer for 5 minutes or precipitated with 0.1 M Glycine (stabilized with 5% 1M Tris-HCl pH 9.5). Then E-selectin-Ig immunoprecipitated (IP) is treated with PNGase F (Sigma-Aldrich) and Neuraminidase from Clostridium perfringens (Roche) enzymes as recommended by the manufacturer. Finally, samples of E-selectin-Ig IP and CD44 IP (non-treated or treated with PNGase F/

Neuraminidase) were subjected to reducing 8% SDS-PAGE, transferred to PVDF membrane, and immunoblotted with mAb HECA-452, CD44-h-pan or E-selectin-Ig.

2.4.3 SDS-PAGE and Western blot

SDS-PAGE: Sample were loaded intoin 4% stacking gel (dH2O, 30% Acrylamide, 0.5 Tris pH6.8, 10% Sodium dodecyl sulphate (SDS) [Invitrogen], Tetramethylethylenediamine (TEMED) [Sigma-Aldrich], 10% Ammonium Persulfate (APS) [Sigma-Aldrich]) and separated into a 8% resolving gel (dH2O, 30% Acrylamide, 1.5 Tris pH8.8, 10% SDS [Invitrogen], TEMED [Sigma-Aldrich], 10% APS [Sigma-Aldrich]) in Running buffer) at 100 V for 30 min followed by 120 V for 1h.

Transfer: Proteins separated by SDS-PAGE proteins were transferred into a polyvinylidene fluoride (PVDF) membrane at 400 mA for 1 hour at 4ºC and using Transfer buffer (25 mM Tris, 192 mM glycine, pH 8.3, 20% methanol). Membranes were blocked with 7,5% non-fat milk powder dissolved in Tris-buffered saline- tween (TBST) 0.1% for 1 hour at RT. Membranes were incubated with HECA-542 mAb (1:1500), anti-CD44-h-pan-2C5(1:1500), anti-CEA – CD66e (1:500), anti-CD15s - CSLEX1(1:500), diluted in TBST 0.1% and E-selectin chimera (1:1000 overnight at 4ºC) TBST 0.1% and 2 mM $CaCl_2$.

Revealing: Primary antibodies were washed 3X for 10 min in TBST 0.1%. Next, membranes were incubated with appropriated Horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in TBST 0.1% for 1 hour at RT (anti-mouse IgG 1:5000, Southern Biotech; anti-rat Ig, 1:1500, Southern Biotech). Chimera antibody, E-selectin was washed TBST 0.1% and 2 mM $CaCl_2$ and stained with anti CD62E (1:1500) (anti-human, rat; clone 68-5H1, BD Pharmigen) for 1h at room temperature. After 3X washes of 10 min, was stained with anti-rat HRP IgG (H+L) (1:1000) (Southern Biotech), diluted TBST 0.1% and 2 mM $CaCl_2$.

Final 3X wash with TBS-t 0.1% was followed by staining with Lumi-light Western Blotting Substrate (Roche) for detection. Revelation was performed exposing the membranes to Amersham Hyperfilm ECL (GE Healthcare Life Sciences) and using Corestream® Kodac® autoradiography GBX developer and Corestream® Kodac® autoradiography GBX fixer (sigma).

2.5 Biomarker glycosylation characterization

2.5.1 Enzyme treatment

The sialidase treatment was performed to remove the sialic acid from structures according to the experimental case, once it had hydrolysed the α2-6 and α2-3 sialic acid linkage. After the first wash, 100µg/µL of immunoprecipitated from E-selectin-Ig (R&D) was mixed with 400 µl of sialidase buffer. A sialidase treated and non-treated control samples were prepared. Sialidase from *Clostridium perfringes*, (Roche) was added to a final concentration of 100 mU/ml. Both samples were incubated at 37ºC, for 90 minutes in continuous mixing.
PNGase F treatment was performed on protein immunoprecipitated from E-selectin-Ig (R&D) to remove N-glycans. Samples and control samples were diluted in 3X denaturing buffer (New England, Biolabs) and denatured at 100˚C for 10 min. following PNGase F reaction was performed following manufacturer's instruction (New England, Biolabs) with 500U/µL at 37˚C in continuous mixing.
Both enzyme reactions were tested by 8% SDS-PAGE and Western Blot analysis, described in section 2.3.3.

2.5.2 Förster Resonance Energy transfer

Förster/Frequency Resonance Energy transfer or FRET is a no-radioactive transfer from a suitable molecule donor to a suitable molecule acceptor. FRET happens when emission wavelength fluorophore of the donor overlaps with the excitation wavelength of the acceptor, and if the two molecules are at the max distance of 9 nm. For the FRET experiment were used three primary antibodies: anti-CD44, rabbit polyclonal (Abcam), anti-CD15s- CSLEX1 mouse (BD Pharmigen), anti-CA19.9-203 mouse (Abcam). Secondary conjugated antibodies were chosen in pairs for each different experiment, as shown below:
1. anti-rabbit FITC IgG 495-519nm (Sigma, 1 µg/mL) --> anti-mouse IgG PE 566-576 nm (Sigma, 1µg/mL)
2. anti-mouse IgG PE 566-576nm (Sigma, 1 µg/mL) --> anti-rabbit Cy5 IgG 649-666 nm (Abcam, 1µg/mL) -->

3. anti-rabbit Cy3 IgG 554-568nm (Abcam, 1 µg/mL) --> anti-mouse Alexa Fluor 633 IgG 631-650 nm (Life Technologies, 1 µg/mL)
Samples were analysed in a Attune Acoustic Focusing Cytometer (Applied Biosystems) with using (?) BL1, BL2 and RL1 channels. Data were analysed with FlowJo v10 and GraphPrims v7.

2.5.3 Liquid Chromatography-MS-MS

Proteins were in-gel digested with trypsin after reduction (dithiothreitol) and alkylation (iodoacetamide). Tryptic peptides were then extracted from the gel slices, lyophilized, dissolved in solvent A (water/0.1 formic acid (FA) v/v), and subsequently analysed by online C18 nano-HPLC MS/MS with a system consisting of an Easy nLC 1200 gradient HPLC system (Thermo, Bremen, Germany) and a LUMOS mass spectrometer (Thermo). Fractions were injected onto a homemade pre-column (100 µm × 15 mm; Reprosil-Pur C18-AQ 3 µm, Dr. Maisch, Ammerbuch, Germany) and eluted via a homemade analytical nano-HPLC column (20 cm × 75 µm; Reprosil-Pur C18-AQ 3 µm) using a gradient from 10 to 40% solvent B (20/80/0.1 water/acetonitrile/FA v/v/v) in 20 min. The nano-HPLC column was drawn to a tip of ∼5 µm and acted as the electrospray needle of the MS source. The LUMOS mass spectrometer was operated in data-dependent MS/MS mode (cycle time 3 seconds) with a normalised collision energy of 32 and recording of the MS2 spectrum in the Orbitrap. In the master scan (MS1) the resolution was 120 000, and the scan range was from *m/z* 400–1500 at an AGC target of 400 000 with maximum fill time of 50 ms. Dynamic exclusion after n = 1 with exclusion duration of 10 s was applied. Charge states 2–5 were included for MS2. For this, precursors were isolated with the quadrupole with an isolation width of 1.2 Da. The MS2 scan resolution was 30 000 with an AGC target of 50 000 with maximum fill time of 60 ms.
MS/MS data were searched against a human protein database (Uniprot, 67915 entries) using the Mascot search algorithm (Matrix Science, London, UK; version 2.2.04). Trypsin was selected as enzyme (up to two missed cleavages were allowed) and the MS and MS/MS tolerance were 10 p.p.m. and 0.02 Da, respectively. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine was specified as a variable modification. Scaffold (version Scaffold_4.8.4, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications applying an FDR of 1% at the peptide level.

2.6 Antibody production

2.6.1 Antigen preparation and isolation

Two different antigens preparations for immunization were made, membrane protein and immunoprecipitated proteins

Membrane proteins were prepared by protein extraction of LS174T membrane proteins using the Mem-PER™ Plus Membrane Protein Extraction Kit (Thermo scientific). Second sample was prepared by isolation and purification of CD44 extracted from LS174T membrane protein lysate. In this case, 500 µg of LS174T membrane protein were purified with 80µg of protein G beads (Invitrogen). Precleared samples were incubated with 15µg of anti-CD44 for 2h and then incubated with Protein G-agarose beads (pre-blocked with 1% BSA) overnight (paragraph 2.4.2) and precipitated with 0.1M Glycine precipitation in a volume of 200 µL. Sample was run in native acrylamide gel 8% (using paragraph methods 2.4.3, avoiding the use of SDS and 2-Mercaptoethanol), in a unique well of 500 µL volume. Controls for western blot revealing were also prepared. Samples run for 100V 1.30h. At end of the run the gel was separated in two parts. One containing the sample designed for immunization (Part A) and the other destined for western blot developing (Part B). Bands from part A were cut following protein ladder (New England, BioLabs) standards and stored at 4˚C, while part B was transferred in PVDF membrane at constant 400 mA for 1h. Membranes were blocked with 7,5% non-fat milk powder dissolved in TBST 0.1% for 1 hour at RT. Membranes were incubated with anti-CD44-h-pan-2C5(1:1500). Primary antibodies were 3X washed for 10 min in TBST 0.1%; and subsequentially membranes were incubated with appropriated Horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in TBST 0.1% for 1 hour at RT (anti-mouse IgG 1:5000, Southern biotech). Final 3Xx3 wash with TBST 0.1% was followed by staining with Lumi-light Western Blotting Substrate (Roche) for detection. Revelation was performed exposing the membranes to Amersham Hyperfilm ECL (GE Healthcare Life Sciences) and using Corestream® Kodac® autoradiography GBX developer and Corestream® Kodac® autoradiography GBX fixer (Sigma).

The band revealing CD44 in part B was selected among the stored bands of part A, diluted in PBS and send for immunization. Maximum yield possible achieved is 0.5 µg/µL.

2.6.2 Immunization

Female Balb/c mice aged 6 weeks (Harlan UK) (Ref: 0421/000/000/2017) were used and fed *ad libitum*. Good animal handling practices as well as good laboratory practices used at IHMT/UNL animal room are according EU Directives (2000/54/CE Exposure to biological agents, 88/320/CE Good laboratory practices, 90/679/CE, 405/98/PT and 1036/98/PT). Efforts were made to minimize suffering, and at the completion of the study, the mice were euthanized by CO_2 followed by cervical dislocation.

MAb production was performed according to the hybridoma technology described by Köhler and colleagues. First group of six-week-old female Balb/c mice (Harlan UK) were immunized intraperitoneally with 100 µg of LS174T membrane extraction emulsified 1:1 (v/v) with complete Freund's adjuvant (Sigma, F5881) followed by 2 additional injections of LS174T membrane proteins emulsified with incomplete Freund's adjuvant (Sigma, F5506) with intervals of 21 days. Second group of six-week-old female Balb/c mice (Harlan UK) were immunized intraperitoneally with 100 µL (supposed 40µg) isolated CD44 from LS174T emulsified 1:1 (v/v) with complete Freund's adjuvant (Sigma, F5881) followed by 2 additional injections of isolated CD44 isolated from LS174T emulsified with incomplete Freund's adjuvant (Sigma, F5506) with intervals of 21 days. The acrylamide was kept in the sample (no extra purification needed) and was used has extra adjuvant. Blood samples were collected from the mice cheek throughout the immunization procedure and collected serum was screened by indirect ELISA, as described below.

2.6.3 Hybridoma technology

When the serum presented the desired and specific immune response, a final boost injection was performed three days before the animals were euthanized and subsequently the spleens were harvested. Splenocytes from the immunized mouse were mixed with Sp2/0-Ag14 myeloma cells (ATCC® CRL-1581™) at a ratio of 3:1 and fused in the presence of polyethylene glycol/dimethylsulphoxide (PEG/DMSO, Hybri-max Sigma-Aldrich, P7306) using a standard protocol. Cells were plated into 96-well flat bottom micro plates (Orange Scientific) and maintained in selection medium, DMEM medium (Gibco) supplemented with HAT medium (1×10^{-4} M hypoxanthine, 4×10^{-7} M aminopterin, 1.6×10^{-5} M thymidine (Sigma), 10% (v/v) FBS (Sigma), 2 mM L-glutamine, 0.2 mg/ml gentamycin (Sigma), 1 mM sodium

pyruvate (Gibco), 1% (v/v) non-essential amino acids (Gibco) and incubated at 37ºC, 95% humidity containing 5% CO_2 for 7-12 days. Hybridoma cells were expanded in HAT medium, screened by indirect ELISA and cloned by the limiting dilution method to obtain stable single clone cell lines. Selected hybridomas were further adapted to grow in medium without HAT supplementation.

Two hybridoma series were created named mice^{LS174T} (from immunization with membrane protein total lysate) and mice^{CD44} (from CD44 extracted from LS174T membrane proteins)

The hybridoma cells were cultured and cloned to produce antibody against SLewis antigens and/or CD44. The hybridoma cells were mainly adherent and were maintained on T25 and/or T75 culture flasks (SARSTEDT) on an incubator at 37ºC, with a humidified atmosphere and 5% CO_2, and cultivated in complete DMEM medium (Gibco). The hybridoma cells were detached from the culture flasks with a scraper whenever their confluence was approximately of 80%. After centrifugation at 1200 RPM for 5 minutes, the supernatant was collected and stored for further analysis. The corresponding pellet was resuspended in complete medium for subculture or resuspended in fetal bovine serum (FBS) (Gibco) with 10% (v/v) of DMSO (Sigma) for storage in liquid nitrogen

2.6.4 Cloning

To select the stable hybridoma clone producing antibodies, 100 hybridoma cells diluted in 60mL of DMEM with the same composition described in 2.1.2 were distributed by five 96-well plates with 200µL each well (calculating 0.5 cell/well). When the cells were confluent the supernatant was collected, and the ELISA assay was performed as described in 2.6.6. The exceed clones were resuspended in complete medium for subculture or resuspended in fetal bovine serum (FBS) (Gibco) with 10% (v/v) of DMSO (Sigma) for storage in liquid nitrogen.

2.6.5 Dot blot

Dot blot using a dot microfiltration apparatus. A volume of 30µl of 5µg/µL of HT29 FUT6, LS174T and SW620 wt cell lysate was dotted on a nitrocellulose membrane (Bio-rad) and vacuum was applied. Next, the membrane was blocked using 3% milk-PBST solution for 1h at RT. After blocking the membrane was washed three times with PBST and further incubated with hybridomas supernatant total supernatant for 1h at RT. After 3X washes with PBST the

membrane was incubated for with anti-mouse IgG (Fc specific) HRP (Sigma-Aldrich) for an additional 1h at RT. 3X washes with PBST were followed by staining with Lumi-light Western Blotting Substrate (Roche) for detection. Revelation was performed exposing the membranes to Amersham Hyperfilm ECL (GE Healthcare Life Sciences) and using Corestream® Kodac® autoradiography GBX developer and Corestream® Kodac® autoradiography GBX fixer (Sigma-Aldrich).

2.6.6 Enzyme-linked immunosorbent assay (ELISA)

Mouse serum titrations and screening of hybridoma supernatants of mice^{LS174T} (immunized with total lysate) were determined by indirect ELISA using as antigens; commercial against Bovine Submaxillary Mucin (BSM), Porcin Submaxillary Mucin (PSM), Sialyl Lewis^{a}, or Sialyl Lewis^{x} from Sigma, or recombinant 36 kDa CD44 (Abcam), or 5 µg/mL membrane lysates of LS174T, HT29FUT6, SW948wt [partially positive for CD44] and, or SW620 wt cell lines.

Mouse serum titrations and screening of hybridoma supernatants of mice^{CD44} (immunized with recombinant human CD44) were determined by indirect ELISA using as antigens: CD44 [Abcam, Recombinant human 36kDa], CEA [Abcam, Recombinant human 76kDa], Sialyl Lewis^{a} [Sigma Aldrich],-Sialyl Lewis^{x}-[Sigma Aldrich] and 5 µg/mL membrane lysates of LS174T, HT29FUT6, SW948wt and SW620 wt

96-well polystyrene plates (Corning) were coated overnight at 4˚C with 50 µl of 5 µg/mL of cell lysates [LS174T, HT29 FUT6, SW948 wt, SW620 wt], dissolved in Carbonate-Bicarbonate Buffer solution pH=9.2; no blocking was followed the coating.

3µg/mL of CEA, sialyl Lewis$^{x/a}$, BSM, PSM and 0.5 µg/mL of CD44 were coated overnight at 4˚C with 50 µl in PBST 0.1%and incubated overnight at 4°C. Blocking was followed with 5% skim milk powder (VWR) in PBST 0.1% for 60 min. Next washing steps, diluted mouse sera or hybridoma supernatants were added to the wells and incubated for 90 min. Further washing steps were performed followed by incubation with HRP-conjugated goat anti-mouse Ig (1:1000) for 60 min. After three additional washing steps, 50 µl of 3,3',5,5'-tetramethylbenzidine (TMB) (Thermo Scientific, 002023) substrate were added to each well, plates were incubated in the dark and the reaction was stopped by adding 50 µl of 1M HCl. Optical Density (OD) was measured at 450 nm on a microplate reader (SpectraMax 190 Microplate Reader, Molecular Devices). The mouse producing the highest titre of antibodies

of interest was selected for splenocyte fusion. To screen the antibody production of hybridoma cells, the same procedure was implemented on the hybridoma cell culture supernatants. Unless otherwise mentioned, the incubation steps were performed at room temperature.

2.6.6.1 Live cell ELISA

Hybridoma supernatants were screened with freshly cultured LS174T and SW948 wt cell lines. Cells were grown in 96-well plates and incubated with mice serum or hybridoma culture supernatants at 25°C for one hour. Plates were washed 3 times in washing buffer PBST 0.05% and the bound antibodies were detected by incubation with HRP-conjugated anti—mouse IgG at 25°C for another hour. After careful washing, 3,3′,5,5′-tetramethylbenzydine (Sigma, Deisenhofen, Germany) was added as substrate. The colour reaction was measured at 450 nm with a Bio-Rad ELISA reader (Richmond, CA).

2.6.7 Flowcytometry of hybridomas

Hybridomas were tested using different cells supernatants against several colorectal cancer cell lines. Methods used followed the indications described in 2.4.1. No dilution of primary antibody was required.

2.7 Statistical analysis

The correlation between data was analysed using Spearman Correlation and unpaired student's t-test was used for the comparison of features.
Statistical analysis was performed using the Student's T-test for paired samples. Differences were considered statistically significant when $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.005$ (***). Statistical analysis was performed using GraphPad Prism 6 (GraphPad Software, Inc).

RESULTS

3.1 PART I: biomarker discovery

The first part of the PhD project aimed to the identification of a potential and suitable biomarker for colorectal cancer diagnosis and/or treatment. First, we evaluated the cell surface expression in six colorectal cancer and normal colon cell lines, which were selected based on differences in stages and grades of the primary tumour (Table 3.1). Thoroughly, stages define tumour size and spreading and grades, the level of differentiation from primary cells, as already mentioned in chapter 1.3.3.

Table 3.1 List of colon cell line selected for in vitro experiments. They differ for grades-level of differentiation from primary cell- and stages -level of invasiveness of tumour mass-. Normal colon was used as control cell line

Cell lines	Cancer Type	GRADES	STAGE
CCD 841 CoN	normal colon	X	X
LS174T	Colon Adenocarcinoma	**Grade I**	*Stage 2*
HT29	Colon Adenocarcinoma	**Grade I**	*Stage 3*
SW948	Colon Adenocarcinoma	**Grade III**	*Stage 2*
SW48	Colon Adenocarcinoma	**Grade III to IV**	*Stage 3*
SW620	Metastasis Colon Adenocarcinoma	**Grade IV**	*Stage 3*

For aim of the project, we have also selected a series of existing antibodies against glycan antigens and glycoproteins for cell lines expression analysis (Table 3.2). The glycans antigens were chosen based on literature data regarding aberrant glycans in colorectal cancer[199,250,258,292]. The final selection included the major glycans overexpressed throughout all the stages of the cancer development, corresponding to Lewis$^{a/x/y}$ (Le $^{a/x/y}$), sialyl Lewis$^{a/x}$ (sLe$^{a/x}$), Thomsen-Friedenreich (T), Thomsen-nouvelle (Tn) and sialyl-Thomsen-nouvelle (sTn)[153]. Two major

glycoproteins have been selected for testing: carcinoembryonic antigen (CEA), used as prognostic factor in CRC, CD44, found overexpressed in CRC studies[127].

Table 3.2 List of glycans and glycoprotein antigens chosen for cells and tissue screening and their respective names used in experiments

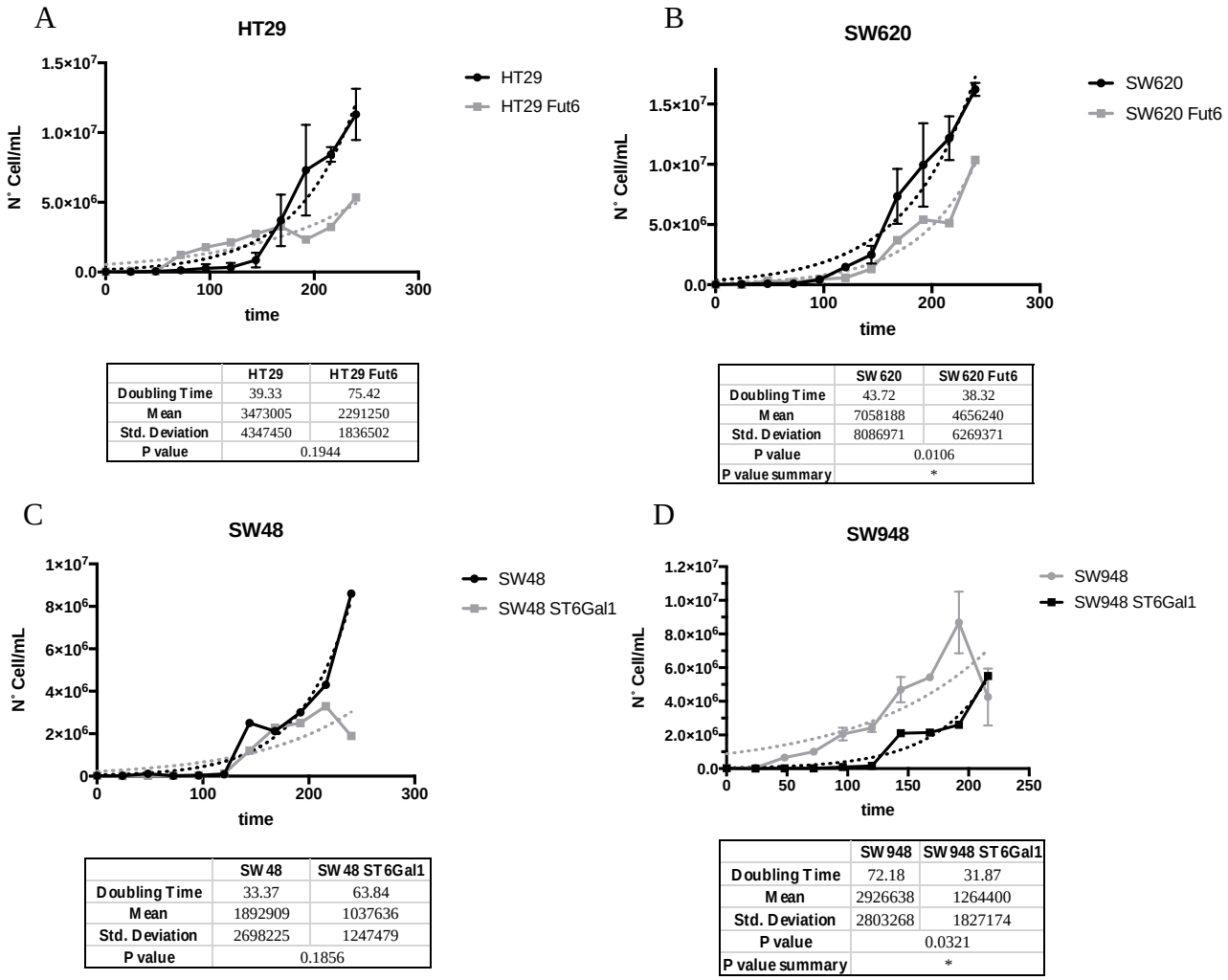
.

ANTIGEN TYPE	ANTIGENS	ANTIBODY NAMES
Glycoproteins	*CD44*	anti-CD44-hpan
	CEA	anti-CD66e
Glycans	*Sialyl Lewis$^{x/a}$*	HECA-425 or E-selectin-Ig
	Sialyl Lewis^{a}	anti-CA19.9
	Sialyl Lewis^{x}	anti-CD15s
	Lewis^{x}	anti-CD15
	Lewis^{y}	anti-Le^{y} (A70/C-C8)
	Sialyl Tn	B.72.3 or LA25 or 3F1
	Tn	5F4
	T	HH8 or PNA

The selected material, cell lines and antibodies, was used in different immunoassays, in order to characterize CRC cells expression and new potential biomarkers. The immunoassay used includes flow cytometry, western blot, immunofluorescence and immunohistochemistry.

3.1.1 Cell growth time glycan and cancer stage dependent

In order to pursue study on glycosylation, four of the CRC cells lines available were transfected or transduced with genes that code two glycosyltransferases, respectively FUT6 and ST6Gal1. FUT6 (Alpha-(1,3)-fucosyltransferase) is a fucosyltransferases responsible for fucosylation of sLewis^{x} antigen and was stably transfected in HT29 and SW620 cell lines (Table 1.2). ST6Gal1 (Beta-galactoside alpha-2,6-sialyltransferase 1) is a sialyltransferase responsible for sialylation on terminal gal in α-2,6 of N-glycan and it was stably transduced in SW948 and SW48 cell lines (Figure 1.8). Rate of cellular growth, depending on glycosylation levels, was calculated by measuring the doubling time of the cell lines (Figure 3.13).



	HT29	HT29 Fut6
Doubling Time	39.33	75.42
Mean	3473005	2291250
Std. Deviation	4347450	1836502
P value	0.1944	

	SW620	SW620 Fut6
Doubling Time	43.72	38.32
Mean	7058188	4656240
Std. Deviation	8086971	6269371
P value	0.0106	
P value summary	*	

	SW48	SW48 ST6Gal1
Doubling Time	33.37	63.84
Mean	1892909	1037636
Std. Deviation	2698225	1247479
P value	0.1856	

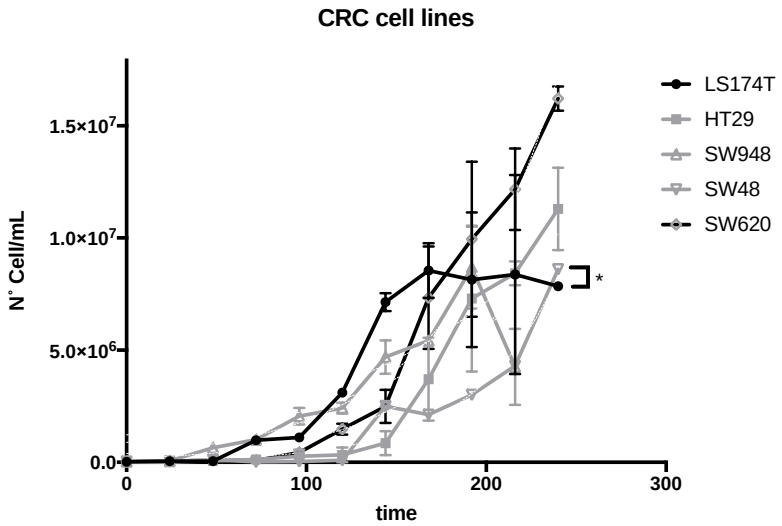
	SW948	SW948 ST6Gal1
Doubling Time	72.18	31.87
Mean	2926638	1264400
Std. Deviation	2803268	1827174
P value	0.0321	
P value summary	*	

Figure 3.13 Colorectal cell lines calculation of doubling time, glycan dependent. Cell lines were equally distributed in culture and analysed at the same time of the day for 10 days. Dots represent day of harvest while continuous dotted lines show trend lines of growth. In A) HT29 and HT29 transfected with FUT6 gene shows an increase of almost 2x of doubling time, contrasted by results in B) were differences between SW620 and SW620 FUT6 are almost null. In C) increase of doubling time from SW48 to SW48 ST6Gal1 is shown, while in D) decrease of cell growth between SW948 and SW948 St6Gal1 has been revealed. The growth curves are representative of n=3 experiment; the unpaired T test found significance in delay of doubling time between SW620 WT/FUT6, and SW948 WT/ST6GAL1.

Scores obtained, revealed contrasting results. HT29 FUT6 shows differences in doubling times of almost 2fold compared to their wild type (wt) (Figure 3.13 A), yet almost no difference in doubling time was shown in SW620 cell lines analysis (Figure 3.13B). Results suggests no influence in cell growth rating from increase of sLe^x antigens.

Increase in time for cell growth rating is also depicted from SW48 to SW48 ST6Gal1(Figure 3.13C); nevertheless, doubling time for SW948 ST6Gal1 is shorter (<h) than SW948 wild type (Figure 3.13D), suggesting that even if increase of sialylation might influence cell growth, further analyses are required.

Differences of cells stage-dependent growth were also pursued. The five CRC cell lines selected (LS174T, HT29, SW948, SW48 and SW620; Table 3.1) were originated from adenocarcinomas of Caucasian patients, collected from differ stages and cancer grades. We investigated doubling cell times in order to analyse differences in cells behaviour stage/grade dependent (Figure 3.14). Comparisons shows a consistency in slower doubling time of LS174T and SW948, of stage II CRC, compared to HT29, SW48 and SW620, of stage III.



	LS174T	SW48	HT29	SW948	SW620
Doubling Time	78.52	33.37	39.33	72.18	43.72
P value	0.0145				
P value summary	*				

Figure 3.14 Comparing of CRC growth rating stage dependent. Analysis show that LS174T and SW948, of stage II, have a slower doubling time compared to HT29, SW48 and SW620 of stage III. The growth curves are representative of n=3 experiment; the unpaired T test found significance in delay of doubling time between LS174T (grade 1) and SW48 (grade 3/4)

The total analysis aimed to study differences in cell growth rate, in order to understand whether increase of glycosylation might influence tumour growth. Moreover, in order to guaranty consistency between cell culture and patient tumour cells, correlation between stage and growth was made. Stage associated cells analysis confirmed an increase of growth in higher staging, while contrasting results on glycosylation analysis on cells behaviour required further investigations with different immunoassays.

3.1.2 Glycan and protein expression -live cells-

In order to find the most expressed and thus, most promising biomarker, a panel of 9 antibodies was selected for flow cytometry live cells analysis. The selected commercial antibodies comprehend 6 glycans cancer specific antigen ($sLe^{x/a}$, Le^{y}, Le^{x}, Tn, sTn), already proven to be

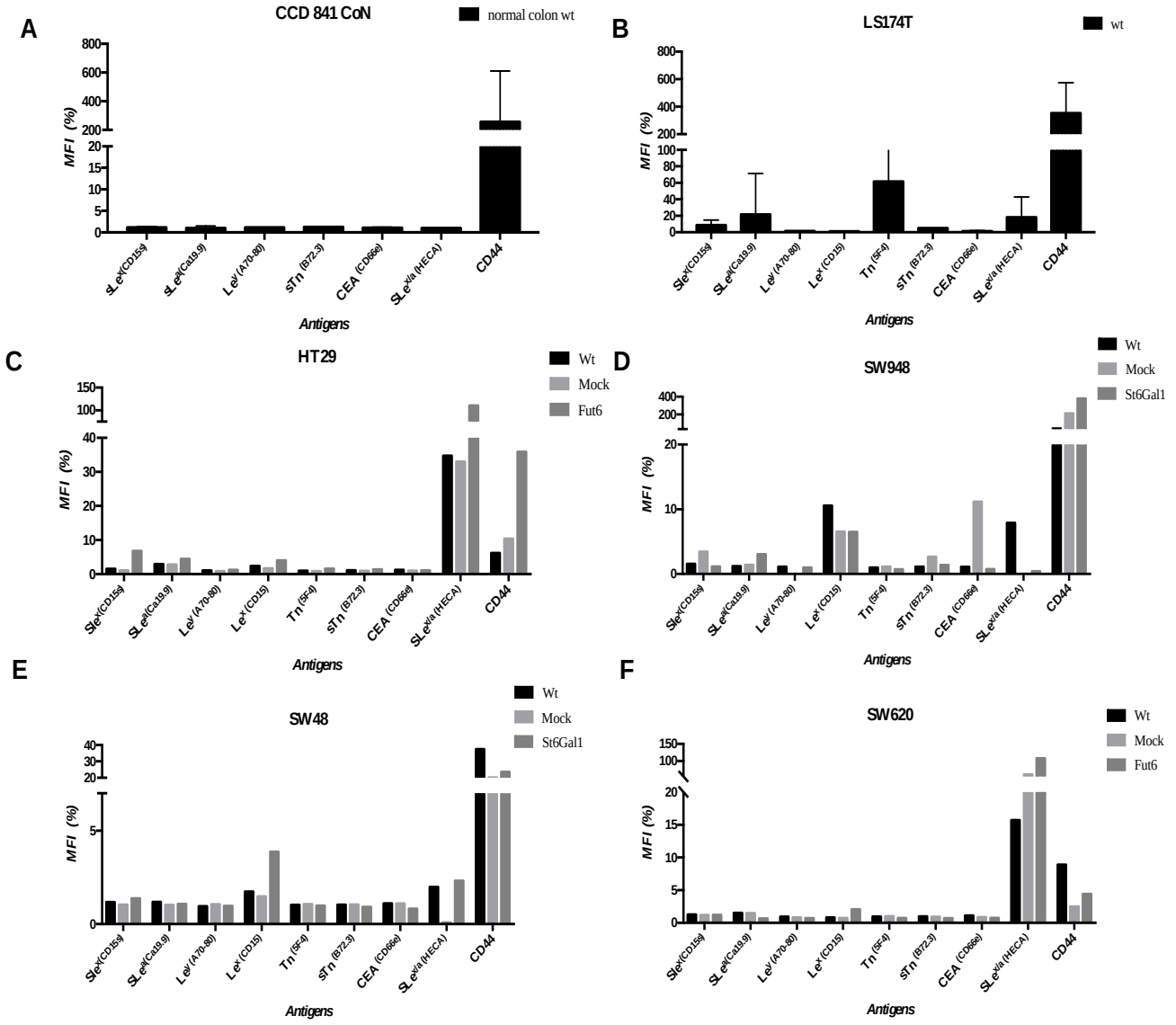


Figure 3.15 Representation of median fluorescence intensity (MFI) of different antigens extracellular expression. The antigen tested are $sLe^{x/a}$, Le^{y}, Le^{x}, Tn, sTn, CEA and CD44. Different colon cell lines (A-F) have been tested in order to look for a potential biomarker and to find correlation between cancer grades and antigens expression. In A) normal colon CCD841Con presents high expression of CD44 but none of aberrant glycosylation. Grade 1 B) LS174T presents high level of CD44, of Tn and $sLe^{x/a}$ antigens. In grade 1 C) HT29and its correlate FUT6 transfected cell lines, is maintained high expression in $sLe^{x/a}$ antigens but a moderate expression of CD44 antigen. In grade 3 D) SW948 and its corresponding ST6Gal1 transduced cell lines, shows high expression in CD44, in Le^{x} and a relative low expression in CEA and $sLe^{x/a}$ antigens. In E) SW48 and the transduced ST6Gal1 maintain a low signal in Le^{x} and relatively low CD44. In F) grade 4 SW620 and its relative FUT6 cell line displayed low CD44 expression contrasted by high $sLe^{x/a}$ expression. The MFI obtained in flow cytometry experiments are representative of n=2 experiment; no statistical analysis was performed on the results.

overexpressed throughout all the stages of colorectal cancer[153] and 2 glycoproteins. The glycoproteins selected were the carcinoembryonic antigen (CEA) and the protein CD44, with the intention to compare what is already in use as a prognostic biomarker with a new potential biomarker, highly discussed in the literature[255,256].

All the colon cell lines were screened including normal colon CCD841Con, grade 1 LS174T, grade 1 HT29 wt (including HT29 mock and HT29 FUT6), grade 2 SW948 wt (including SW948 mock and SW948 ST6Gal1), grade 3/4 SW48 wt (including SW48 mock and SW48 ST6Gal1) and last grade 4 SW620 wt (including SW620 mock and SW620 FUT6) (Figure 3.15).

As result of the extracellular screening, the negative control CCD841Con showed no aberrant expression of glycans, no expression of CEA antigen but high expression in CD44 (Figure 3.15A). Likewise, LS174T highly expresses CD44 glycoprotein, but increase in $sLe^{x/a}$ and Tn antigens is also depicted (Fig. 3.3B). Similar conditions are found in the HT29 cell lines, with an overall considerate expression in $sLe^{x/a}$ and CD44 antigens (Figure 3.15C). In SW948 and SW48 a common increase of Le^{x} with reduction of $sLe^{x/a}$ antigens is presented (Figure 3.15D). CD44 remains highly expressed in SW948 while in SW48 is detected reduction in signal, until almost completed lost in expression in SW620 cell lines, which shows high expression of $sLe^{x/a}$ antigens (Figure 3.15F).

Further analysis comparing the different stages related cell lines expression confirmed sLewis antigens as the most expressed, especially in the LS174T (Figure 3.16A). The sLewis antigens were studied either with the antibody HECA-452 specific for Cutaneous lymphocyte antigen (CLA, a carbohydrate epitope of sialic acid ligand for E-selectin, P-selectin, and L-selectin) or with sLe^{x}-antiCD15s and sLe^{a}-anti-CA19.9 carbohydrate specific antibodies. The antigens Le^{x} and Le^{y} tested with specific monoclonal antibodies, respectively anti-CD15 and anti-A70-80, did not reveal any unique nor intense MFI expression (Figure 3.16D-E). Therefore, sLewis antigens were selected as the main target for further downstream analysis.

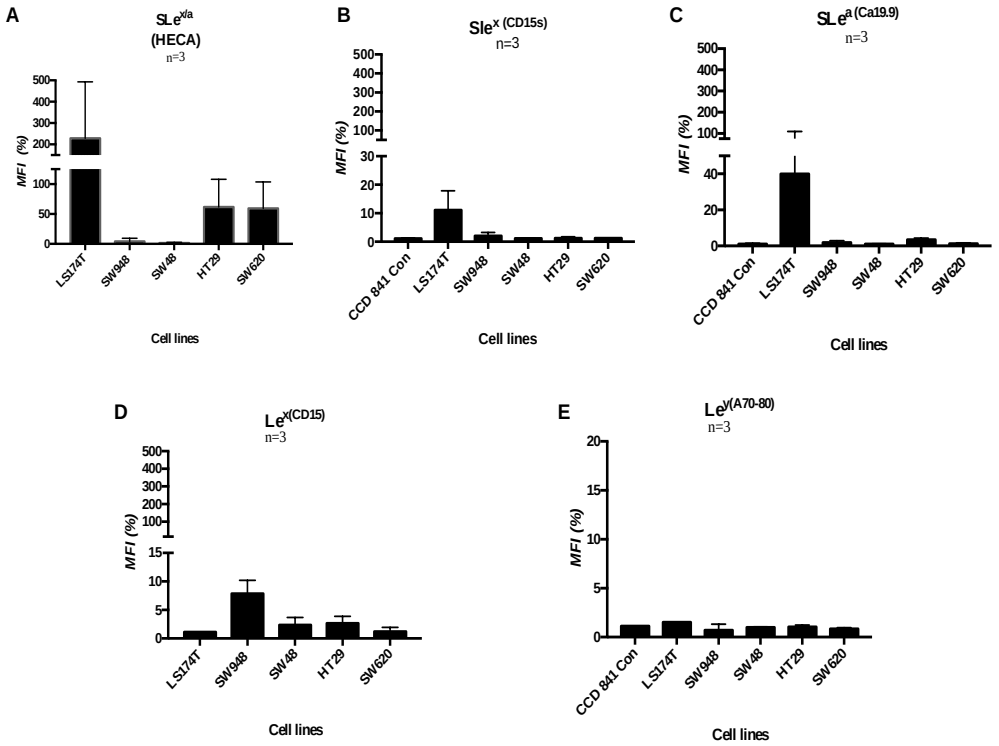


Figure 3.16 Correlation of (s)Lewis antigen expression in different stage-correlated cell lines. A-C) High expression of $sLe^{x/a}$ is presented in most of the cell line studied LS174T (stage 2), HT29 and SW620 (stage 3). In D) lower expression of Le^x is presented in SW948 (stage 2), while in E) no cell line responded to Le^y signals. . The MFI obtained in flow cytometry experiments are representative of n=3 experiment; no significance was found with impaired T test.

After evaluation and selection of aberrant glycosylation, also the two-glycoprotein selected (CEA, CD44) were compared in different cancer stages (Figure 3.17).

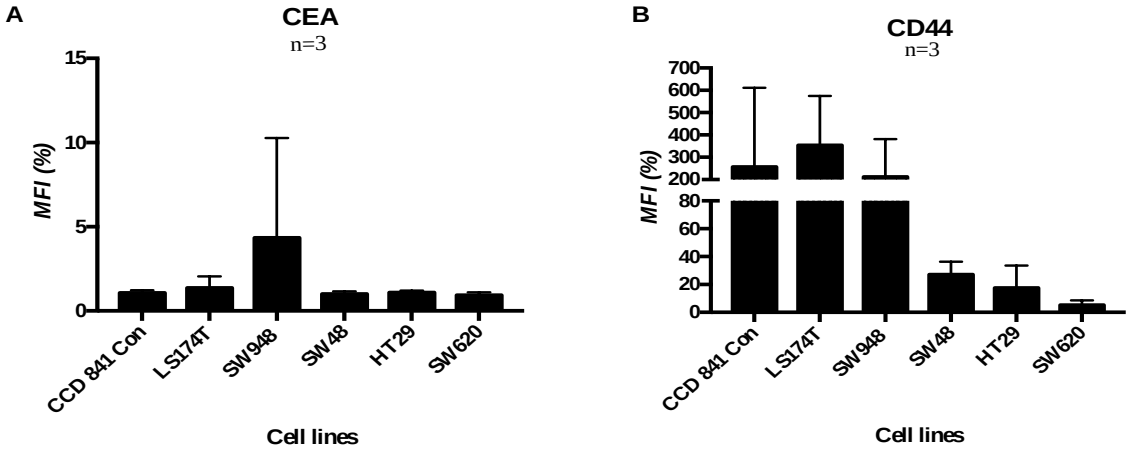


Figure 3.17 Differences in glycoprotein expression in different cancer stages. In A) SW948 shows a moderate-low expression of the CEA protein, while in B) the expression of CD44 is very high in CCD 841 Con and stage 2, LS174T and SW948, and is reduced in stage 3, SW48-HT29- SW620, cell lines. The MFI obtained in flow cytometry experiments are representative of n=3 experiment; no significance was found with impaired T test.

CEA analysis revealed expression of the glycoprotein in SW948 cell line but no presence in the other 5 cell lines, including normal colon, negative control, is shown (Figure 3.17A). On the contrary, CD44 appeared to be highly expressed in normal colon CCD 841 Con, with a progressive lost throughout the cancer staging (Figure 3.17 B). CD44 is an adhesion protein of different isoforms and highly glycosylated, and the results suggested possible implication of the protein in cancer progression.

Moreover, an interesting correlation between increase of FUT6, thus in sLe^x antigens, and CD44 expression in HT29 cell lines has been revealed, suggesting a possible association between aberrant glycosylation in cancer progression and increase of protein expression (Figure 3.18).

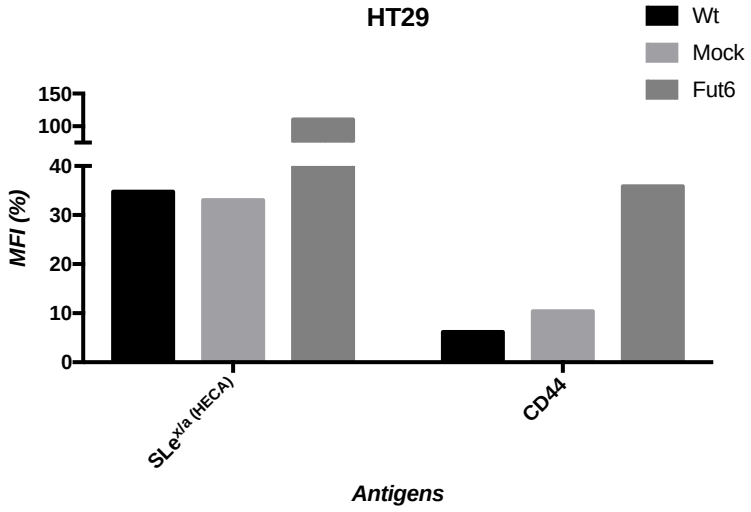


Figure 3.18 study of sLewis$^{x/a}$ antigens expression correlated with the CD44 glycoprotein. In the wt and mock HT29, sLex/a antigens are moderately expressed while the protein CD44 is almost absent. In the HT29 FUT6 transfected cell line, an expected increase of SLe^x antigens is correlated to a substantial increase of CD44 expression. The MFI obtained in flow cytometry experiment are representative of unique experiment; no statistical experiments have been performed on the results obtained

3.1.3 Glycan and protein expression -protein extraction-

Following cells membrane expression analysis, cell total proteins were extracted and analysed by western blot. The analysis had the purpose to confirm observation made by flow cytometry, to characterize aberrant glycosylation and glycoprotein isoforms and to validate antibodies affinity. Four antibodies were selected for the screening:

1. HECA-452: to confirm $sLe^{x/a}$ as the most expressed in LS174T wt, in HT29 (wt, mock and FUT6) and in SW620 (wt, mock and FUT6) (Figure 3.19);
2. anti-CD44-hpan: to confirm high CD44 expression in LS174T, in SW948 (wt, mock, ST6Gal1); change in expression between HT29 (wt, mock) and HT29 FUT6; lower

expression in SW48 (wt, mock, ST6Gal1) and almost no expression in SW620 (wt, mock and FUT6) (Figure 3.20);

3. anti-CD66e: to compare expression of CEA protein to CD44, and verify its potential as biomarker at intramembrane and cytoplasmatic level (Figure 3.21);
4. anti-CD15: to compare at the intracellular level of CD15 antigens in the CRC cell lines level (Figure 3.22).

Western blot analysis of HECA-452 (1:1000) of total lysate extraction, revealed to be in agreement with flow cytometry results. In detail, LS174T presents two major bands, at >245kDa and 180>x>135, with intensive signal for $sLe^{x/a}$ antigens (Figure 3.19A); HT29 FUT6 cell lines confirmed a substantial increase of $sLe^{x/a}$ signal at ~135kDa and ~90 kDa, compared to the almost missing signals in wt and mock controls (Figure 3.19B). SW948 and 48 (wt, mock) present smaller signals at ~100 kDa, at 75>x>63 kDa; no increase in signals of ST6Gal1 transduced cell lines has been noticed (Figure 3.19C-D). SW620 FUT6 presents increase of expression at ~200 kDa, ~170 kDa and at ~90 kDa, and disappear of bands present at 100 kDa in wt and mock.

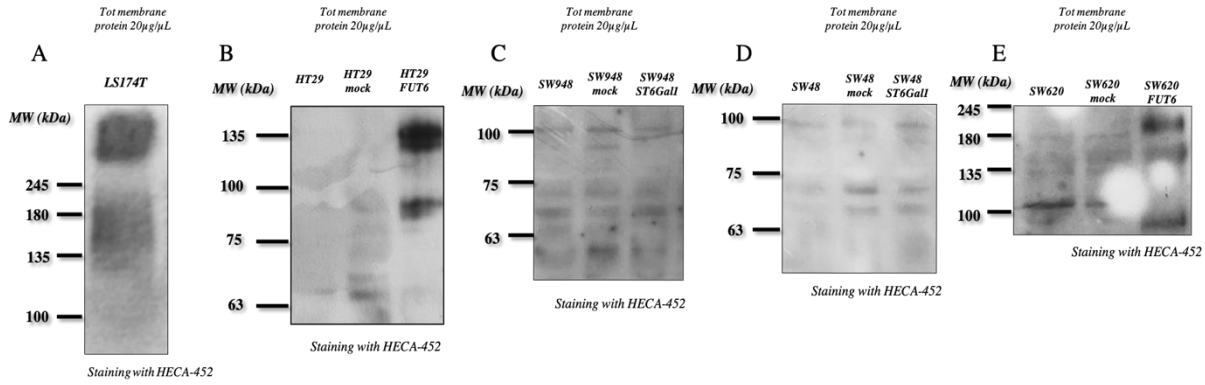


Figure 3.19 western blots total lysate extraction of CRC cell lines (20µg/µL) analysed by HECA-452 (1:1000). In A) LS174T presents two major bands, at >245kDa and 180>x>135; in B) HT29 FUT6 present increase of $sLe^{x/a}$ signal at ~135kDa and ~90 kDa and no almost signals is detected in wt and mock controls; in C) SW948 and 48 (wt, mock,St6Gal1) present smaller signals at ~100 kDa and at 75>x>55 kDa; in D) SW 48 (wt, mock, ST6Gal1) present smaller signals at ~100 kDa, at 75>x>63 kDa; in E) SW620 wt and mock present the only band at ~100 kDa, while FUT6 presents increase of expression at ~200 kDa, ~170 kDa and at ~90 kDa..Western blots shown are representative of 2 experiments.

~

CD44 is a glycoprotein presenting alternative splicing that give origin to different isoforms. Eight isoforms have been already recognized; their molecular weight depends on the level of glycosylation of each correspondent exons and it could vary from 250 kDa, down to 16kDa (see paragraph *1.7.5 CD44*)[256].

Western blot analysis revealed the most intense signals in LS174T, HT29 (wt, mock, FUT6) and SW948 (wt, mock, ST6Gal1). LS174T investigation presents two major bands at >135kDa

(probably corresponding to glycosylated form of CD44v2-10/ CD44v3-10) and at ~90kDa (probably corresponding to glycosylated form of isoform CD44s); same bands are found with less intense signal in HT29 (wt, mock, FUT6) (Figure 3.20A). Of particular remark, is the appearance of new band at ~130kDa (probably corresponding to glycosylated form of CD44v2-10/ CD44v3-10/CD44v8-10) that with the increase of sLe^x (as revealed in HECA-452 analysis-Figure 3.20B) could correspond to an increase or change in glycosylation.

SW48 (mock, ST6Gal1) revealed only two bands at ~90kDa and ~75kDa (probably corresponding to glycosylated form of CD44s) (Figure 3.20 B); same pattern has been revealed in SW948 cells with the exception of a consistent isoform at ~135kDa (probably corresponding to glycosylated form of CD44v2-10/ CD44v3-10) in wt, mock and ST6Gal1 (Figure 3.20B). Although, the analysis of SW948 presents similarity with LS174T and HT29 (wt, mock, FUT6), there is no variation in molecular weight in SW948 ST6Gal1 cell line, suggesting that no modification of sialylation occurred on CD44 from the wt.

SW620 (wt, mock) present only one band at ~50kDa (might corresponding to the glycosylated form of the isoform CD44v8-10 or CD44v10) (Figure 3.20C).

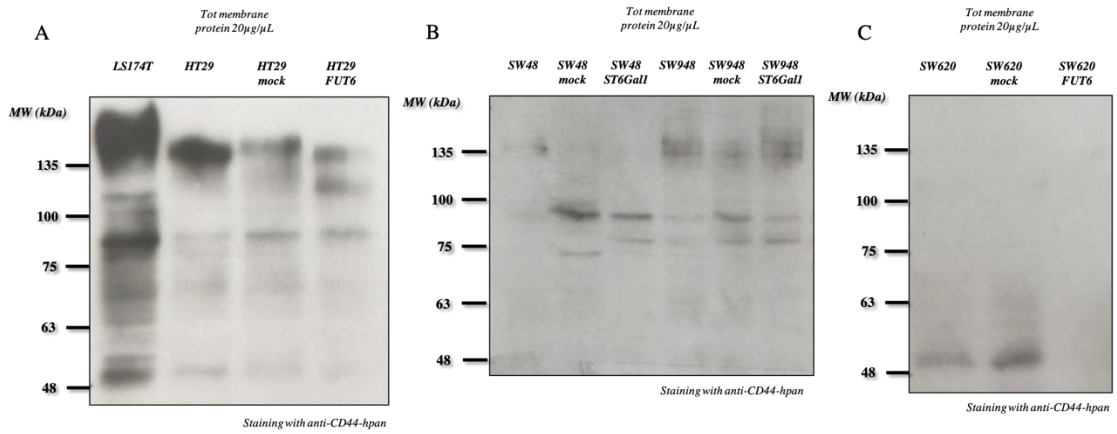


Figure 3.20 western blots total lysate extraction of CRC cell lines (20µg/µL) analysed by anti-CD44-hpan (1:1500. In A) LS174T presents two major bands at >135kDa and at ~90kDa, probably two glycosylated form. Same bands are found in HT29 wt, mock and FUT6. However, in HT29 FUT6 is presented a new band at ~130kDa, probably due to differences in glycosylation. In B) SW48 (wt, mock, ST6Gal1) reveal two bands at ~90kDa and ~75kDa, with lower signal intensity. Same pattern is revealed in SW948 (wt, mock, ST6Gal1) with the exception of a consistent isoform at ~135kDa. In D) SW620 (wt, mock) presents a band at ~50kDa probably no glycosylated. No bands are revealed in SW620 FUT6. Western blots shown are representative of 2 experiments

CEA is a glycoprotein present in different isoforms. Glycosylation can influence greatly its molecular weight, with full-length CEA of 180–200kDa passing to 76.8 kDa when no-glycosylated. New isoforms, no glycosylated, have also been detected at ~60 and ~40 kDa[293].

Western blot screening presents two mayor band in LS174T (~200kDa) and in HT29 (wt, mock) (~180kDa), probably glycosylated (Figure 3.21A-B). Cell lines SW948 and 48 (wt, mock, ST6Gal1) present an almost identical pattern with double bands between ~48-40 kDa, suggesting the presence of no-glycosylated form of CEA (Figure 3.21 C-D). SW620 (wt, mock) present a major band ~60 kDa, but no signal was revealed in SW620 FUT6.

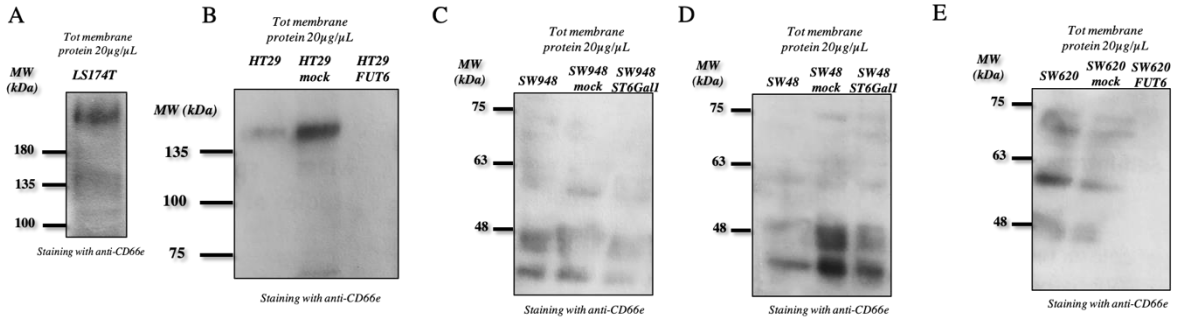


Figure 3.21 western blots total lysate extraction of CRC cell lines (20µg/µL) analysed by anti-CD66e (1:1000). In A) LS174T presents one major band at ~200kDa, probably glycosylated. In B) HT29 wt, mock one band at ~180kDa has been revealed but no signal is found in FUT6. In C) andD) SW948 and SW48 (wt, mock, ST6Gal1) reveal two bands between ~48-40 kDa, with lower signal intensity. In E) SW620 (wt, mock) presents a band at ~60kDa probably no glycosylated. No bands are revealed in SW620 FUT6. Western blots shown are representative of 2 experiments

CD15 (Le^x) was also tested, as a tumour -associated glycan and being abundantly expressed in SW948 cell line. In analysis, contrasting to flow cytometry results, LS174T shows high intensity smeared band suggesting abundant presence of CD15 antigen in the protein lysate (Figure 3.22A). Same results were encountered in SW948 (wt, mock, ST6Gal1) cell lysates with evident smeared band at 250>x>100 kDa (Fig. 3.10C). However, HT29 wt and mock presented more defined bands <48 kDa. One band at ~48 kDa is shown in HT29 FUT6 (Figure 3.22 B). SW48 cell lysates (wt, mock, ST6Gal1) at ~180kDa (Figure 3.22 D). No staining was found on SW620 cells lines (Figure 3.22 E)

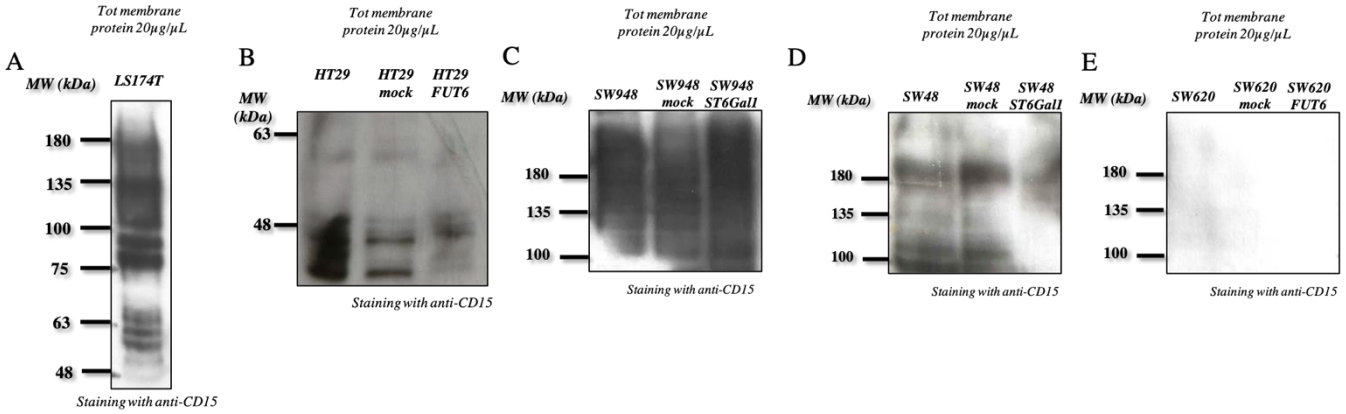


Figure 3.22 western blots total lysate extraction of CRC cell lines (20µg/µL) analysed by anti-CD15 (1:1500). In A) LS174T presents smeared band from 180kDa down to ~50kDa. In B) HT29 (wt, mock) two bands at <48 kDa and one band at ~48kDa in found in FUT6. In C) SW948 and smeared band at 250>x>100 kDa are presented, validating results obtained by flowcytometry. In D) SW48 (wt, mock, ST6Gal1) reveal two bands at ~180kDa and ~100kDa, with medium-high signal intensity. In E) SW620 (wt, mock, FUT6) no bands are revealed. Western blots shown are representative of 2 experiments

In conclusion, two main glycan antigen, $sLe^{x/a}$ and Le^{x}, and two major glycoproteins, CD44 and CEA, have been selected on previous studied of the cell membrane screening, obtained by flow cytometry. The total lysate analysis has been important for identification of specific targets and to test potential function of the commercial antibodies. CD44 and $sLe^{x/a}$ have been revealed to be the most promising target because of the high molecular weight presented in LS174T, HT29 (wt, mock, FUT6) and SW948 (wt, mock, St6Gal1), suggesting elevated influence of post-translational modification. Moreover, increase in glycosylation of sLe^{x} in HT29 FUT6 correspond not only to an increase of HECA-452 signal at ~135kDa, but also to the formation of a different molecular weight of the protein CD44, suggesting a correlation between the two antigens. CEA analysis show abundancy in high molecular weight protein isoforms in LS174T and in HT29 (wt, mock), but lower molecular weight in other cell lines. Moreover, its glycosylation level did not seem dependent to variation of presence of glycotransferases. Le^{x}, in contrast to flow cytometry results, showed intense signals not only in SW948 but also in LS174T cell lines. However, this might suggest higher presence at the intracellular level and for this reason was not selected as a good candidate as a biomarker and antibody production.

3.1.4 Glycan and protein expression -Colorectal cancer patients-

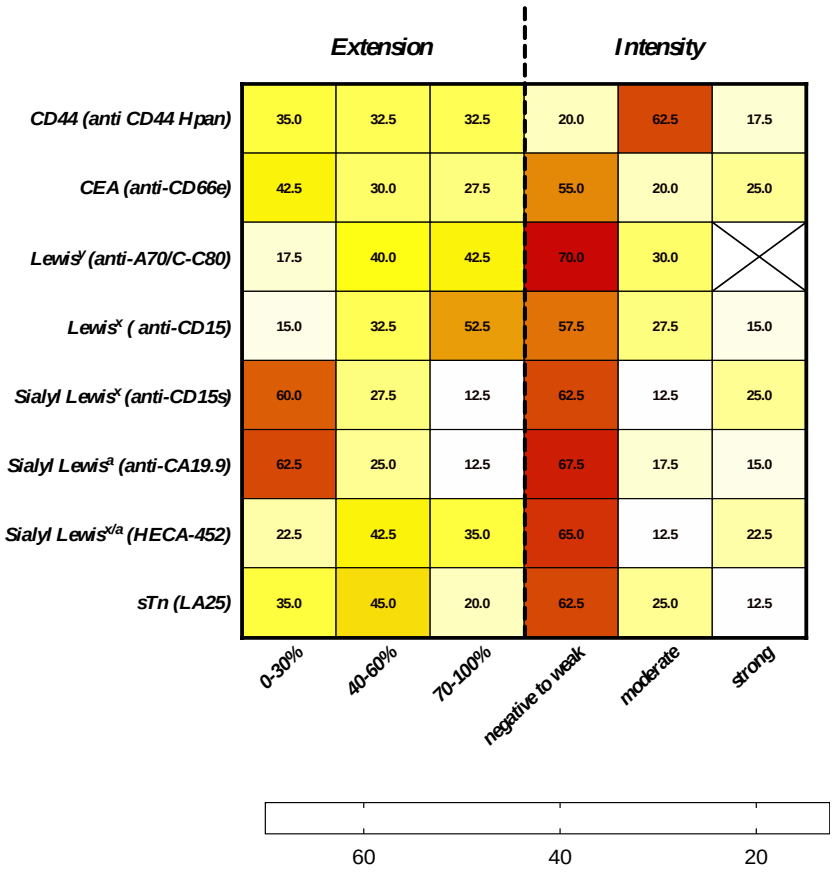
In order to understand the diagnostic and therapeutic potential of the different glycosylated antigens, CRC tissues from a cohort of 40 patients were selected from different stages of CRC and different locations, to optimize variability. The analysis performed by immunohistochemistry served, also, for validation and translation of the results obtained by cell lines screening, previously described (paragraphs 3.1.1/3.1.3/3.1.4).

The antibodies used for the screening and their relative antigens were the following:

- CEA: anti-CD66e
- CD44: anti-CD44-hpan (all isoforms)
- Le^x: anti-CD15
- Le^y: anti-A70/C-C80
- sLe^x: anti-CD15s
- sLe^a: anti-CA19.9
- $sLe^{x/a}$: HECA-425
- sTn: LA25

Tissue staining in a cohort of 40 CRC patients, revealed a positive staining for all antibodies in approximately 90 to 95% of cases. Staining for $sLe^{x/a}$ was detected in 90% to 98% of the cases, with similar lower intensity, independently of the antibodies used. Yet staining with HECA-452 antibody showed higher extension covered, as compared with the other antibodies (anti-CD15s and anti-CA19-9).
The CEA protein, which is actually used in the clinics as a prognostic marker, was detected in nearly 90% of the samples, while the signal intensity was moderate to strong in almost 50% of the samples. On the other hand, CD44 was also detected in nearly 90% of the samples, showed a general homogeneous extension, and 62.5% of the cases gave moderate intensity of the signals (Table 3.3).

. Table 3.3. Percentage representation of signal intensity and extension of the reactivity of antibodies in a cohort of 40 different CRC tissues with different tumour stages. The legend represents the percentages for each group, from 0% (white) up to 70% (red). On the left, it is presented the extension of signal, which is well distributed in the glycoprotein. On the contrary, glycan such as Le^x showed high tumour extension, while SLe^x and SLe^a antibodies exhibit a low extension in ~ 60% of the cases. However, HECA-425 displayed an increased extension suggesting different antibodies specificity for the same antigens. On the right, the intensity staining is homogeneously negative to weak in the glycans alone, but there is an increase of CD44 staining suggesting an improved the signal specificity.



In detail, in normal cells, CD44 is found in leukocytes and nerve plexuses, but is not expressed in muscle and adipose tissue and the positive control used was the skin, which shows intense membrane marking. CD44 labelling is cytoplasmic and membrane. Its signal intensity was based on the strong membrane labelling present on the lymphocytes / leukocytes of each cut (internal control). Informative data (labelling site, positive normal tissues, among others) and optimizations of this antibody were based on the expression of this marker in the "The Human Protein Atlas" database. CD44 expression is found in the cytoplasm and membrane of tumour cells and normal colonic epithelial tissue has no labelling.

CEA marking presents cases with low extension, that are located at the secretory pole of the tumour gland structures and is also present in the secretion outside the tumour cells. CEA mark

on the secretion of the Lieberkühn crypts and is not expressed on any other structure that constitutes normal colon / rectal epithelial tissue.

The expression of CD15s is mainly located in the secretory pole of the characteristic glandular structures of adenocarcinomas and in the secretion produced by the tumour. The type of labelling is variable and may be in both cytoplasm and membrane. In cases with normal epithelium representativeness, the existence of CD15s is not evident, except in some cases, marking in areas of mucus (not inside the cells that constitute Lieberkühn crypts). Even in some cases negative for CD15s, positive mucus is found for CD15s.

The expression of CA19.9 is localized in the tumour cells and staining in secretion produced by the tumour is noticeable. In the normal epithelium, in some cases can be observed the presence of marking in the goblet cells (cytoplasmic) as well as in the mucus produced by the epithelium. CA19.9 labelling extends from the cytoplasm to the membrane. Even in cases negative for CA19.9 in tumour cells, we can verify the presence of positive secretion for CA19.9. The positive control used was a previously tested bladder cancer cut.

HECA expression is found primarily in the cytoplasm of tumour cells, with few cases of membrane labelling. Additionally, it has been noticed presence of positive marking in the secretion produced by the mucus. Regarding normal tissue labelling, HECA marks on immune system cells (to be designated) and goblet cells that produce the mucus responsible for the protection of the colon. HECA expression is also found in some cells of the base of the colonic epithelium. Even in cases negative for HECA in tumour cells, we can observe positive secretion. The positive control used was the same cut as the CA19.9 immunohistochemistry.

L2A5 expression is mostly cytoplasmic, with two or three cases of enhanced labeling in the cell membrane region. Staining on the secretion produced by the tumor was visible and it was also possible to see the L2A5 label on the goblet cells (mucus producing) that make up the normal mucosa. Even in negative cases (in the tumor cell) for L2A5, it is visible presence of STn positive mucus.

CD15 expression is mostly cytoplasmic, with few cases of enhanced labeling in the cell membrane region. It is noticed in secretion produced by the tumor and mostly weak cytoplasmic labeling of CD15 in the enterocytes that make up the normal mucosa. The positive control used was the internal positive control given by leukocytes present in all cases. Intravascular or tissue leukocytes have strong membrane labeling for CD15. In 5 cases a very specific labeling pattern, revealed that cells with greater labeling intensity on membranes are located in front of tumor invasion.

Lewis^{y} expression is exclusively cytoplasmic, mostly weak and exclusively tumoral. Lewis Y expression is not immunohistochemically visualized in immune system cells, muscle, blood vessels, as well as in most secretion products (mucus). The marking pattern is not defined in all cases, however, there seems to be a definite Lewis Y marking pattern in the globality of the analysis: sites on the invasion front have higher marking intensity. In normal epithelial tissue Lewis^{y} appears weak in cytoplasmic of goblets cells and enterocytes.

In general, all antibodies showed reactivity with CRC tissue. As represented in Figure 3.23A, it was possible to observe overlaps of positive signals of CEA and CD44 with (s)Lewis antigens which likely indicated the co-localization of the glycoproteins and the glycans. Le^{x}, sLe^{x} and sLe^{a} showed higher extension of the staining (Figure 3.23B). Moreover, the presence of CD44 in mucus was observed and taken a positive staining for a potential soluble biomarker (Figure 3.23C). For this reason, a first attempt of proximity ligation assay (PLA) between anti-CD44 and anti-CD15s/CA19.9 and HECA was tried. However, no remarkable signal was observed suggesting further optimization and better tissue selection was needed, before taking major conclusions.

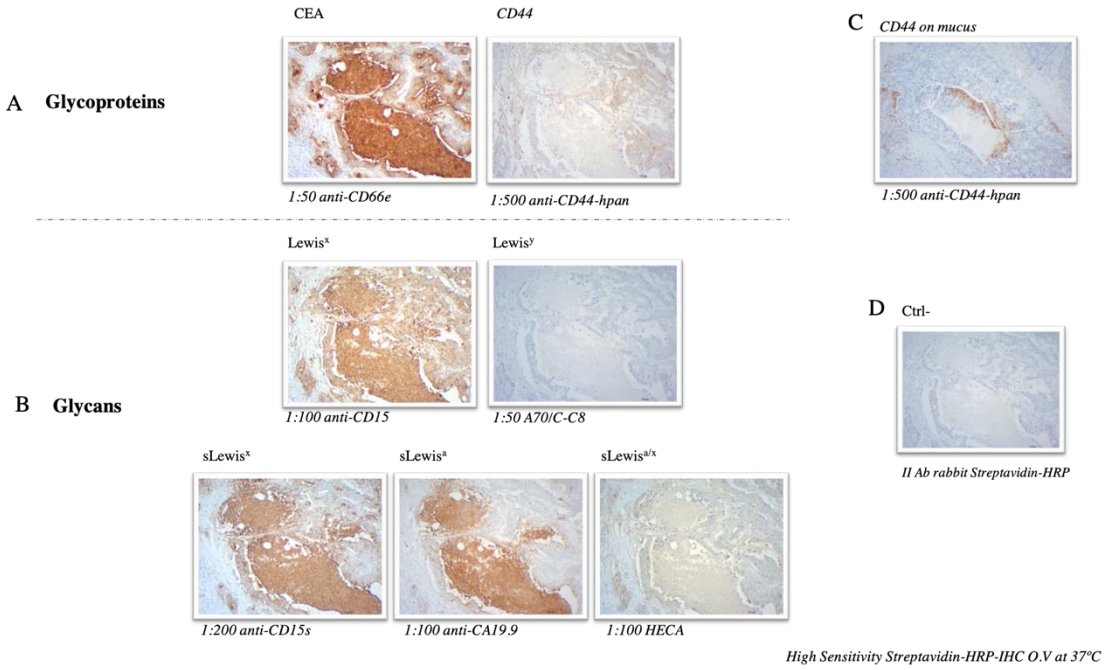


Figure 3.23 Example of immunohistochemistry staining with different antibodies tested on same tissue cut from CRC patient. In tissue section in A) CEA presents high intensity and extension andCD44 present an intense membranal staining. In tissue section in B) Le^{x}, sLe^{x} and sLe^{a} presents high extension and intensity, overlapping signals in A). In B) Le^{y} and HECA-452 (sLe$^{x/a}$) present, respectively none and lower staining

3.1.4.1 Correlations in immunohistochemistry results

Further analysis of immunohistochemistry results had the purpose to reveal linkage between antigens expression. Correlation between antigens expression scorings have been performed (Figure 3.24). Interpolation of the data revealed main significant correlation between

- CD15s and CA19.9 (**), HECA-452 (**)
- CD15 and CD15s (**), HECA-452 (**)
- CD44 and Le^y (*)
- CEA and CD15s (**), HECA-452 (**)
- CA19.9 and CD15s (**), HECA-452 (**)
- HECA and CD15s (**), CD15 (**)
- sTn and CD15s (**), CD15 (**)
- Le^y and CD44(**) and HECA-452 (**)

Correlations

		CD15s_score	CD15_score	CD44_score	CEA_score	CA19_9_score	HECA_score	STn_score	LeY_score
CD15s_score	Pearson Correlation	1	,557**	,187	,607**	,764**	,787**	,586**	,329*
	Sig. (2-tailed)		,000	,248	,000	,000	,000	,000	,041
	N	40	40	40	40	40	40	40	39
CD15_score	Pearson Correlation	,557**	1	,093	,492**	,483**	,640**	,550**	,176
	Sig. (2-tailed)	,000		,569	,001	,002	,000	,000	,284
	N	40	40	40	40	40	40	40	39
CD44_score	Pearson Correlation	,187	,093	1	,196	,021	,150	,098	,343*
	Sig. (2-tailed)	,248	,569		,226	,898	,356	,546	,033
	N	40	40	40	40	40	40	40	39
CEA_score	Pearson Correlation	,607**	,492**	,196	1	,441**	,612**	,336*	,200
	Sig. (2-tailed)	,000	,001	,226		,004	,000	,034	,222
	N	40	40	40	40	40	40	40	39
CA19_9_score	Pearson Correlation	,764**	,483**	,021	,441**	1	,496**	,446**	,086
	Sig. (2-tailed)	,000	,002	,898	,004		,001	,004	,601
	N	40	40	40	40	40	40	40	39
HECA_score	Pearson Correlation	,787**	,640**	,150	,612**	,496**	1	,464**	,367*
	Sig. (2-tailed)	,000	,000	,356	,000	,001		,003	,022
	N	40	40	40	40	40	40	40	39
STn_score	Pearson Correlation	,586**	,550**	,098	,336*	,446**	,464**	1	,178
	Sig. (2-tailed)	,000	,000	,546	,034	,004	,003		,279
	N	40	40	40	40	40	40	40	39
LeY_score	Pearson Correlation	,329*	,176	,343*	,200	,086	,367*	,178	1
	Sig. (2-tailed)	,041	,284	,033	,222	,601	,022	,279	
	N	39	39	39	39	39	39	39	39

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 3.24 Antigens scoring inter-correlations obtained by immunohistochemistry results of cohort of 40 CRC patients. Main significant correlations are found among CD15s, CA19.9 and HECA (**)

Further analyses were performed in order to reveal correlation between antigens expression and cancer staging and invasion. Cancer staging, defined by I (early-stage cancer), II (cancer growth in tissues), III (lymph nodes invasion) and IV (metastasis)[132], has been divided in mostly non-invasive profile (I-II) and invasive profile (III-IV) and correlated with antigens extensions/intensity. Analysis revealed a significative increase of CD44 protein expression (Figure 3.25A) and significative extension of Le^y expression (Figure 3.25B) in invasive stages.

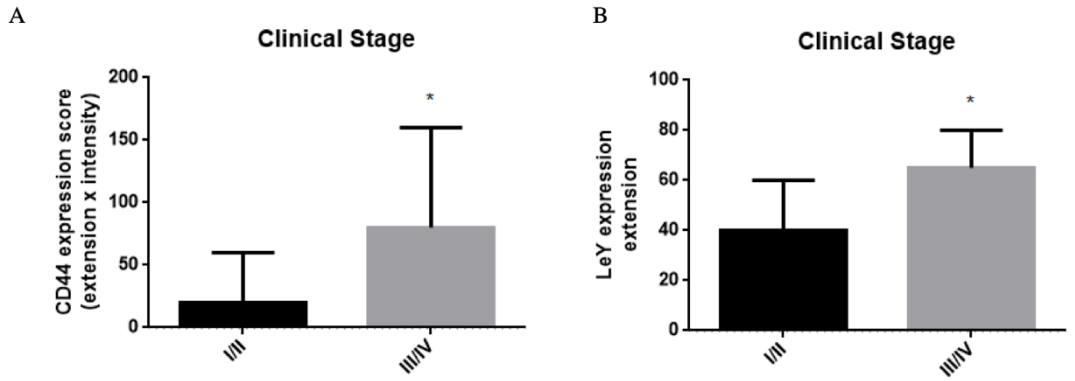


Figure 3.25 Correlation studies between cancer staging and antigen expression. In A) increase in CD44 expression (extension x intensity) is correlated with significant (*) increase of cellular invasiveness (stages III/IV). Equally, in B) increase in Le^y extension in expression is significantly (*) correlated with a more malignant phenotype. Immunochemistry antigen signals were measured by the pathologist and analysed with unpaired T-test.

Studies on antigen characterization in lymph node cancer invasion were also performed. In the actual analyses, invasion (defined in TNM system by the symbol N (node)) has been divided in N=0, no invasion, and N>1 if the cancer has spread to the lymph nodes. SLe^x (scored by anti-CD15s staining) shows a significative increase in expression in cancer cell invasion (Figure 3.26A). Likewise, Le^y extension expression as a significant increase in lymph node invasion (Figure 3.26 B)

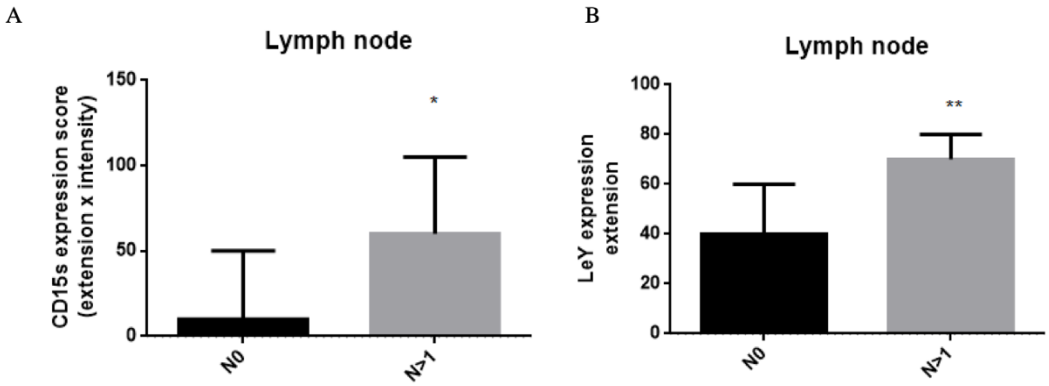


Figure 3.26 Correlation studies between cancer invasiveness (N) and antigen expression. In A) increase in CD15s (sLe^x) expression (extension x intensity) is correlated with significant (*) increase of cellular invasiveness (N>1). Equally, in B) increase in Le^y extension in expression is significantly (**) correlated with a more malignant phenotype, in agreement with the aforementioned staging results. . Immunochemistry antigen signals were measured by the pathologist and analysed with unpaired T-test.

Crosstab analysis were also applied in order to understand correlation of antigen expression in ganglion metastasis (N=0 no invasive and N>0 invasive phenotype). HECA-452 ($sLe^{x/a}$) signals, defined as no-present by ",00" and present by "1,00", revealed to be highly expressed in ganglion metastasis (65,5%), while its absence has been correlated with a non-invasive cellular behaviour (72.7%) (Figure 3.27A). Equally, analysis with only CD15s (sLe^{x}) revealed that absence of antigen corresponds to the 81.8% of non-invasive tissue sections, while presence of the antigens increases ganglion metastasis up to 62.1% (Figure 3.27B).

A

Crosstab

			N_01		
			N0	N>1	Total
HECA_P50	,00	Count	8	10	18
		% within N_01	72,7%	34,5%	45,0%
	1,00	Count	3	19	22
		% within N_01	27,3%	65,5%	55,0%
Total		Count	11	29	40
		% within N_01	100,0%	100,0%	100,0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,713[a]	1	,030		
Continuity Correction[b]	3,294	1	,070		
Likelihood Ratio	4,797	1	,029		
Fisher's Exact Test				,040	,034
Linear-by-Linear Association	4,595	1	,032		
N of Valid Cases	40				

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,95.

b. Computed only for a 2x2 table

B

Crosstab

			N_01		
			N0	N>1	Total
CD15s_P50	,00	Count	9	11	20
		% within N_01	81,8%	37,9%	50,0%
	1,00	Count	2	18	20
		% within N_01	18,2%	62,1%	50,0%
Total		Count	11	29	40
		% within N_01	100,0%	100,0%	100,0%

Chi-Square Tests

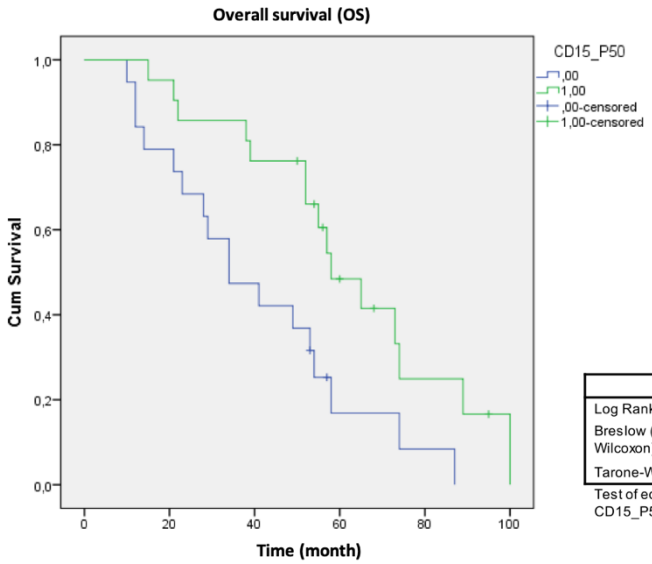
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6,144[a]	1	,013		
Continuity Correction[b]	4,514	1	,034		
Likelihood Ratio	6,525	1	,011		
Fisher's Exact Test				,031	,015
Linear-by-Linear Association	5,991	1	,014		
N of Valid Cases	40				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 5,50.

b. Computed only for a 2x2 table

Figure 3.27 Crosstab correlation between antigen expression and percentage of ganglion metastasis. In A) HECA-452 signal, which recognize $sLe^{x/a}$, is associated with an increase of ganglion metastasis. In B) analysis of CD15s (sLe^{x}) confirmed increases of invasive behaviour in 62.1% of the cases. Chi-square tested confirmed significance of the test with A) p=0.03 (*) and B) p=0.013 (**)

Overall survival (OS) patient's percentage, dependent on different antigens expression, was also measured (Figure 3.28). Analysis revealed that the increase in CD15 expression (Le^{x}) correspond to increase of 8-years OS, with a significance of p=0.013 (**)



Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	6,201	1	,013
Breslow (Generalized Wilcoxon)	6,330	1	,012
Tarone-Ware	6,394	1	,011

Test of equality of survival distributions for the different levels of CD15_P50.

Figure 3.28 8-years overall patient survival. Expression of CD15 increase chances of OS with a significance of p=0.013 (**)

As a conclusion we have shown correlation in expression between sLe^x and sLe^a in different cancer staining, obtained with anti-CD15s, anti-CA19.9 and HECA-452. Moreover, studies of all the antigen revealed significant correlation among $sLe^{x/a}$, Le^y and CD44 with invasiveness behaviour in tissue scoring. On the contrast, expression of Le^x showed an increase of 8-years overall survival, suggesting that sialylation level of Lewis antigens is involved in malignancy profiles.

Being the purpose of the project to find biomarkers that could detect CRC throughout the cancer staging and to build antibodies as diagnostic and therapeutic tool, $sLe^{x/a}$ antigens, as aberrant glycans and CD44, as glycoprotein, were selected for further studies and biomarker characterization.

3.1.5 Biomarkers anti-tumour effect

In order to validate the potential anti-tumour effect of commercial antibodies in vitro and to validate the potential efficacy of sLewis and CD44 antigens as novel biomarkers, scratch wound assay was performed. The migration potential of the cells was measured in the presence or absence of antibody against CD44 and SLe^x, in order to see their potential in influencing cells migration, with both CD44 and $sLe^{x/a}$ antigens involved in cells adhesion and

migration[222,294]. Three cell lines were selected by their antigen expression levels: LS174T (high in $sLe^{x/a}$ and CD44), SW48 (medium-low in $sLe^{x/a}$ and CD44), SW620 ((medium-low in $sLe^{x/a}$ and null in CD44). With the purpose to maintain low variability, the antibody selected were all mouse binding human IgG type: anti-mouse II, anti-CD15s and anti-CD44-hpan. Antibody were kept in motion with concentration of 0.5 µg/mL, renewed every 3 days, given the stability of antibodies in culture. Cells were cultivated in Ibidi culture insert and maintained in culture up to full wound closure of the negative control (anti-mouse II IgG) or cell death.

In Figure 3.29 example of the scratch wound assay is shown. The positive control showed a confluency of more than 90% in the scratched fragment at day 7, while the cells stained with antibodies against SLe^x exhibited a delay in closure. Moreover, with sLe^x being more expressed than CD44, the sLe^x treatment resulted in less migration on day 7 than the anti-CD44 antibody. These preliminary results suggest the involvement of sLe^x antigens in cell migration and adhesion, and the effect of commercial antibodies on cell migration in vitro.

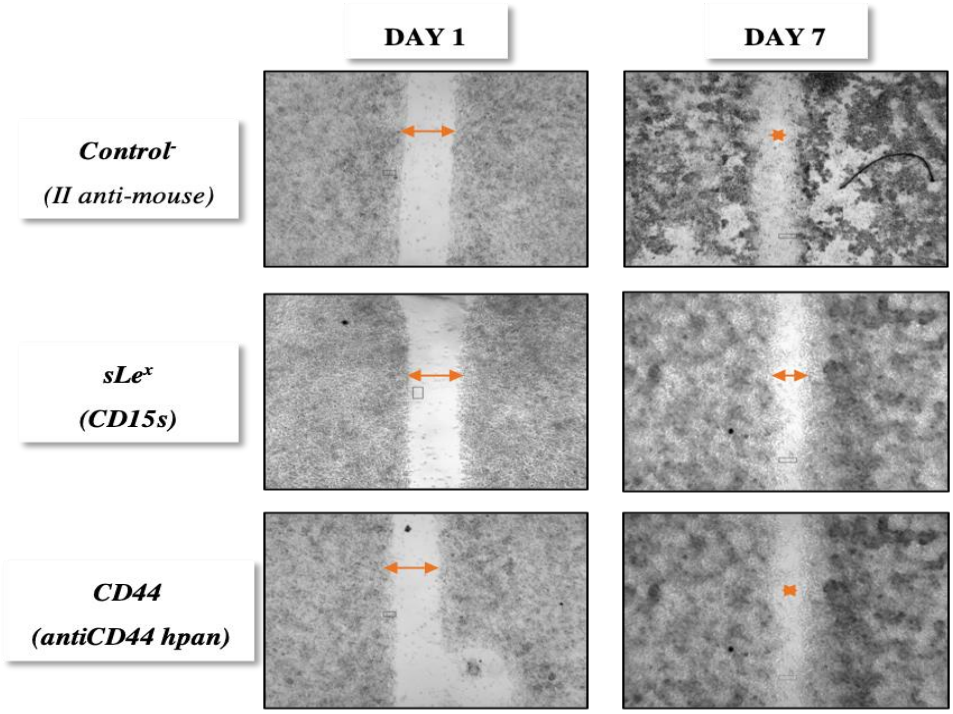


Figure 3.29 Example of scratch wound assay performed with SW620 wt cell lines. The cells are positive for sLe antigens but have low expression of CD44 SW620 shows a most fully closure at the 7th day, while wounds maintained with anti-CD15s show delayed in stability. CD44 being mostly not expressed by the SW620, show similar closure to the negative

LS174T were maintained in culture for four days. Wounds with anti-CD15s and anti-CD44 shows almost no closure, compared to the linear growth of the negative control (Figure 3.30 A1-2). Delay in closure have been noticed in SW48 where, after 8 days, samples containing anti-CD15s and anti-CD44 appear to have substantial slower migration compared to the negative control (Figure 3.30 B1-2). On the contrary, SW620 maintained with anti-CD44 revealed almost identical closure of the control, while samples co-cultivated with anti-CD15s shows differences in closure at the 7th day (Figure 3.30 C1-2).

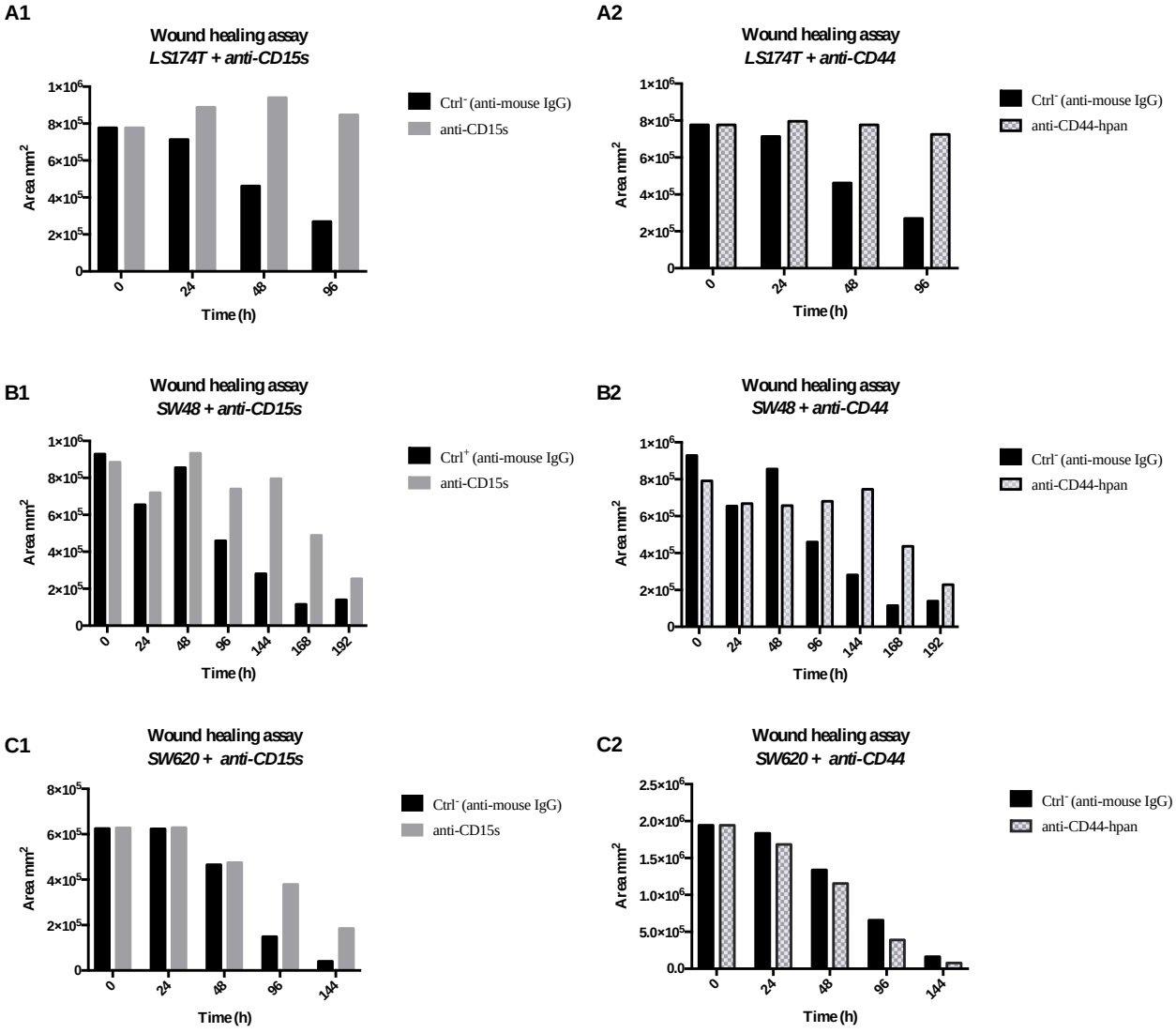


Figure 3.30 Comparison of wound closure of LS174T, SW48 and SW620, co-cultured with anti-mouse II (Ctrl$^-$), anti-CD15s (sLe^x) and anti-CD44-hpan (CD44). Antibodies were maintained in motion at 0.5µg/mL, until full wound closure or cell death. In A) LS174T, high in sLe and CD44, at the 4th day show almost no closure compare to the negative control. In B) SW48 medium-low in sLe$^{x/a}$ and CD44, show clear delay in closure after the 8th day. In C) SW620, low in sLe, has a delayed wound repair, compared not only to the negative control, but also to the samples co-cultured with anti-CD44. In fact, this antibody appears not to affect cell migration when CD44 is mostly absent from cell expression. The values obtained by the scratch wound assay experiment are representative of unique experiment; no statistical experiments have been performed on the results obtained

Further preliminary testing with LS174T co-cultured with anti-CA19.9 (sLe^a) and E-selectin-Ig (chimera protein binding $sLe^{x/a}$) revealed similar behaviour in wound healing delay. In fact, not only the antibody anti-CA19.9 confirmed difference in closure at the 4th day, but also cells in contact with E-selectin -Ig increase cells distance (Figure 3.31 A-B).

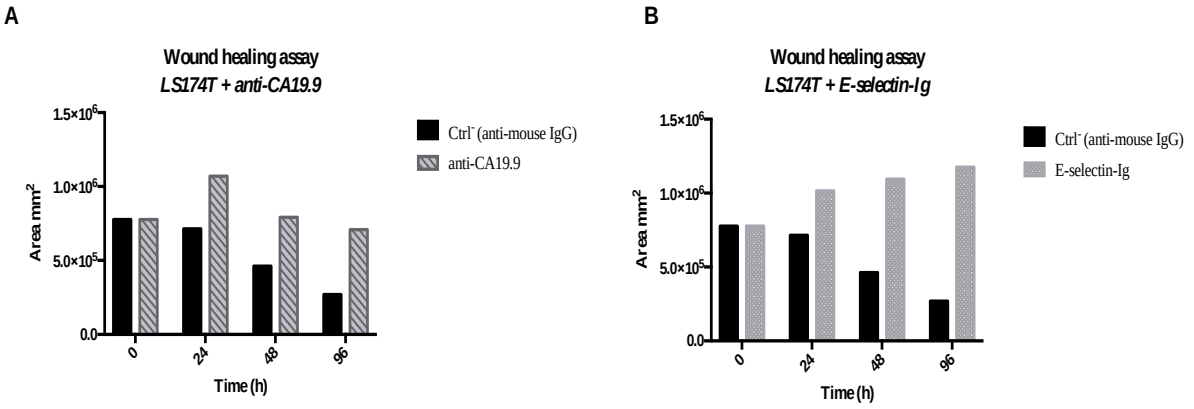


Figure 3.31 Preliminary results of wound closure of LS174T co-cultured with anti-mouse II ($Ctrl^-$), anti-CA19.9 (sLe^a) and E-selectin-Ig ($sLe^{a/x}$) antibodies were maintained in motion at 0.5µg/mL, until full wound closure or cell death. In A) LS174T, high in sLe, at the 4th day show almost no closure compare to the negative control. In B) LS174T co-cultured with E-selectin --Ig revealed proportional increase in wound distance, suggesting an active role of E-selectin in blocking cell migration.The values obtained by the scratch wound assay experiment are representative of unique experiment; no statistical test have been performed on the results obtained.

This preliminary result, suggest that E-selectin when used as decoy for $sLe^{x/a}$ antigens, may participated in cancer cell toxicity. However, further experiments should be applied before any conclusion could be made.

Thus, given the difference of antigen expression of the three cell lines, it is possible to notice a proportional delay in migration in all cell lines. Of particular interest is the cell behaviour of SW620 with anti-CD44, not only because wound closure is proportional to the control, but also because validates the antibody effects in LS174T and SW48 wound (Figure 3.30 C2). Increase in wound space in presence of E-selectin-Ig have also been considered. Data suggest that not only sLe^x and CD44 could play major in cell migration, but also the production of specific monoclonal antibody against sLe^x and/or CD44 might have good potential if used as therapeutic tool. For this reason, CD44 and $sLe^{x/a}$ were selected as new potential biomarkers.

3.2 PART II: biomarker isolation and characterization

Biomarker discovery has been performed by cell lines screenings, by analysis cohort of 40 CRC patients FFPE tissue and by antibody *in vitro* efficacy. Experiments confirmed overexpression of aberrant glycosylation compared to healthy tissue. Moreover, increase in CEA (used as a prognostic biomarker) and CD44 (adhesion protein involved in tumour malignancy) has also been observed. In detail, $sLe^{x/a}$ revealed to be involved in most cancer stages, also when translated into cancer line analysis; it also shows to play major role in ganglion metastasis. On the other hand, CD44 have shown higher expression in HT29 when transfected with FUT6, glycotransferase involved in core-fucosylation of sLe^{x}, and in LS174T, which expresses high levels of $sLe^{x/a}$ antigens. It has been also displayed its involvement in cell migration in tissue analysis and migration assay (wound scratch assay). For all these reasons, $sLe^{x/a}$ antigens and CD44 glycoprotein have been selected for further characterization and eventual mice immunization (Figure 3.32). $SLe^{x/a}$ and CD44 are also known as E-selectin ligands. The role of E-selectin in cancer progression and metastasis has been reinforced by a number of basic researches[69,203], thus studies of E-selectin ligands behaviour could be of great utility in biomarker research. Furthermore, because an ideal biomarker should be cancer-specific and that both, $sLe^{x/a}$ and CD44, can be found in normal tissue, it has been decided to study specific change in glycosylation on CD44 surface. In fact, glycoprotein carrying a specific aberrant glycosylation could be the key for the selection of new cancer- specific biomarker.

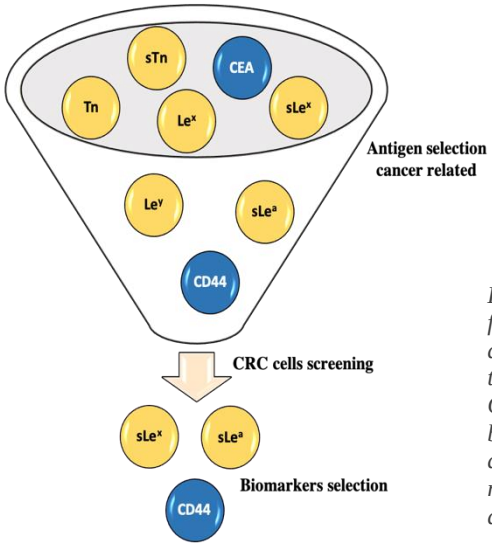


Figure 3.32 Representative funnelling structure for antigen selection. In yellow are represented the different aberrant cancer-related glycans, in blue the glycoproteins involved in cancer profiling. CRC cells were tested by flow cytometry, western blot, immunohistochemistry and wound scratch assay. $SLe^{x/a}$ and CD44 have been selected as the most promising biomarkers involved throughout different cancer stages.

3.2.1 Immunoprecipitation and isolation of antigens

In order to further characterize CD44 and sLe$^{x/a}$ as E-selectin ligands, we have performed protein isolation. Immunoprecipitation (IP) technique was selected and applied on LS174T and HT29 FUT6 membrane protein extraction, as both cell lines show high expression of CD44 and sLe$^{x/a}$ antigens (Fig 3.21).

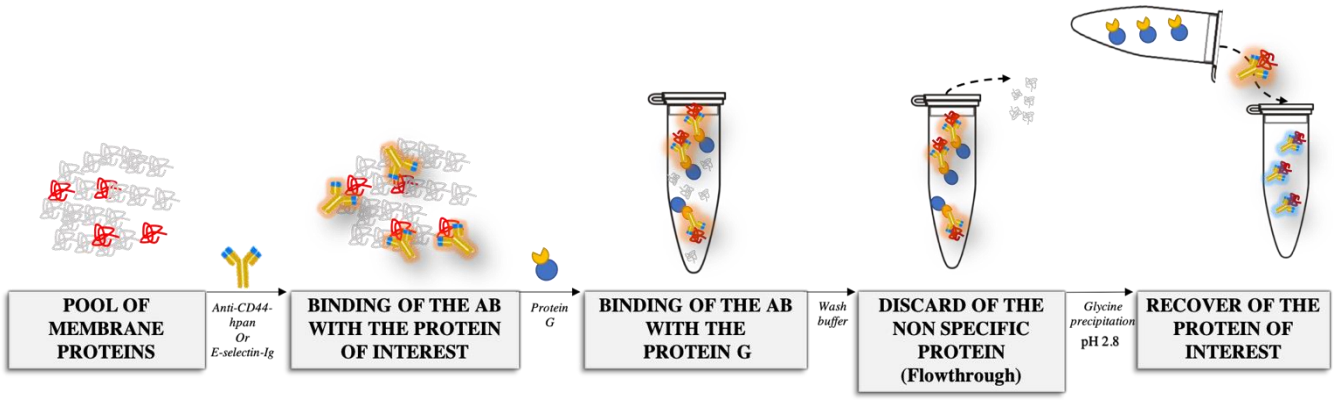


Figure 3.33 Schematic representation of immunoprecipitation (IP) technique. The different steps are described in figure. Final sample is recovered with glycine-low pH where both antibody and desired protein precipitate. In the final sample, pH is balanced by Tris-base solution at pH6.8

Two antibodies were selected for immunoprecipitation: anti-CD44-hpan and E-selectin-Ig chimera (3µg of antibodies for 100 µg of protein). Example of IP with HT29 FUT6 cell membrane extract is shown in figure (Figure 3.34 A-B). Both precipitate show positive staining for CD44 and E selectin ligands with major bands at ~135kDa.

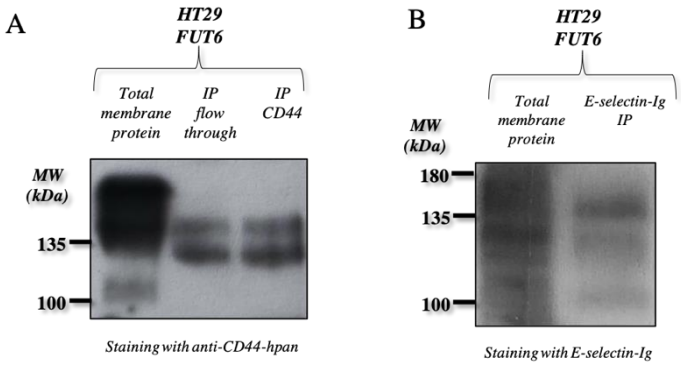


Figure 3.34 Example of IP results for HT29 FUT6 membrane protein extract. In A) precipitate with anti-CD44-hpan revealed two major bands in ~135kDa. IP flow through shows did not occur that recovery of all CD44 proteins. In B) E-selectin IP revealed two major bands at ~135kDa and a minor at ~100 kDa. Western blots shown are representative of 18 experiments

.

3.2.2 E-selectin ligands analysis

Precipitate of LS174T and HT29 FUT6 (duplicate) by E-selectin-Ig were analysed through liquid chromatography- mass spectrometry (LC-MS/MS). Samples were separated in gel in 7 bands each (Figure 3.35A) and proteins were in-gel digested for LC-MS/MS analysis. In the four sample (duplicates of both cell lines) CD44, backbone 82kDa, was identified with 100% confidence (Figure 3.35 B-C). Identification of the protein do not include post-translational modification thus the possible CD44 isoform might be CD44v2-10[256], that could be found glycosylated up to 250 kDa.

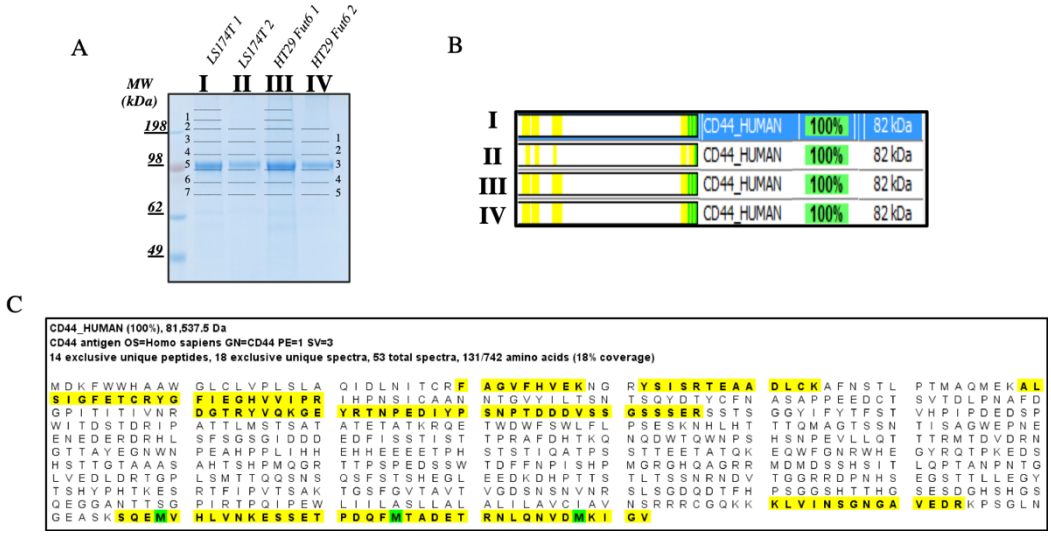


Figure 3.35 LC-MS/MS of LS174T and HT29FUT6 membrane protein E-selectin-Ig IP. In A) gel separation of the protein and division in 7 bands. In B) confidence of 100% CD44 human. Backbone of 82kDa. In C) correspondent sequence of peptides confidence.

3.2.2.1 CD44 as E-selectin ligand in CRC

Immunoprecipitation of HT29 FUT6 with anti-CD44-hpan and E-selectin-Ig were tested and compared in immunoblots. This experiment served as a validation of CD44 carrier of $sLe^{x/a}$.
In Figure 3.36 is shown western blot revelation, presenting in sequence, positive and negative controls (HT29 FUT6 total membrane proteins, IP of CD44 flow through), IP of CD44 and positive control (IP of E-selectin ligands). Blot has been stained with E-selectin-Ig and one band appeared at the expected ~135kDa, confirming CD44 as carrier or E-selectin ligand.

Validating CD44 as glycoprotein carrying the sLe$^{x/a}$ antigens, we have planned to produce antibody combining its double specificity for CD44 and sLe$^{x/a}$ antigens, to increase specificity and sensitivity in biomarkers detection.

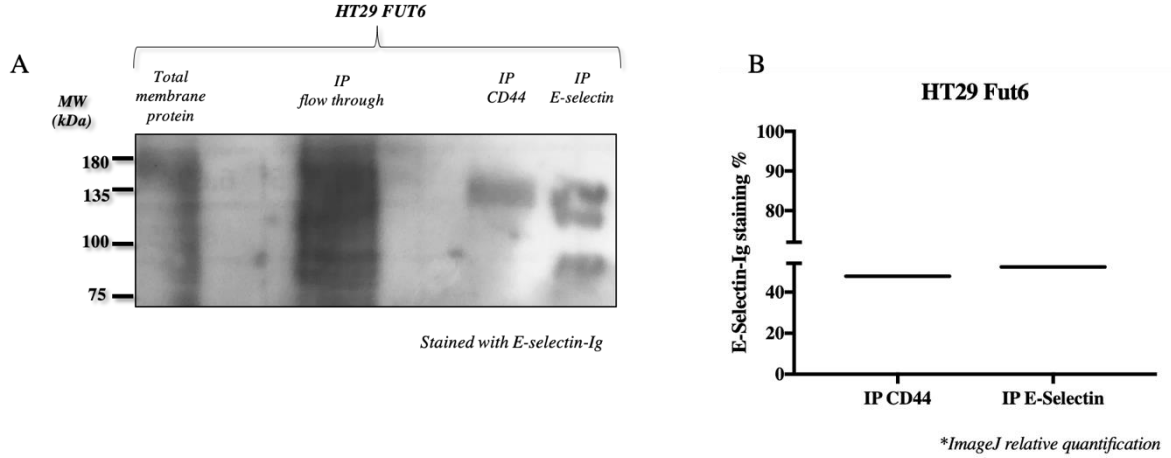


Figure 3.36 Validation of CD44 IP as E-selectin ligand. In A) blot is stained with E-selectin Ig, and both CD44 IP and E-selectin-Ig IP presents signal at ~135 kDa. E-selectin-Ig IP has been used as a positive control. In B) qualitative signal intensity has been measured by imageJ. 135 kDa bands in IP CD44 and IP E-selectin, show similar intensity percentage. Western blots shown are representative of 2 experiments

The CD44 immunoprecipitated was also tested by enzyme-linked immunosorbent assay (ELISA). Plate were coated with a max yield of 3µg/mL and tested by anti-CD44-hpan (positive control) and HECA-452, as E-selectin substitute, recognizing sLe$^{x/a}$. Absorbance (450nm) revealed positive signal for sLe$^{x/a}$ antigens for CD44 IP (Figure 3.37). Cross-testing was useful to confirm CD44 as a carrier of or/and E-selectin ligand.

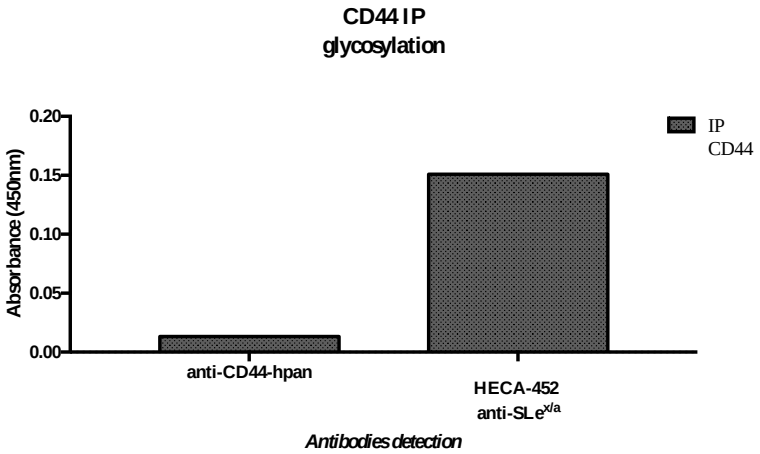


Figure 3.37 ELISA testing of CD44 IP. Although weak signal is revealed in the positive control (anti-CD44-pan), positive signal for HECA-452 (sLe$^{x/a}$) is shown. Negative controls, respectively II anti-mouse-HRP IgG for anti-CD44 and II anti-rat-HRP IgM, already subtracted from the absorbance intensity. The ELISA experiment is representative of a unique experiment.

3.2.3 Förster resonance energy transfer between CD44 and sLe$^{x/a}$

Förster resonance energy transfer (FRET) is a mechanism that describes energy transfer between two fluorophores. Energy transfers happens when a donor molecule is at a max distance of 9nm from an acceptor molecule and the donor-emission-wavelength is overlapping with the excitation-wavelength of the acceptor. For this reason, the choice of the proper fluorophores was critical. FRET was performed by flow cytometer on LS174T cell lines. In case of positive FRET was expected median fluorescence intensity (MFI) increase in the acceptor channel. Three primary antibodies have been selected: anti-CD44 (rabbit polyclonal), anti-CD15s and anti CA19.9 (mouse monoclonal). Combination of secondary antibodies was chosen based on fluorophore qualities and on coupling: PE (anti-mouse) --> Cy5 (anti-rabbit) [glycan donors, CD44 acceptor] and FITC (anti-rabbit) --> PE (anti-mouse) [CD44 donor,

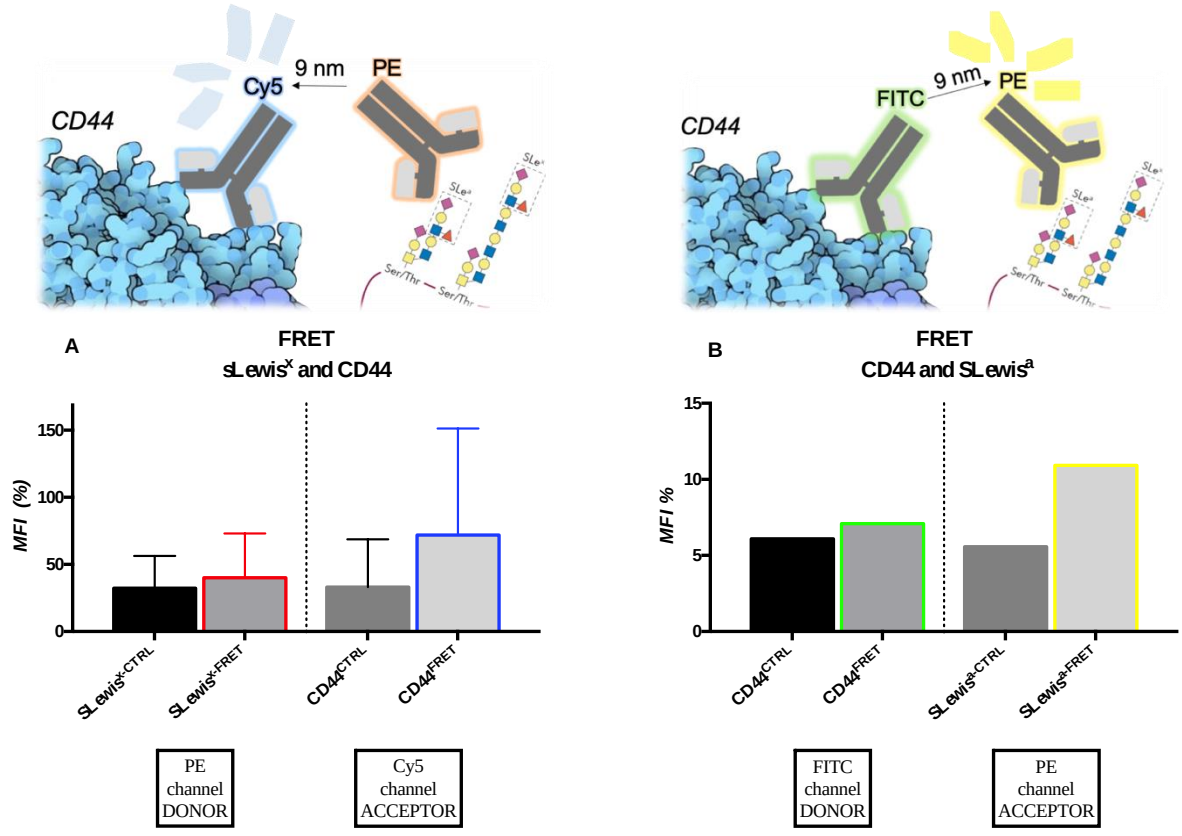


Figure 3.38 FRET representation between sLewis antigens and CD44 on LS174T cell line. In A) sLe^x donor-PE signal remained equal, while substantial increase is seen in CD44 Cy5-acceptor, suggesting presence of sLe^x on protein surface. In B) CD44-PE (donor) and sLe^a-PE (acceptor), showed increase in MFI PE-acceptor channel, indicating presence of sLe^a on the glycoprotein. MFI is normalized. The MFI measurement obtained by flow cytometry experiment are representative of n=2 expeeriments

glycans acceptors]. As a result of the combinations, visible increase of signal in CD44 (Cy5-channel acceptor) from sLe^x donor suggest presence of the glycan on the protein surface (Figure 3.38A). Combination between CD44-FITC (donor) and sLe^a-PE (acceptor), showed increase in MFI PE-acceptor channel, indicating presence of sLe^a on the glycoprotein expressed on LS174T surface (Figure 3.38B).

3.2.4 CD44 type of glycosylation

Sialyl $Lewis^{x/a}$ can be found either as O- or N-glycans. Thus, in order to characterize the type of glycosylation, IP of E-selectin-Ig from HT29 FUT6 total membrane protein was treated with two enzymes. PNGase F is the most effective enzymatic method for removing almost all N-linked oligosaccharides from glycoproteins; neuramidase (Sialidase) cleaves terminal sialic-acid residues that are α2,3-, α2,6-, or α2,8-linked to Gal, GlcNAc, GalNAc, AcNeu, GlcNeu, oligosaccharides, glycolipids, or glycoproteins.

PNGase F treatment of E-selectin-Ig IP was stained with anti-CD44-hpan and as E-selectin-Ig as cross-control. As a result, drop of signal and of molecular weight is visible in the blot A, from ~150 kDa down to ~120 kDa (Figure 3.39A) and lost in signal with E-Selectin-Ig (Figure

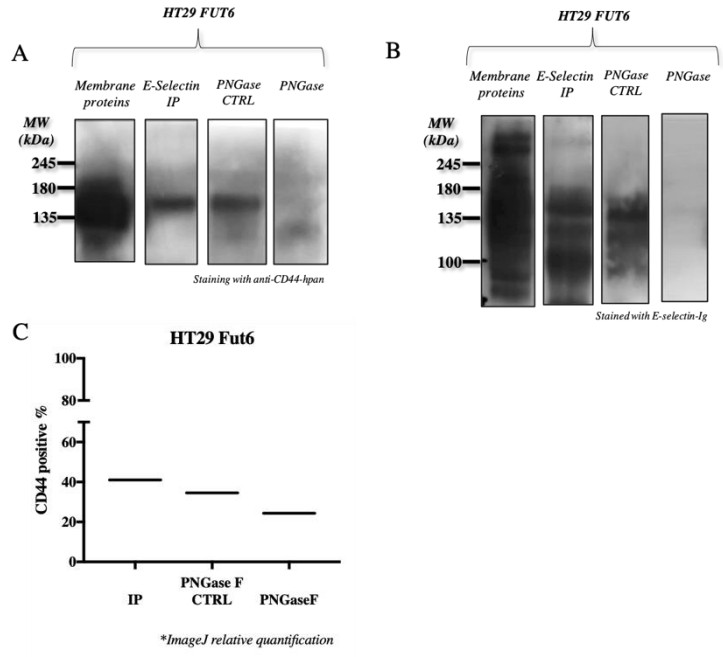


Figure 3.39 PNGase treatment of E-selectin-Ig IP from HT29 FUT6. In A) blot is stained with anti-CD44-hpan and is possible to visualize a drop in signal and molecular weight from ~150 kDa to ~120 kDa. In B) twin blot is stained with E-selectin-Ig and lost in signal is visible, confirming the correct enzymatic reaction. In C) relative quantification of CD44 signal is visible. Drop in intensity may be due to loss of glycosylation and thus antibody affinity. Western blots shown are representative of a unique experiment.

3.39B). Drop in intensity of signal may be due to glycan influencing anti-CD44-hpan antibody binding (Figure 3.39C)

Neuramidase treatment of E-selectin-Ig IP was also stained with anti-CD44-hpan and as E-selectin-Ig as cross-control. As a result, increase of signal of CD44 is visible in figure A and decrease in signal with E-Selectin-Ig in B (Figure 3.40). Increase of CD44 signal (Figure 3.40C) may confirm a role of glycosylation in anti-CD44-hpan binding and affinity. Moreover, it suggests that the sialidase reaction was partially completed.

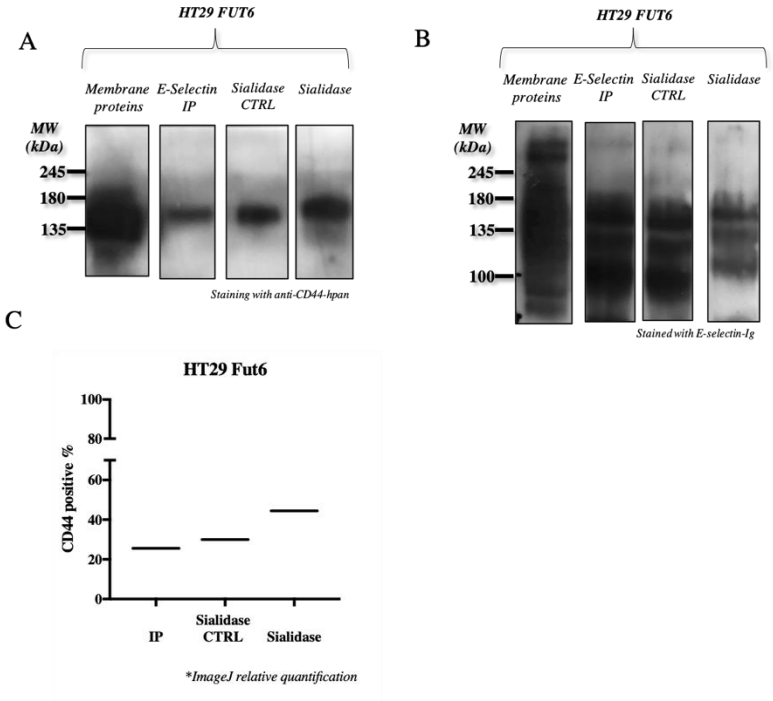


Figure 3.40 Sialidase treatment of E-selectin-Ig IP from HT29 FUT6. In A) blot is stained with anti-CD44-hpan and is possible to increase in signal intensity. In B) twin blot is stained with E-selectin-Ig and decrease in signal is visible. In C) relative quantification of CD44 signal is showed. Increase in intensity may be due to loss of glycosylation and thus antibody affinity. Western blots shown are representative of a unique experiment.

Using enzymatic reaction, it was possible to prove that CD44 in HT29 FUT6 is decorated majorly with N-glycans and that E-selectin ligands ($sLe^{x/a}$) are bound to N-glycan chains. Reactions with neuramidase suggested partial removal of sialic acid and role of glycosylation in influencing anti-CD44-hpan binding and affinity. Since no enzyme for O-glycan cleavage with the same function of the PNGase F is available, it was no possible to study directly O-type of glycosylation on CD44 on HT29 FUT6.

3.2.5 CD44 as a soluble biomarker

Study of glycosylation of CD44 suggested its potential as a cancer-specific biomarker. However, a good biomarker should be detectable with the use of non-invasive techniques. For this reason, the next question raised was to discover if CD44 is soluble in the serum. The aforementioned IHC experiment revealed presence of CD44 in tumoral mucus. Nonetheless, further studies aimed to test the presence of CD44 in LS174T cell line supernatant. Being the presence or the concentration of free CD44 unknown, supernatant was concentrated 6x and 9x times. Western blot revealed proportional increase in signal of bands at ~150kDa and ~100 kDa, in accordance with LS174T total membrane positive control (Figure 3.42). Results show that in CRC LS174T cell line culture, CD44 can be found as soluble form, increasing its potential as good biomarker.

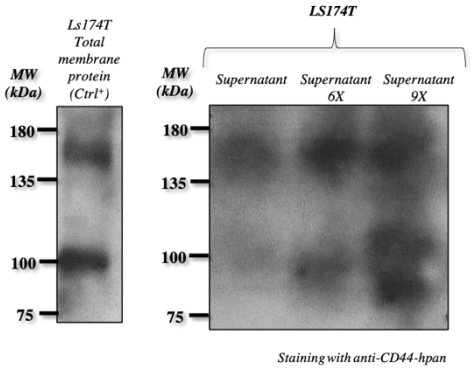


Figure 3.41 Testing of LS174T supernatant for CD44 as soluble protein. Increase of supernatant concentration, correspond to a proportional increase in signal, correspondent to same molecular weight of the bands used as a control of total membrane protein LS174T. Bands are revealed at ~150 kDa and ~100 kDa.. Western blots shown are representative of a unique experiment.

3.2.6 Antigen isolation and preparation

As a consequence of the results obtained, CD44 and $sLe^{x/a}$ were selected for mouse immunization.

However, in order to increase antibody production specificity a certain purification of CD44 is required. LS174T was selected as cell line with highest expression of high molecular weight of CD44.

IP of CD44 was performed and in order to separate the protein precipitate with the antibody anti-CD44-hpan (that co-precipitated during isolation), sample was subject to a native gel PAGE (Figure 3.42A) with the main). interest to stimulate antibody production against native form of the protein. 3/4 of native gel was then cut in different bands (A-D) and stored in PBS. 1/4 of gel was used for western blot analysis of the protein bands. Protein ladder was used as standard and mark.

After revelation, band containing CD44 (with maximum yield possible achieved is 0.5 µg/µL) was extracted and injected into the mice. Presence of polyacrylamide was used as an extra immunization adjuvant. Mice after immunization of CD44 will be named mice^{CD44}.

A second immunization series was performed in order to produce sLe$^{x/a}$ specific antibodies.

For this immunization was chosen to inject directly into the mice 10µg (per inoculation) of LS174T total membrane protein, high in sLe$^{x/a}$ antigen expression (Figure 3.42B). These mice will be referred as mice^{LS174T}.

.

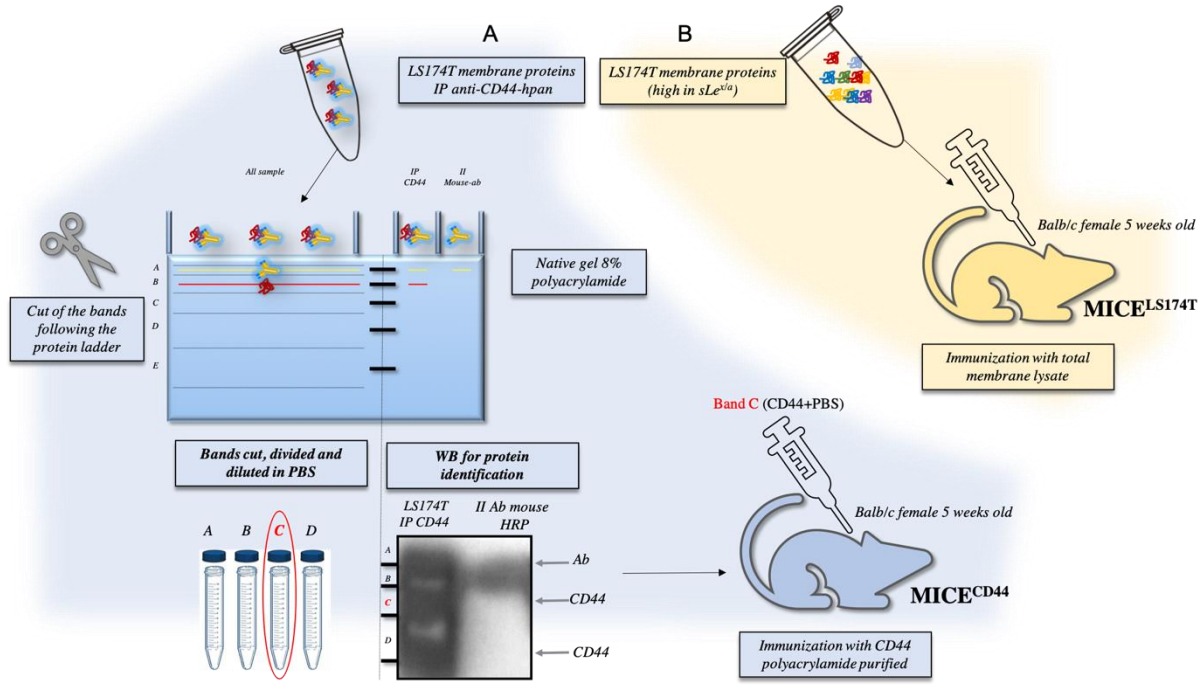


Figure 3.42Schematic representation of mice antigen immunizations. In A) CD44 IP of LS174T was run in native gel 8% in order to separate the protein from the antibody anti-CD44-hpan, co-precipitated. western blot revelation was needed for identification of the correct band containing CD44. samples in polyacrylamide, diluted in PBS 1x, were directly injected in the mice, afterwards named mice^{CD44} (n=3). in B) direct injection of LS174T total membrane lysate in the mice. antigen was selected based on the high expression of sLe$^{x/a}$ glycans on Ls174T membrane. mice after immunization have been referred as mice^{LS174T}(n=2)

3.3 PART III: immunization and serum testing

CRC cells and tissue screening lead to the identification of CD44 glycoprotein and $sLe^{x/a}$ antigens as a good potential therapeutic biomarker. Moreover, CD44 showed to be a soluble protein that could be useful in early CRC diagnosis. Thus, two series of mice were immunized with CD44 isolated from LS174T ($mice^{CD44}$) and LS174T total membrane lysate high in $sLe^{x/a}$ antigens ($mice^{LS174T}$). This section is dedicated to immunization timelines and techniques adopted in order to check antibody serum titer and its antigen specificity.

3.3.1 Immunization types and timeline

Following the previous results, it was decided to pursue with immunizations of the mice with CD44 immunoprecipitated from LS174T cells or the LS174T total lysates. The immunoprecipitated containing CD44 glycoprotein was obtained by co-precipitation with the antibody anti-CD44-hpan and subsequently run on native gel for separation. The polyacrylamide band corresponding to CD44 protein was cut and dissolved in PBS. The mix was blended with adjuvants and directly injected into the mice. A theoretical yield (based on max antibody capture efficiency) was calculated, but because of the impossibility to quantify the protein, a total of 5 immunizations were performed (Figure 3.43A). A second type of injection in different mice was performed with LS174T total membrane extraction (Figure 3.43B).

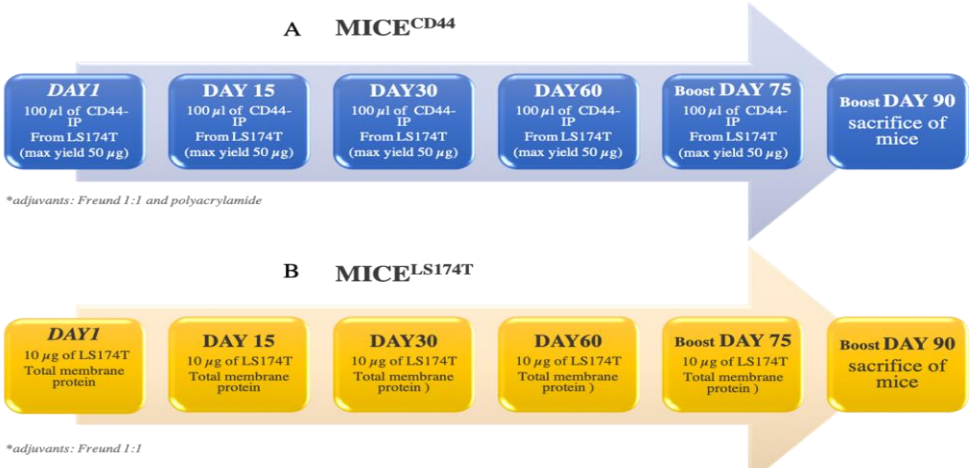
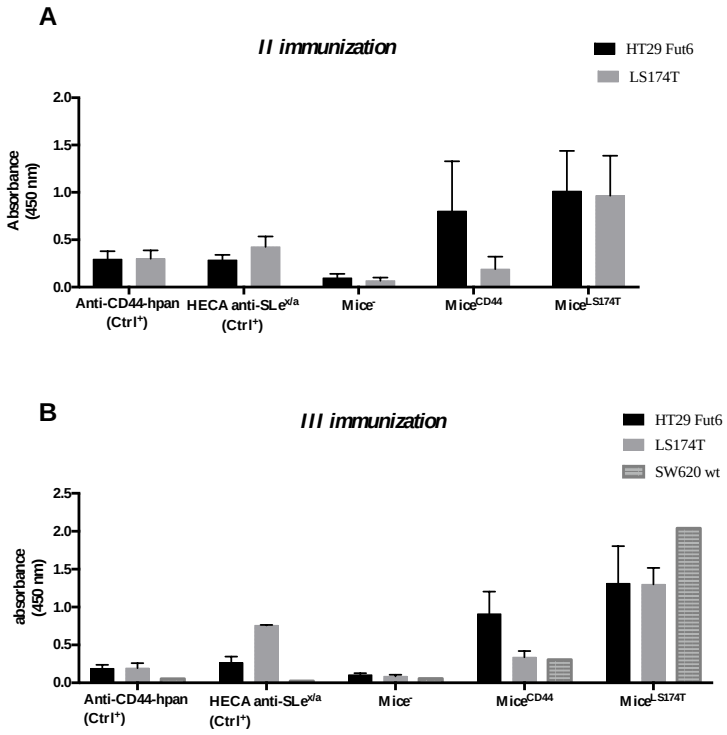


Figure 3.43 Timeline representing the different injections performed on mice during the immunization protocol. Two parallel protocols were performed with CD44 (A) or total membrane protein LS174T cells (B). Mice were immunized with CD44 extracted from LS174T mixed with polyacrylamide and adjuvants or immunized with total membrane extraction protein from LS174T.

3.3.2 Serum testing by ELISA

The antibody titre in serum of immunized mice was evaluated mainly by ELISA. It has been developed specific protocols to allow a more efficient screening that comprises ELISA plates coated with membrane proteins extracted from cell lines previously tested: LS174T (high expression of CD44 and sLe^x), SW620wt (not expressing CD44) and HT29 FUT6 (high expression of CD44). The coated wells were incubated with serum from immunized mice and a non-immunized mouse (negative control). Other controls were performed to test the performance of the method which included the anti-CD44-hpan antibody and HECA-452 antibody positive for E-selectin ligands (Figure 3.44). The analysis of antibody titre was checked regularly after the 2nd immunization.

The results showed high reactivity between HT29 FUT6 total lysate and serum from mice immunized with CD44 ($Mice^{CD44}$). All three coatings (LS174T (high expression of CD44 and sLe^x), SW620wt (not expressing CD44) and HT29 FUT6 (high expression of CD44) showed high reactivity with serum of $Mice^{LS174T}$. In order to prevent tolerance reactions with loss of antibodies titre, due to prolonged immunizations, the mice were sacrificed at the 5th immunization. The spleens from mice showing antigen reactivity were then used to obtain hybridoma cells by fusion with myeloma cells, as part of the workflow (Figure 3.44).



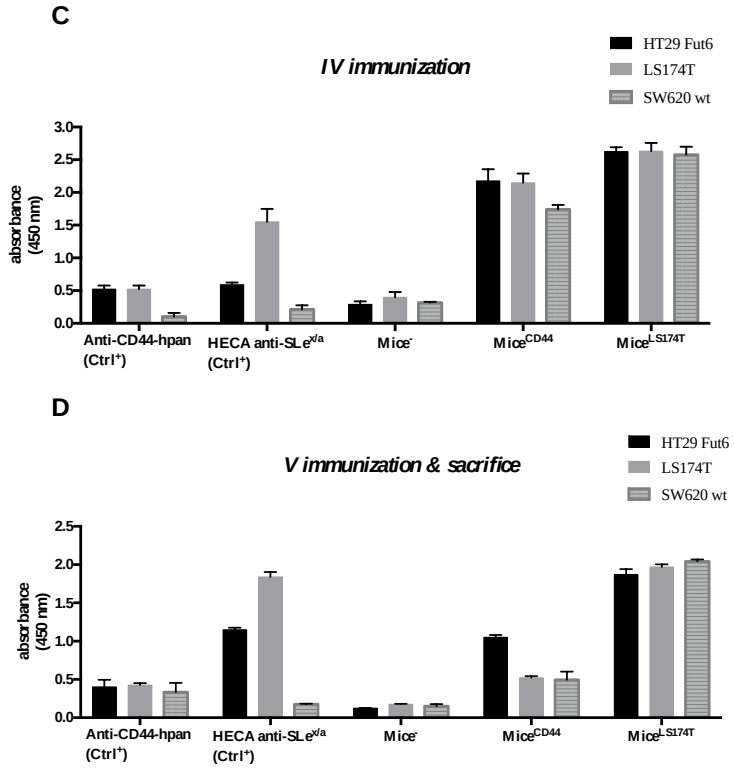


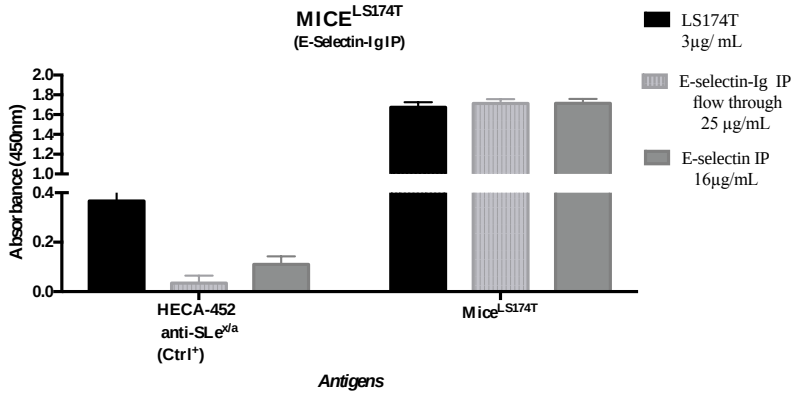
Figure 3.44 Results of mice serum tested for reactivity against CRC cell lysates. The presence of antibody titres of mouse sera was tested against the antigens coated and evaluated by ELISA. During the immunization three mouse groups were tested: Mice^{-} (control mice with no immunization), Mice^{CD44} obtained from the mice immunized with CD44 extracted from LS174T cell line, and Mice^{LS174T} immunized with LS174T total membrane protein. The coating was performed with LS174T (high in CD44 and E Selectin ligands), HT29 FUT6 (high in CD44) and SW620 wt (negative for CD44). During the different immunizations a high affinity for sera of Mice^{CD44} with HT29FUt6 and high signals with Mice^{LS17T} were detected for the three coatings. Mice were sacrificed after the 5th immunization. In A) serum check at day 30; in B) serum check at day 60; in C) serum check at day 75; in D) serum check at day 90 and sacrifice. ELISA experiment shown are representative of n=5 experiment. No significance was found with one way ANOVA test.

3.3.3 E-selectin testing by ELISA

Mice^{LS174T} serum presented clear positivity in signals by staining with three different CRC cell lines proteins. However, no specificity or difference in reaction was found among the three samples, probably due to the high titre of polyclonal produced by the immunization. Thus, in order to double-check specificity against E-selectin ligands, IP with E-selectin-Ig of LS174T total protein was tested against mice^{LS174T} serum (Figure 3.45). Cross-control HECA-452 (anti-sLe$^{x/a}$) was used to check IP reaction. Results shows a clear high binding in LS174T control, IP flow through (probably with specific proteins) and E-selectin-IP. Although, many non-specific bounds are present, positive signal in E-selectin-IP wells were positive indicating the use of this mice to obtain hybridomas.

3.3.4 Serum check on live cells

In order to check whether mice $^{CD44/LS174T}$ could recognize antigen on live cells, flow cytometry against LS174T (high expression of CD44 and sLe^x) and HT29 FUT6 (high expression of CD44) was tested. Serum was checked at the 4th immunization (day 60). Results show low intensity signal recognition of both cell lines by serum of mice^{LS174} (Figure 3.46A+B), but no binding was detected in mice^{CD44} (Figure 3.46A). Although recognition on membrane of CD44 have been preferable, antibody detecting CD44 in serum for diagnostic proposed was also desired. Thus, we decide to pursue flow cytometry experiments only with mice^{LS174T} in order to obtain future hybridomas.



Source of Variation	% of total variation	P value	P value summary
Interaction	1.01	<0.0001	****
Row Factor	98.08	<0.0001	****
Column Factor	0.5413	0.0013	**

Figure 3.45 ELISA testing of mice^{LS174T} on E-selectin IP of LS174T cell line. Signal show positive reactivity against LS174T total membrane lysate, protein flow through and IP, suggesting presence of polyclonal specificity and aspecificity for E-selectin ligands. mice were sacrificed in order to detect monoclonal specific for E-selectin ligands. ELISA experiment shown are representative of n=3 experiment. Significance was found with two way ANOVA test.

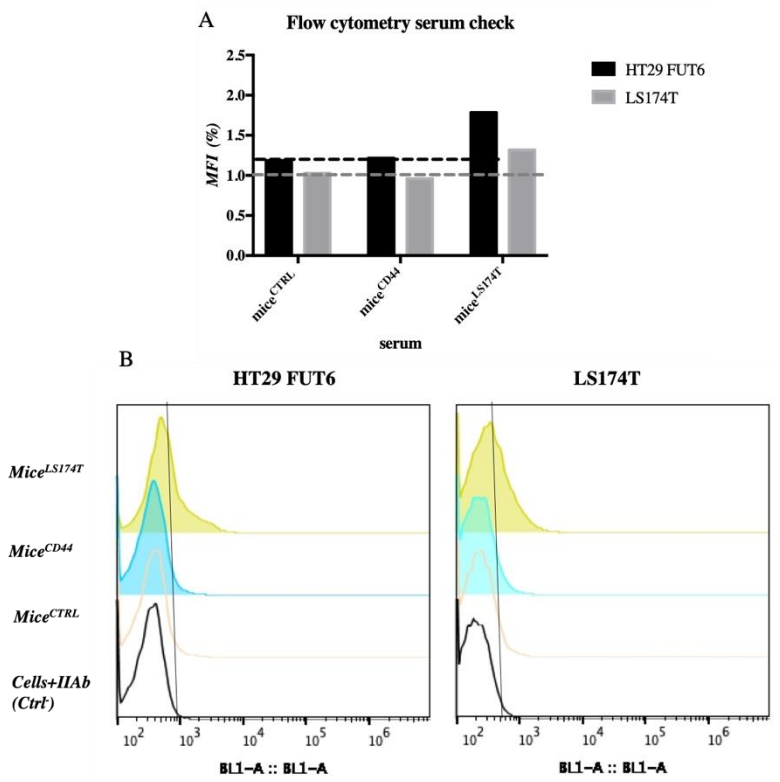


Figure 3.46 representation of flow cytometry analysis of HT29 FUT6 and LS174T cell lines by serum of mice$^{LS174T/CD44}$. In A) relative MFI have been plotted and shows positive reactivity by mice^{LS174T}, but no visible staining was noted by serum of mice^{CD44}. In B) flow cytometry histogram representation of difference in signal intensity by serum of mice$^{LS174T/CD44}$. only polyclonal serum of mice^{LS174T} shows partial binding. Flow cytometry experiment shown are representative of n=2 experiment.

3.3.5 Serum check of mice^{CD44} by dot blot

As a consequence of no specific binding of mice^{CD44}, cross check with dot-blot analysis were required. Different concentrations of total membrane protein (1.5/3/6 µg/µL) of LS174T (high expression of CD44 and sLe^{x}), SW620wt (not expressing CD44) and HT29 FUT6 (high expression of CD44) were bound on nitrocellulose membrane. Dot-blot showed proportional reactivity for LS174T and HT29 FUT6, with increase of signal in LS174T samples. No staining was shown in SW620 cell lines, neither non-specific binding of secondary antibody was detected. Results suggested presence of specific antibody against CD44 present in serum of mice^{CD44}.

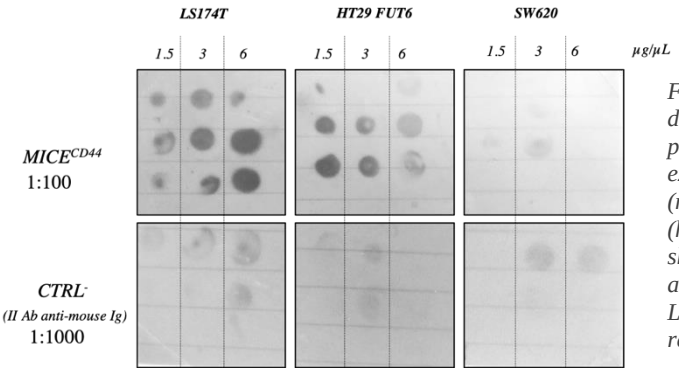


Figure 3.47 Dot blot of mice^{CD44} against different concentration of total membrane protein (1.5/3/6 µg/µL) of: LS174T (high expression of CD44 and sLe^x), SW620wt (not expressing CD44) and HT29 FUT6 (high expression of CD44). Dot-blot showed proportional reactivity for LS174T and HT29 FUT6, with increase of signal in LS174T samples. Dot blot experiment is representative of n=4 experiments

3.4 PART IV: hybridomas cloning and antibody selection

Two series of hybridomas were produced from spleen extracted from, respectively, mice^{CD44} and mice^{LS174T}. As previously described in chapter 1.8.1.1, hybridomas are fusion cells obtained from fusion of collected B cells from immunized mice and myeloma cells producing each one a specific monoclonal antibody. Each clone is tested and selected against the desired antigen[271,272,295]. The cells of the two hybridomas lines produced were mainly tested by ELISA coated with total membrane lysate of different CRC cell lines or commercial antigen as coating for the plate ELISA well. The selection process of the two hybridoma lines will be discussed in different paragraphs.

3.4.1 Mice^{CD44} clone's selection process

The hybridomas obtained from mice^{CD44} were cloned and a group of 5 different clones was evaluated. Cloning is the technique required to isolate each fused cell from the pool, in order to isolate B cell producing monoclonal antibody. Moreover, repetition of cloning helps to select the most stable hybridoma clone producing the antibody with the desired specificity and affinity.

The supernatant was tested against wells coated with LS174T (high in CD44 and E selectin ligand) and SW948 (low in CD44 and moderate in E-selectin ligands by ELISA), as total membrane lysate or live cells. To confirm binding specificity, Mice^{CTRL}, which had not been immunized, were also used (Figure 3.48). Results show that all clones have a low binding to LS174T and SW948 extracts like serum control. However, a higher binding to the LS174T live cells antigens was observed by the with clone CD44.4 supernatant relative to the other 4 clones, presenting simultaneously, a minimum signal against SW948 live cells antigens. Clone CD44.5 supernatant show no binding to LS174 live cells antigens but with the highest binding to the SW948 relative to the other clones.

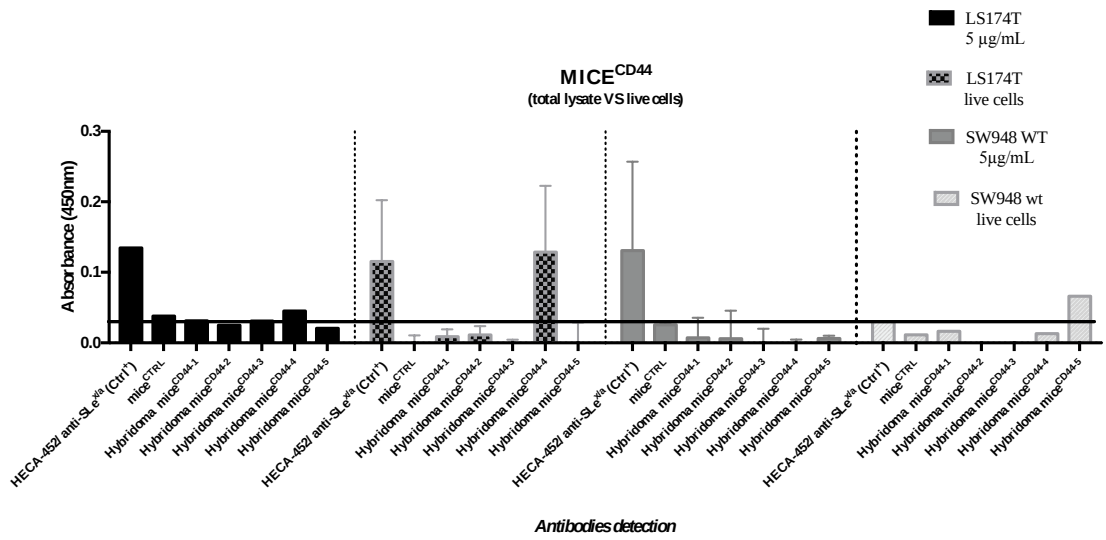


Figure 3.48 testing of 5 different clones obtained from mice^{CD44}hybridoma fusion. The clones have been tested against LS174T (high in CD44 and E selectin ligand) and SW948 (low in CD44 and moderate in E-selectin ligands by ELISA). the wells were coated either by total membrane lysate (5 µg/mL) or with 90% confluency live cells. Membrane lysate were blocked 1% BSA; no blocking was performed on live cells. Results show higher binding of CD44.4 in LS174T and CD44.5 in SW948 live cells. ELISA experiment shown are representative of n=2 experiment.

Thus, clone CD44.4 was selected for 2nd cloning of hybridomas, due to the signal intensity and specificity. In fact, SW948 wt is expressing minor isoforms of CD44 that might carry lower glycosylation levels, which would not be useful for the purpose. Rather, an antibody with high affinity for CD44 and detection of aberrant glycosylation, such as sLe$^{x/a}$ antigens, is needed for diagnostic and treatment scopes.

A secondary cloning was performed for CD44.4 (Figure 3.49). The sub-clones were tested against different cell membrane lysates: LS174T (high expression of CD44 and sLe^x), HT29 FUT6 (high expression of CD44), SW948 (low in CD44 and moderate in E-selectin ligands) and SW620wt (not expressing CD44) (Figure 3.49A). Clones were also tested against commercial tetramers of sialyl Lewis^x (sLe^x) and sialyl Lewis^a (sLe^a) (Figure 3.49B).

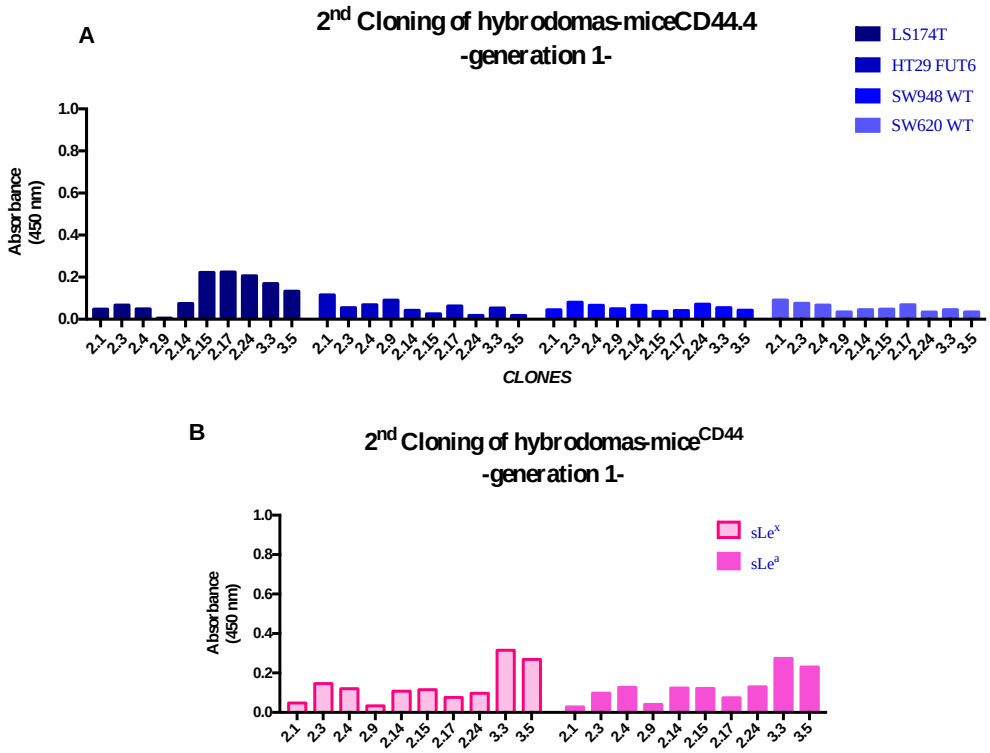


Figure 3.49 testing of the 10 more stable clones against membrane protein lysates and commercial sLe$^{x/a/}$. In A) cell membrane lysates LS174T (high expression of CD44 and sLe^{x}), HT29 FUT6 (high expression of CD44), SW948 (low in CD44 and moderate in E-selectin ligands) and SW620wt (not expressing CD44) have been tested. Clones 2.15, 2.17, 2.24, 3.3, 3.5 show higher signal on LS174T membrane lysate coating rather the almost absent binding with the other cell lines. In B) clones 2.3, 2.4, 2.14, 2.15, 3.3, 3.5 clones displayed binding with sLe^{x} and sLe^{a}.. ELISA experiment shown are representative of n=4 experiment. No significance was found with one way ANOVA test.

The results suggest that the supernatants of clones 2.15, 2.17, 2.24, 3.3 and 3.5 show higher signal against LS174T membrane lysate extracts (high expression of CD44 and sLe^{x}) rather the almost absent binding with the other cell lines. Signals towards signals HT29 FUT6 (high expression of CD44) and SW620 (not expressing CD44) was low, suggesting that clones 2.15, 2.17, 2.24, 3.3 and 3.5 do not recognize CD44 but probably sLe^{x} or other proteins.

On the other hand, screening of all 10 clones against commercial sLe$^{x/a/}$ Figure 3.497) suggest that clones 2.3, 2.4, 2.14, 2.15, 3.3 and 3.5 displayed binding with sLe^{x} and sLe^{a}.

Clones3.3 and 3.5, having in common binding with LS174T and sLe$^{x/a}$ antigens, were selected for further expansion and analysis. Clones 2.3, 2.4 (positive for sLe$^{x/a}$) and 2.15, 2.17 and 2.24 (positive for LS174T) were also selected for further validation.

After expansion, supernatants were tested against the same antigens and the recombinant form of CD44 (36 kDa), produced in human cell lines. However, after expansion most of the signals were lost exception made for clones 2.17, 2.24 and 3.3 for LS174T (Figure 3.50A) and 2.3 for sLe^{x} (Figure 3.50B). Moreover, almost no signal was detected from the clones against CD44.

Although the use of a minor isoform (instead of 82 kDa) does not guaranty the presence of the CD44 epitope against which the antibodies were produced, no clones gave inter-crossing positive signals. Moreover, signal intensity was overall low, suggesting little hybridomas stability.

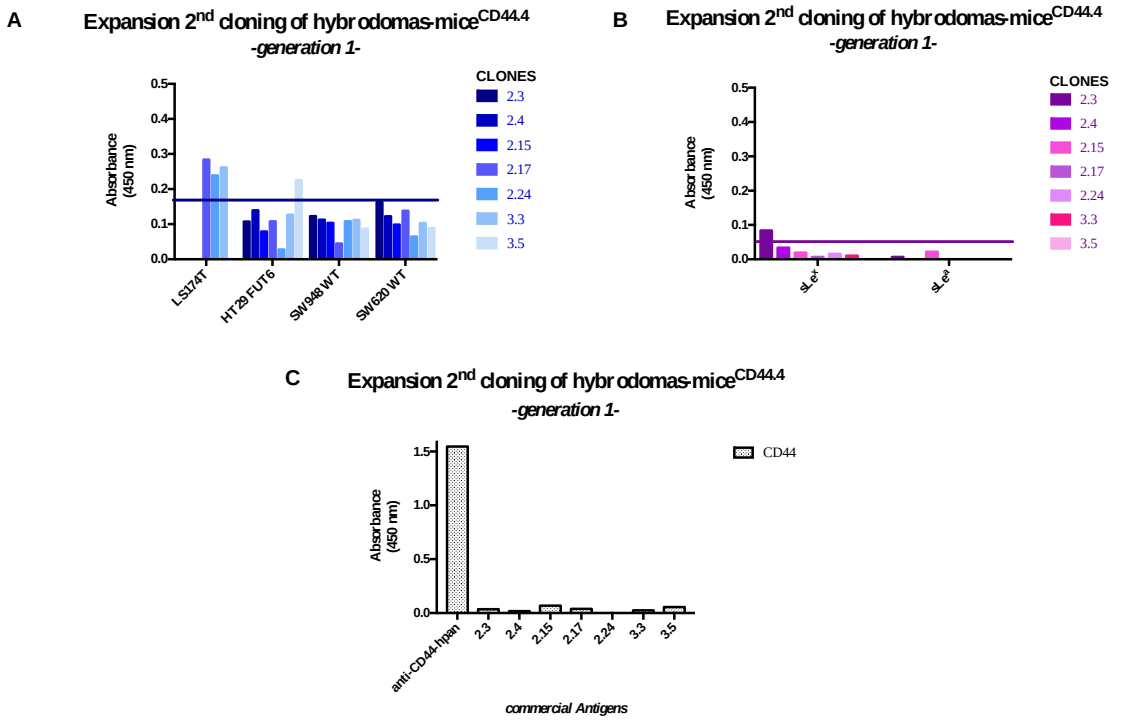


Figure 3.50 Expansion in vitro of the selected clones: 2.15, 2.17, 2.24, 3.3, 3.5 positive for LS174T and .3, 2.4, 2.14, 2.15, 3.3, 3.5 positive for sLe$^{x/a}$. In A) cell membrane lysates LS174T (high expression of CD44 and sLe^{x}), HT29 FUT6 (high expression of CD44), SW948 (low in CD44 and moderate in E-selectin ligands) and SW620wt (not expressing CD44) have been tested. clones 2.17, 2.24, 3.3 displayed specificity for LS174T. 3.5 show binding for HT29 FUT6. In B) only 2.3 clone show binding for sLe^{x}. In C) no clone show affinity for CD44 36kDa recombinant compared to the control used with anti-CD44-hpan. ELISA experiment shown are representative of n=2 experiment. No statistical test was performed

For this reason, we took the decision to proceed with a new mice immunization with CD44 extracted. From new immunization mice^{CD44}, a 2nd hybridoma generation was obtained.

Mice^{CD44} serum mice tested against the cell membrane lysates LS174T (high expression of CD44 and sLe$^{x/a}$), HT29 FUT6 (high expression of CD44), SW948 (low in CD44 and moderate in E-selectin ligands) and SW620wt (not expressing CD44) (Fig.3.40). Serum was also tested against commercial tetramers of sialyl Lewis^{x} (sLe^{x}) and sialyl Lewis^{a} (sLe^{a}) (Figure 3.51). Results showed high binding of the serum against LS174T, HT29 FUT6 and SW948 and a lower signal towards SW620 cell membrane lysates. Moreover, the serum resulted positive for

both sLe^x and $sLe^{a.}$ Low binding to all the cell membrane lysates were observed for control serum ($mice^{CTRL}$) (non-immunized) as expected.

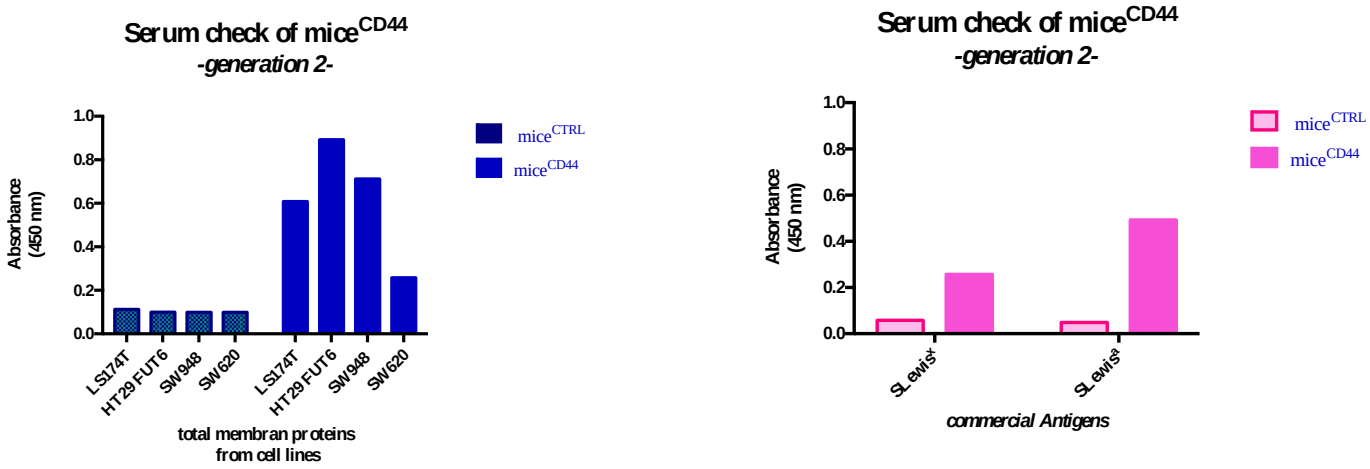


Figure 3.51 testing of the serum of $mice^{CD44}$-2 generation. On the left, cell membrane lysates LS174T (high expression of CD44 and sLe^x), HT29 FUT6 (high expression of CD44), SW948 (low in CD44 and moderate in E-selectin ligands) and SW620wt (not expressing CD44) have been tested. Proportional binding of the serum is matching with CD44 expression in each cell lines. Morover, negative control $mice^{CTRL}$ serum shows no binding. On the right, serumt show intense binding with sLe^x and higher for sLe^a. Negative control $mice^{CTRL}$ serum shows no binding ELISA experiment shown are representative of n=3 experiment. No statistical test was performed

Given the positive results obtained, a 2nd fusion was done and cloned in 20 different clones. Selection of positive clones was done using as antigens: commercial CD44 recombinant (36 kDa), membrane total lysate LS174T and commercial sLe^x and sLe^a (Figure 3.52). The serum from $mice^{CD44}$-2nd generation- was used as positive control. From the 20 clones, only clone 2C2 for CD44 and clone2B1 for LS174T gave a positive reaction (Figure 3.52A). No binding was detected for $sLe^{x/a}$ (Figure 3.52 B).

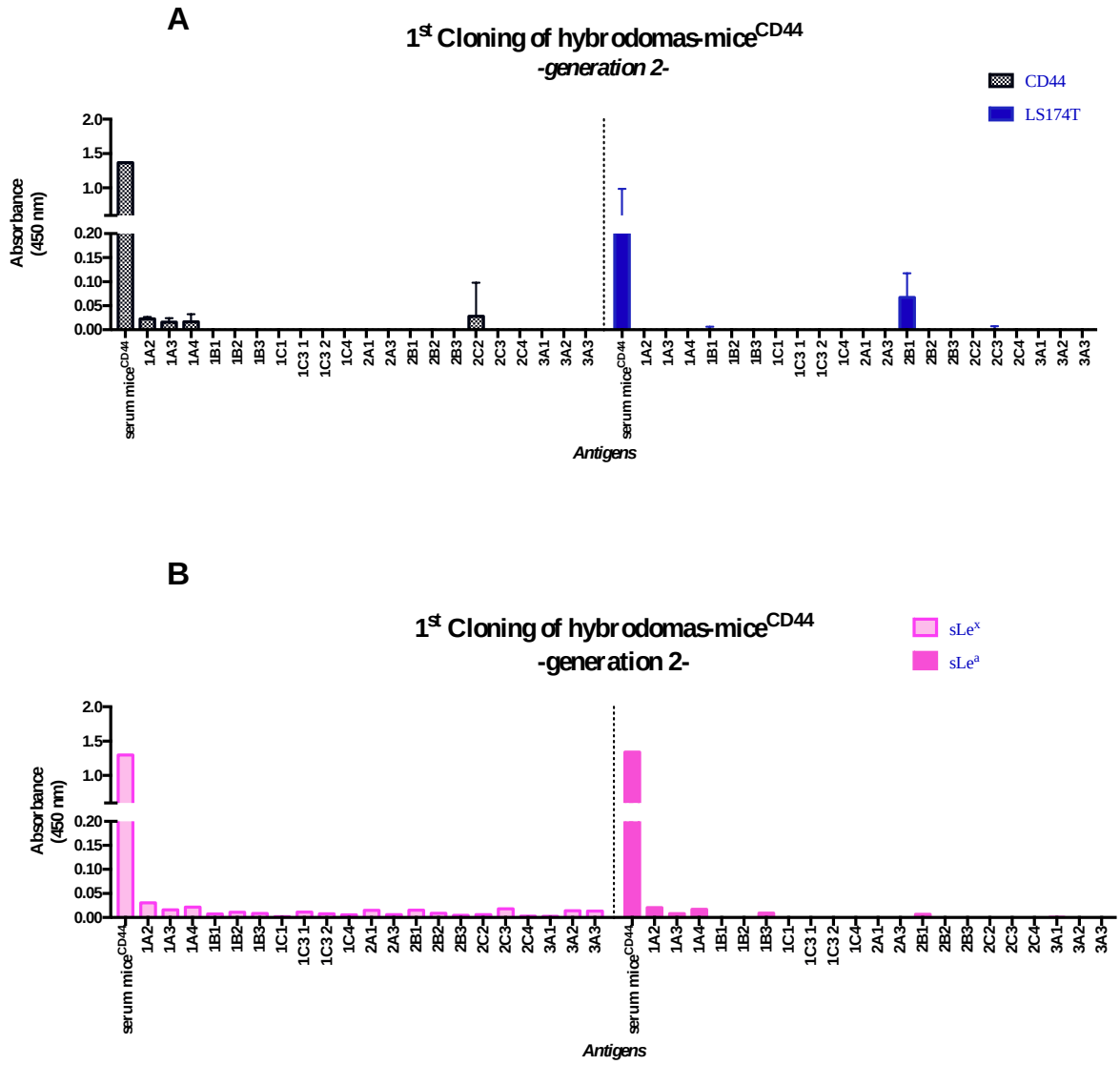


Figure 3.52 Testing of clones obtained from mice^{CD44} -generation 2- towards CD44 and sLe$^{x/a}$. In A) clone 2C2 show positive reaction for CD44 human recombinant (36 kDa) and clone 2B1 for LS174T (high in CD44 [82kDa in the no glycosylated form]). In B) almost no recognition of sLe$^{x/a}$ is showed. ELISA experiment shown are representative of n=2 experiment. No statistical test was performed

Although, the low signal obtained and the no recognition of sLe$^{x/a}$ antigens, clones 2C2 and 2B1 were selected for cells expansion and further testing against all the antigens used up to the moment: cell lysates membrane proteins (LS174T, HT29 FUT6, SW948, SW620), commercial sLe$^{x/a}$ and human recombinant CD44 (36kDa) (Figure 3.53)

ELISA shows interesting binding of 2B1 supernatant for LS174T (confirming results obtained in Figure 3.52), sLe^{x} and minor binding for sLe^{a} and CD44 recombinant. On the other hand,

clones 2C2, revealed minor binding for sLe^x and CD44 recombinant. Thus, clone 2B1 has been selected for further antibody characterization.

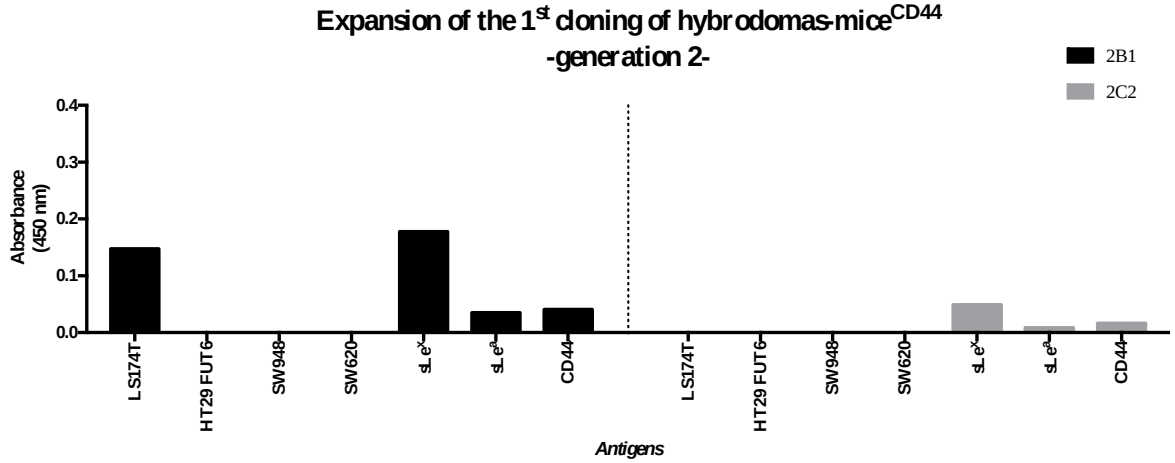


Figure 3.53 ELISA assay for 2B1 and 2C2 clones, against all the antigens previously tested. 2B1 shows recognition of LS174T, sLe$^{x/a}$ and CD44 recombinant (36kDa). Clone 2C2, show minor binding for sLe^x and CD44 recombinant. 2B1have been selected for further characterization. ELISA experiment shown are representative of unique experiment of triplicates. No statistical test was performed

3.4.2 Mice^{LS174T} clone's selection process

The second series of hybridomas produced with the injection of total membrane lysate of LS174T cell lines presented two main focuses. The first objective had the purpose to select hybridomas producing antibodies against sLe$^{x/a}$ antigens. For this reason, hybridomas were tested against LS174T (high expression of sLe$^{x/a}$), HT29 FUT6 (moderate expression of sLe$^{x/a}$), SW948 (low/moderate in E-selectin ligands) and SW620wt (moderate expression of sLe$^{x/a}$); supernatants were also tested directly against commercial tetramers of sialyl Lewis^x (sLe^x) and sialyl Lewis^a (sLe^a). Moreover, given the difference in sialic acid expression, clones were also tested against Mucin from bovine submaxillary glands (BSM), highly O-glycosylated and high in sialic acid, and mucin from porcine stomach (PSM), with a content of bound sialic acid between 0.5-1.5%. The combo BSM-PSM was used to validate specific recognition against sialic acid, that was very important as a first step selection of clones recognising sLe$^{x/a}$. The

second aim of this hybridoma series was to check for antibodies recognizing the two mains glycoprotein studied, CEA and CD44.

Firstly, the serum was tested against wells coated with LS174T (moderate in CEA, high in CD44 and E selectin ligand) and SW948 (high in CEA, low in CD44 and moderate in E-selectin ligands by ELISA), as total membrane lysate or live cells. HECA-452 was used as a positive control of $sLe^{x/a}$ on the antigen surfaces, while to confirm binding specificity $Mice^{CTRL}$ serum which had not been immunized, was also used (Figure 3.54)

Results show serum recognition of LS174T cell membrane lysate as well as live cell of LS174T and SW948, suggesting a successful immunization.

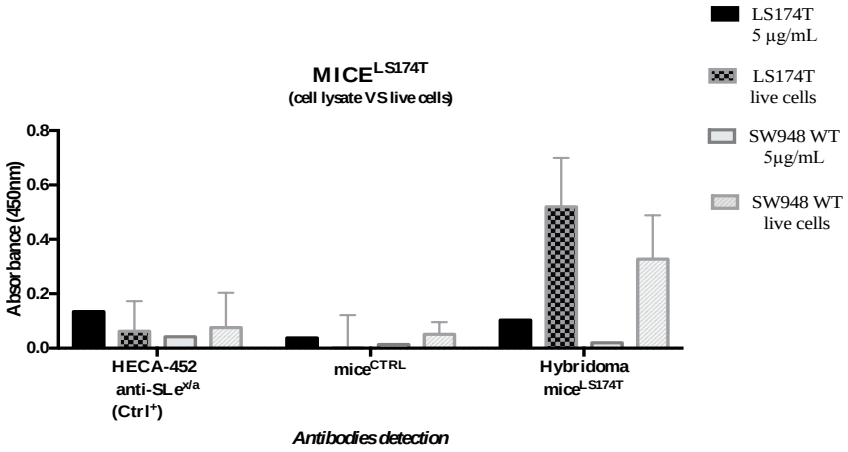
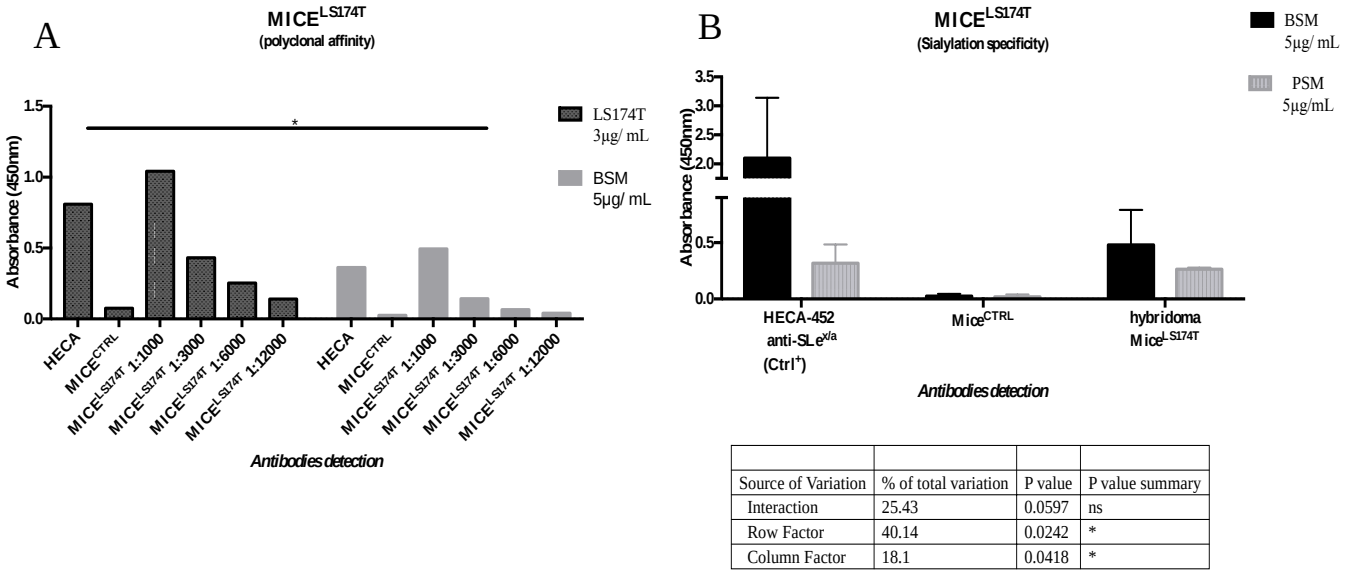


Figure 3.54 testing of serum obtained from mice^{LS174T}hybridoma fusion. The serum has been tested against LS174T (moderate in CEA, high in CD44 and E selectin ligand) and SW948 (high in CEA, low in CD44 and moderate in E-selectin ligands by ELISA). The wells were coated either with total membrane lysate (5 µg/mL) or with 90% confluency live cells. Membrane lysate were blocked 1% BSA; no blocking was performed on live cells. Results show serum recognition of LS174T cell membrane lysate as well as live cell of LS174T and SW948. ELISA experiment shown are representative of n=3 experiment. No significance was found with one way ANOVA test.

High binding to live cells compared to the HECA-452, $sLe^{x/a}$ positive control, suggested presence of no-specific antibodies in the serum. In order to prove the presence of specific antibodies against $sLe^{x/a}$ or sialic acid, serum was tested again in different dilution against LS174T and BSM (Figure 3.55A). As cross control, ELISA against BSM and PSM was also tested (Figure 3.55B). Staining HECA-452 is used as positive control and $Mice^{CTRL}$ as negative. As a result, the serum displayed proportional binding to both LS174T and BSM, suggesting the presence of specific antibodies against $sLe^{x/a}$ and O-glycans. Cross control, depicted highest binding for BSM rather than PSM, confirm production of antibodies specific for O-glycans and/or sialic acid.



Source of Variation	% of total variation	P value	P value summary
Interaction	25.43	0.0597	ns
Row Factor	40.14	0.0242	*
Column Factor	18.1	0.0418	*

Figure 3.55 The serum of mice^{LS174T} series testing affinity for specific glycosylation. HECA-452 is used as positive control and mice^{CTRL} as negative for sLe$^{x/a}$ and sialic acid. In A) both reactions show proportional reduction in signal intensity, suggesting the presence of specific antibodies against sLe$^{x/a}$ and/or sialic acid. In B) mice$^{LS174T\ serum}$ displayed highest intensity signal for BSM, confirming presence of antibodies specific for O-glycans and or/ sialic acid. ELISA experiment shown are representative of n=3 experiment. Significance was found performing two way ANOVA test

The supernatants of hybridoma clones obtained by fusion of this mice have been tested against all the antigens described at the beginning of this paragraph: LS174T, HT29 FUT6, SW948, SW620wt (total membrane cell lysates), commercial tetramers of sialyl Lewis^x (sLe^x) and sialyl Lewis^a (sLe^a), BSM, PSM CEA (Fig. 3.45). No CD44 recombinant was available at the time of the cells cloning.

The results showed that clone 23 presents the highest binding for LS174T, HT29 FUT6, SW948, moderate for SW620 (Figure 3.56A);; moreover, it showed highest affinity for BSM and PSM, with preference of binding for BSM(Figure 3.56C), and for CEA(Figure 3.56D).. On the other hand, clone 13 showed preference in binding for sLe$^{x/a}$ (Figure 3.56B). For this reason, clones 13 and 23 were chosen for cloning.

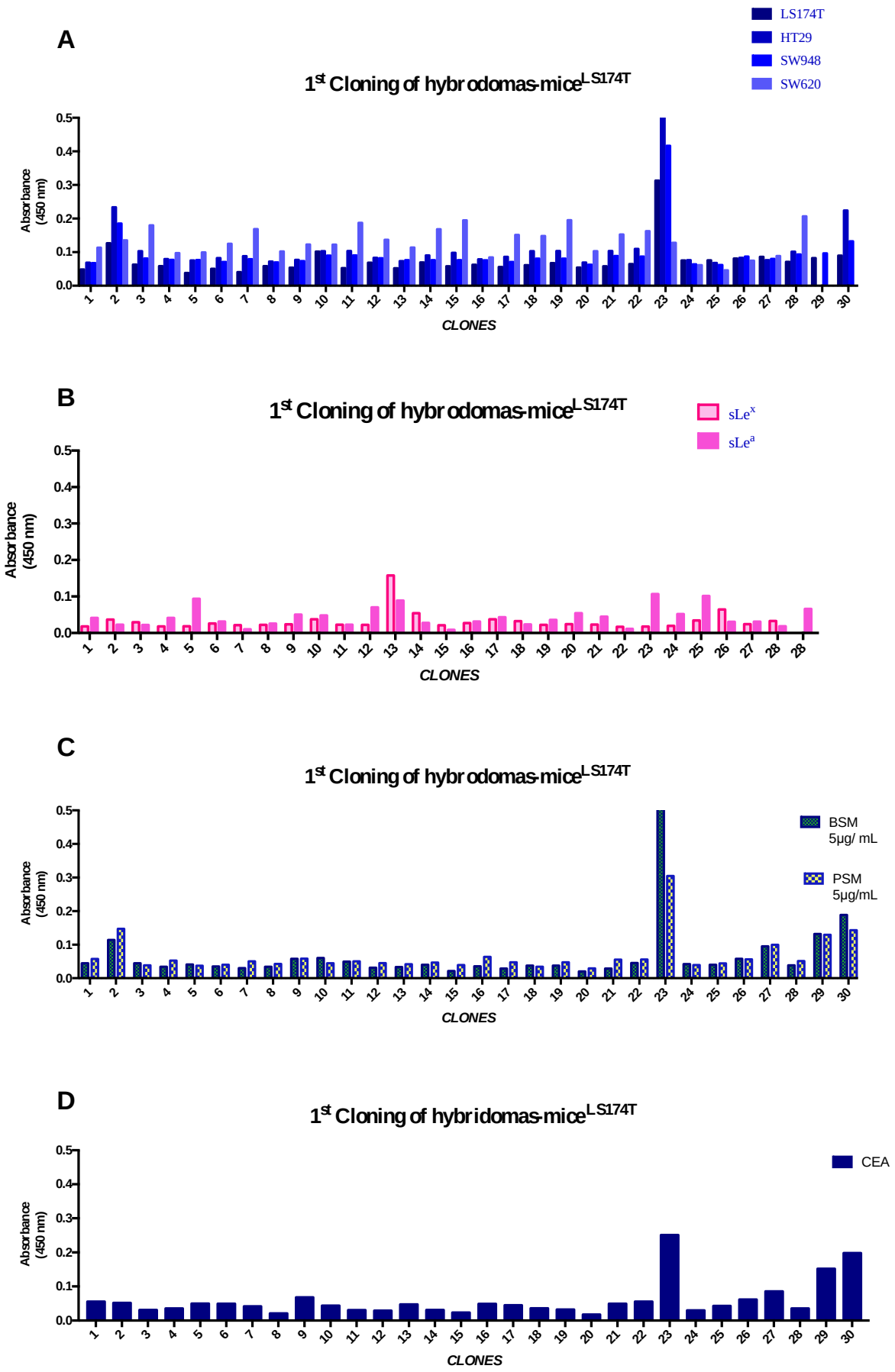


Figure 3.56 ELISA testing of the 30 clones obtained from the fusion of mice^{LS174T}. In A) clone 23 followed by weaker signals of clones 2, 29, 30 show high binding of LS174T (high expression of sLe$^{x/a}$), HT29 FUT6 (moderate expression of sLe$^{x/a}$), SW948 (low/moderate in E-selectin ligands) and low signals against SW620wt (moderate expression of sLe$^{x/a}$). In B) clone 13 depicted intense binding of sLe$^{x/a}$. weaker signals of clones 5,23,25, 28 show binding for sLe^{a}. In C) clone 23 show preferential binding for BSM rather than PSM. weaker binding is also displayed by clones 2, 29,30. In D) CEA binding with clone 23 is visible. weaker signals are also obtained with clones 29 and 30. ELISA experiment shown are representative of n=4 experiment. No statistical test was performed

Clones 13 and 23 were tested against all the aforementioned antigens, in order to pursue validation consistency: LS174T, HT29 FUT6, SW948, SW620wt (total membrane cell lysates), commercial tetramers of sialyl Lewis^{x} (sLe^{x}) and sialyl Lewis^{a} (sLe^{a}), BSM, PSM CEA (Figure 3.57). No CD44 recombinant was available at the time of the assay.

Supernatants of the clones 13 and 23, show a marked decreased in signals probably due to hybridoma instability may have occurred during the cloning and maintenance of the cells. However, two clones 23.2 and 23.11 tested against total membrane lysate depicted moderate binding against LS174T, HT29 FUT6 and SW948 (Figure 3.57A). All the other antigens showed almost no signal for sLe^{x} and sLe^{a} (Figure 3.57B) or BSM, PSM (Figure 3.57C), exception made for clone 13.29 which showed binding for CEA (Figure 3.57D). However, no correspondent signal is detected against sLe$^{x/a}$. Being detection of glycosylation of the antibody produced as the first purpose, we decide to proceed with a 3rd subcloning only of clones 23.2 and 23.11, in order to check hybridoma stability and to increase antibody production.

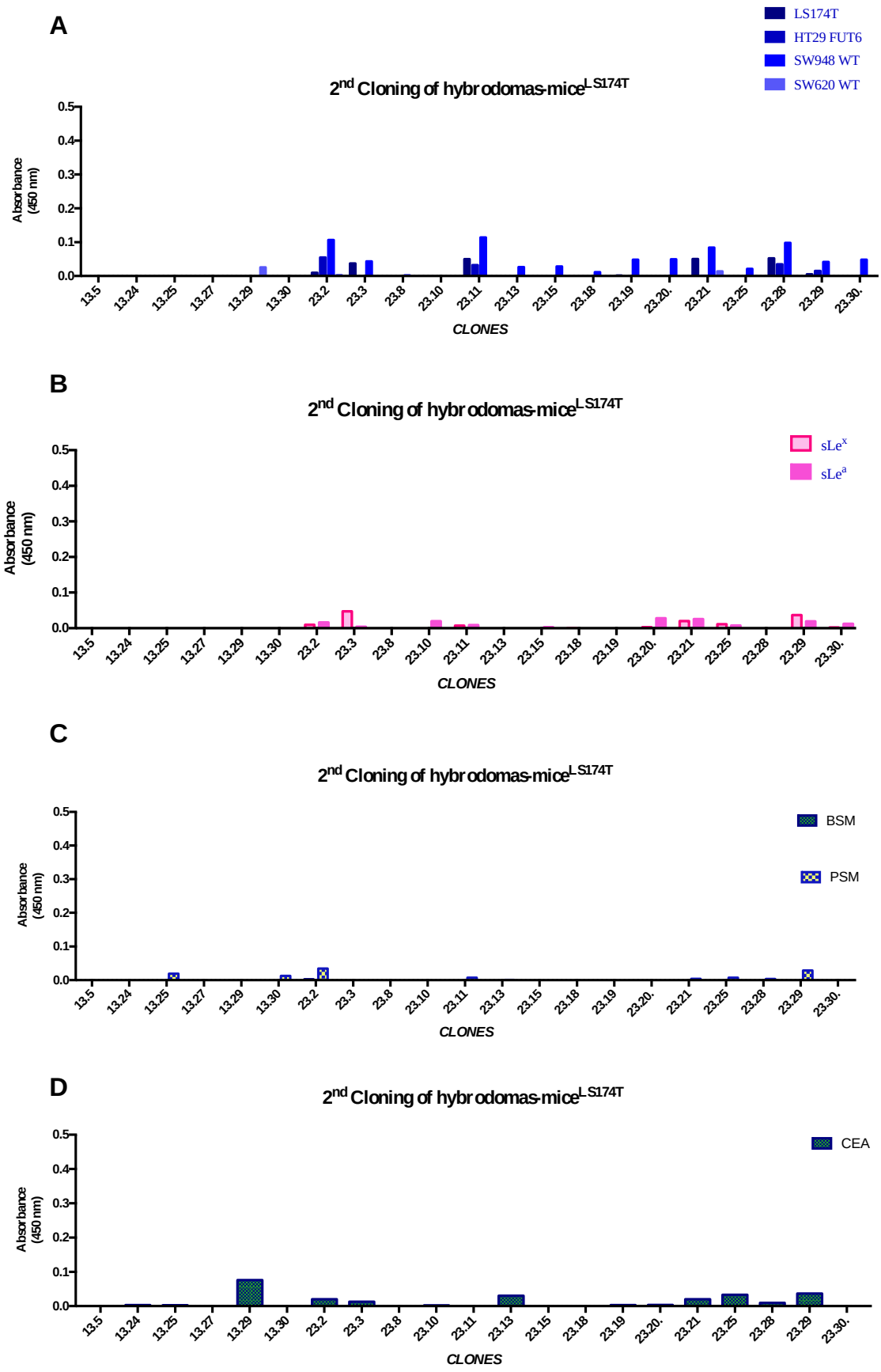


Figure 3.57 ELISA testing of the clones obtained from the 2nd cloning of clones 13 and 23 of mice^{LS174T}. In A) clone 23.2 and 23.11 depicted moderate binding against LS174T (high expression of sLe$^{x/a}$), HT29 FUT6 (moderate expression of sLe$^{x/a}$), SW948 (low/moderate in E-selectin ligands) and no signals against SW620wt (moderate expression of sLe$^{x/a}$). In B) clone 23. 3, 23.29, 23.30 13 depicted low binding of sLe$^{x/a}$. In C) almost no clone binding for BSM rather than PSM. In D) CEA binding with clone 13.29. ELISA experiment shown are representative of n=3 experiments. No statistical test was performed

The 2rd cloning of clones 23.2 and 23.11 was performed against all the aforementioned antigens, in order to pursue testing consistency: LS174T, HT29 FUT6, SW948, SW620wt (total membrane cell lysates), commercial tetramers of sialyl Lewis^{x} (sLe^{x}) and sialyl Lewis^{a} (sLe^{a}), BSM, PSM CEA and CD44 (Figure 3.58).

Subclones 2A4 and 3A1, showed the highest binding for all the antigens tested, suggesting the presence of antibody binding specific glycosylation. However, positive signal also for PSM, CEA and CD44 may suggest presence of polyclonal production in the supernatant. In order to better understand antibody binding, we have decided to test clones 2A4 and 3A1 on live cells, using flow cytometry technique (Figure 3.59). Clones were tested against HT29 wt (negative control), HT29 FUT6 and LS174T (Figure 3.59A). Results showed only binding of the clone 3A1. Moreover, the clone showed specific binding only on HT29 FUT6, although also in LS174T was expected.

Thus, clone have been now sent for sequencing and we are waiting for results. In fact, antibody sequencing and cloning, will be useful for further characterize potentiality of this novel antibody.

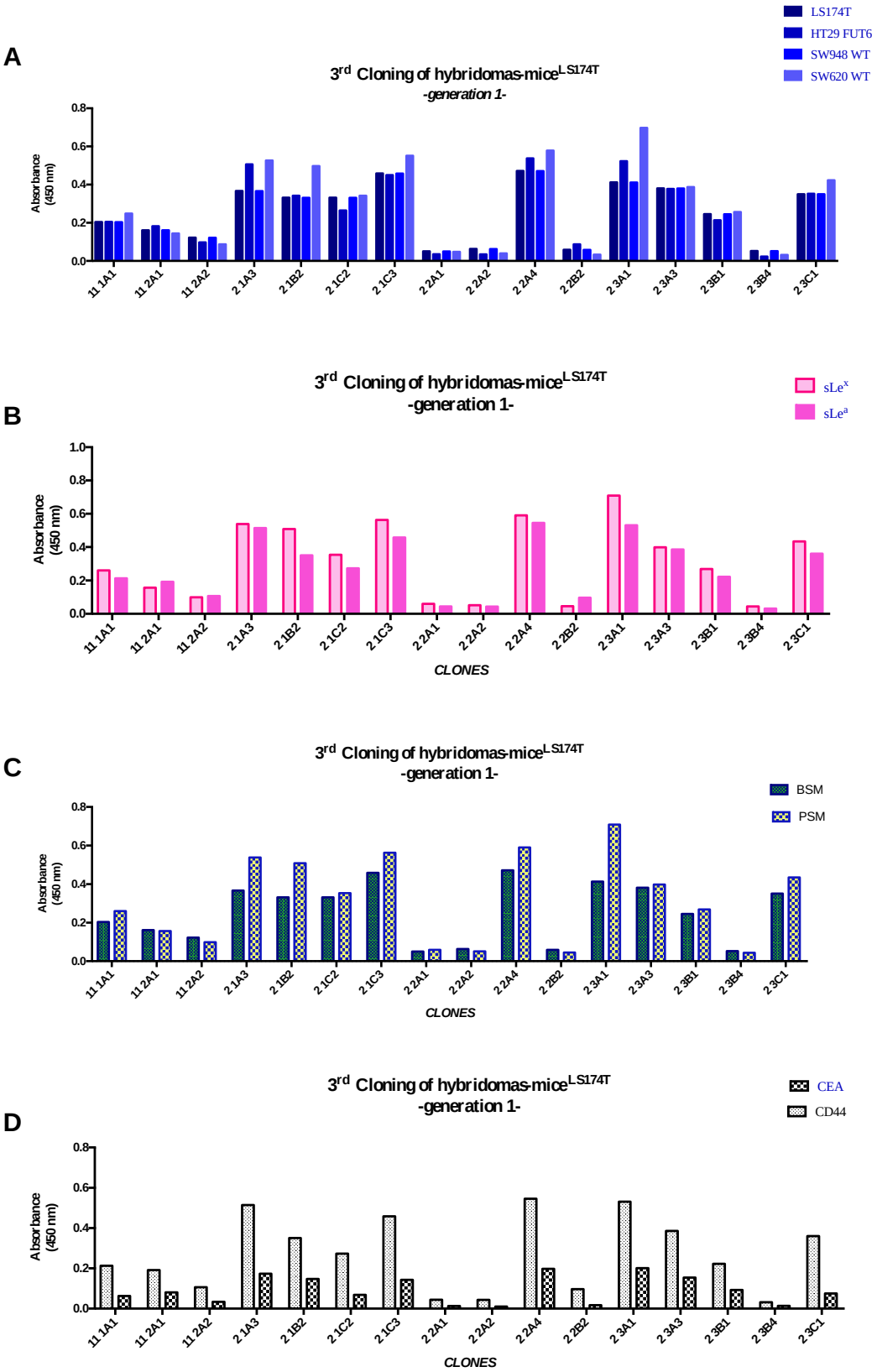


Figure 3.58ELISA testing of the clones obtained from the 3rd cloning of clones23.2 and 23.11 of miceLS174T. In A) clone 2/3A4 and 2/3A1 show highest binding against LS174T (high expression of sLex/a), HT29 FUT6 (moderate expression of sLex/a), SW948 (low/moderate in E-selectin ligands) and SW620wt (moderate expression of sLex/a). In B) 2/3A4 and 2/3A1 depicted the highest binding of sLex/a. In C) almost all clones show preference in binding of PSM rather than BSM. 2/3A4 and 2/3A1 depicted the highest binding In D) CD44 and CEA are recognized by most of clones, however 2/3A4 and 2/3A1 depicted the highest binding of CD44 and CEA. ELISA experiment shown are representative of n=3 experiments. No statistical test was performed

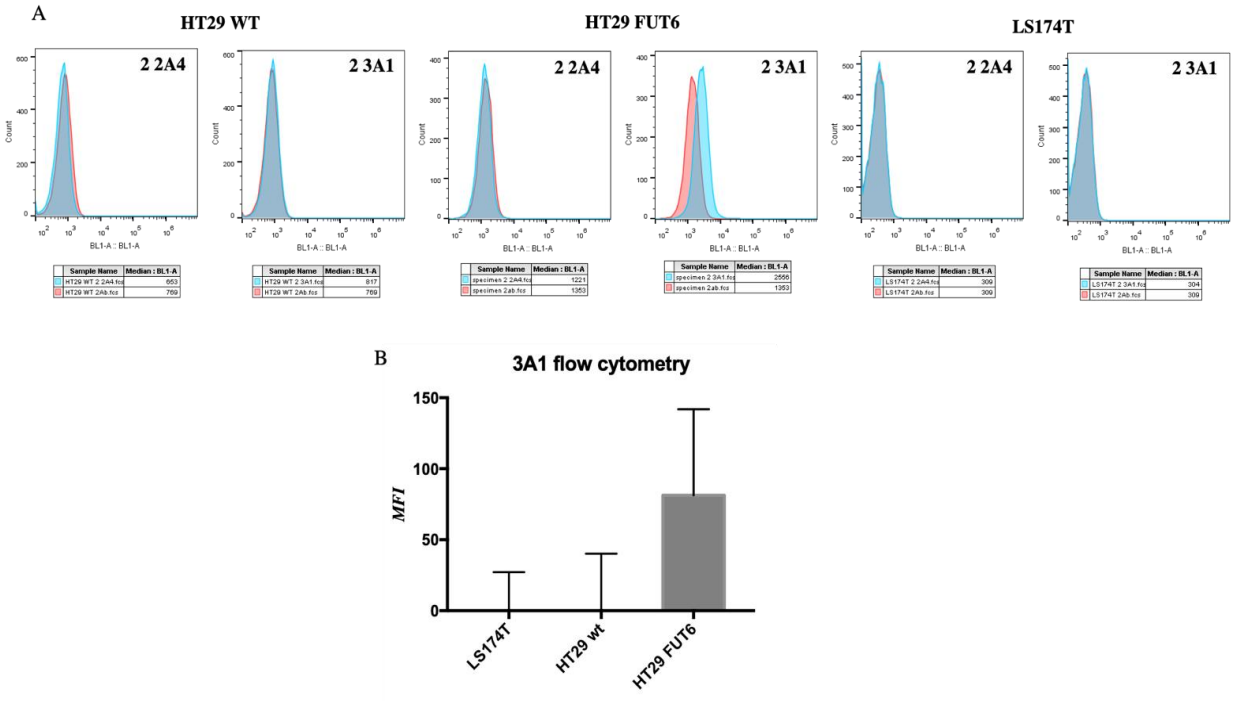


Figure 3.59 Flow cytometry staining of the depicted clones 23.2/ 2A4 and 23.2/ 3A1. In A) clones were tested against HT29 wt (negative control), HT29 FUT6 and LS174T. Flow cytometry intensity are represented. Only clone 3A1 show staining with HT29 FUT6. In B) histogram MFI intensity of clone 3A1, shows only binding for HT29 FUT6 cell line. The MFI obtained in flow cytometry experiments are representative of n=2 experiment; no statistical analysis was performed on the results.

3.5 PART V: side experiments

In this paragraph are listed all the preliminary side experiments performed during the PhD project development. Although incomplete works, the following experiments shows promising results that could be deepened in future projects.

3.5.1 Cell confluency influencing glycosylation expression

Reliability of glycosylation in cell culture is and has been a central focus. Although working with cell lines is fundamental to understand glycosylation behaviour and expression, it remains hard to completely rely on consistency of the results. In order to reduce the variables at the minimum, two cell lines expressing different glycans and glycoprotein were selected for further study on cell membrane expression. Being cell-to-cell communication an important influencer of cellular behaviour, LS174T (high in $sLe^{x/a}$ and CD44) and SW948 (moderately high in Le^{x} and high in CD44) were tested at 20%, 40%, 60%, 80% and 100% confluency. The cells were then harvested and analysed at the same time, in order to see if confluency could induce changes in glycan and/or glycoprotein expression. As results, in LS174T (Figure 3.60A) CD44 expression increases at high cellular confluency while $sLe^{x/a}$ expression with HECA-452, anti-CD15s and anti-CA19.9 diminish from 20% confluency and remains lower but stable from 60% to100% confluency.

Consequently, cells analysis and protein extractions have been performed following the results. For instance, if $sLewis^{x/a}$ antigen were to be analysed, cells would have been harvested at low cellular density, while for analysis of CD44 proteins were extracted at 80-100% confluency.

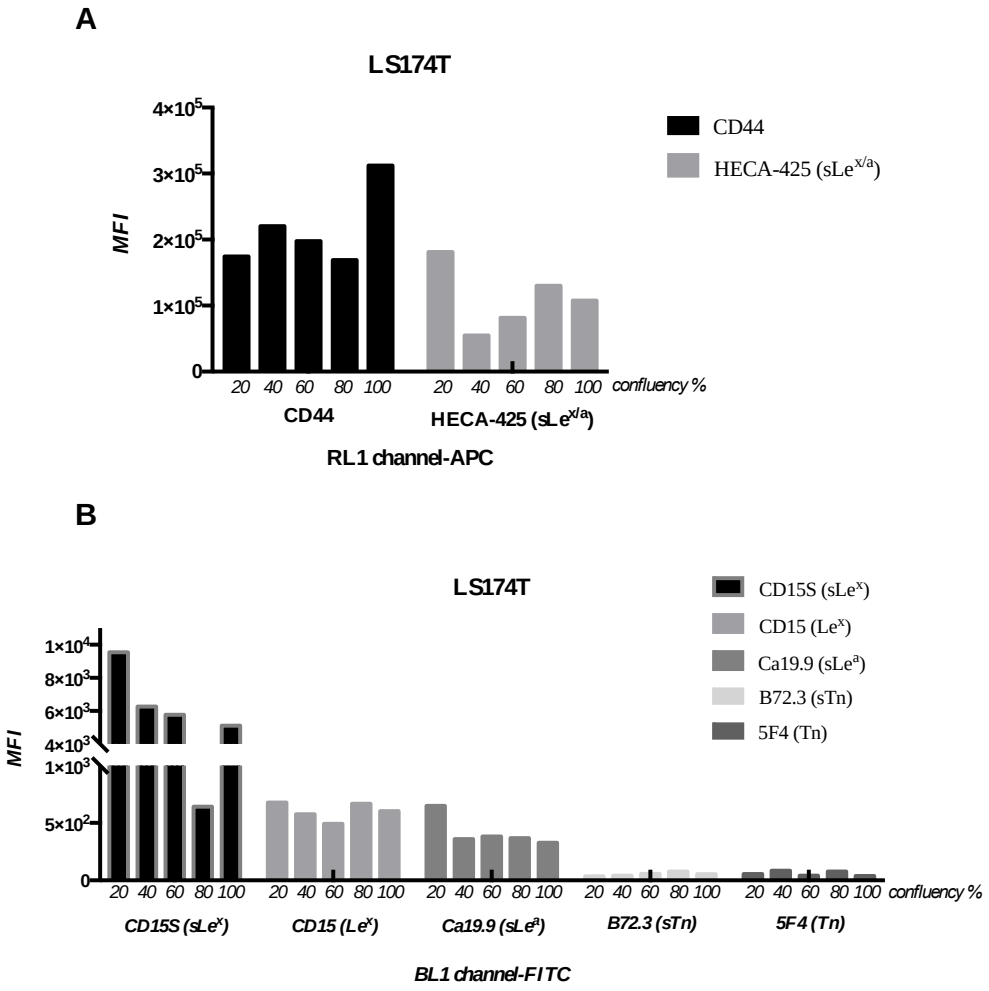


Figure 3.60 LS174T expression variability, confluency dependent. In A) comparing of CD44 and HECA-452 signals, showing a substantial increase in CD44 expression and decrease in sLe$^{x/a}$ antigens from 20%-100% confluency. In B) sLe^{x}, Le^{x}, sLe^{a}, sTn and Tn antigen were tested. separated analysis of sLe$^{x\ and\ a}$ confirm decreasing in expression from 20%, previously reported with HECA-452 antibody. no signal of sTn nor Tn has been observed. The MFI obtained in flow cytometry experiments are representative of n=2 experiment; no statistical analysis was performed on the results.

On the other hand, analysis of SW948 did not show changes neither in glycan nor CD44 expression, exception made for the antigen sTn which and Tn which starts their expression at 100% confluency (Figure 3.61).

Although, the confluence related expression experiment may not appear crucial for cell analysis, results suggest that it could serve as cellular behaviour standardization, very important to reduce variables and to work at the best cellular potential.

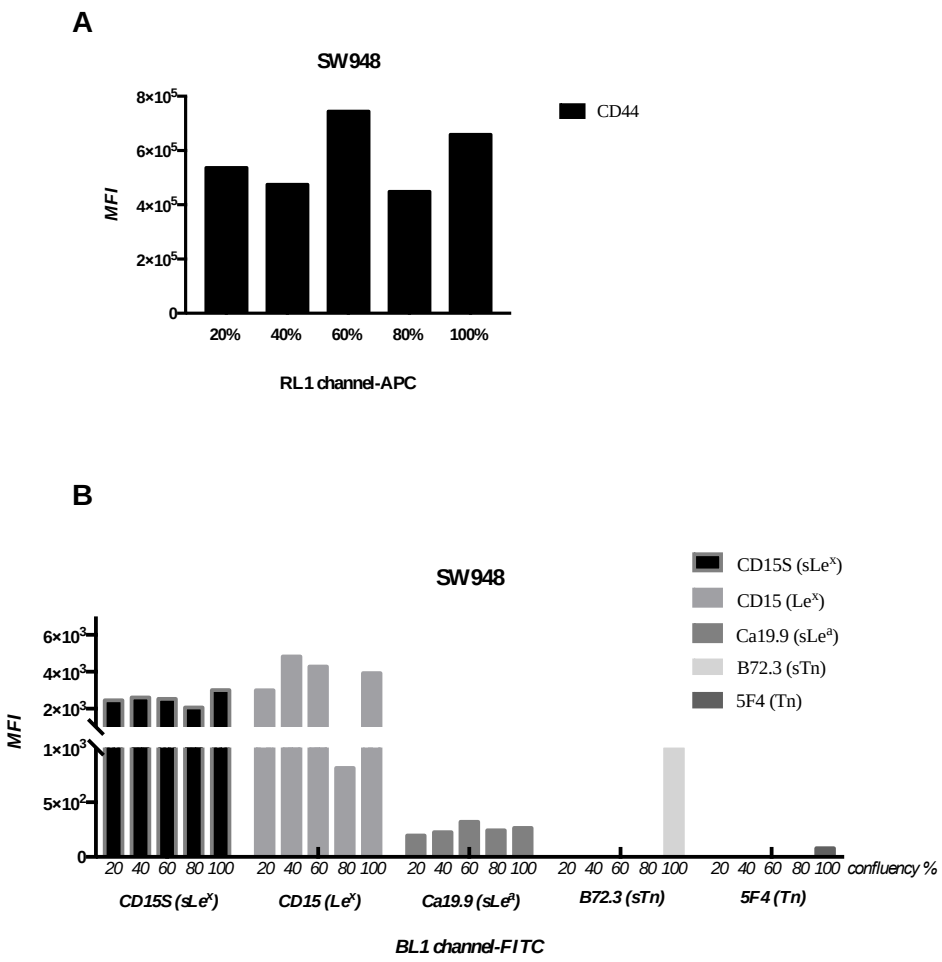


Figure 3.61 SW948 expression variability, confluency dependent. In A) comparing MFI at different confluency of CD44 do not show particular changes in expression. In B) sLe^x, Le^x, sLe^a, sTn and Tn antigen were tested. Le^a confirm higher stable expression, as in sLe$^{x/a}$. Signal of sTn nor Tn appear at 100% confluency. The MFI obtained in flow cytometry experiments are representative of n=2 experiment; no statistical analysis was performed on the results.

3.5.2 Colorectal cancer CMS

In the paragraph 1.3.4, it was explained how Guinney *et al*. (2015) regrouped data on CRC by forming four consensus molecular subtypes (CMS) and how Trinh et al. (2016) validated the prognostic value of CRC subtyping by identifying a robust IHM assay for different markers, that help to identify molecular CRC subtypes. Five markers (CDX2, FRMD6, HTR2B, ZEB1 and KER) were used. CSM1 was selected by the MSI status and the CSM2/3 by CDX2, a marker highly expressed in epithelial-like tumour. CSM4 were selected with HTR2B and FRMD6 expressed in mesenchymal like cells. To confirm the ETM transition, the marker ZEB1 and an anti-cytokeratin (AE1/AE3) was used to normalize for epithelial content. The

experiment permitted to confirm the stratification of CRC groups and verify the therapeutic benefit for each subtype[135].

For this reason, we decided to reproduce the experiment on the 40 patient's cohort CRC tissue, in order to pave the way for a possible association between each CSM and altered glycosylation (Table 3.3). However, no MSI of the cohort was traceable and in order to obtained significant results, more sample were needed. Nonetheless, preliminary results showed consistency with results obtained by Trinh *et al*. (2016) and we were able to show high tissue diversity used for IHC experiment. Prediction probability, showed in table 3.3, was calculated by the online mini-IHC-classifier based on the automated image analysis of Trinh *et al.* (2016) which uses a random forest classifier to determine patient subtypes. Comparison between the automated and (online) semiquantitative classifiers showed strong concordance, highlighting that the simplified classifier can assign subtype classes in the absence of continuous quantitative information. Further IHC testing on CRC tissue is highly recommended.

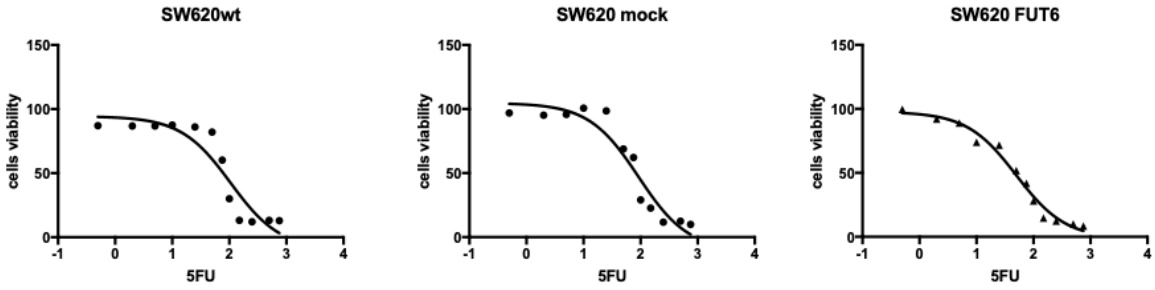
Table 3.3 Representation of the subtypes and prediction probability of CSM for the cohort of 40 patients' tissue IHC, following online mini-IHC-classifier based on the automated image analysis of Trinh et al. (2016)

CASES	SUBTYPE	PREDICTION PROBABILITY	MICROSATELLITE INSTABILITY
2042/12 A5	Mesenchymal	74.0%	Without information
2064/07 A3	Epithelial	86.4%	Without information
3813/11 C5	Mesenchymal	50.6%	Negative
5183/11 A6	Mesenchymal	61.8%	Without information
3716/11 A1	Epithelial	87.4%	Without information
3041/12 B5	Mesenchymal	59.0%	Without information
6381/08 3	Epithelial	87.4%	Without information
3038/12 A4	Mesenchymal	61.4%	Without information
5894/06 B2	Mesenchymal	55.8%	Without information
4199/12 C5	Epithelial	82.8%	Without information
4900/08 B5	Epithelial	94.8%	Without information
6963/12 B5	Epithelial	89.6%	Negative
4139/09 3	Mesenchymal	67.6%	Without information
1224/12 A5	Epithelial	87.6%	Negative
7104/10 A4	Epithelial	87.0%	Without information
6474/12 D6	Epithelial	99.4%	Negative
4590/09 C3	Epithelial	87.2%	Without information
7253/12 B5	Epithelial	87.4%	Without information
11621/11 A5	Epithelial	87.4%	Without information
2461/12 A4	Epithelial	63.6%	Without information
3410/11 A7	Mesenchymal	64.8%	Negative

11024/09 A2	Epithelial	86.4%	Without information
10552/10 8	Epithelial	82.2%	Negative
11260/11 A4	Mesenchymal	60.8%	Without information
5317/11 C3	Mesenchymal	65.4%	Without information
11166/09 3	Epithelial	86.4%	Without information
5581/09 B4	Epithelial	98.0%	Negative
9201/06 1	Epithelial	99.8%	Without information
10160/12 A3	Mesenchymal	57.0%	Without information
10086/12 A3	Mesenchymal	60.0%	Without information
11987/10 A2	Epithelial	99.2%	Without information
6016/12 A7	Epithelial	87.4%	Without information
9638/07 3	Epithelial	85.8%	Without information
11315/08 A2	Epithelial	74.4%	Without information
4968/06 1	Mesenchymal	58.8%	Without information
9989/10 B3	Epithelial	83.2%	Negative
7030/08 1	Epithelial	63.6%	Without information
2500/12 4	Epithelial	74.4%	Without information
1033/11 A6	Epithelial	84.6%	Without information
2344/05 C3	Epithelial	95.4%	Without information

3.5.3 5FU IC50 in highly glycosylated cells

Biomarker research have been applied on CRC cell lines and CRC tissue (cohort of 40 patients) All the CRC tissue originates from patients treated with chemotherapeutic and in particular all treated by 5fluorouracile (5FU). Thus, we decided that it would have been interesting to create continuity on the cell lines available, by testing the half maximal inhibitory concentration (IC50) on CRC cell lines; moreover, we wanted to see if glycosylation may influence on drug absorption and thus drug efficiency. Preliminary results were obtained by testing SW620 cell line (wt, mock and FUT6) with different concentration of 5FU. Cell viability was then measured by MTT reaction (Figure 3.62). Results show that SW620 FUT6 (thus higher expressing sLe^x) has half maximal inhibitory concentration (IC50) ~50% lower when compared to the wt and mock (low in sLe^x). This suggest that presence of sLe^x glycans can increase absorption and efficacy of 5FU and thus, reinforcing the idea that study of aberrant glycosylation can be useful as prognostic and predictive marker.



SW620	wt	mock	fut6
IC50	105.3	91.89	50.01

Figure 3.62 IC50 of 5FU concentrations on SW620 wt, mock and FUT6 Results show that SW620 FUT6 (thus higher expressing sLe^x) has half maximal inhibitory concentration (IC50) of ~50% less when compared to the wt and mock (low in sLe^x). The cytotoxic assay is representative of n=3 experiments.

3.5.4 Luciferase assay

The final goal of antibody production was to test in vivo antibody efficacy by creating xenografts of human cell expressing the luciferase gene in mice. The luciferase expression is fundamental for cell tracking and antibody efficacy quantification. Two cell lines have been chosen. HT29 (mock and FUT6) and SW948 (mock and ST6GAL1) were the mock were used as negatives and the transfected/transduced have higher expression of respectively, sLe^x and α-2,6-sialic acid. Lentiviral transduction with Luciferase gene was performed on cell lines and cells were tested for Luciferase expression through bioluminescence emission after cell lysis and substrates addition. The lysates were tested in *Chemidoc* with a 50 seconds exposition period in a 96 well plate where the previous cells.

However, as aforementioned, stability of glycosylation *in vitro* can be variable and subject to changes, especially after further transduction with Luciferase gene. For this reason, antigen screening has been performed of both cell lines with main antigens CD44 and sLe$^{x/a}$.

Cells were harvested at same confluency rates; however, results show differences in antigens expression before and after luciferase lentiviral transduction (Fig. 3.52).

In detail, SW948 mock and SW948 mock luciferase, do not show any changes in antigens expression, while the ST6GAL1 shows block of expression for both antigens.

On the other hand, HT29 mock luciferase when compared to mock, show a general decrease in both antigens. No remarkable changes are instead noticed between HT29 FUT6 and HT29 FUT6 luciferase. In general, change in antigen expression is noticed, and while HT29 FUT6 cells show almost no difference, SW948 ST6GAL1 must be newly transduced with the luciferase gen.

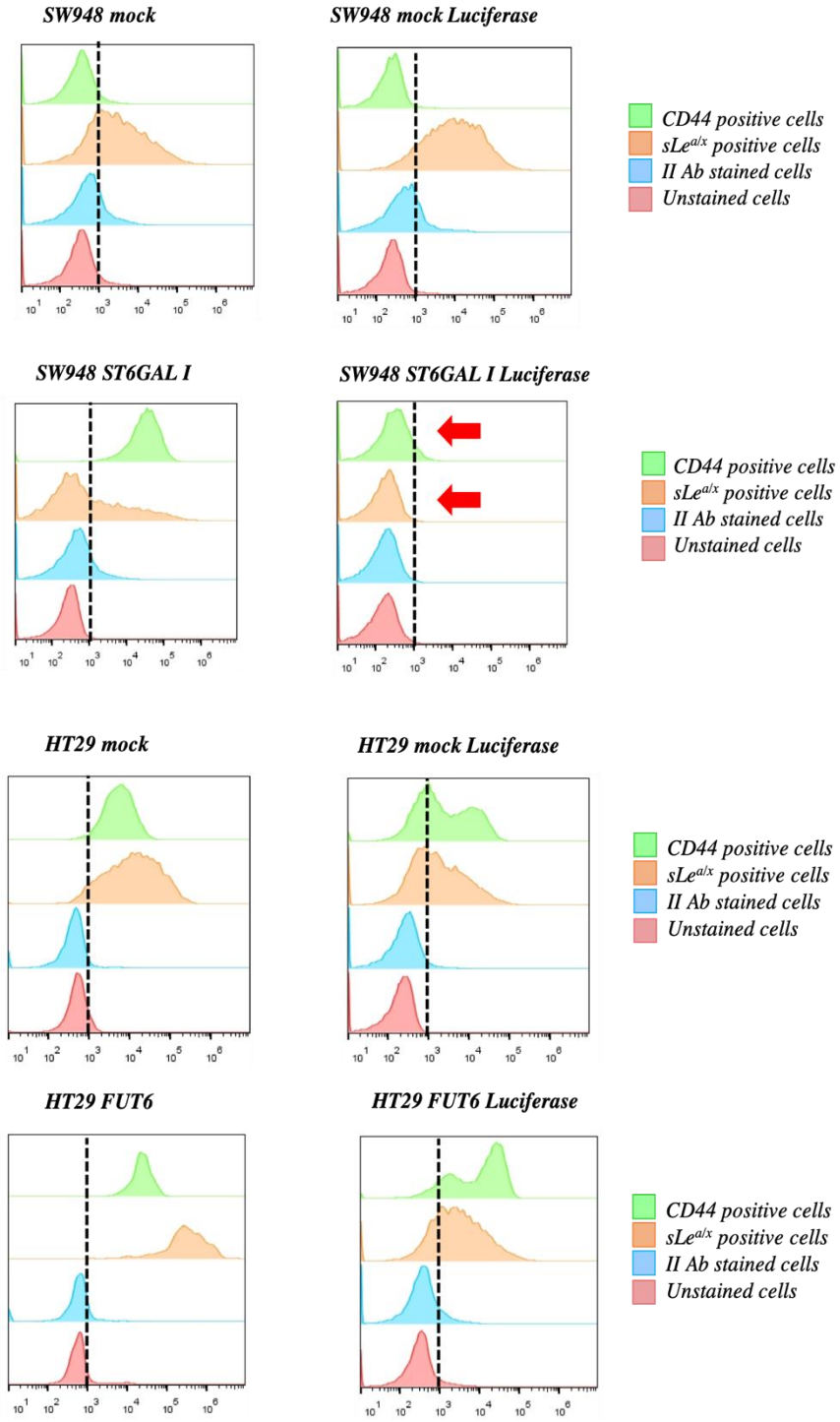


Figure 3.63 screening of cell lines SW948 mock and ST6GAL1 and HT29 mock and FUT6, before and after luciferase transduction. The MFI obtained in flow cytometry experiments are representative of n=2 experiment; no statistical analysis was performed on the results.

DISCUSSION & CONCLUSION

4.1 Biomarker research and material selection

The first part of the project aimed to the identification of the most promising biomarker for colorectal cancer detection, among aberrant glycans and overexpressed glycoproteins. The determination of the possible new biomarker was necessary to produce novel monoclonal antibodies. Thus, we were looking for biomarkers with specific characteristics. Initially, the biomarker should detect most of the stages of cancer, with preference for CRC early stages detection. Indeed, as introduced in chapter 1.3.1, tracing cancer at first stages can raise 5 years survival probability up to 90%[145].

Secondly, a good biomarker should be identified with no-invasive techniques. Blood sampling is considered a minimal invasive technique, yet it is generally considered less invasive than colonoscopy and tissue specimen collection[128]. Thus, a biomarker detectable in patients' serum would be of easier access both for patient and clinician.

In addition, a good biomarker should be cancer specific. Alterated glycosylation have been largely proven to be present specifically in cancer tissue; for instance sTn has been found overly expressed in bladder cancer and on the mucin 1 type (MUC1) surface[211,296]. More studies, on aberrant glycosylation on CRC cancer, revealed the overall presence of aberrantly expressed sialyl Lewis (sLe) antigens, throughout all the stages of tumour formation.

Lastly, the biomarker should be stable, in order to reduce false positive or negative during serum testing. Proteins, among all the biomarker in use have been shown to be highly stable and to function as good diagnostic and prognostic biomarkers[236,249,253]. For instance, carcinoembryonic antigen (CEA) is in use as a prognostic biomarker after surgery resection, where scoring is expecting to lower in case of a successful operation[138,139]. On top of that, glycosylation covers the surface of glycoproteins and it is possible to trace aberrant glycosylation presence by studying the different carrier,

For all this reason, we have decided to screen for two types of antigens: aberrant glycans and glycoproteins, already proven to be overexpressed in cancer or in colorectal cancer.

To increase the translational power of this project, we have chosen to screen both for cell lines and patients' tissue specimens. The use of cell lines in translational project has been very discussed, as very useful for experimental set up and protocols optimizations, but unsure in human specimen reproducibility. To aggravate, as briefly mentioned in paragraph 3.5.1, glycosylation pathways may be lost or altered in cell lines, as a consequence of the immortalization process and cell lines passages in culture.

However, the use of cell lines remains crucial in the first steps of the project and the possibility to transduce/transfect cell lines with glycoenzymes, helps defining roles and function of aberrant glycosylation in cancer development.

For project consistency, we have chosen cell lines harvested form colorectal patients in different stages and grades of the cancer (Table 3.1). The cell lines were tested for differences in doubling times, with first stages having higher doubling time of the last stages, as expected (Fig. 3.2). Alteration in cell growth, depending on differences in glycosylation expression, was also studied: we tested enzyme FUT6 expression, translated in HT29 and SW620 cell lines and enzyme ST6GAL1 expression, transfected in SW948 and SW48 cell lines. In detail, FUT6 is responsible of fucosylation of sLe^x antigens and ST6GAL1 for increase of $\alpha-2,6-$ sialic acid. Contrasting results show that modification in glycosylation may influence cells growth, yet it has not been possible to define how doubling time is dependent from increase sLe^x nor sialic acid expression (Fig. 3.1). Nonetheless, we have decided to pursue with the screening for antigen expression of all cell lines wt, mock and transfected/transduced, in order to study similarities and differences in expression.

4.2 Colorectal cancer cells lines overexpressing CD44 and $sLe^{x/a}$

Different immunoassays were planned in order to screen for antigen expression and to test commercial antibody specificity, by using different antibodies clones against the same antigens (Table 3.2).

Firstly, live cell screening was performed by flow cytometry technique and as a results two major antigens have shown major expression: $sLe^{x/a}$ and the glycoprotein CD44. In order to validate the results obtained, screening has been performed not only on CRC cell lines, but also in normal colon CCD 841 Con. $sLe^{x/a}$ showed the maximum expression in LS174T, stage 2 and grade I, followed by HT29 and SW620, respectively, stage 3 grade I and stage 3 grade IV. The antigen Le^x have been also taken under consideration, given the expression in cell line SW948, stage 2 grade III.

CD44 was highly expressed, revealing a progressive lost in expression from normal colon down to grade IV in CRC lines. The CD44 antigen is a cell-surface glycoprotein involved in cell–cell interactions, cell adhesion and migration thus expression in normal colon was expected. However, studies of alteration in isoform expression and/or glycosylation suggests CD44 as suitable biomarker candidate. In fact, further analysis on HT29 FUT6 cell lines, did

not only show an increase of sLe^x, but also an important increase in CD44 expression, suggesting a correlation between the antigens (Fig. 3.6). On the other hand, unanticipated results revealed lower expression of the CEA antigen only in SW948. In fact, CEA is used a prognostic factor for CRC in clinics and it was expected to show an overall higher expression in cancer cell lines[297].

In the second part of biomarker discovery, we aimed to study the stability of the proteins and glycans after extraction from the membrane. Western blot technique was applied in order to validate results previously obtained by flow cytometry, to characterized molecular weight expression of the antigens and to analyse antibody sensitivity on antigen detection. $SLe^{x/a}$, Le^x, CEA and CD44 were selected (see paragraph 3.1.4). In general, blots run with the total membrane lysate of all cells in analyses, revealed high expression in $sLe^{x/a}$ and CD44 antigens. Moreover, increase of CD44 correlated to increase of sLe^x in HT29 FUT6 cell was also noticed. Curiously, blot revelation with HECA-452 ($sLe^{x/a}$) and anti-CD44hpan depicted increase of signal in both staining of a band at 135 kDa in HT29 FUT6. CEA isoforms were mainly detected at low molecular weight, suggesting low presence of post-translational modifications while Le^x revealed increase in signals, compared to live cells analyses. However, because of the difficulty in obtaining specific molecular weight for Le^x and complete absence of the antigen in SW620 cell lines, we decided to not consider further Le^x for future experiments.

Thus studies obtained on cell lines, highly suggested to consider $sLe^{x/a}$ and CD44 glycoprotein as potential future biomarker, supporting others literature results[153,255,298].

4.3 CD44 and $sLe^{x/a}$ involved in ganglion metastasis

Validation of the translational potential of the results obtained required further analysis on human specimen. Therefore, we have decided to screen human specimens by immunohistochemistry (IHC) (paragraph 3.1.5). A cohort of 40 CRC patients was selected to ensure significant results of the screening. Analyses showed an overall presence of aberrant glycosylation on tissues and high intense staining of the marker CD44 (Table 3.3).

$SLe^{x/a}$ and CD44 were also stained on colon cell mucus, suggesting its potential as no-invasive detection biomarker (Fig. 3.11) Moreover, further analyses depicted high correlation among the antigen's expression, suggesting that in tumoral environments, glycotransferase disfunction may influence other expression pathways. In general, CD44 and Le^y are found more expressed in cancer stages III/IV and sLe^x and Le^y highly expressed in tissue with ganglionic metastasis

(Fig. 3.13/3.14). Values were confirmed by crosstabs obtained by CD15s and HECA-452 antibodies scoring, that revealed higher expression of the antigens sLe^x (and maybe sLe^a) in ganglion metastasis (Fig.3.15). Nonetheless, overall survival (OS) analysis revealed that presence of Le^x is associated with an increase of the OS. This support the idea that the level of sialylation may play a central role in cancer invasion and metastasis[153]. Consequently, we managed to validate the results obtained in cell culture screening and to choose $sLe^{x/a}$ as new potential target.

In tissue analyses CEA shows higher expression than the results obtained in cell lines analyses (Table 3.3). However, CD44 not only show higher intensity signals, but also correlation with staging progression and tissue metastasis. Thus, we have confirmed CD44 as glycoprotein candidate biomarker.

Lastly, in order to understand the role of sLe^x and CD44 in cell migration and to validate the potential of antibodies against these antigens, we have performed the scratch wound assay. Thoroughly, cells were cultivated in inserts at 90% confluence or higher and after insert removal, we kept the cells in gentle motion in presence of anti-CD15s (sLe^x) and anti-CD44-Hpan (Fig. 3.18). To study cell migration, we compared the closure timing of the gap. Results shows a general delay in closure in presence of anti-CD15s (sLe^x) of the three cell lines selected [LS174T (high in $sLe^{x/a}$ and CD44), SW48 (medium-low in $sLe^{x/a}$ and CD44), SW620 (medium-low in $sLe^{x/a}$ and null in CD44)]. The results confirm that $sLe^{x/a}$ play a role in cell migration. Moreover, LS174T, maintained in culture with anti-CD44-hpan, display almost no cell migration and SW48 show delay in closure, while SW620 in presence of anti-CD44-hpan depicted no differences in migration of the control sample. The results not only confirmed the probable role of sLe^x and CD44 in cell invasion, but the effect of the antibody revealed a potential beneficial use of antibodies in preventing migration.

4.2 CD44 as E-selectin ligand

CD44 and $sLe^{x/a}$ have been selected as possible candidates for CRC biomarkers. Moreover, both antigens are considered E-selectin ligand. E-selectin, also known as CD62 antigen-like family member E (CD62E), is a selectin cell adhesion molecule expressed only on endothelial cells activated by cytokines[69]. E-selectin studies have showed that E-selectin binding to colon cancer cells correlates with increasing metastatic potential, and that cancer cells bind E-selectin using glycoprotein or glycolipid ligands normally expressed on immune cells[69]. Therefore, we

decided to further characterize CD44 glycosylation, by extraction from LS174T and HT29 FUT6 cell lines. In fact, LS174T presented high expression of high molecular weight of CD44 (>135 kDa), while HT29 FUT6 showed a parallel increase of expression of sLe^x and CD44 antigen, suggesting correlation between the two antigens. Moreover, as presented by Azevedo et al. (2018)[256], CD44 presents different isoforms and on the spliced sites are presented many O- and N-glycosylation sites. This may support the theory, that a less un-spliced form of CD44 presents a higher post-translational modification and thus, in cancer cell, could express higher aberrant glycosylation on its surface[256]. The highest isoform back bone not glycosylated correspond to 82 kDa that can be found glycosylated up to 250 kDa. In order to identify the backbone of CD44 expressed by the two cell lines, we immunoprecipitated the sample with E-selectin-Ig chimera and sent to analyse through liquid chromatography mass spectrometry (LC MS/MS). Four separated samples (duplicated for each cell lines) confirmed the immunoprecipitation of CD44 82 kDa as E-selectin ligand (Fig. 3.23). Samples were also cross analysed in western blot by precipitating E-selectin ligands and CD44 (with anti-CD44-hpan) of HT29 FUT6. Both signals were revealed at 135 kDa and resulted positive for E-selectin-Ig in blot staining, confirming the presence of $sLe^{x/a}$ on CD44 immunoprecipitated.

At this point, it was important to understand whether CD44 was preferentially glycosylated by sLe^x or/and sLe^a. Thus, foster resonance energy transfer (FRET) of CD44/sLe^x and CD44/sLe^a was applied. Energy transfers happens when a donor molecule is at a max distance of 9nm from an acceptor molecule and the donor-emission-wavelength is overlapping with the excitation-wavelength of the acceptor. For this reason, the choice of the proper fluorophores was critical. FRET was performed by flow cytometer on LS174T cell lines. In case of positive FRET was expected median fluorescence intensity (MFI) increase in the acceptor channel. Using different fluorophore combinations, we revealed positive fret for both antigens, although sLe^x seem to have higher expression when compare to sLe^a (Fig. 3.26). Nonetheless, both sLe^x and sLe^a can be present as O- and N- glycosylation, and in order to understand the type of glycosylation, we have used two enzymes to cleavage N-glycans or sialic acid, respectively with PNGase F and neuramidase. The reaction with PNGase F depicted a successful result, with CD44 protein losing signal against E-selectin-Ig and reducing its molecular weight from 150 kDa to 120 kDa with anti-CD44-hpan (Fig 3.27). This confirmed the presence of N-glycans on CD44 surface. Treatment with neuramidase results in increase of signal of CD44 and decrease of E-selectin Ig staining. This suggested the role that glycosylation have in influencing antibody binding and affinity.

All the results obtained on CD44 glycosylation, imply that even though CD44 may be found expressed on normal tissue, the use of CD44 aberrantly glycosylated can be used as cancer specific biomarker. However, before proceeding to antigen isolation, we have decided to check the presence of CD44 as a soluble marker on LS174T supernatant. In fact, both CD44 and $sLe^{x/a}$ have been found in the cancer cell mucus on IHC staining and we aimed to compare the result in cell culture. Being the presence or quantity of CD44 in the cell culture supernatant unknown, we have used different concentration of supernatant and compared to the positive control total membrane lysate of LS174T. Two bands at ~150 and ~100 kDa, match the positive control signals with increase in intensity (Fig. 3.29). The results suggest, that not only CD44 can be found soluble, but also the major glycosylated forms can be release outside the cell, probably after cell death.

Therefore, we decided to isolate CD44 from LS174T and we injected in the mice for immunization. Being the extraction of CD44 a difficult process (Fig. 3.30) were most of the immunoprecipitant can be lost, we have also decided to immunize a second series of mice with LS174T total membrane protein, having a high concentration of $sLe^{x/a}$ and CD44 antigen.

4.3 Serum testing and antibody specificity

The final aim of the project was to obtain antibody that could recognize CD44/$sLe^{x/a}$ either as soluble marker in blood analysis or in live cells, to be used as therapeutic tool. Initially, testing of mice serum have been conducted mainly by enzyme-linked immunosorbent assay (ELISA). We performed a total of 5 immunizations and while we were aware of the concentration of LS174T total membrane lysate injections, we had calculated the quantity of CD44 isolated by theoretic best yield of extraction. The calculation in based on the maximum efficiency of antibody of protein pulled down. Serum titer was checked using membrane lysate of different cell lines expressing CD44 and $sLe^{x/a}$ in different concentrations. For this we have used LS174T (high expression of CD44 and sLe^{x}), SW620wt (not expressing CD44, moderate in $sLe^{x/a}$) and HT29 FUT6 (high expression of CD44, moderate in $sLe^{x/a}$). Controls were selected as HECA-452 (anti-$sLe^{x/a}$) anti-CD44-hpan and $mice^{CTRL}$, not immunized serum.

Mice series immunized by CD44 ($Mice^{CD44}$), shows from the second immunization preferential binding for HT29 FUT6 and a reduce signals for LS174T and SW620. Although we were expecting major signals in also in LS174T cell lines, specificity of binding for HT29 FUT6 suggested recognition of a particular CD44 glycosylated form, and thus we proceed to mice

sacrifice (Fig 3.33). On the other hand, mice immunized with LS174T total membrane lysate (mice^{LS174T}), shows high intense signals for the three cell lines. Being immunized with total membrane lysate, many aspecific signal were expected (Fig. 3.33). Thus, we decided to proceed with mice sacrifice and to select the monoclonal after hybridomas creation. Further analyses of serum mice on flow cytometry, revealed the recognition of cells membrane cells of LS174T and HT29 FUT6 by serum mice^{LS174T}. On the contrary mice^{CD44} did not show any signal, thus we proceed to hybridoma formation, with purpose to look for a monoclonal antibody that can be used for detection techniques (i.e. diagnostic testing).

4.4 Hybridoma maintenance and clones' selection

Both hybridomas series, were selected mainly by ELISA screening. Being the immunizations coming from cells products, hybridomas often results difficult to select. In fact, hybridoma cells are highly sensitive to culture conditions and can lack of stability: thus, the selection process was hard to handle with the many antigens needed for selection. For instance, the coating of ELISAs well with total membrane lysate could not guaranty repetition of antigen pool variation in each well and each clone tested may differ in antibodies production from week to week. However, antigen signals were often positive for more than one antigen which, added to different replicant, have often cross-validated antibody affinity. For instance, in the second cloning of mice^{CD44} hybridoma (Fig. 3.38) both clones 3.3 and 3.5 were positive for LS174T (high in E-selectin ligands) and commercial sLe$^{x/a}$ tetramers. The result suggested probable clone specificity for E-selectin ligands.

Cloning process was used to select hybridoma clones, to expand them and increase specificity and titre of antibody produced. Cloning was also advantageous by increasing the selection of stable hybridomas. Mycoplasma contamination have affected also clones selection, with loss of positive clones during the selection procedure[299]. Although, mycoplasma infection is often a burden in mammalian cells culture, most of the clones were treated on time by mycoplasma-antibiotic treatment[300]. This led to a temporary loss of antibody production induced both by mycoplasma and antibiotic. Thus, it is very important to establish routine mycoplasma tests until final clones are selected and kept at -80˚C.

4.5 Clone 2B1 from mice^{CD44}

The aim of monoclonal production of mice^{CD44} was to obtain a monoclonal antibody recognize CD44 82 kDa in its cancer glycosylated form. Contrasting results were obtained with the recognition of the polyclonal titre against live cells. In fact, while from a side, there was not recognition of serum mice^{CD44}, of HT29 FUT6 nor LS174T cell lines (Fig. 3.35), from the other side hybridomas CD44.4 showed positive reactivity against live cell ELISA of LS174T and SW948 (Fig. 3.37). Nonetheless, negative results of the mice serum on flow cytometry narrowed the initial production to a possible monoclonal for protein detection tests. Selection occurred mainly by ELISA against different membrane protein of the CRC cell lines [LS174T (high expression of CD44 and sLe^x), HT29 FUT6 (high expression of CD44), SW948 (low in CD44 and moderate in E-selectin ligands) and SW620wt (not expressing CD44)] and sLe$^{x/a}$ commercial tetramers. Commercial CD44 protein was used as final testing, because the only human recombinant version available was of a 36 kDa and it was probably missing most of the post-translational modification required (Fig. 3.39). Loss of signal at the 3rd cloning ???,, especially of sLe$^{x/a}$ antigens and no signal found on CD44, pushed towards the creation of a second generation mice^{CD44}. The second generation showed since the beginning more consistency with the signal expectations, with strong recognition of LS174T, HT29 FUT6, sLe$^{x/a}$ antigens and lower signal in SW620 (Fig. 3.40). However, on the second cloning, most of the signals were lost exception made for clones 2C2 recognizing CD44 and 2B1 recognizing LS174T. At the end of expansion 2B1 revealed an interesting positive signal for LS174T, sLe$^{x/a}$ and CD44, while 2C2 lost most of the signals (Fig. 3.42). Thus, it has been selected for sequencing and further characterization in vitro and future in vivo.

4.6 Clone 2 3A1 from mice^{LS174T}

mice^{LS174T} was created with the intention to produce antibodies mainly against sLe$^{x/a}$, but we have also decided to screen for the two main glycoproteins that we have analysed: CEA and CD44. Hybridomas were tested mainly by ELISA against LS174T (high expression of sLe$^{x/a}$), HT29 FUT6 (moderate expression of sLe$^{x/a}$), SW948 (low/moderate in E-selectin ligands) and SW620wt (moderate expression of sLe$^{x/a}$); against commercial tetramers of sLe^x and sLe^a; against Mucin from bovine submaxillary glands (BSM), highly O-glycosylated and high in sialic acid, and mucin from porcine stomach (PSM), with bound sialic acid only at 0.5-1.5%.

Thus, while mice^{CD44} went straight forward to specific anti-CD44 monoclonal production (from protein to antibody), mice^{LS174T} aimed to broad antigenic reaction, and attempt clones' selections after hybridomas formation (from CRC cells to best antigenic reaction). However, as shown in serum of the mice^{LS174T} tested, high levels of aspecific bound (for our final purpose) revealed intense signal in LS174T, HT29 FUT6 and SW620 (Fig. 3.33). Therefore, firstly we analysed the level of specificity against sialic acid and probable sLe$^{x/a}$ specific recognition by serials dilution of the hybridomas against LS174T and BSM; we have also tested the hybridomas against BSM and PSM to check preference in sialic acid binding (Fig. 3.44). Both results suggest the presence of monoclonal specify and affinity for sialic acid.

Hybridomas were then cloned three times and clones 2.2A4 and 2.3A1 have been selected as showing highest binding against most of the antigens tested; however, the high positive signals for most of the antigens suggests that more than one clone is present in 2.2A4 and 2.3A1 cultures and further cloning is needed. Moreover, in the 3rd cloning most of the clones were more positive for PSM than to BSM, suggesting antibody preference for mucin sequences rather than the glycosylation on their surface (Fig. 3.48).

Great reactivity against CD44 was taken into consideration and we have selected clones 2.2A4 and 2.3A1 for further characterization on live cells. Flow cytometry performed against HT29 wt (control) HT29 FUT6 and LS174T revealed binding of the antibody(ies) on HT29 FUT6 surface. Although binding on LS174T was also expected, we have decided to select clone 2.3A1 for antibody sequencing and further characterization in vitro and in vivo.

4.7 Importance of glycosylation standardization

All the side works presented in chapter 3.5, aimed to understand how glycosylation may influence or be influenced in different condition.

Glycosylation is affected by many factors in cell culture[301] and although many advances in research have been made by analysing glycans with different techniques (including immunoassay and mass spectrometry[153]), research should be able to standardize culture condition and glycosylation expression. Thus, as a first attempt, we have tried to see how much cell confluency can affect glycosylation expression. Two cell lines were chosen with the most diverse antigens expression, LS174T and SW948. Results shows that sLe$^{x/a}$ antigens are highly expressed during low confluency (20%) while the adhesion protein CD44 is mostly expressed at high confluency (Fig. 3.49). Moreover, SW948 start expressing sTn only at 100%

confluency (Fig. 3.50). Thus, although in most of the cases glycan expression is maintained, we think that standardization of screening procedure in experiment that involves glycosylation expression would be highly desirable. Following the same concept, we tested glycosylation levels variation after transduction of luciferase gene into the cell lines. General variation of signal is noticed, however SW948 ST6GAL1 luciferase lost signals in CD44 and $sLe^{x/a}$ antigens. Thus, during transduction, the gene could have been inserted in critical transcribing sequences interfering with main pathways of expression. Thus, new transduction is necessary to perform experiments glycan related in vivo.

In paragraph 3.5.2, we have tried to validate the classifier proposed by Trinh et al. (2016)[134,302]. Comparison between the automated and (online) semiquantitative classifiers showed strong concordance, highlighting that the simplified classifier can assign subtype classes in the absence of continuous quantitative information. However, in order to associate to each CSM glycan expression, it is necessary to test a higher patient's cohort in order to guaranty significant results. Moreover, all the tissues were treated with 5FU and in order to understand glycan influence on the treatment, we performed a half maximal inhibitory concentration (IC50) between SW620 wt and FUT6 transfected. As a result, higher expression of sLe^{x} antigen show lower IC50 of 5FU absorption when compare to the wt and mock (low in sLe^{x}). This suggest that most of the tissue tested, may have a lower glycan expression, when compared to the tumour before 5FU treatment. This study should be continued for further mechanism comprehension.

5.1 Conclusions

Colorectal is considered the second deathliest cancer-type and one of the most common human malignant diseases. Due to its localization in the body, it is still very difficult to be detected in the early stages, when treatment is more effective. CRC detected at late stages is more likely to develop recurrences, even after treatment, leading to the necessity to create a new system to early stages detection. Glycosylation is one of the most common and important post-translational modification that involves key biology processes such as cell-cell interaction, cell growth, cell adhesion etc. Several studies on cancer-associated glycosylation revealed that aberrant glycosylation is a universal feature in various steps of malignant transformation and tumour progression.

Thus, the current project aimed to the identification of aberrant glycans as new possible biomarkers and to the production of specific monoclonal antibodies for their recognition. The work has been divided in two main part: biomarker discovery and antibody production.

Colorectal cancer cell lines and patients' tissue have been screened for different glycans overexpressed in tumour and for carcinoembryonic antigen (CEA) and CD44 glycoproteins. As a result of the overall screening performed with immunoassay techniques, sialyl Lewis$^{x/a}$ ($sLe^{x/a}$) and CD44 are the most expressed antigen. Data suggest that not only sLe^{x} and CD44 could play major in cell migration, but also the production of specific monoclonal antibody against sLe^{x} and/or CD44 might have good potential if used as therapeutic tool.

Moreover, correlation with increase of sLe^{x} and CD44 expression have been found. Both antigens are E-selectin ligands, where E-selectin is a selectin cell adhesion molecule expressed only on endothelial cells activated by cytokines[69]. The role of E-selectin in cancer progression and metastasis has been reinforced by a number of basic researches[69,203]. Therefore, E-selectin ligands have been precipitated from LS174T and HT29 FUT6 and analysed by LC-MS/MS and enzymatic reactions.[69]. The role of E-selectin in cancer progression and metastasis has been reinforced by a number of basic researches[69,203]. Therefore, E-selectin ligands have been precipitated from LS174T and HT29 FUT6, and analysed by LC-MS/MS and enzymatic reactions. CD44 expressed from both cell line is an e-selectin ligand, presenting $sLe^{x/a}$ on its surface. Thus, in order to obtain the best hybridomas clones, we decided to perform two different immunization. The first immunization preparation was obtained by isolation of CD44 by LS174T (mice$^{CD44)}$, while the second by total membrane protein of LS174T (mice^{LS174T}). LS174T cell lines was selected because of their high expression of both antigens.

Immunization titer have been tested mainly by ELISA using different extract from cell lines [(LS174T (high expression of CD44 and sLe^x), SW620wt (not expressing CD44, moderate expression of $sLe^{x/a}$) and HT29 FUT6 (high expression of CD44, moderate expression of $sLe^{x/a}$)]. Flow cytometry experiments have also been performed on LS174T and HT29 FUT6 with the serum of the two mice series. However, only $mice^{LS174T}$ serum was able to bind both cell lines. Thus, hybridomas from $mice^{CD44}$ were tested only for antibody serving a purpose in immunoassay testing, while $mice^{LS174T}$ were also tested for live cell binding, and thus for potential therapeutic tool.

From $mice^{CD44}$, clone 2B1 has been selected as positive for CD44 and $sLe^{x/a}$ commercial antigens; from $mice^{LS174T}$, clone 2.3A1 has been selected not only for positive binding against most of the antigen tested, but also for binding of HT29 FUT6 in flow cytometry. Both clones will be sequenced and further characterized and validated as potential clinical tools.

5.2 Future work

The project presents an important experiment set up, from cell lines screening to antibody production. Cell line expression and translation on human specimen is still to be defined and for this reason different side experiment have shown good preliminary results.

Glycosylation can be influenced in culture by many factor, and as future work, standardization of material handling is fundamental. On the same line, translation between cell lines and human specimen remains very important, and would be interesting to continue validating the IC50 of 5FU in different cell lines overexpressing different aberrant glycans; in fact, most of the tissue are harvested after 5FU treatment and further analyses of glycans influencing 5FU absorption would help understanding better the samples and the results obtained. For the same reason, we find very interesting the correlation study between glycan expression and the four consensus molecular subtypes (CMS) classification, proposed by Guinney *et al*. (2015).

Biomarker research in cell line and tissues showed mostly agreement in antigens expression. However, further testing of glycoprotein expression in serum of CRC patients' needs to be validated before taking major conclusions.

Monoclonal clones produced show promising recognition of CD44 and $sLe^{x/a}$ antigens, thus it is necessary to characterize the antibody isotype and sequence the variable fragments in order to work with more stable products. Glycans microarray and affinity binding will be very important testing.

OTHER CONTRIBUTIONS

PAPER_REVIEW WITH METANALYSIS
SUBMITTED TO Cancer Epidemiology, Biomarkers & Prevention
EPI-19-1276

Colorectal cancer in-between clinical application and translational research: where do we stand and what can be improved?

R. Zoppi[1], F. Deschepper[1], P. A. Videira[1*]
1 UCIBIO, Department of Life Sciences, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

Running title. Colorectal cancer in clinics and research

Keywords. Colorectal cancer, diagnosis, heterogeneity, biomarkers, glycosylation

Additional information.

- The researchers RZ, FD and PAV have received funding from the European Commissions Horizon 2020 program under grant agreement number 676421 (Glycocan) at Faculdade de Ciências e Tecnologia (FCT) Lisbon, PT.
- *Corresponding author at: Departamento Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal. Email: p.videira@fct.unl.pt
- The authors declare that there is no conflict of interest'.
- Word count: 5680 excluding Figure/Title/References
- Total number of figures:2
- Total number of Tables:2

Colorectal cancer in-between clinical application and translational research: where do we stand and what can be improved?

R. Zoppi[1], F. Deschepper[1], P. A. Videira[1*]

1 UCIBIO, Department of Life Sciences, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

Corresponding author at: Departamento Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal.
Email: p.videira@fct.unl.pt

Abstract

Colorectal cancer remains the second deadliest type of cancer with many causes resulting in a severe outcome. It is well recognized the higher level of cellular heterogeneity of colorectal cancer respect to any other type of cancer, which plays a big role in colorectal diagnosis, prognosis and treatment.

Colorectal cancer is a curable disease when detected in early phases, up to 90% when detected in stage I, but the absence of symptoms makes the diagnosis a difficult process. Thus, the understanding of the tumor dynamics, cancer genetics and the expression of specific tumour biomarkers is crucial for the cancer early detection. Furthermore, parallel studies demonstrated the determinant role of post-translational modification in cancer formation and progression. The aim of this review is to resume and combine all the different aspects involved in colorectal cancer malignancy, important for clinicians and researchers to understand where we currently stand, and which improvements are required.

1.Introduction

Cancer is still considered one of the major death-related problem worldwide.

In 2015, in Europe, out of the 5,217,376 deaths registered, 1,062,112 were caused by cancer, reaching roughly the 20% of the population.

Colorectal cancer (**CRC**) dwells from the second to the third position as the most diagnosed malignancy and leading cancer related-death, in both sexes. In 2012, just in Europe, 447,000 new cases have been registered with 215,000 deaths [1]. It has been considered a major public concern in the Western world (Fig.1), but its incidence is also highly increasing in the developing countries. Genetic mutations, comprehensive of previous or family history of

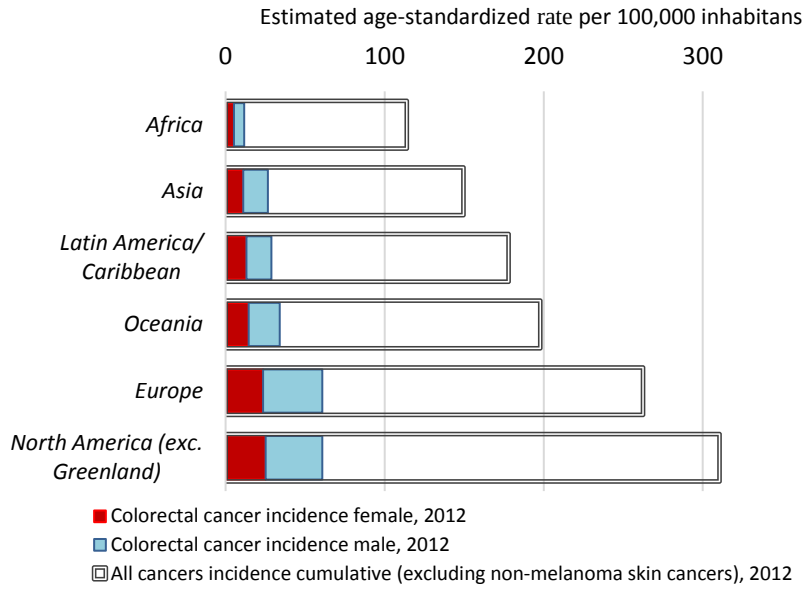
polyps-formation, and environmentally-induced epigenetic factors derived from food habits, physical activity and/or smoking are the main responsible of CRC incidence and increase [2–6].

In the advanced stages, current therapies are not curative, with the necessity of continuous treatments, which is becoming an important burden in the society[7].

At the metastatic stages, chemo-radiotherapy is the main therapy, which improves the treatment for the local advanced staged cancer, but is not curative [1,8].

Thus, the greatest challenge of CRC, is the necessity to early stage detection methods. The identification of adenomas (benign tumour) before carcinomas formation (malignant tumour) can help treatment, thus reducing cancer incidence and increasing overall survival (OS)[7].

Figure 4 Comparison of all cancer incidences (transparent), excluding non-melanoma skin cancer with colorectal cancer incidence in males (blue) and females (red) on different continents. Western countries resulted in both cancer incidence and CRC formation, as a consequence of a consumerism life style. The data are retrieved from project GLOBOCAN 2012.



2. The role of CRC heterogeneity in the clinical decision

Investigating cancer heterogeneity has been one of the central focus in the last two decades. Cellular heterogeneity is often responsible of misleading diagnosis and/or prognosis associated with inefficient classification and reduced treatment efficacy[9].

CRC displays a high level of heterogeneity and thus of high molecular complexity.

From the pathological point of view, CRC histological analysis is essential for a correct diagnosis and prognosis. As shown in the table 1, carcinomas account for 99.5% of all CRC of which 92% are adenocarcinomas, originated from the normal epithelium of colorectal

mucosa [6,7]. In the remaining 8%, in a highly heterogenous population, sarcomas, originating from the connective tissue, and other specific or unspecific carcinomas could be identified (Table 1).

The study of the tissues by histological analysis is also essential for determination of the stage and the grade of the tumour. The stages determine the level of extension of the tumour resulted from the uncontrolled growth, while the grade measures the level of differentiation of the tumour cells and is a stage-independent measurement[10,11].

Two of the staging classifiers in use aim to classify tumour progression in different groups: i) in situ or stage 0, defines the presence of abnormal tissue, ii) localized or stage I where the cancer is present, iii) regional or stages -II /-III with the tumour spreading and distant or stage IV with a spread into distant parts of the body. On the other hand, the TNM staging helps to define the tumour in greater detail and is useful to determine the grade of the cancer, with T describing the size of the tumour, N the nearby lymph nodes and M eventual metastasis present[5]. Although the identification of the correct cancer stage and grade is fundamental to assign the correct therapy, the full understanding of the origin of CRC heterogeneity remains crucial for choosing the most efficient and effective treatment. Many theories have been postulated to explain the role, the origin and the function of the molecular heterogeneity in the tumour microenvironment [12].

In the first theory, the cancer stem model, it is proposed that a minor group of cells, called stem-cells, is responsible of tumour initiation and progression, since they often carry specific proteins capable of inducing metastasis and tumour progression. In the second model of clonal evolution, a more classical Darwinian model has been pursued, where the mutations are acquired upon previous mutations and

Table 1 Percent Distribution by Histology among Histologically Confirmed Cases, both sexes, USA, from 2011-2015.Different types (highlighted in darker-grey) and respective subtypes (light grey) of CRC carcinoma. Data collected by SEER 18 areas, adapted from SEER Cancer Statistics Review 1975-2015

Adenocarcinoma	92.10%
Adenocarcinoma, NOS	63.20%
Adenocarcinoma in adenomatous polyps	8.80%
Mucinous adenocarcinoma	6.30%
Adenocarcinoma in villous adenoma	1.90%
Signet ring adenocarcinoma	1.10%
Mucin-producing adenocarcinoma	0.70%
Other adenocarcinomas	10.20%
Other specific carcinomas	5.90%
Neuroendocrine carcinoid	4.40%
Others	1.50%
Epidermoid carcinomad	0.70%
Squamous cell carcinoma	0.70%
Unspecified, Carcinoma, NOS	0.80%

the fittest survives upon the other clones. Finally, in the third model of the big bang, the mutations responsible for cancer development and progression happen mainly in early CRC formation[13].

These models have been theorised for a better comprehension of tumour progression and instead of been contradictory, they can be considered complementary solutions to the problem.

In general, heterogeneity can be subdivided in inter-tumours variety, among different primary tumours of the same type in the same patient, or intra-tumour variety within the same tumours. It can also be sub grouped into spatial and temporal heterogeneity, defining respectively, the diversity derived from different areas of the same neoplasm and the variability occurred throughout the time[14].

CRC among all cancers is considered one of the most heterogeneous, with differences in molecular expression, cellular subtypes and cancer location. Diversities among left, right colon and rectum have been investigated underlying genetic and immunological differences of colorectum in metastatic colorectal cancer[15]. As consequence, the creation of a new classification, including all the variables that contributes to diversity, has become fundamental for impact on colorectal cancer screening and therapy.[13].

3. Factors inducing heterogeneity and new classification

The origin of the complex heterogenetic tumour environment can be related both to genetic and to epigenetic factors, such as post-translational modifications, epigenetic regulation and differential tumour environment.

Among the various genetic modifications, chromosomal instability (CIN), microsatellite instability (MSI), CpG island methylator phenotype (CIMP) and somatic copy number alterations (SCNAs) are considered the major cause for alterations in signalling pathways involved in CRC. In particular, the Adenomatous polyposis coli (APC) tumour suppressor protein, activator of the WNT protooncogene pathway, the RAS-MAPK cascade (with NRAS, BRAF and PIK3), the MYC transcription factor, the transforming growth factor (TGF)-β and the p53 pathways, are all responsible for cells proliferation and uncontrolled growth [6,16,17]. For instance, the cause for around the 90% of sporadic colon cancers is the aberrant activation of the Wnt/β-catenin signalling pathway led by APC mutations. On the other hand, mutations in the epidermal growth factor receptor (EGFR) is responsible for upregulating the oncogenic PTEN/PI3K/Akt and RAS/RAF/MEK/ERK signalling pathways. However, the upregulated genes may differ between tumour types, underlying the need for mutation profiling for all CRC tumours[7]. Moreover, heterogeneity cannot be explained just by cell alteration pathways since it is well known that other aspects of host-tumour interaction are crucial for the tumour microenvironment progression such as extracellular matrix, supporting stromal cells, metabolism and immune cells interactions[6,18]. As a consequence, the predictive value of the

staging, has resulted limited for estimation of the outcome and implementation of precision was needed.

Guinney et.al. (2015) after evaluating six subtyping algorithms, that combined all the factors contributing to cancer heterogeneity, suggested a new sub classification for CRC cancer types, denominated consensus molecular subtypes (CSM). It is consisted of 4 defined sub groups:

CSM1 (14%), *MSI immune*, characterized by an overexpression of genes involved with immune infiltrate combined with a strong activation of immune evasion pathways associated with higher MSI.

CSM2 (37%), *the canonical*, shows an elevated SCNA with an upregulation of WNT and MYC protooncogenes activation.

CSM3 (13%), *the metabolic*, with mutations in KRAS, presents an heterogenous MSI associated with high metabolic activity.

CSM4 (23%), *the mesenchymal*, with the worst outcome, presents activation of TGF-β, angiogenesis and an upregulation of factors involved in the epithelial-to-mesenchymal transition (EMT) necessary for migration and infiltration[19]. The CSM classification is recognized as an important tool, that combines the different variables for a better understanding of the complex tumour dynamics. Nevertheless, a 5th subgroup containing the 13% of unclassified cases of CRC was later identified, suggesting the necessity of ulterior refinement of the CRC-CSMs, including new variables such as post-transcriptional and epigenetic factors[19].

4. CRC detection

The current diagnostic and prognostic process of CRC involve very complex procedures that involves patients' symptoms, colonoscopy (on the left or right colon) and biomarkers [20].

It remains therefore a major burden for the health system and patient quality of life.

One major problem is the identification of early symptoms, since lower pain can be a result of many intestinal diseases, most of the time not being cancer related. As presented in the review by Vega et al.[20], many meta data analysis have been performed in order to find some implications between symptoms and CRC. Among the different signs and symptoms, correlation is very poor. Yet, in some cases some signs and symptoms show high specificity but low sensitivity, such as weight loss, iron deficiency anaemia and all bleeding (dark blood) (85%-92%), in other cases, age (>50) family history and faecal occult blood test appeared

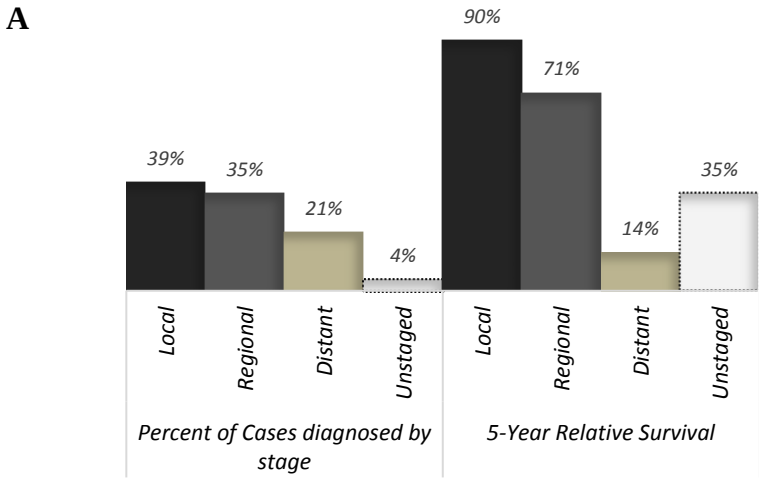
more sensitive (75%-91%) than any other symptoms. However, the presence of one or another symptoms could not be directly correlated to CRC[20–23]. Moreover, left-colon (LC) and the right colon (RC) have disclosed different behaviour in the symptoms, microbiome, clinical-, chromosomal- and molecular characteristics, with RC having a worse prognosis and suggesting different tumour entities[15,24]. In case of symptoms suspect of CRC, many screening programs have been developed, especially in the western countries, to help early diagnosis[20]. The main instrument used for CRC is colonoscopy, an endoscope used not only to detect polyps formation in the intestine, but also to cut the abnormal mass in situ for tissue biopsy and diagnosis. In particular sigmoidoscopy is a colonoscopy that is performed only in the lower part of the colon, including the rectum, and is less complicated of a full colonoscopy, with no need for sedation. Besides being an invasive technique, the availability of the instruments for colonoscopy is limited and it has associated complications such as bleeding, perforation and bacterial infections including antibiotic-resistant bacteria such as *Escherichia coli* and *Klebsiella*[25]. Additionally, CRC endoscopy has shown to have a limited value as diagnostic tool in symptomatic patience, having high sensitivity but lacking specificity in recognizing cancer cells[20]. Therefore, great efforts have been put in the investigation and funding of biomarkers accessible by non-invasive techniques, as blood and stools. These biomarkers include genetic, proteomic and cellular fragments that can be release into the circulation of the patient by the cancer cells. They can be detected through tests like the faecal immunohistochemistry test (FIT) and blood or faecal haemoglobin (FOBT)[7]. Current detection of liquid biomarkers, found in no-solid biological tissue, include also micro-RNA (miRNA), circulating tumour DNA (ctDNA) and protein such as the carcinoembryonic antigen (CEA)[26]. Liquid biomarkers used for diagnostic and prognostic purposes are indeed contributing to the development of the personalized treatment and targeted therapy.

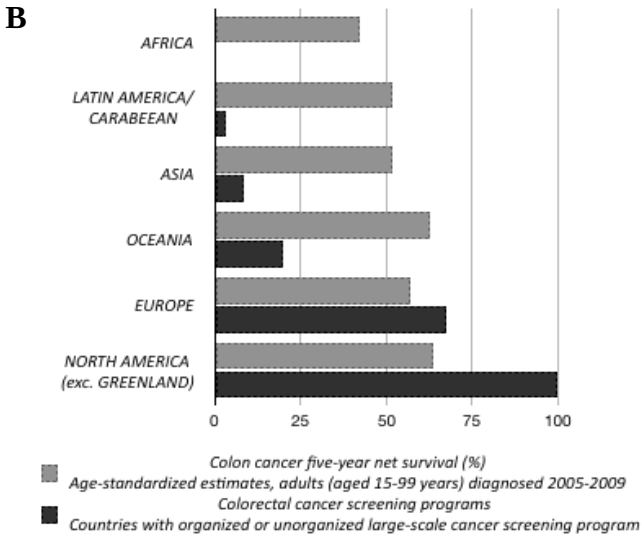
5. Improving CRC screening

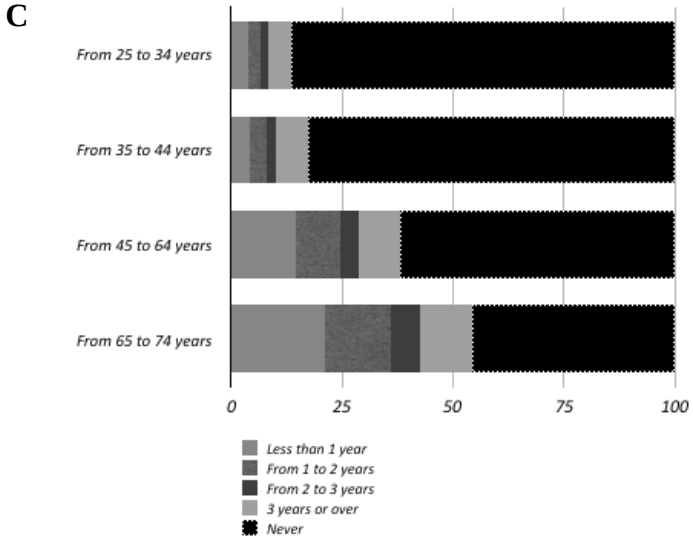
One of the greatest challenges is to find a good screening program, accessible and less invasive, with different diagnostic tools that would allow physicians to detect the disease at early stages. Advanced stages are not curable and as shown in figure 2A, with data obtained by the *Surveillance, Epidemiology, and End Results* (SEER) US cancer program: early diagnosis of CRC increases the chances of survival, from 90% when detected in stage I or local, with a drop to 14% when detected in metastasis, or distant phase [7,27]. However, data retrieved by the

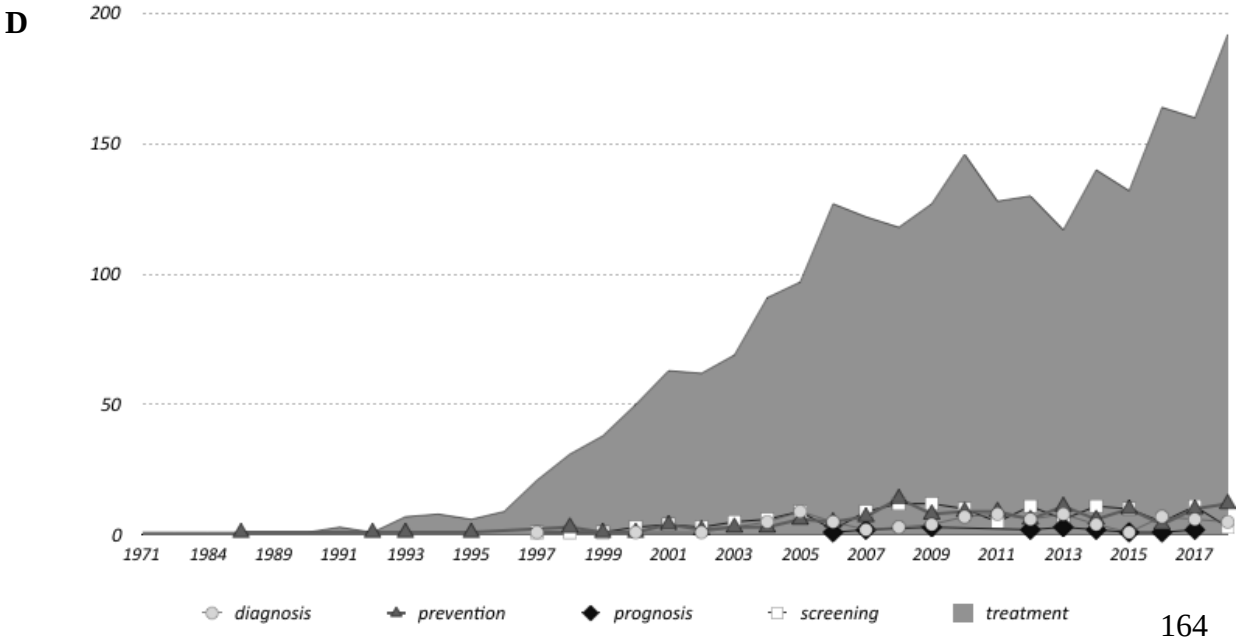
International Agency for Research on Cancer and the project CONCORD-2, show that up to 2009 the screening programs were not able to improve the 5-years survival rate, suggesting that instruments and biomarkers for early detections were still lacking sensitivity or specificity (Fig. 2B) [28,29]. Moreover, as researched in the statistical study of 2014, the invasiveness feature of the CRC screening, it is not well accepted by the population, who avoid volunteering for the test. The study, entitled "self-reported last colorectal cancer screening test" showed that more of than 50% of the population enquired, in-between 25 and 64 years, did not performed any colorectal cancer screening test up to this date[30](Fig.2C). Concerning data result also from a meta-analysis performed on clinical trials acquired from *ClinicalTrials.gov*: the analysis showed that from 1971 to 2018 most of the trials presented, focused on the different treatments for CRC but few efforts were made on the discovery of new tools for diagnosis, prognosis, prevention and screening (Fig. 2D). Although finding the cure is and still remains a very important objective, we must not forget the impact that prevention and early diagnosis would have in colorectal cancer treatment and statistics.

Figure 2 Statistical analysis of CRC in clinics. ***A)*** *Cancer of the Colon and Rectum 5-Year SEER Conditional Relative Survival and 95% Confidence Intervals Probability of surviving the next 5 years given the cohort has already survived 0, 1, or 3 years 1998-2014 by stage at diagnosis.* ***B)*** *Correlation between colorectal cancer program screening and colon cancer five-year net survival worldwide 2005-2009* ***C)*** *Self-reported last colorectal cancer screening test by sex, age and educational attainment level.2014. Data by Eurostat.* ***D)*** *Trends of CRC clinical trials research types: From 1971 to 2018, most of the efforts throughout the 4 clinical trial phases have been put on treatment investigation. Study for new methods of prevention, diagnosis, screening and prognosis has slightly increase in the last 15 years. Data retrieved from ClinicalTrials.gov on the 28th October of 2018: presented 2331 of the 4911 cases downloaded.*









6. Biomarkers in use

Nowadays, the discovery of new biomarkers and the validation of the existing ones is crucial to facilitate the diagnostic process that leads to the understanding of a correct prognosis and treatment. Biomarkers or tumour markers can be genetic (DNA, RNA and miRNA), proteins (such as antibodies or glycoproteins) or may be the results of epigenetic factor (i.e. glycosylation and aberrant carbohydrate formation)[31]. In general, they are referred as tumour associated antigens (TAAs) and are expressed by the cancer cells in quantity proportionated to the number or mass of the neoplasm formation. The concentration of the TAAs in the body fluids can help in early diagnosis stages, but also to monitor the treatment and to help during the follow-ups[32].

Among the different biomarkers, DNA, RNA and proteins have been mainly tested. Common DNA markers used for detection are KRAS, TP53, APC and MSI. Faecal and serum testing, such as the BEAMing technology, have

shown high specificity in CRC detection, in detecting circulating tumour DNA (ctDNA), up to 98%[32,33]. RNA and micro RNA (miRNA, non-coding oligonucleotides) have been investigated as stool and blood biomarkers: MMP7, PTGS2,TP53 and MYBL2 (cancer specific genes) isolated from colonocytes are lacking of sensitivity for CRC detection, due to the fact that RNA is particularly subjected to degradation in the stool[34].

Protein biomarkers are more stable and seem to be a more promising approach. However, on date, only few protein biomarkers are in use. Carcinoembryonic antigen (CEA), a glycoprotein overexpressed in many tumours, cannot distinguish malignant form benign neoplasm but is in use as a prognostic tool: its centration after R0 resection surgery, the complete removal of the tumour mass, is useful for understand the success of the treatment, since its level decline after 6 weeks. In case of continuous high concentrations post-surgery, suggests metastasis and infiltration [32,35]. CA19.9 is a carbohydrate antigen, often associated with mucins, high molecular weight glycoproteins, and is also used as a prognostic factor: high levels determine a poor prognosis. However, it is less sensitive of CEA. Further studies, that combined an algorithm with CA19.9, CEA and other two carbohydrates antigens, CA72-4 and CA242, showed and improved diagnostic power[36].

Further promising marker are the circulating tumour cells (CTCs). It is known that tumour cells detach from the primary tumour into the blood stream in the early phases of progression. It could become a significant prognostic factor detecting the gravity of tumour stage. Until now, CTCs are isolated through positive selection for an epithelial cell surface marker—, epithelial

cell adhesion marker (EpCAM), overexpressed in cancer cells. However, when the cells undergo to epithelial to mesenchymal (ETM) the EpCAM expression drops and the sensitivity of the detection, decreases. [26,37,38].

Nevertheless, as presented in a study of Neves et al. (2019), CTCs can also present another specific tumour antigen, named STn (sialylated Thomsen-nuovelle antigen) which is an aberrant O-glycan. It decorates glycoproteins and has been recognized to be overexpressed in advanced bladder and colon cancer. Using a size-based microfluidic device, they have shown that the majority of CTCs from blood of patients with metastatic bladder and colorectal cancers (>90%) are correlated with high sTn expression[39].

Tumour markers other than diagnostic and prognostic tools can be used as predictive tools. While with prognosis determines the overall outcome of the disease, independent of the treatment, the predictive marker allows to distinguish between different therapies that will be effective[40,41]. The role of the predictive biomarkers has gained importance for personalized treatment and targeted therapy. MSI, chromosome 18q loss of heterozygosity (18qLOH), p53, KRAS, BRAF, NRAS, PIK3CA mutations, PTEN expression were all studied in combination with Adjuvant therapy that may include chemotherapy (5-fluorouracil, oxaliplatin or irinotecan),radiation therapy, hormone therapy, targeted therapy (anti-VEGF or anti-EGFR: cetuximab, panitumumab, or bevacizumab) or biological therapy[42]..

It has been described the role of RAS^{mut} as negative predictive biomarker for anti-EGFR treatment, and the reduced efficiency of the therapy with $RAS^{wt}/BRAF^{mut}$ patients.

$BRAF^{mut}$, near always mutated in position V600E, is mostly located in the right-side colon and are negative prognostic for metastatic CRC (mCRC). MSI has a predictive values in immune checkpoint therapies[1]. MSI or microsatellite stability (MSS) and the presence or absence of 18q chromosome deletion are needed to be tested for therapies based on 5-FU[32].

However, most of the studies showed contradictory results, excluding few established cases, such as the role of KRAS gene in anti-EGFR therapy. Further work on function definition and adequate validation remains to be done (Table 2)[42].

Table 2 Genetic tumour markers original function, mutation and consequences of alterated pathway. Preferences of colon location and predictive potential are also listed. Data adapted from different publications, mainly the reviews of G.Zarkavelis et al. and M.E. Salem et al. [25,80–86].

MARKER	FUNCTION	MUTATION	CONSEQUENCE	COLORECTAL LOCATION	PROGNOSTIC AND PREDICTIVE VALUE
KRAS/NRAS	proto-oncogene encodes a	KRAS: exon 2 (codons 12 and 13) or exon 3 o	permanent activation of the RAS	NRAS more present in the left-side	Resistance to anti-EGFR

	GTPase protein (KRAS)	NRAS: exons 2, 3 and 4 of	(RAS/RAF/MAPK) pathway		
BRAF	encodes serine threonine kinase proteins	8% of CRC carry the distinct BRAF V600E mutation	direct downstream target of KRAS	More present in the right side associated with high MSI	resistance to BRAF inhibitors
DNA MMR (mismatch repair) genes /MSI Microsatellites instability	MMR gene ability to fix DNA errors occurring during replication	inactivated as a result of sporadic MLH1 promoter hypermethylation or germline mutations in MLH1, MSH2, MSH6 and PMS2 genes	MSI are originated from MMR deficient or proficient tumours and thus genomic instability	Higher in right-side of the colon	High MSI doesn't benefit of 5FU but have better outcomes in stage II
CpG island methylator phenotype (CIMP)	CpG island are regions with a high frequency of CpG site (i.e., cytosine residues preceding guanines)	Aberrant methylation of CpG islands	hypermethylated genes, such as SLC5A8, ITGA4, SFRP2, CDKN2A, HLTF, and MGMT	High in the right side associated with KRAS wt and BRAF mut	Worse survival outcomes and poorer response to anti-EGFR
EGFR/HER family EGFR	Receptor of epidermal growth factor (EGF) that stimulates cell growth and differentiation	Increased gene copies	Uncontrolled cell growth when constantly activated	Mostly Right sided	HER2 could be responsible for EGFR resistant pathway
TP53	TP53 protein has role in cells growth arrest DNA repair	Point mutation in codon 72	Promotes malignant process and carcinogenesis	Mostly rectal	TP53 high mutation suggest worse survival
APC/β-catenin	APC (adenomatous polyposis) tumour suppressor protein	promoter hypermethylation and somatic mutations	Mutations of APC activates the Wnt pathway and inactivates glycogen synthase-kinase-3β and β-catenin	Mostly rectal	APC hypermethylation for early diagnosis
miRNA	microRNA are highly stable structures with a hairpin-loop shape and small size	Expression associated with different mutations of other markers	microRNAs, play a key role in tumour suppression or growth	Depends on miRNAs. i.e. miRNA146a and 147b higher in the left side	1. miR-31: present with $BRAF^{mt}$. potential predictive biomarker 2. miR-99a and miR-125b: good prognosis with $KRAS^{wt}$

					responding to EGFR therapies 3. miR-181 poor prognosis with KRAS wt responding to EGFR therapies 4. miR-622: poor responders to radiation therapy in rectal cancer
18qLOH	Chromosome loci	Loss of heterozygosis At D18S58 and D18S61 loci	Allelic imbalance	Distal colon, Left sided	greater survival rate when treated with capecitabin
PIK3CA	Encodes for human p110α protein is encoded by the *PIK3CA* gene	GLU542, GLU545, and HIS1047	Highly oncogenic	Decreasing mutation from right to left to rectal	Implications with anti-EGFR treatment resistance and increased benefit from aspirin therapy
PTEN	Phosphatase and tensin homolog (PTEN) Is a tumor suppressor protein	Mutation and deletion of PTEN	Increase cell proliferation and reduce cell death	Rectal tumour mainly but also right and left side	Loss of expression may induce anti-EGFR resistance
MYC	protooncogene	Constitutively expressed	Cell proliferation	Left sided	Unfavourable prognostic
WNT	Group on signal pathways protooncogene	Overexpression of WNT ligands	Involved with ETM	Left sided	Early stages of carcinogenesis
TGF-β	Transforming growth factor cytokine, stop cell cicle at G1	Inhibition of TGF-beta	Proliferation, angiogenesis and immunosuppression	Left sided	TGF-beta activity can induce High risk of CRC relapse upon treatment

7. The role of CSM classification in prognosis and treatment

At the present, the only curative treatment is surgery with the R0 resection that consist of the complete removal of the tumour mass, to be distinguished from the R1 or R2 tumour-positive

residual resections[43]. Nevertheless, the CSM classification is becoming the most important tool for researchers and clinician to combine and correlate the role of the different tumour marker for a correct CRC prognosis and treatment. The classification takes into account the different outcomes of the therapies and the proper therapies are associated with the biomarker's expression: ESMO guidelines have already adopted and adapted the therapies indication following the genetic, epigenetic and proteomic profile, associated to the different CSM classification[1,19].

Trinh et al.[44] validated the prognostic value of CRC subtyping using five additional markers (CDX2, FRMD6, HTR2B, ZEB1 and KER).

Using immunohistochemistry technique on four independent CRC patients' cohorts, they have created a tool for CRC classification based on semi-quantitative pathology scoring. The CSM1 subtype is identified by the MSI status, the CSM2/3 were selecting by CDX2 a marker highly expressed in epithelial-like tumour. CSM4 subtype is selected with HTR2B and FRMD6 expressed in mesenchymal like cells.

To confirm the ETM transition, of each subtype, the marker ZEB1 and an anti-cytokeratin (AE1/AE3) ware used to normalize for epithelial content. The experiment permitted to confirm the stratification of CRC groups and verify the therapeutic benefit for each subtype[45]. It was proven an increase of OS of epithelial like tumour $KRAS^{wt}/BRAF^{wt}$ patients in treatment of anti-EGFR (cetuximab) to a standard regimen of capecitabine, oxaliplatin and anti-VEGF (bevacizumab), in advanced CRC patients. Same therapy on CSM4 mesenchymal-like, with $KRAS^{mt}/BRAF^{mt}$ was not beneficial. On the down side this robust IHC classifier can be used more as a tool for the stratification than a prognostic or predictive marker with a missing marker that allows to distinguish CSM2 from CSM3 epithelial-like tumour.

8. Glycosylation as the missing link

Genomic studies and their correlation with tumour development and treatment success has been the major focus of the last decade. The significance of these studies in colorectal cancer research is undeniable in diagnosis and prognosis. However, as previously discussed, the heterogeneity of CRC is still responsible of misleading diagnosis. For this reason, more studies on tumour microenvironment have been equally in-depth, for understanding of cancer dynamics. The importance of epigenetic factors influencing post-translational modification cells mechanisms has emerged. Among these, aberrant glycosylation has shown to be highly

correlated with CRC phenotype and the translation from bench to clinic may results in new improvement not only in treatment but also in diagnosis and in discovering new biomarkers[31,46].

Glycosylation is one of the most common and important post-translational modification that involves key biology processes such as cell-cell interaction, cell growth, cell adhesion etc[47].

Glycosylation occurs both on proteins and lipids: on the former, the sugar structures can be borne on asparagine (N-Glycans) or on serine/threonine (O-Glycans) residues while on the latter, the major class is recognized mostly in the glycosphingolipids class (GLS)[48]. Glycans are not encoded by the genome and their biosynthesis depends on metabolism signal transduction and the cellular status[49]. Several studies on cancer-associated glycosylation revealed that aberrant glycosylation is a universal feature in various steps of malignant transformation and tumour progression[50]. Aberrant glycosylation leads to dysregulation of essential cellular processes Additionally, it induce the novel biomarkers that can distinguish healthy from malignant tissue, characteristic very important in therapeutic approach[51]. For instance, the expression of the STn antigen, has been validated as promising tumour marker. In bladder cancer it is expressed specifically in tumour tissue and it has been reported overexpressed in 75% of high-grade bladder tumours presenting elevated proliferation rates and high risk of recurrence/progression expressed STn. Thus, targeting sTn could be the new valid therapy approach[52]. Likewise, aberrant glycosylation in CRC stages have been shown. General increase in N-glycan β1,6-branching, (poly-)N-acetyllactosamine extensions and (truncated) high-mannose has been proven, as well as, higher levels of core 1 glycans, (sialyl) T-antigen, (sialyl) Tn-antigen, and a generally higher density of O-glycans. Increase in sialylation and fucosylation results in a high expression of Lewis antigens and their sialylated derivates[31]. All these alterations can be considered a valid target for personalized therapy, being involved in main tumour cell functions such as tumorigenesis, metastasis, modulation of immunity, and resistance to antitumor therapy.

Two are the main mechanisms involved in glycan structure changes: incomplete synthesis and neosynthesis[48]. Examples of incomplete synthesis can be attributed to the misregulation or silencing of glycosyltransferases and glycosidases, key enzymes responsible for the glycosylation processes, which occur in the formation and accumulation of aberrant O-glycans such as SLewis ^{X} and STn(sialylated-Thomsen-nouvelle antigen)[53]. Neosynthesis can originate from altered expression of enzymes involved in the glycosylation pathways, possibly related to hypoxic conditions in tumors[48]. A normal colon mucosa presents bisecting N-

acetylglucosamines (GlcNAc) on N-glycans as well as core 3 and core 4 O-glycans, globo-type glycosphingolipid (GSL) glycans, and disialylated gangliosides[31].

Many modifications have been found and investigated throughout the stages of CRC. Lewis antigens and its sialylated derivatives (Le^x/sLe^x and Le^a/sLe^a) are the most prominent cancer-associated epitopes on both glycoproteins and glycolipids, since overexpression is related to CRC malignant transformations and may lead to increased tumour cell adhesion and motility, thereby resulting in metastasis[54]. Some alterations, such as gangliosides GD3 and GM2 as well as the globo-type GSL Gb3 are specifically related to angiogenesis[55]. Many carriers, such as the mucin MUC1, CD44v6, and CEA have been identified carrying T- sT- (sialylated/Thomsen-Friedenreich (T)-antigen) and sTn-antigens (sialylated-Thomsen-nouvelle (Tn)-antigen), tumour specific antigens[31]. Mucins, for instance, are major secretory products of the colon and are hyper glycosylated in colorectal cancer, due to the overexpression of the β1,3-galactosyltransferase. This aberrant glycosylation is associated with poor survival, cancer progression, and metastasis[56].

Nowadays, the glycoproteins carcinoembryonic antigen (CEA) and carbohydrate antigen sLe^a (CA19.9), as already mentioned, are the most widely applied serum biomarkers in clinics. Increased serum level of CEA or sLe^a indicates the presence of CRC[57]. Other biomarkers such as, highly fucosylated haptoglobin (Fuc-Hpt) are elevated in certain colorectal cancer patients, in relation to the proximity of the tumour to the liver and distance metastasis. The study of combination of Fuc-Hpt and CEA, has shown to be a promising novel prognostic marker in colorectal cancer[58].

A different approach was adopted by Rho et al (2014). With a novel microarray paired with specific antibodies anti- sLe^x/sLe^a, it was possible to discover new biomarkers that can distinguish between stages III, IV and healthy control in CRC [59]. Croce *et al.(*2005) have also investigated the α1-acid glycoprotein (AGP) which has been identified as a putative carrier of sLe^x antigen in colorectal carcinoma[60].

The role of aberrant glycosylation has shown to have an effective correlation with colorectal cancer cells formation and has disclosed a great potential in non-invasive biomarker discovery. However, a characterisation between aberrant glycosylation and cancer stages is still needed in order to depict the right candidates for early diagnosis.

9.The role of glycosylation in therapy

The tumour-associated carbohydrate antigens (TACAs) are identified among the altered glycans expression, frequently associated with cancer results. The absence in normal healthy tissues, facilitates TACAs to be disease-specific immunotherapeutic targets and, thereby, reduce the potential risk of side effects, playing critical roles in cancer cell biology[48]. Due to this, many monoclonal antibodies (mAbs) have been and are being produced against these type of antigens, employing phage display and monoclonal antibodies techniques[47].
Important characteristics of mAbs are affinity, which measures the strength of interaction between an epitope and an antibody's antigen binding site and the avidity, which gives a measure of the overall strength of an antibody-antigen complex: these two parameters are often exploited to selectively target tumours. IgG mAbs production is generally preferred because of their high affinity, but also IgMs isotype is used as a strong inducer of complement-dependent cytotoxicity (CDC). Antibody-dependent cell cytotoxicity (ADCC) and CDC mechanisms have been proved to be very important in the clinical efficacy of antiganglioside mAbs[48].

Sterner et al. (2016) generated a database for Anti-Glycan Reagent (DAGR) where currently, 1120 unique monoclonal antibody entries have been classified[61]. Many mAbs have been produced for Lewis antigen detection, in particular of Lewis^{y} (Le^{y}). They have been created with different degrees of specificity and cross reaction with other closely related Lewis antigens and tested either alone or as a targeting agent coupled to a toxin or drugs[62].
Among these are BR96, a mouse or chimeric human anti-Le$^{y/x}$ and B3, a mouse anti-Le^{y} mAb: the former has been studied coupled either to toxins or chemotherapeutic agents, while the latter has been tested coupled to Pseudomonas aeruginosa exotoxin A. BR96 resulted as a powerful immunoconjugate with high efficiency in CRC xenograft studies[63]. Also B3, showed significant clinical activity, with responses in colon cancer[64]. Durrant *et al.* (2006) produced SC104, a mouse IgG1 that recognizes sialyltetraosylceramide, that can cause tumour cell death by ADCC, CDC and apoptosis. *In vivo*, it induced potent tumour rejection in colorectal xenograft models[65]. Other two mAbs anti sLe^{a} (5B1, IgG1 and 7E3, IgM) kill target cells by CDC and/or ADCC, thereby prolonging the survival of mice in a colorectal xenograft model[66].
RAV12, a chimeric IgG1 mAb which recognize a sugar, RAAG12, overexpressed on adenocarcinomas of colorectal has been taken into account as a targeting approach. RAV12 has demonstrated to be cytotoxic *in vitro* against the colorectal cell line COLO205 and induced antitumor activity *in vivo* in athymic mice bearing human colon[67].

Other immune-conjugates have been created against protein biomarkers in CRC investigation. For instance PR1A3, a humanized IgG1 mAb anti-CEA linked to the immunotoxin n-succinimidyl-3-(2-pyridyldithio)-propionate was demonstrated to be equally effective in the treatment of CRC[68]. Edrecolomab, is a murine monoclonal antibody against EpCAM, which is a pan-epithelial differentiation antigen that is expressed in 90% of CRC cases. It is considered to be a potential alternative of 5-FU-based chemotherapy as it shows quantitative expression differences between normal and cancerous tissues in CRC[69]. CD24 is also expressed in ~ 90% of colorectal adenomas and adenocarcinomas, and anti-CD24 mAb (SWA11) has been not-covalently linked to the immunotoxin ZZ-PE38; this immunoconjugate showed improved cytotoxity *in vitro*[70].

Loureiro et al. (2018) developed a monoclonal antibody against sTn-antigen, by hybridoma technique. The antibody, named LA25, has shown high affinity and specificity anti-sTn and additional ability to bind 2–6-linked sialyl core-1 probes. This shows its great potential as therapeutic and diagnostic tool in clinics, being sTn a cancer specific aberrant glycan [71].

Conventional mAbs however can be recognized as 'non-self' molecules by the immune system and can be carrier of risk of immune reactions such as acute anaphylaxis and serum sickness. Hence, in order to minimize the possibility of these events, antibody technology researcher have generated single-chain variable fragments: these engineered antibodies conserve the function of the variable part (Fv) but lack the constant Fc region, that is considered to be highly immunogenic[72]. Lutterbuese *et al.* (2009), have constructed a series of bispecific single-chain antibodies (bscAb) that combine various single-chain variable fragments, recognizing human CEA and human CD3, a T-cell co-receptor that helps to activate the cytotoxic T-Cell. Treatment with CEA/CD3-bscAbs showed human T cell killings of tumour and prevented the growth of human CRC cell lines in a severe combined immunodeficiency mouse[73].

In this contest, existing monoclonal antibodies gained attention for the production of bscAb, where the bi-specificity can be directed simultaneously against both protein biomarkers, such as CEA and TACAs.

10. Glycosylation and new therapies

The urge for innovative treatments has led many research groups to investigate new ways to exploit aberrant glycosylation in cancer.

A promising approach is the carbohydrate-based anticancer vaccine, which have been ongoing under clinical trials. The idea is based on breaking the immune self-tolerance for tumour associate glycans, since many aberrant glycans are recognized as 'self' molecules: the rare immunogenicity is due to the presence in embryonic tissues or low expression levels in normal tissues[74]. The anticancer vaccines link multiple copies of the aberrant glycan to the immunogenic carrier protein keyhole limpet hemocyanin (KLH)[74]. Theratone, a phase III KLH-STn conjugate, tested in metastatic breast cancer has shown to increase the OS of the patients if injected with hormone therapy. However, this vaccine fails in its efficacy when administered alone as it fails to activate T-cell-mediated immune response [75].

Further studies have tried to mimic the cell surface to stimulate B-cell and T-cell-mediate immune responses. A tri-genic vaccine containing globo H, Le^y and Tn was produced from Danieshefky *et al.* (2000) and showed to recruit the T-cell help against each of the aberrant glycan[76]. Finally, higher immunogenic vaccines have been produced by incorporating chemically modified sialic acid residues with unnatural N-acyl side chains[51] into KLH conjugates.

Esko et al.(2005) proposed to use decoys as anti-metastatic drugs though metabolic inhibition of sLe^x in metastasis[77]. In a parallel, Shirota *et al.* investigated the GSC-150 an analogue of sLe^x that binds to selectin, adhesion receptors which promote interactions of tumour cells with host platelets, leukocytes and endothelial cells. The data suggested that the sLe^x –selectin interaction is a contributing factor to metastasis with selectin inhibitors having a potential role in cancer treatment[78]. In addition to this theory the clinically approved anticoagulant drug heparin as a selectin inhibitor, resulted in the suppression of tumour metastasis in experimental animal models and has shown beneficial effects in human clinical trials of colon cancer[79].

Posey et *al.* (2016) used genetically modified T cells that express chimeric antigen receptor (CARs) targeting cancer-associated Tn glycoform of MUC1. The CARs displayed target-specific cytotoxicity and could successfully control tumour growth in xenografts models of T cell leukaemia and pancreatic cancer[80].

Finally, important developments have been made in the field of nano-glycomics. For instance, glyconanoparticles (GNP) have been designed to mimic the carbohydrates present in target cells. These glyco-decoy control the biological adhesion processes and competing with interactions at the host cell surface. Nanoparticles (NPs) could be functionalized to deliver glycan-based galectin inhibitors or glycan ligands to sites of tumour growth, as already shown in pancreatic cancer tissue[81].

Danhier *et al.(*2015) have performed local administration of chitosan lipid Nano capsules (LNCs) containing the anti-epidermal growth factor receptor (EGFR) and anti-GAL-1 siRNAs, which prolonged survival of nude mice bearing orthotropic U87MG glioblastoma cells[82].

11. Conclusion

In the last 10 years, many advances have been made in CRC research and many new technologies have been developed helping the clinics to increase the average of the overall survival of CRC and mCRC patients. High cancer heterogeneity makes the progression towards a personalized treatment more difficult. Early diagnosis can improve the OS of CRC patients but the absence of symptoms and of specific early stages liquid biomarkers leads to late detection. Most of the screening program proposed did not result correlated with overall 5 years of survival, suggesting the urge to improve the existing program with different and more specific test. Moreover, in the last 30 years the clinical trials have been focused to find the most effective treatment, but few efforts have been put in looking for new early diagnostic methods.

Genetic biomarkers have been studied in order to understand their role as diagnostic, prognostic and predictive biomarkers. Except for some established cases such as such the role of KRAS in anti-EGFR therapies, further validation of other biomarkers has to be improved. The classification of the sub molecular consensus type (CSM) in colorectal cancer, contributed to the understanding of cancer heterogeneity by considering other aspects of host-tumour interaction, i.e. the immune system. Besides the DNA- and Protein-based therapies, also post-translational modifications, such as glycosylation, has been proven to play a crucial role in tumour microenvironment and for this, it has been investigated. Aberrant glycans expressed mainly on protein surfaces suggested high potential as diagnostic tool in liquid biopsy; the same targets could be exploited in future targeted therapies.

Antibodies and related products are the fastest growing class of therapeutic agents. Monoclonal and bispecific short fragments antibodies have been demonstrated to have promising effects in specific cancer cytotoxicity. However, the presence of side effects requires new ideas to combine functionality of monoclonal antibodies, epitopes and stealth 'self' molecular behaviour. As alternative techniques cancer-based vaccines and CAR-T cells have been proposed and studied.

In conclusion, colorectal cancer still remains the second deadliest kind of cancer and although the many steps forward made in understanding cancer heterogeneity and improving personalized treatment, it is still missing a focus on early diagnosis. The field of glycobiology

appears interesting for the creation of new therapeutic and diagnostic target, where the challenge is the understanding of the biological context of the glycoprotein and glycolipid targets on cancer cells.

Acknowledgements

We thank Anindita Chakraborty and Athanasios Blanas for revision and helpful discussions.This work was supported in part by funding from the European Commissions Horizon 2020 program under grant agreement number 676421 (GlyCoCan)

References

1. Van Cutsem, E. *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* **27**, 1386–1422 (2016).
2. Bianchini, F., Kaaks, R. & Vainio, H. Overweight, obesity, and cancer risk. *Lancet. Oncol.* **3**, 565–74 (2002).
3. Dong, Y. *et al.* Abdominal obesity and colorectal cancer risk: systematic review and meta-analysis of prospective studies. *Biosci. Rep.* **37**, (2017).
4. Limsui, D. *et al.* Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. *J. Natl. Cancer Inst.* **102**, 1012–22 (2010).
5. Sagaert, X., Vanstapel, A. & Verbeek, S. Tumor Heterogeneity in Colorectal Cancer: What Do We Know So Far? *Pathobiology* **85**, 72–84 (2018).
6. Blanco-Calvo, M., Concha, Á., Figueroa, A., Garrido, F. & Valladares-Ayerbes, M. Colorectal cancer classification and cell heterogeneity: A systems oncology approach. *Int. J. Mol. Sci.* **16**, 13610–13632 (2015).
7. Van Cutsem, E. *et al.* Improving outcomes in colorectal cancer: Where do we go from here? *Eur. J. Cancer* **49**, 2476–2485 (2013).
8. Van Cutsem, E., Cervantes, A., Nordlinger, B. & Arnold, D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **25**, iii1–iii9 (2014).
9. Sikic, B. I. Natural and Acquired Resistance to Cancer Therapies. in *The Molecular Basis of Cancer* 583–592 (Elsevier, 2008). doi:10.1016/B978-141603703-3.10049-4
10. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma:

Pathologic aspects. *J. Gastrointest. Oncol.* **3**, 153–73 (2012).

11. Duraiyan, J., Govindarajan, R., Kaliyappan, K. & Palanisamy, M. Applications of immunohistochemistry. *J. Pharm. Bioallied Sci.* **4**, S307-9 (2012).
12. Shlush, L. I. & Hershkovitz, D. Clonal Evolution Models of Tumor Heterogeneity. *Am. Soc. Clin. Oncol. Educ. B.* **35**, e662–e665 (2015).
13. Rosa, S. La *et al.* Tumor Heterogeneity in Primary Colorectal Cancer and Corresponding Metastases. Does the Apple Fall Far From the Tree? *Front. Med.* | *www.frontiersin.org* **1**, 234 (2018).
14. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2017).
15. Yang, S. Y., Cho, M. S. & Kim, N. K. Difference between right-sided and left-sided colorectal cancers: from embryology to molecular subtype. *Expert Rev. Anticancer Ther.* **18**, 351–358 (2018).
16. Colussi, D., Brandi, G., Bazzoli, F. & Ricciardiello, L. Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int. J. Mol. Sci.* **14**, 16365–85 (2013).
17. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
18. Broussard, E. K. & Disis, M. L. TNM staging in colorectal cancer: T is for T Cell and M is for memory. *J. Clin. Oncol.* **29**, 601–603 (2011).
19. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
20. Vega, P., Valentín, F. & Cubiella, J. Colorectal cancer diagnosis: Pitfalls and opportunities. *World J. Gastrointest. Oncol.* **7**, 422–33 (2015).
21. Jellema, P. *et al.* Value of symptoms and additional diagnostic tests for colorectal cancer in primary care: systematic review and meta-analysis. *BMJ* **340**, c1269–c1269 (2010).
22. Ford, A. C. *et al.* Diagnostic utility of alarm features for colorectal cancer: systematic review and meta-analysis. *Gut* **57**, 1545–1553 (2008).
23. Astin, M., Griffin, T., Neal, R. D., Rose, P. & Hamilton, W. The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. *Br. J. Gen. Pract.* **61**, e231–e243 (2011).
24. Zarkavelis, G. *et al.* Current and future biomarkers in colorectal cancer. *Ann. Gastroenterol.* **30**, 613–621 (2017).
25. Wang, P. *et al.* Rates of infection after colonoscopy and osophagogastroduodenoscopy

in ambulatory surgery centres in the USA. *Gut* **67**, 1626–1636 (2018).

26. Norcic, G. Liquid biopsy in colorectal cancer-current status and potential clinical applications. *Micromachines* **9**, (2018).
27. SEER. No Title. *Both Sexes by Race. U. S. National Institutes of Health, National Cancer Institute.* Available at: https://seer.cancer.gov/csr/1975_2015/results_merged/topic_histologic_tabs.pdf#search=percent distribution and counts.
28. Allemani, C. *et al.* Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* **385**, 977–1010 (2015).
29. Allemani, C. *et al.* Colorectal cancer survival in the USA and Europe: a CONCORD high-resolution study. *Open* **3**, 3055 (2013).
30. eurostat-european statistics. Self-reported last colorectal cancer screening test by sex, age and educational attainment level - Eurostat. Available at: https://ec.europa.eu/eurostat/web/products-datasets/-/hlth_ehis_pa5e. (Accessed: 13th December 2018)
31. Holst, S., Wuhrer, M. & Rombouts, Y. *Glycosylation characteristics of colorectal cancer*. *Advances in Cancer Research* **126**, (Elsevier Inc., 2015).
32. Lech, G., Słotwiński, R., Słodkowski, M. & Krasnodębski, I. W. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J. Gastroenterol.* **22**, 1745–55 (2016).
33. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics NIH Public Access Author Manuscript. *Nat Med* **14**, 985–990 (2008).
34. Koga, Y. *et al.* Detection of colorectal cancer cells from feces using quantitative real-time RT-PCR for colorectal cancer diagnosis. *Cancer Sci.* **99**, 1977–83 (2008).
35. Locker, G. Y. *et al.* ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **24**, 5313–27 (2006).
36. Carpelan-Holmström, M., Louhimo, J., Stenman, U. H., Alfthan, H. & Haglund, C. CEA, CA 19-9 and CA 72-4 improve the diagnostic accuracy in gastrointestinal cancers. *Anticancer Res.* **22**, 2311–6
37. Normanno, N., Cervantes, A., Ciardiello, F., De Luca, A. & Pinto, C. The liquid biopsy in the management of colorectal cancer patients: Current applications and future scenarios. *Cancer Treat. Rev.* **70**, 1–8 (2018).
38. Woo, D. & Yu, M. Circulating tumor cells as “liquid biopsies” to understand cancer

metastasis. *Transl. Res.* **201**, 128–135 (2018).

39. Neves, M. *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.* **49**, 77–87 (2019).
40. *Understanding Prognostic versus Predictive Biomarkers*.
41. Oldenhuis, C. N. A. M., Oosting, S. F., Gietema, J. A. & de Vries, E. G. E. Prognostic versus predictive value of biomarkers in oncology. *Eur. J. Cancer* **44**, 946–953 (2008).
42. Deschoolmeester, V., Baay, M., Specenier, P., Lardon, F. & Vermorken, J. B. A review of the most promising biomarkers in colorectal cancer: one step closer to targeted therapy. *Oncologist* **15**, 699–731 (2010).
43. Hermanek, P. & Wittekind, C. The Pathologist and the Residual Tumor (R) Classification. *Pathol. - Res. Pract.* **190**, 115–123 (1994).
44. Trinh, A. A. *et al.* Colorectal Cancer Subtypes by immunohistochemistry Keywords: Colorectal cancer subtypes, anti-EGFR therapy, immunohistochemistry, image analysis, prognostic Funding Sources: (2016). doi:10.1158/1078-0432.CCR-16-0680
45. Trinh, A. *et al.* Practical and Robust Identification of Molecular Subtypes in Colorectal Cancer by Immunohistochemistry. *Clin. Cancer Res.* (2016).
46. Lauc, G., Vojta, A. & Zoldoš, V. Epigenetic regulation of glycosylation is the quantum mechanics of biology. *Biochim. Biophys. Acta - Gen. Subj.* **1840**, 65–70 (2014).
47. Loureiro, L. R. *et al.* Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules* **5**, 1783–809 (2015).
48. Rabu, C., McIntosh, R., Jurasova, Z. & Durrant, L. Glycans as targets for therapeutic antitumor antibodies. *Futur. Oncol.* **8**, 943–960 (2012).
49. Dennis, J. W., Nabi, I. R. & Demetriou, M. Metabolism, cell surface organization, and disease. *Cell* **139**, 1229–41 (2009).
50. Holst, S. *et al.* Investigations on aberrant glycosylation of glycosphingolipids in colorectal cancer tissues using liquid chromatography and matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS). *Mol. Cell. Proteomics* **12**, 3081–93 (2013).
51. Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488 (2005).
52. Ferreira, J. A. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol. Oncol.* **7**, 719–731 (2013).
53. Julien, S., Videira, P. A. & Delannoy, P. Sialyl-tn in cancer: (how) did we miss the

target? *Biomolecules* **2**, 435–66 (2012).

54. Durrant, L. G., Noble, P. & Spendlove, I. Immunology in the clinic review series; focus on cancer: glycolipids as targets for tumour immunotherapy. *Clin. Exp. Immunol.* **167**, 206–215 (2012).
55. Birklé, S., Zeng, G., Gao, L., Yu, R. . & Aubry, J. Role of tumor-associated gangliosides in cancer progression. *Biochimie* **85**, 455–463 (2003).
56. Hung, J.-S. *et al.* C1GALT1 overexpression promotes the invasive behavior of colon cancer cells through modifying O-glycosylation of FGFR2. *Oncotarget* **5**, 2096–106 (2014).
57. Thirunavukarasu, P. *et al.* C-stage in Colon Cancer: Implications of Carcinoembryonic Antigen Biomarker in Staging, Prognosis, and Management. *JNCI J. Natl. Cancer Inst.* **103**, 689–697 (2011).
58. Kirwan, A., Utratna, M., O'Dwyer, M. E., Joshi, L. & Kilcoyne, M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *Biomed Res. Int.* **2015**, 1–16 (2015).
59. Rho, J. *et al.* Discovery of sialyl Lewis A and Lewis X modified protein cancer biomarkers using high density antibody arrays. *J. Proteomics* **96**, 291–9 (2014).
60. Croce, M. V., Sálice, V. C., Lacunza, E. & Segal-Eiras, A. α1-acid glycoprotein (AGP): A possible carrier of sialyl lewis X (slewis X) antigen in colorectal carcinoma. *Histol. Histopathol.* **20**, 91–97 (2005).
61. Sterner, E., Flanagan, N. & Gildersleeve, C. of a Community Resource Database. doi:10.1021/acschembio.6b00244
62. Sharkey, R. M. & Goldenberg, D. M. Targeted Therapy of Cancer: New Prospects for Antibodies and Immunoconjugates. *CA. Cancer J. Clin.* **56**, 226–243 (2006).
63. Carter, P. Improving the efficacy of antibody-based cancer therapies. *Nat. Rev. Cancer* **1**, 118–129 (2001).
64. Stuhler, G. (Gernot) & Walden, P. *Cancer immune therapy : current and future strategies*. (Wiley-VCH, 2002).
65. Durrant, L. G., Harding, S. J., Green, N. H., Buckberry, L. D. & Parsons, T. A New Anticancer Glycolipid Monoclonal Antibody, SC104, which Directly Induces Tumor Cell Apoptosis. *Cancer Res.* **66**, 5901–5909 (2006).
66. Sawada, R. *et al.* Human Monoclonal Antibodies to Sialyl-Lewisa (CA19.9) with Potent CDC, ADCC, and Antitumor Activity. *Clin. Cancer Res.* **17**, 1024–1032 (2011).
67. Zhang, G. *et al.* Suppression of human prostate tumor growth by a unique prostate-

specific monoclonal antibody F77 targeting a glycolipid marker. *Proc. Natl. Acad. Sci.* **107**, 732–737 (2010).

68. Conaghan, P. *et al.* Targeted killing of colorectal cancer cell lines by a humanised IgG1 monoclonal antibody that binds to membrane-bound carcinoembryonic antigen. *Br. J. Cancer* **98**, 1217–25 (2008).
69. Punt, C. J. *et al.* Edrecolomab alone or in combination with fluorouracil and folinic acid in the adjuvant treatment of stage III colon cancer: a randomised study. *Lancet* **360**, 671–677 (2002).
70. Shapira, S., Lisiansky, V., Arber, N. & Kraus, S. Targeted immunotherapy for colorectal cancer: monoclonal antibodies and immunotoxins. *Expert Opin. Investig. Drugs* **19 Suppl 1**, S67-77 (2010).
71. Loureiro, L. R. *et al.* Novel monoclonal antibody L2A5 specifically targeting sialyl-Tn and short glycans terminated by alpha-2–6 sialic acids. *Sci. Rep.* **8**, 1–16 (2018).
72. Hansel, T. T., Kropshofer, H., Singer, T., Mitchell, J. A. & George, A. J. T. The safety and side effects of monoclonal antibodies. *Nat. Rev. Drug Discov.* **9**, 325–338 (2010).
73. Lutterbuese, R. *et al.* Potent Control of Tumor Growth by CEA/CD3-bispecific Single-chain Antibody Constructs That Are Not Competitively Inhibited by Soluble CEA. *J. Immunother.* **32**, 341–352 (2009).
74. Danishefsky, S. J. & Allen, J. R. From the Laboratory to the Clinic: A Retrospective on Fully Synthetic Carbohydrate-Based Anticancer Vaccines. *Angew. Chemie Int. Ed.* **39**, 836–863 (2000).
75. Holmberg, L. A. & Sandmaier, B. M. Theratope® vaccine (STn-KLH). *Expert Opin. Biol. Ther.* **1**, 881–891 (2001).
76. Zhu, J., Warren, J. D. & Danishefsky, S. J. Synthetic carbohydrate-based anticancer vaccines: the Memorial Sloan-Kettering experience. *Expert Rev. Vaccines* **8**, 1399–413 (2009).
77. Fuster, M. M. & Esko, J. D. THE SWEET AND SOUR OF CANCER : GLYCANS AS NOVEL THERAPEUTIC TARGETS. **5**, 526–542 (2005).
78. Nitti, D. *et al.* Final results of a phase III clinical trial on adjuvant intraportal infusion with heparin and 5-fluorouracil (5-FU) in resectable colon cancer (EORTC GITCCG 1983–1987). *Eur. J. Cancer* **33**, 1209–1215 (1997).
79. Varki, N. M. & Varki, A. Heparin Inhibition of Selectin-Mediated Interactions during the Hematogenous Phase of Carcinoma Metastasis: Rationale for Clinical Studies in Humans. *Semin. Thromb. Hemost.* **28**, 53–66 (2002).

80. Posey, A. D. *et al.* Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity* **44**, 1444–1454 (2016).
81. Wang, X. *et al.* Dynamic light scattering as an efficient tool to study glyconanoparticle–lectin interactions. *Analyst* **136**, 4174 (2011).
82. Danhier, F., Messaoudi, K., Lemaire, L., Benoit, J.-P. & Lagarce, F. Combined anti-Galectin-1 and anti-EGFR siRNA-loaded chitosan-lipid nanocapsules decrease temozolomide resistance in glioblastoma: In vivo evaluation. *Int. J. Pharm.* **481**, 154–161 (2015).

11th -13th Jun 2018 Immunoshape International Symposium on Glycoimmunology at Sam Sebastian, Spain.

Development and validation of novel antibodies against Sialylated antigens in ColoRectal Cancer

R. Zoppi[1], F. Deschepper[1], C. Novo[1,2],L. L Santos[3], P. A. Videira[1]

[1] UCIBIO, Department of Life Sciences, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

[2] IHMT, Higiene e Medicina Tropical da Universidade Nova de Lisboa

[3]IPO, Instituto Português de Oncologia (**IPO**) Porto

Colorectal cancer (CRC) is considered the second deathliest cancer-type and one of the most common human malignant diseases. Due to its localization in the body, it is still very difficult to be detected in the early stages and it is essential to develop new sensitive, low cost, minimally invasive and accurate test to improve CRC early detection, prognosis and treatment (1).

Several studies revealed that aberrant glycosylation is a universal feature in various steps of malignant transformation and tumour progression (2). Among all the tumoral glycans, the Lewis antigens, the Tn antigens and their sialylated derivates are probably the most prominent cancer-associated epitopes in CRC, since their overexpression is related to malignant transformations and may lead to increased tumour cell adhesion and motility and, thereby, result in metastasis (3).

This project aims to the development of novel antibodies against -CRC associated glycans, for improve of CRC detection. To achieve this, an initial validation of glycans and candidate protein scaffolding the glycans has been performed using existing antibodies. Secondly, protein extracts enriched from CRC cell lines expressing the identified glycan/protein were used in immunization schedule to develop novel antibodies, which are now being characterized for their fine specificity.

Three main screening methods have been adopted in order to look for new CRC biomarkers: flowcytometry, western blot and immunohistochemistry. All the screenings have been conducted in comparison with a panel of different existing antibodies mainly against Sialyl Lewis antigens and glycoproteins.

As a result of the screening, using existing antibodies, CD44 and Sialyl Lewis$^{x/a}$ have been selected has potential antigens for new antibodies productions. We then immunized mice with protein extracts of cancer cell line LS174T high in sLewis antigens or with CD44

immunoprecipitated from cell lysate of HT29 transfected with FUT6, responsible for sLewis^x expression increase. Following immunization and hybridoma technology, different hybridoma clones were tested for production of antibodies against, Slewis and/or CD44, The first hybridomas show reactivity against CRC total lysates, CRC live cells and protein rich in O-glycan such as bovine serum mucin, further screening are ongoing.

The selected clones will be tested on a cohort of 300 tissues of CRC patient, comprehensive of all the different stages of the cancer and finally tested in vivo experiment, in order to see they interaction with colon cancer cells in the microenvironment of xenograft mice.

Antibody/ies produced will be expected to help improving the clinical diagnosis and prognosis of CRC, with a high potential to be used also as treatment tools.

1.Holst, S., Wuhrer, M., & Rombouts, Y. (2015). Chapter Six - Glycosylation Characteristics of Colorectal Cancer. In R. D. Richard & E. B. Lauren (Eds.), *Advances in Cancer Research* (Vol. Volume 126, pp. 203-256): Academic Press.

2. Hung, J. S., Huang, J., Lin, Y. C., Huang, M. J., Lee, P. H., Lai, H. S., et al. (2014).C1GALT1 overexpression promotes the invasive behavior of colon cancer cells through modifying O-glycosylation of FGFR2. Oncotarget, 5(8), 2096–2106.

3. Durrant, L. G., Noble, P., & Spendlove, I. (2012). Immunology in the clinic review series; focus on cancer: Glycolipids as targets for tumour immunotherapy. Clinical and xperimental Immunology, 167(2), 206–215.

8th-9th Jun 2017 XLI Genetics Portuguese days at Aveiro, Portugal.

Genetic cell models to assess the role of colorectal cancer-associated glycans

R. Zoppi[1], C. Amaral[2], F. Deschepper[3], J. Ramalho[4], P. Videira[5]

[1,2,3,5] UCIBIO, Department of Life Sciences, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

[4]CEDOC, Nova Medical School from Universidade Nova de Lisboa, Portugal

r.zoppi@campus.fct.unl.pt

Cancer is one of the most prevalent diseases worldwide and its metastasization is a major cause of cancer related deaths.

Some carbohydrates attached to proteins or lipids, i.e. glycans, are aberrantly expressed in cancer cells and this altered glycosylation pattern is highly associated with cancer malignancy and tumour progression what makes them promising candidates as novel cancer biomarkers and therapeutic targets. Overexpression of certain glycosyltransferases (e.g. sialyltransferase and fucosyltransferases) is the main mechanism behind tumour associated glycans(1). Our group is especially interested in the role of sTn, Tn, sLex, sLea and Ley antigens as cancer specific glycans, with absent or minimal expression in normal cells (2).

Colorectal cancer (CRC) is the second most common cancer type and it is also correlated with an altered glycosylation pattern. Although this correlation is well known, none of the proposed glycan biomarkers has already reached clinical application. This can be ascribed to the scarce understanding of the molecular mechanisms and the lack of relevant in vitro and in vivo models to address them.

Therefore, the main goal of this study is to establish experimental models based on CRC cell lines expressing altered glycans and to evaluate their behaviour in vitro and in vivo.

To assess it, we analysed CRC cell lines, SW48, SW948, SW620 and HT29 that express relevant glycans and that were previously transfected and transduced with sialyl and fucosyltransferases, the ST6GalI responsible for increased alpha 2-6 sialylation, and FUT VI responsible for increased sialyl Lewis X/A/Y expression. These cell lines were genetically modified to express luciferase (luc) gene as a tracking mechanism by bioluminescence.

After a successful transduction with pLenti6/V5 directional TOPOvector containing the luc gene and an antibiotic selection, the luciferase expression was tested by bioluminescence and the cell lines were characterized in terms of glycosylation profiles, comparing them with their negative controls, MOCK cell lines, once previous studies showed that it's no differences between MOCK and WT.

At present, we are analysing tumour growth of the newly transduced cells lines. In the future progression and metastasis formation in animal models, will be also assessed.

1. Trinchera M, Malagolini N, Chiricolo M, Santini D, Minni F, Caretti A, et al. The biosynthesis of the selectin-ligand sialyl Lewis x in colorectal cancer tissues is regulated by fucosyltransferase VI and can be inhibited by an RNA interference-based approach. Int J Biochem Cell Biol [Internet]. 2011 Jan [cited 2017 Apr 30];43(1):130–9.

- 2nd Jul 2018 Ciência 2018 - Science and Technology in Portugal Summit, at the Lisbon Congress Centre, **poster presentation**.
-
- 12th -13th Mar 2018 Glycocan annual meeting at University Medical Center - LUMC (Leiden, The Netherlands). **Oral presentation** of PhD project.
-
- 22nd -23rd Jun 2017 Glycocan annual at FCT-UNL (Caparica, Portugal). **Oral presentation** of PhD project.
-
- 25th April 2016 Committee evaluation from FCT-UNL. **Oral presentation** of PhD project and future perspectives.
-

Bibliography

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
2. Parkin, D. M. The role of cancer registries in cancer control. *International Journal of Clinical Oncology* (2008). doi:10.1007/s10147-008-0762-6
3. Allemani, C. *et al.* Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* **385**, 977–1010 (2015).
4. Maule, M. & Merletti, F. Cancer transition and priorities for cancer control. *Lancet Oncol.* (2012). doi:10.1016/s1470-2045(12)70268-1
5. Bray, F., Jemal, A., Grey, N., Ferlay, J. & Forman, D. Global cancer transitions according to the Human Development Index (2008-2030): A population-based study. *Lancet Oncol.* (2012). doi:10.1016/S1470-2045(12)70211-5
6. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2018).
7. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–74 (2011).
8. Phng, L. K. & Gerhardt, H. Angiogenesis: A Team Effort Coordinated by Notch. *Developmental Cell* (2009). doi:10.1016/j.devcel.2009.01.015
9. Geretti, E., Shimizu, A. & Klagsbrun, M. Neuropilin structure governs VEGF and semaphorin binding and regulates angiogenesis. *Angiogenesis* (2008). doi:10.1007/s10456-008-9097-1
10. Bedell, V. M. *et al.* roundabout4 is essential for angiogenesis in vivo. *Proc. Natl. Acad. Sci.* (2005). doi:10.1073/pnas.0408318102
11. Pasquale, E. B. Eph receptors and ephrins in cancer: Bidirectional signalling and beyond. *Nature Reviews Cancer* (2010). doi:10.1038/nrc2806
12. Cross, M. J. & Claesson-Welsh, L. FGF and VEGF function in angiogenesis: Signalling pathways, biological responses and therapeutic inhibition. *Trends in Pharmacological Sciences* (2001). doi:10.1016/S0165-6147(00)01676-X
13. Fagiani, E. & Christofori, G. Angiopoietins in angiogenesis. *Cancer Letters* (2013).

doi:10.1016/j.canlet.2012.08.018

14. Korc, M. & Friesel, R. The Role of Fibroblast Growth Factors in Tumor Growth. *Curr. Cancer Drug Targets* (2009). doi:10.2174/156800909789057006
15. Pietras, K. & Östman, A. Hallmarks of cancer: Interactions with the tumor stroma. *Experimental Cell Research* (2010). doi:10.1016/j.yexcr.2010.02.045
16. Singh, A. & Settleman, J. EMT, cancer stem cells and drug resistance: An emerging axis of evil in the war on cancer. *Oncogene* (2010). doi:10.1038/onc.2010.215
17. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* (2009). doi:10.1016/j.cell.2009.11.007
18. Bergfeld, S. A. & DeClerck, Y. A. Bone marrow-derived mesenchymal stem cells and the tumor microenvironment. *Cancer and Metastasis Reviews* (2010). doi:10.1007/s10555-010-9222-7
19. Räsänen, K. & Vaheri, A. Activation of fibroblasts in cancer stroma. *Experimental Cell Research* (2010). doi:10.1016/j.yexcr.2010.04.032
20. Qian, B. Z. & Pollard, J. W. Macrophage Diversity Enhances Tumor Progression and Metastasis. *Cell* (2010). doi:10.1016/j.cell.2010.03.014
21. Gubin, M. M. *et al.* Tumour-Specific Mutant Antigens. *Nature* (2014). doi:10.1038/nature13988
22. Coffelt, S. B. *et al.* Elusive identities and overlapping phenotypes of proangiogenic myeloid cells in tumors. *American Journal of Pathology* (2010). doi:10.2353/ajpath.2010.090786
23. Swanton, C. Intratumor heterogeneity: Evolution through space and time. *Cancer Research* (2012). doi:10.1158/0008-5472.CAN-12-2217
24. Fouad, Y. A. & Aanei, C. *revisiting the hallmarks of cancer*. *Am J Cancer Res* **7**, (2017).
25. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
26. Rosa, S. La *et al.* Tumor Heterogeneity in Primary Colorectal Cancer and Corresponding Metastases. Does the Apple Fall Far From the Tree? *Front. Med.* | *www.frontiersin.org* **1**, 234 (2018).
27. Sottoriva, A. *et al.* A Big Bang model of human colorectal tumor growth. *Nat. Genet.* (2015). doi:10.1038/ng.3214
28. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
29. Van Roosbroeck, K. & Calin, G. A. Cancer Hallmarks and MicroRNAs: The Therapeutic Connection. in 119–149 (2017). doi:10.1016/bs.acr.2017.06.002

30. Souho, T., Lamboni, L., Xiao, L. & Yang, G. Cancer hallmarks and malignancy features: Gateway for improved targeted drug delivery. *Biotechnol. Adv.* **36**, 1928–1945 (2018).
31. Moses, C., Garcia-Bloj, B., Harvey, A. R. & Blancafort, P. Hallmarks of cancer: The CRISPR generation. *Eur. J. Cancer* **93**, 10–18 (2018).
32. Munkley, J. & Elliott, D. J. *Hallmarks of glycosylation in cancer*. *Oncotarget* **7**, (2016).
33. Vajaria, B. N. & Patel, P. S. Glycosylation : a hallmark of cancer ? *Glycoconj. J.* (2016). doi:10.1007/s10719-016-9755-2
34. Witsch, E., Sela, M. & Yarden, Y. Roles for growth factors in cancer progression. *Physiology (Bethesda).* **25**, 85–101 (2010).
35. Pu, Y.-S. *et al.* Characterization of membranous and cytoplasmic EGFR expression in human normal renal cortex and renal cell carcinoma. *J. Biomed. Sci.* **16**, 82 (2009).
36. Fernández-Medarde, A. & Santos, E. Ras in cancer and developmental diseases. *Genes Cancer* **2**, 344–58 (2011).
37. Jiang, B. & Liu, L. Chapter 2 PI3K/PTEN Signaling in Angiogenesis and Tumorigenesis. in *Advances in cancer research* **102**, 19–65 (2009).
38. Sudarsanam, S. & Johnson, D. E. Functional consequences of mTOR inhibition. *Curr. Opin. Drug Discov. Devel.* (2010).
39. Desgrosellier, J. S. & Cheresh, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nat. Rev. Cancer* **10**, 9–22 (2010).
40. Collado, M. & Serrano, M. Senescence in tumours: Evidence from mice and humans. *Nature Reviews Cancer* (2010). doi:10.1038/nrc2772
41. Francis, S. M. *et al.* A functional connection between pRB and transforming growth factor beta in growth inhibition and mammary gland development. *Mol. Cell. Biol.* **29**, 4455–66 (2009).
42. Xu, J., Lamouille, S. & Derynck, R. TGF-B-induced epithelial to mesenchymal transition. *Cell Research* (2009). doi:10.1038/cr.2009.5
43. Matsushita, M. *et al.* Down-regulation of TGF-β receptors in human colorectal cancer: implications for cancer development. *Br. J. Cancer* **80**, 194–205 (1999).
44. Amin, A. R. M. R. *et al.* Evasion of anti-growth signaling: A key step in tumorigenesis and potential target for treatment and prophylaxis by natural compounds. *Semin. Cancer Biol.* **35**, S55–S77 (2015).
45. Zuo, L. *et al.* Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat. Genet.* **12**, 97–99 (1996).
46. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad

Network and the Transcriptional Control of Cell Behavior. *Annu. Rev. Cell Dev. Biol.* **16**, 653–699 (2000).

47. Morin, P. J. *et al.* Activation of beta -Catenin-Tcf Signaling in Colon Cancer by Mutations in beta -Catenin or APC. *Science (80-.).* **275**, 1787–1790 (1997).
48. Curto, M., Cole, B. K., Lallemand, D., Liu, C. H. & McClatchey, A. I. Contact-dependent inhibition of EGFR signaling by Nf2/Merlin. *J. Cell Biol.* (2007). doi:10.1083/jcb.200703010
49. Adams, J. M. & Cory, S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* (2007). doi:10.1038/sj.onc.1210220
50. Junttila, M. R. & Evan, G. I. P53 a Jack of all trades but master of none. *Nature Reviews Cancer* (2009). doi:10.1038/nrc2728
51. Pitti, R. M. *et al.* Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* (1998). doi:10.1038/25387
52. Cantley, L. C. & Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl. Acad. Sci.* (1999). doi:10.1073/pnas.96.8.4240
53. Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* (2008). doi:10.1016/j.cell.2007.12.018
54. White, E., Karp, C., Strohecker, A. M., Guo, Y. & Mathew, R. Role of autophagy in suppression of inflammation and cancer. *Current Opinion in Cell Biology* (2010). doi:10.1016/j.ceb.2009.12.008
55. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, Inflammation, and Cancer. *Cell* (2010). doi:10.1016/j.cell.2010.01.025
56. Jafri, M. A., Ansari, S. A., Alqahtani, M. H. & Shay, J. W. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med.* **8**, 69 (2016).
57. Bernal, A., Moltó-Abad, M., Domínguez, D. & Tusell, L. Acute telomere deprotection prevents ongoing BFB cycles and rampant instability in $p16^{INK4a}$-deficient epithelial cells. *Oncotarget* (2018). doi:10.18632/oncotarget.25502
58. Park, J. Il *et al.* Telomerase modulates Wnt signalling by association with target gene chromatin. *Nature* (2009). doi:10.1038/nature08137
59. Ribatti, D. & Crivellato, E. 'Sprouting angiogenesis', a reappraisal. *Developmental Biology* (2012). doi:10.1016/j.ydbio.2012.09.018
60. Singh, R. K. *et al.* Interferons alpha and beta down-regulate the expression of basic

fibroblast growth factor in human carcinomas. *Proc. Natl. Acad. Sci.* (2006). doi:10.1073/pnas.92.10.4562

61. Ferrara, N. Vascular Endothelial Growth Factor. in *Brenner's Encyclopedia of Genetics: Second Edition* (2013). doi:10.1016/B978-0-12-374984-0.01620-X
62. Ribatti, D. Endogenous inhibitors of angiogenesis. A historical review. *Leukemia Research* (2009). doi:10.1016/j.leukres.2008.11.019
63. Ravi, R. *et al.* Regulation of tumor angiogenesis by p53-induced degradation of hypoxia- inducible factor 1α. *Genes Dev.* (2000).
64. Ferrara, N. Pathways mediating VEGF-independent tumor angiogenesis. *Cytokine Growth Factor Rev.* (2010). doi:10.1016/j.cytogfr.2009.11.003
65. Cavallaro, U. & Christofori, G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nature Reviews Cancer* (2004). doi:10.1038/nrc1276
66. Varner, J. A. & Cheresh, D. A. Integrins and cancer. *Curr. Opin. Cell Biol.* (1996). doi:10.1016/S0955-0674(96)80115-3
67. Stetler-Stevenson, W. G. Matrix metalloproteinases in angiogenesis: A moving target for therapeutic intervention. *Journal of Clinical Investigation* (1999). doi:10.1172/JCI6870
68. Cummings, R. D. & McEver, R. P. *C-type Lectins*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2009).
69. Mann, A. P. & Tanaka, T. E-selectin: Its Role in Cancer and Potential as a Biomarker Biological Role of E-selectin. (2011). doi:10.4172/2161-1025.S1-002
70. Sawada, R., Tsuboi, S. & Fukuda, M. Differential E-selectin-dependent adhesion efficiency in sublines of a human colon cancer exhibiting distinct metastatic potentials. *J. Biol. Chem.* (1994).
71. Fogar, P. *et al.* Neural cell adhesion molecule (N-CAM) in gastrointestinal neoplasias. *Anticancer Res.* (1997).
72. Micalizzi, D. S., Farabaugh, S. M. & Ford, H. L. Epithelial-mesenchymal transition in cancer: Parallels between normal development and tumor progression. *Journal of Mammary Gland Biology and Neoplasia* (2010). doi:10.1007/s10911-010-9178-9
73. Peinado, H. Snail and E47 repressors of E-cadherin induce distinct invasive and angiogenic properties in vivo. *J. Cell Sci.* (2004). doi:10.1242/jcs.01145
74. Karnoub, A. E. *et al.* Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* (2007). doi:10.1038/nature06188
75. Demicheli, R., Retsky, M. W., Hrushesky, W. J. M., Baum, M. & Gukas, I. D. The

effects of surgery on tumor growth: A century of investigations. *Annals of Oncology* (2008). doi:10.1093/annonc/mdn386

76. Barkan, D., Green, J. E. & Chambers, A. F. Extracellular matrix: A gatekeeper in the transition from dormancy to metastatic growth. *Eur. J. Cancer* (2010). doi:10.1016/j.ejca.2010.02.027
77. Teng, M. W. L., Swann, J. B., Koebel, C. M., Schreiber, R. D. & Smyth, M. J. Immune-mediated dormancy: an equilibrium with cancer. *J. Leukoc. Biol.* (2008). doi:10.1189/jlb.1107774
78. Kenific, C. M., Thorburn, A. & Debnath, J. Autophagy and metastasis: Another double-edged sword. *Current Opinion in Cell Biology* (2010). doi:10.1016/j.ceb.2009.10.008
79. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology* (2010). doi:10.1038/nrm2858
80. Mittal, D., Gubin, M. M., Schreiber, R. D. & Smyth, M. J. New insights into cancer immunoediting and its three component phases--elimination, equilibrium and escape. *Curr. Opin. Immunol.* **27**, 16–25 (2014).
81. Smyth, M. J., Dunn, G. P. & Schreiber, R. D. Cancer Immunosurveillance and Immunoediting: The Roles of Immunity in Suppressing Tumor Development and Shaping Tumor Immunogenicity. *Advances in Immunology* (2006). doi:10.1016/S0065-2776(06)90001-7
82. Jiang, X. *et al.* Role of the tumor microenvironment in PD-L1/PD-1-mediated tumor immune escape. *Mol. Cancer* **18**, 10 (2019).
83. Mougiakakos, D. Regulatory T cells in colorectal cancer: From biology to prognostic relevance. *Cancers* (2011). doi:10.3390/cancers3021708
84. Yang, L., Pang, Y. & Moses, H. L. TGF-β and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends in Immunology* (2010). doi:10.1016/j.it.2010.04.002
85. DeNardo, D. G., Andreu, P. & Coussens, L. M. Interactions between lymphocytes and myeloid cells regulate pro-versus anti-tumor immunity. *Cancer and Metastasis Reviews* (2010). doi:10.1007/s10555-010-9223-6
86. Granja, S., Pinheiro, C., Reis, R. M., Martinho, O. & Baltazar, F. Glucose Addiction in Cancer Therapy: Advances and Drawbacks. *Curr. Drug Metab.* **16**, 221–42 (2015).
87. Warburg, O. On the origin of cancer cells. *Science (80-.).* (1956). doi:10.1126/science.123.3191.309

88. Warburg, O. Injuring of Respiration the Origin of Cancer Cells. *Science (80-.).* (1956).
89. DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. *Cell Metabolism* (2008). doi:10.1016/j.cmet.2007.10.002
90. Altman, B. J., Stine, Z. E. & Dang, C. V. From Krebs to clinic: glutamine metabolism to cancer therapy. *Nat. Rev. Cancer* (2016). doi:10.1038/nrc.2016.71
91. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
92. Van Blarigan, E. L. & Meyerhardt, J. A. Role of physical activity and diet after colorectal cancer diagnosis. *J. Clin. Oncol.* **33**, 1825–34 (2015).
93. Moss, A. & Nalankilli, K. The Association Between Diet and Colorectal Cancer Risk: Moving Beyond Generalizations. *Gastroenterology* **152**, 1821–1823 (2017).
94. Limsui, D. *et al.* Cigarette smoking and colorectal cancer risk by molecularly defined subtypes. *J. Natl. Cancer Inst.* **102**, 1012–22 (2010).
95. Larsson, S. C. & Wolk, A. Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies. *Int. J. Cancer* (2006). doi:10.1002/ijc.22170
96. Binefa, G., Rodríguez-Moranta, F., Teule, À. & Medina-Hayas, M. Colorectal cancer: From prevention to personalized medicine. *World J Gastroenterol* **20**, 6786–6808 (2014).
97. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* (1990). doi:10.1016/0092-8674(90)90186-I
98. Church, J. Molecular genetics of colorectal cancer. *Semin. Colon Rectal Surg.* (2016). doi:10.1053/j.scrs.2016.04.013
99. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* (2006). doi:10.1038/ng1834
100. Van Cutsem, E., Cervantes, A., Nordlinger, B. & Arnold, D. Metastatic colorectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **25**, iii1–iii9 (2014).
101. *GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 v1.0.*
102. Konrad, G. Dietary interventions for fecal occult blood test screening: Systematic review of the literature. *Canadian Family Physician* (2010).
103. Burch, J. A. *et al.* Diagnostic accuracy of faecal occult blood tests used in screening for

colorectal cancer: a systematic review. *J. Med. Screen.* **14**, 132–137 (2007).

104. Vilkin, A. *et al.* Performance Characteristics and Evaluation of an Automated-Developed and Quantitative, Immunochemical, Fecal Occult Blood Screening Test. *Am. J. Gastroenterol.* **100**, 2519–2525 (2005).
105. Klaassen, C. H. W. *et al.* Quantification of human DNA in feces as a diagnostic test for the presence of colorectal cancer. *Clin. Chem.* **49**, 1185–7 (2003).
106. Imperiale, T. F. *et al.* Fecal DNA versus Fecal Occult Blood for Colorectal-Cancer Screening in an Average-Risk Population. *N. Engl. J. Med.* **351**, 2704–2714 (2004).
107. Calistri, D. *et al.* Fecal DNA for Noninvasive Diagnosis of Colorectal Cancer in Immunochemical Fecal Occult Blood Test-Positive Individuals. *Cancer Epidemiol. Biomarkers Prev.* **19**, 2647–2654 (2010).
108. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics NIH Public Access Author Manuscript. *Nat Med* **14**, 985–990 (2008).
109. Tamkovich, S. N. *et al.* Circulating DNA and DNase activity in human blood. in *Annals of the New York Academy of Sciences* (2006). doi:10.1196/annals.1368.026
110. Van Cutsem, E. *et al.* Improving outcomes in colorectal cancer: Where do we go from here? *Eur. J. Cancer* **49**, 2476–2485 (2013).
111. Bo, B. L. *et al.* Aberrant methylation of APC, MGMT, RASSF2A, and Wif-1 genes in plasma as a biomarker for early detection of colorectal cancer. *Clin. Cancer Res.* (2009). doi:10.1158/1078-0432.CCR-09-0111
112. Lofton-Day, C. *et al.* DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin. Chem.* (2008). doi:10.1373/clinchem.2007.095992
113. Grützmann, R. *et al.* Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. *PLoS One* (2008). doi:10.1371/journal.pone.0003759
114. Li, Y. *et al.* Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* (2015). doi:10.1038/cr.2015.82
115. Hamfjord, J. *et al.* Differential expression of miRNAs in colorectal cancer: Comparison of paired tumor tissue and adjacent normal mucosa using high-throughput sequencing. *PLoS One* (2012). doi:10.1371/journal.pone.0034150
116. Huang, Z. *et al.* Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int. J. Cancer* (2010). doi:10.1002/ijc.25007
117. Lanza, G. *et al.* mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. *Mol. Cancer* (2007). doi:10.1186/1476-4598-6-54
118. Hardt, P. D., Ngoumou, B. K., Rupp, J., Schnell-Kretschmer, H. & Kloer, H. U. Tumor

M2-pyruvate kinase: a promising tumor marker in the diagnosis of gastro-intestinal cancer. *Anticancer Res.* **20**, 4965–8

119. Karl, J. *et al.* Improved Diagnosis of Colorectal Cancer Using a Combination of Fecal Occult Blood and Novel Fecal Protein Markers. *Clin. Gastroenterol. Hepatol.* **6**, 1122–1128 (2008).
120. Kawahara, M. *et al.* Detection of sialylated lewisX antigen in cancer sera using a sandwich radioimmunoassay. *Int. J. Cancer* **36**, 421–425 (1985).
121. Kobold, S., Luetkens, T., Cao, Y., Bokemeyer, C. & Atanackovic, D. Prognostic and Diagnostic Value of Spontaneous Tumor-Related Antibodies. *Clin. Dev. Immunol.* **2010**, 1–8 (2010).
122. Babel, I. *et al.* Identification of Tumor-associated Autoantigens for the Diagnosis of Colorectal Cancer in Serum Using High Density Protein Microarrays. *Mol. Cell. Proteomics* **8**, 2382–2395 (2009).
123. Normanno, N., Cervantes, A., Ciardiello, F., De Luca, A. & Pinto, C. The liquid biopsy in the management of colorectal cancer patients: Current applications and future scenarios. *Cancer Treat. Rev.* **70**, 1–8 (2018).
124. Woo, D. & Yu, M. Circulating tumor cells as "liquid biopsies" to understand cancer metastasis. *Transl. Res.* **201**, 128–135 (2018).
125. Norcic, G. Liquid biopsy in colorectal cancer-current status and potential clinical applications. *Micromachines* **9**, (2018).
126. Neves, M. *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.* **49**, 77–87 (2019).
127. Vega, P., Valentín, F. & Cubiella, J. Colorectal cancer diagnosis: Pitfalls and opportunities. *World J. Gastrointest. Oncol.* **7**, 422–33 (2015).
128. Wang, P. *et al.* Rates of infection after colonoscopy and osophagogastroduodenoscopy in ambulatory surgery centres in the USA. *Gut* **67**, 1626–1636 (2018).
129. Rosman, A. S. & Korsten, M. A. Meta-analysis Comparing CT Colonography, Air Contrast Barium Enema, and Colonoscopy. *American Journal of Medicine* (2007). doi:10.1016/j.amjmed.2006.05.061
130. Winawer, S. J. *et al.* A Comparison of Colonoscopy and Double-Contrast Barium Enema for Surveillance after Polypectomy. *N. Engl. J. Med.* (2002). doi:10.1056/nejm200006153422401
131. Blanco-Calvo, M., Concha, Á., Figueroa, A., Garrido, F. & Valladares-Ayerbes, M.

Colorectal cancer classification and cell heterogeneity: A systems oncology approach. *Int. J. Mol. Sci.* **16**, 13610–13632 (2015).

132. Greene, F. L. & Sobin, L. H. The Staging of Cancer: A Retrospective and Prospective Appraisal. *CA. Cancer J. Clin.* (2008). doi:10.3322/ca.2008.0001
133. Broussard, E. K. & Disis, M. L. TNM staging in colorectal cancer: T is for T Cell and M is for memory. *J. Clin. Oncol.* **29**, 601–603 (2011).
134. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21**, 1350–1356 (2015).
135. Trinh, A. A. *et al.* Colorectal Cancer Subtypes by immunohistochemistry Keywords: Colorectal cancer subtypes, anti-EGFR therapy, immunohistochemistry, image analysis, prognostic Funding Sources: (2016). doi:10.1158/1078-0432.CCR-16-0680
136. Zhang, X., Sun, X.-F., Shen, B. & Zhang, H. Potential Applications of DNA, RNA and Protein Biomarkers in Diagnosis, Therapy and Prognosis for Colorectal Cancer: A Study from Databases to AI-Assisted Verification. *Cancers (Basel).* **11**, 172 (2019).
137. Hermanek, P. & Wittekind, C. The Pathologist and the Residual Tumor (R) Classification. *Pathol. - Res. Pract.* **190**, 115–123 (1994).
138. Carpelan-Holmström, M., Louhimo, J., Stenman, U. H., Alfthan, H. & Haglund, C. CEA, CA 19-9 and CA 72-4 improve the diagnostic accuracy in gastrointestinal cancers. *Anticancer Res.* **22**, 2311–6
139. Lech, G., Słotwiński, R., Słodkowski, M. & Krasnodębski, I. W. Colorectal cancer tumour markers and biomarkers: Recent therapeutic advances. *World J. Gastroenterol.* **22**, 1745–55 (2016).
140. Schmoll, H. J. *et al.* ESMO Consensus Guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making. (2012). doi:10.1093/annonc/mds236
141. Van Cutsem, E. *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* **27**, 1386–1422 (2016).
142. Zhang, X. *et al.* CBD: A biomarker database for colorectal cancer. *Database* (2018). doi:10.1093/database/bay046
143. Tol, J. & Punt, C. J. A. Monoclonal antibodies in the treatment of metastatic colorectal cancer: A review. *Clin. Ther.* **32**, 437–453 (2010).
144. Arora, S. P. & Mahalingam, D. Immunotherapy in colorectal cancer: for the select few or all? *J. Gastrointest. Oncol.* **9**, 170–179 (2018).
145. Noone AM, Howlader N, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z,

Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2015, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2, A. 2018. No Title.

146. Taylor, M. Introduction to glycobiology. *Choice Rev. Online* (2003). doi:10.5860/choice.40-6408
147. Reitsma, S., Slaaf, D. W., Vink, H., van Zandvoort, M. A. M. J. & oude Egbrink, M. G. A. The endothelial glycocalyx: composition, functions, and visualization. *Pflugers Arch.* **454**, 345–59 (2007).
148. Tarbell, J. M. & Cancel, L. M. The glycocalyx and its significance in human medicine. *Journal of Internal Medicine* (2016). doi:10.1111/joim.12465
149. van Kooyk, Y., Kalay, H. & Garcia-Vallejo, J. J. Analytical tools for the study of cellular glycosylation in the immune system. *Front. Immunol.* **4**, 451 (2013).
150. Wiederschain, G. Y. Essentials of glycobiology. *Biochem.* (2009). doi:10.1134/s0006297909090156
151. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867 (2006).
152. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **15**, 540–555 (2015).
153. Holst, S., Wuhrer, M. & Rombouts, Y. *Glycosylation characteristics of colorectal cancer*. *Advances in Cancer Research* **126**, (Elsevier Inc., 2015).
154. Stanley, P., Taniguchi, N. & Aebi, M. *N-Glycans*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2015). doi:10.1101/GLYCOBIOLOGY.3E.009
155. Brockhausen, I. & Stanley, P. *O-GalNAc Glycans*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2015). doi:10.1101/GLYCOBIOLOGY.3E.010
156. Varki, A. *et al.* O-Glycans. (1999).
157. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80**, 825–858 (2011).
158. Crocker, P. R. & Varki, A. Siglecs in the immune system. *Immunology* (2001). doi:10.1046/j.0019-2805.2001.01241.x
159. Haltiwanger, R. S. *et al.* Biological roles of glycans.
160. Trinchera, M. *et al.* The biosynthesis of the selectin-ligand sialyl Lewis x in colorectal cancer tissues is regulated by fucosyltransferase VI and can be inhibited by an RNA interference-based approach. *Int. J. Biochem. Cell Biol.* **43**, 130–139 (2011).

161. Ashwell, G. & Harford, J. Carbohydrate-Specific Receptors of the Liver. *Annu. Rev. Biochem.* (1982). doi:10.1146/annurev.bi.51.070182.002531
162. Partridge, E. A. *et al.* Regulation of cytokine receptors by golgi N-glycan processing and endocytosis. *Science (80-.).* (2004). doi:10.1126/science.1102109
163. Ohtsubo, K. *et al.* Dietary and genetic control of glucose transporter 2 glycosylation promotes insulin secretion in suppressing diabetes. *Cell* (2005). doi:10.1016/j.cell.2005.09.041
164. Vajaria, B. N. & Patel, P. S. Glycosylation: a hallmark of cancer? *Glycoconj. J.* **34**, 147–156 (2017).
165. Wang, L., Liu, Y., Wu, L. & Sun, X. L. Sialyltransferase inhibition and recent advances. *Biochim. Biophys. Acta - Proteins Proteomics* (2016). doi:10.1016/j.bbapap.2015.07.007
166. Loureiro, L. R. *et al.* Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules* **5**, 1783–809 (2015).
167. Shah, M. H. *et al.* E-Cadherin truncation and Sialyl Lewis-X overexpression in oral squamous cell carcinoma and oral precancerous conditions. *Neoplasma* (2009). doi:10.4149/neo_2009_01_40
168. Thiery, J. P. Epithelial-mesenchymal transitions in development and pathologies. *Current Opinion in Cell Biology* (2003). doi:10.1016/j.ceb.2003.10.006
169. Lu, Z., Ghosh, S., Wang, Z. & Hunter, T. Downregulation of caveolin-1 function by EGF leads to the loss of E-cadherin, increased transcriptional activity of β-catenin, and enhanced tumor cell invasion. *Cancer Cell* (2003). doi:10.1016/S1535-6108(03)00304-0
170. Büll, C., Stoel, M. A., Den Brok, M. H. & Adema, G. J. Sialic acids sweeten a tumor's life. *Cancer Research* (2014). doi:10.1158/0008-5472.CAN-14-0728
171. Ponnusamy, M. *et al.* Emerging Role of Mucins in Epithelial to Mesenchymal Transition. *Curr. Cancer Drug Targets* (2014). doi:10.2174/15680096113136660100
172. Peixoto, A., Relvas-Santos, M., Azevedo, R., Santos, L. L. & Ferreira, J. A. Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. *Front. Oncol.* **9**, 380 (2019).
173. Caldwell, S. A. *et al.* Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1. *Oncogene* (2010). doi:10.1038/onc.2010.41
174. Itkonen, H. M. *et al.* O-GlcNAc transferase integrates metabolic pathways to regulate

the stability of c-MYC in human prostate cancer cells. *Cancer Res.* (2013). doi:10.1158/0008-5472.CAN-13-0549

175. Ju, T., Aryal, R. P., Kudelka, M. R., Wang, Y. & Cummings, R. D. The Cosmc connection to the Tn antigen in cancer. *Cancer Biomarkers* (2014). doi:10.3233/CBM-130375
176. Lau, K. S. & Dennis, J. W. N-Glycans in cancer progression. *Glycobiology* (2008). doi:10.1093/glycob/cwn071
177. English, N. M., Lesley, J. F. & Hyman, R. Site-specific de-N-glycosylation of CD44 can activate hyaluronan binding, and CD44 activation states show distinct threshold densities for hyaluronan binding. *Cancer Res.* (1998).
178. Paszek, M. J. *et al.* The cancer glycocalyx mechanically primes integrin-mediated growth and survival. *Nature* (2014). doi:10.1038/nature13535
179. Yang, W. H. *et al.* Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat. Cell Biol.* (2006). doi:10.1038/ncb1470
180. Wells, L., Slawson, C. & Hart, G. W. The E2F-1 associated retinoblastoma-susceptibility gene product is modified by O-GlcNAc. *Amino Acids* (2011). doi:10.1007/s00726-010-0709-x
181. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: Signal transduction and O-GlcNAc. *Science* (2001). doi:10.1126/science.1058714
182. Rao, X. *et al.* O-GlcNAcylation of G6PD promotes the pentose phosphate pathway and tumor growth. *Nat. Commun.* (2015). doi:10.1038/ncomms9468
183. Yi, W. *et al.* Phosphofructokinase 1 glycosylation regulates cell growth and metabolism. *Science (80-.).* (2012). doi:10.1126/science.1222278
184. Fukumori, T. *et al.* Endogenous galectin-3 determines the routing of CD95 apoptotic signaling pathways. *Cancer Res.* (2004). doi:10.1158/0008-5472.CAN-04-0336
185. Swindall, A. F. & Bellis, S. L. Sialylation of the Fas death receptor by St6Gal-I provides protection against Fas-mediated apoptosis in colon carcinoma cells. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.211375
186. Li, C. *et al.* BCMab1, a monoclonal antibody against aberrantly glycosylated Integrin α3β1, has potent antitumor activity of bladder cancer in vivo. *Clin. Cancer Res.* (2014). doi:10.1158/1078-0432.CCR-13-3397
187. Thamilselvan, V., Craig, D. H. & Basson, M. D. FAK association with multiple signal proteins mediates pressure-induced colon cancer cell adhesion via a Src-dependent PI3K/Akt pathway . *FASEB J.* (2007). doi:10.1096/fj.06-6545com

188. Ma, Z., Vocadlo, D. J. & Vosseller, K. Hyper-O-GlcNAcylation is anti-apoptotic and maintains constitutive NF-κB activity in pancreatic cancer cells. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.470047

189. Chou, T. Y., Hart, G. W. & Dang, C. V. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J. Biol. Chem.* (1995). doi:10.1074/jbc.270.32.18961

190. Iozzo, R. V. & Sanderson, R. D. Proteoglycans in cancer biology, tumour microenvironment and angiogenesis. *Journal of Cellular and Molecular Medicine* (2011). doi:10.1111/j.1582-4934.2010.01236.x

191. Fuster, M. M. & Wang, L. Endothelial heparan sulfate in angiogenesis. in *Progress in Molecular Biology and Translational Science* (2010). doi:10.1016/S1877-1173(10)93009-3

192. Croci, D. O., Cerliani, J. P., Pinto, N. A., GMorosi, L. & Rabinovich, G. A. Regulatory role of glycans in the control of hypoxia-driven angiogenesis and sensitivity to anti-angiogenic treatment. *Glycobiology* (2014). doi:10.1093/glycob/cwu083

193. Lynch, T. P. *et al.* Critical role of O-linked β-N-acetylglucosamine transferase in prostate cancer invasion, angiogenesis, and metastasis. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M111.302547

194. Croci, D. O. *et al.* Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* (2014). doi:10.1016/j.cell.2014.01.043

195. Cole, C. L., Rushton, G., Jayson, G. C. & Avizienyte, E. Ovarian cancer cell heparan sulfate 6-O-sulfotransferases regulate an angiogenic program induced by heparin-binding epidermal growth factor (EGF)-like growth factor/EGF receptor signaling. *J. Biol. Chem.* (2014). doi:10.1074/jbc.M113.534263

196. Benedito, R. *et al.* The Notch Ligands Dll4 and Jagged1 Have Opposing Effects on Angiogenesis. *Cell* (2009). doi:10.1016/j.cell.2009.03.025

197. Takeuchi, H. & Haltiwanger, R. S. Significance of glycosylation in Notch signaling. *Biochemical and Biophysical Research Communications* (2014). doi:10.1016/j.bbrc.2014.05.115

198. Seidenfaden, R., Krauter, A., Schertzinger, F., Gerardy-Schahn, R. & Hildebrandt, H. Polysialic Acid Directs Tumor Cell Growth by Controlling Heterophilic Neural Cell Adhesion Molecule Interactions. *Mol. Cell. Biol.* (2003). doi:10.1128/mcb.23.16.5908-5918.2003

199. Loureiro, L. R. *et al.* Novel monoclonal antibody L2A5 specifically targeting sialyl-Tn and short glycans terminated by alpha-2–6 sialic acids. *Sci. Rep.* **8**, 1–16 (2018).
200. Pinho, S. S. *et al.* The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin. *Hum. Mol. Genet.* (2009). doi:10.1093/hmg/ddp194
201. Rho, J. *et al.* Discovery of sialyl Lewis A and Lewis X modified protein cancer biomarkers using high density antibody arrays. *J. Proteomics* **96**, 291–9 (2014).
202. Kawashima, H. *et al.* N-acetylglucosamine-6-O-sulfotransferases 1 and 2 cooperatively control lymphocyte homing through L-selectin ligand biosynthesis in high endothelial venules. *Nat. Immunol.* (2005). doi:10.1038/ni1259
203. Numahata, K. *et al.* Sialosyl-Lex expression defines invasive and metastatic properties of bladder carcinoma. *Cancer* (2002). doi:10.1002/cncr.10268
204. Murugaesu, N. *et al.* An in vivo functional screen identifies ST6GalNAc2 sialyltransferase as a breast cancer metastasis suppressor. *Cancer Discov.* (2014). doi:10.1158/2159-8290.CD-13-0287
205. Ito, K. *et al.* Galectin-1 as a potent target for cancer therapy: Role in the tumor microenvironment. *Cancer and Metastasis Reviews* (2012). doi:10.1007/s10555-012-9388-2
206. Zhong, J. *et al.* Quantitative phosphoproteomics reveals crosstalk between phosphorylation and O-GlcNAc in the DNA damage response pathway. *Proteomics* (2015). doi:10.1002/pmic.201400339
207. Miura, Y., Sakurai, Y. & Endo, T. O-GlcNAc modification affects the ATM-mediated DNA damage response. *Biochim. Biophys. Acta - Gen. Subj.* (2012). doi:10.1016/j.bbagen.2012.06.013
208. Sevigny, M. B., Graham, K., Ponce, E., Louie, M. C. & Mitchell, K. Glycosylation of human cyclooxygenase-2 (COX-2) decreases the efficacy of certain COX-2 inhibitors. *Pharmacol. Res.* (2012). doi:10.1016/j.phrs.2012.01.001
209. Padró, M. *et al.* Regulation of glycosyltransferases and Lewis antigens expression by IL-1β and IL-6 in human gastric cancer cells. *Glycoconj. J.* (2011). doi:10.1007/s10719-011-9327-4
210. Takamiya, R., Ohtsubo, K., Takamatsu, S., Taniguchi, N. & Angata, T. The interaction between Siglec-15 and tumor-associated sialyl-Tn antigen enhances TGF-β secretion from monocytes/macrophages through the DAP12-Syk pathway. *Glycobiology* (2013). doi:10.1093/glycob/cws139

211. Carrascal, M. A. *et al.* Sialyl Tn-expressing bladder cancer cells induce a tolerogenic phenotype in innate and adaptive immune cells. *Mol. Oncol.* **8**, 753–65 (2014).
212. Beatson, R. *et al.* The mucin MUC1 modulates the tumor immunological microenvironment through engagement of the lectin Siglec-9. *Nat. Immunol.* (2016). doi:10.1038/ni.3552
213. Li, C. W. *et al.* Glycosylation and stabilization of programmed death ligand-1 suppresses T-cell activity. *Nat. Commun.* (2016). doi:10.1038/ncomms12632
214. Gringhuis, S. I. *et al.* Fucose-based PAMPs prime dendritic cells for follicular T helper cell polarization via DC-SIGN-dependent IL-27 production. *Nat. Commun.* (2014). doi:10.1038/ncomms6074
215. Balog, C. I. A. *et al. N* -glycosylation of Colorectal Cancer Tissues. *Mol. Cell. Proteomics* **11**, 571–585 (2012).
216. Zhao, Y. Y. *et al.* Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Science* (2008). doi:10.1111/j.1349-7006.2008.00839.x
217. Muinelo-Romay, L. *et al.* Expression and enzyme activity of α(1,6)fucosyltransferase in human colorectal cancer. *Int. J. Cancer* (2008). doi:10.1002/ijc.23521
218. Demetriou, M., Nabi, I. R., Coppolino, M., Dedhar, S. & Dennis, J. W. Reduced contact-inhibition and substratum adhesion in epithelial cells expressing GlcNAc-transferase V. *J. Cell Biol.* (1995). doi:10.1083/jcb.130.2.383
219. Ichikawa, T. *et al.* Expression of N-acetyllactosamine and β1,4-galactosyltransferase (β4GalT-I) during adenoma-carcinoma sequence in the human colorectum. *J. Histochem. Cytochem.* (1999). doi:10.1177/002215549904701211
220. Hung, J.-S. *et al.* C1GALT1 overexpression promotes the invasive behavior of colon cancer cells through modifying O-glycosylation of FGFR2. *Oncotarget* **5**, 2096–106 (2014).
221. Posey, A. D., Clausen, H. & June, C. H. Distinguishing Truncated and Normal MUC1 Glycoform Targeting from Tn-MUC1-Specific CAR T Cells: Specificity Is the Key to Safety. *Immunity* **45**, 947–948 (2016).
222. Naor, D., Sionov, R. V & Ish-Shalom, D. CD44: structure, function, and association with the malignant process. *Adv. Cancer Res.* **71**, 241–319 (1997).
223. Shimodaira, K. *et al.* 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.* **57**, 5201–6 (1997).
224. Brockhausen, I. Mucin-type O-glycans in human colon and breast cancer:

Glycodynamics and functions. *EMBO Reports* (2006). doi:10.1038/sj.embor.7400705

225. Holst, S. *et al.* Investigations on aberrant glycosylation of glycosphingolipids in colorectal cancer tissues using liquid chromatography and matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS). *Mol. Cell. Proteomics* **12**, 3081–93 (2013).
226. Higashi, H. *et al.* Characterization of N-Glycolylneuraminic Acid-containing Gangliosides as Tumor-associated Hanganutziu-Deicher Antigen in Human Colon Cancer. *Cancer Res.* (1985).
227. Geßner, P., Riedl, S., Quentmaier, A. & Kemmner, W. Enhanced activity of CMP-NeuAc:Galβ1-4GlcNAc:α2,6-sialyltransferase in metastasizing human colorectal tumor tissue and serum of tumor patients. *Cancer Lett.* (1993). doi:10.1016/0304-3835(93)90056-F
228. Itai, S. *et al.* Tissue distribution of 2- 3 and 2- 6 sialyl lewis A antigens and significance of the ratio of two antigens for the differential diagnosis of malignant and benign disorders of the digestive tract. *Cancer* (1991). doi:10.1002/1097-0142(19910315)67:6<1576::AID-CNCR2820670620>3.0.CO;2-2
229. Miyoshi, E., Moriwaki, K. & Nakagawa, T. Biological function of fucosylation in cancer biology. *Journal of Biochemistry* (2008). doi:10.1093/jb/mvn011
230. Taniuchi, F., Higai, K., Tanaka, T., Azuma, Y. & Matsumoto, K. Transcriptional Regulation of *Fucosyltransferase 1* Gene Expression in Colon Cancer Cells. *Sci. World J.* **2013**, 1–9 (2013).
231. Ito, H. *et al.* Altered mRNA expression of specific molecular species of fucosyl- and sialyl-transferases in human colorectal cancer tissues. *Int. J. Cancer* (1997). doi:10.1002/(SICI)1097-0215(19970516)71:4<556::AID-IJC9>3.0.CO;2-T
232. SEER. No Title. Available at: https://seer.cancer.gov/statfacts/html/colorect.html.
233. Kim, N. H. *et al.* Serum CEA and CA 19-9 Levels are Associated with the Presence and Severity of Colorectal Neoplasia. *Yonsei Med. J.* **58**, 918 (2017).
234. Farooqui, M. S. *et al.* Improved diagnostic accuracy of pancreatic diseases with a combination of various novel serum biomarkers - Case control study from manipal teaching hospital, Pokhara, Nepal. *Asian Pacific J. Cancer Prev.* (2012). doi:10.7314/APJCP.2012.13.5.2171
235. Wei, H. *et al.* Glycoprotein screening in colorectal cancer based on differentially expressed Tn antigen. *Oncol. Rep.* **36**, 1313–1324 (2016).
236. Gong, Y. *et al.* TIMP-1 Promotes Accumulation of Cancer Associated Fibroblasts and

Cancer Progression. *PLoS One* (2013). doi:10.1371/journal.pone.0077366

237. Holten-Andersen, M. N. *et al.* Total levels of tissue inhibitor of metalloproteinases 1 in plasma yield high diagnostic sensitivity and specificity in patients with colon cancer. *Clin. Cancer Res.* (2002).
238. Kim, Y. S. *et al.* Functional proteomics study reveals that N-acetylglucosaminyltransferase V reinforces the invasive/metastatic potential of colon cancer through aberrant glycosylation on tissue inhibitor of metalloproteinase-1. *Mol. Cell. Proteomics* (2008). doi:10.1074/mcp.M700084-MCP200
239. Betge, J. *et al.* MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. *Virchows Arch.* **469**, 255 (2016).
240. Tanida, S. *et al.* Binding of the sialic acid-binding lectin, siglec-9, to the membrane mucin, MUC1, induces recruitment of β-catenin and subsequent cell growth. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.471318
241. Yu, L. G. *et al.* Galectin-3 interaction with Thomsen-Friedenreich disaccharide on cancer-associated MUC1 causes increased cancer cell endothelial adhesion. *J. Biol. Chem.* (2007). doi:10.1074/jbc.M606862200
242. Carlos, C. A. *et al.* Human Tumor Antigen MUC1 Is Chemotactic for Immature Dendritic Cells and Elicits Maturation but Does Not Promote Th1 Type Immunity. *J. Immunol.* (2005). doi:10.4049/jimmunol.175.3.1628
243. Itzkowitz, S. H. & Itzkowitz, S. H. Mucins Bearing the Cancer-associated Sialosyl-Tn Antigen Mediate Inhibition of Natural Killer Cell Cytotoxicity. *Cancer Res.* (1992).
244. Zeng, Y. *et al.* MUC1 Predicts Colorectal Cancer Metastasis: A Systematic Review and Meta-Analysis of Case Controlled Studies. *PLoS One* **10**, e0138049 (2015).
245. T., K. *et al.* MUC1 vaccine for individuals with advanced adenoma of the colon: A cancer immunoprevention feasibility study. *Cancer Prev. Res.* (2013).
246. Duffy, M. J. Carcinoembryonic antigen as a marker for colorectal cancer: Is it clinically useful? *Clinical Chemistry* (2001).
247. Saeland, E. *et al.* Differential glycosylation of MUC1 and CEACAM5 between normal mucosa and tumour tissue of colon cancer patients. *Int. J. Cancer* (2012). doi:10.1002/ijc.26354
248. Japink, D., Leers, M. P. G., Sosef, M. N. & Nap, M. CEA in activated macrophages. New diagnostic possibilities for tumor markers in early colorectal cancer. *Anticancer Res.* (2009).
249. Thirunavukarasu, P. *et al.* C-stage in Colon Cancer: Implications of Carcinoembryonic

Antigen Biomarker in Staging, Prognosis, and Management. *JNCI J. Natl. Cancer Inst.* **103**, 689–697 (2011).

250. Duffy, M. J. *Carcinoembryonic Antigen as a Marker for Colorectal Cancer: Is It Clinically Useful?* (2001).
251. Kuppusamy, P., Govindan, N., Yusoff, M. M. & Ichwan, S. J. A. Proteins are potent biomarkers to detect colon cancer progression. *Saudi J. Biol. Sci.* **24**, 1212–1221 (2017).
252. Zhao, Q. *et al.* Glycan analysis of colorectal cancer samples reveals stage-dependent changes in CEA glycosylation patterns. *Clin. Proteomics* **15**, 9 (2018).
253. Tanaka, T. *et al.* Biomarkers for Colorectal Cancer. *Int. J. Mol. Sci.* **11**, 3209–3225 (2010).
254. Ponta, H., Sherman, L. & Herrlich, P. A. CD44: From adhesion molecules to signalling regulators. *Nat. Rev. Mol. Cell Biol.* **4**, 33–45 (2003).
255. Naor, D., Sionov, R. V. & Ish-Shalom, D. CD44: Structure, Function and Association with the Malignant Process. in 241–319 (1997). doi:10.1016/S0065-230X(08)60101-3
256. Azevedo, R. *et al.* CD44 glycoprotein in cancer: a molecular conundrum hampering clinical applications. *Clin. Proteomics* **15**, 22 (2018).
257. Wang, Z. *et al.* The Prognostic and Clinical Value of CD44 in Colorectal Cancer: A Meta-Analysis. *Front. Oncol.* **9**, 309 (2019).
258. Xia, P. & Xu, X.-Y. Prognostic significance of CD44 in human colon cancer and gastric cancer: Evidence from bioinformatic analyses. *Oncotarget* **7**, 45538–45546 (2016).
259. Basakran, N. S. CD44 as a potential diagnostic tumor marker. *Saudi Med. J.* **36**, 273–9 (2015).
260. Katoh, S., McCarthy, J. B. & Kincade, P. W. Characterization of soluble CD44 in the circulation of mice. Levels are affected by immune activity and tumor growth. *J. Immunol.* **153**, 3440–9 (1994).
261. Videira, P. A., Silva, M., Martin, K. C. & Sackstein, R. Ligation of the CD44 Glycoform HCELL on Culture-Expanded Human Monocyte-Derived Dendritic Cells Programs Transendothelial Migration. *J. Immunol.* (2018). doi:10.4049/jimmunol.1800188
262. Dockrell, H. M. Roitt's essential immunology. *Trans. R. Soc. Trop. Med. Hyg.* (2002). doi:10.1016/s0035-9203(02)90267-1
263. GODING, J. W. Antibody Structure and Function. in *Monoclonal Antibodies* (1996). doi:10.1016/b978-012287023-1/50053-5
264. Kellner, C., Derer, S., Valerius, T. & Peipp, M. Boosting ADCC and CDC activity by Fc engineering and evaluation of antibody effector functions. *Methods* (2014).

doi:10.1016/j.ymeth.2013.06.036

265. Liu, J. K. H. The history of monoclonal antibody development - Progress, remaining challenges and future innovations. *Ann. Med. Surg.* **3**, 113–6 (2014).
266. Loureiro, L. *et al.* Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules* **5**, 1783–1809 (2015).
267. Nelson, A. L. Antibody fragments: hope and hype. *MAbs* **2**, 77–83 (2010).
268. Presta, L. G. Molecular engineering and design of therapeutic antibodies. *Current Opinion in Immunology* (2008). doi:10.1016/j.coi.2008.06.012
269. Sanz, L., Cuesta, Á. M., Compte, M. & Álvarez-Vallina, L. Antibody engineering: Facing new challenges in cancer therapy. *Acta Pharmacologica Sinica* (2005). doi:10.1111/j.1745-7254.2005.00135.x
270. Acchione, M., Kwon, H., Jochheim, C. M. & Atkins, W. M. Impact of linker and conjugation chemistry on antigen binding, Fc receptor binding and thermal stability of model antibody-drug conjugates. *MAbs* (2012). doi:10.4161/mabs.19449
271. Hnasko, R. M. & Stanker, L. H. Hybridoma technology. *Methods Mol. Biol.* (2015). doi:10.1007/978-1-4939-2742-5_2
272. Pandey, S. Hybridoma technology for production of monoclonal antibodies. *Int. J. Pharm. Sci. Rev. Res.* (2010).
273. Pandey, S. *HYBRIDOMA TECHNOLOGY FOR PRODUCTION OF MONOCLONAL ANTIBODIES*. *International Journal of Pharmaceutical Sciences Review and Research* **1**,
274. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for therapy. *Nature* **332**, 323–327 (1988).
275. Steinitz, M. Production of human monoclonal antibodies by the epstein-barr virus method. *Methods Mol. Biol.* (2014). doi:10.1007/978-1-62703-586-6_6
276. KOZBOR, D., RODER, J. C., CHANG, T. H., STEPLEWSKI, Z. & KOPROWSKI, H. Human Anti-Tetanus Toxoid Monoclonal Antibody Secreted by EBV-Transformed Human B Cells Fused with Murine Myeloma. *Hybridoma* (2011). doi:10.1089/hyb.1.1982.1.323
277. Kubota, T., Matsushita, T., Niwa, R., Kumagai, I. & Nakamura, K. Novel anti-Tn single-chain Fv-Fc fusion proteins derived from immunized phage library and antibody Fc domain. *Anticancer Res.* (2010).
278. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat. Rev. Cancer* **12**, 278–287 (2012).

279. Brahmer, J. R. *et al.* Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: Safety, clinical activity, pharmacodynamics, and immunologic correlates. *J. Clin. Oncol.* (2010). doi:10.1200/JCO.2009.26.7609
280. Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science (80-.).* (1996). doi:10.1126/science.271.5256.1734
281. Scott, A. M., Allison, J. P. & Wolchok, J. D. Monoclonal antibodies in cancer therapy. *Cancer Immun.* **12**, 14 (2012).
282. Blidner, A. G., Mariño, K. V. & Rabinovich, G. A. Driving CARs into Sweet Roads: Targeting Glycosylated Antigens in Cancer. *Immunity* **44**, 1248–1250 (2016).
283. June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S. & Milone, M. C. CAR T cell immunotherapy for human cancer. *Science* (2018). doi:10.1126/science.aar6711
284. Cummings, R. D. & Etzler, M. E. *Antibodies and Lectins in Glycan Analysis*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2009).
285. Durrant, J. X. C. and L. Monoclonal Antibodies Against Tumour-Associated Carbohydrate Antigens. *Long-Haul Travel Motiv. by Int. Tour. to Penang* **i**, 13 (2017).
286. Yu, A. L. *et al.* Anti-GD2 Antibody with GM-CSF, Interleukin-2, and Isotretinoin for Neuroblastoma. *N. Engl. J. Med.* (2010). doi:10.1056/nejmoa0911123
287. Ross, H. J. *et al.* A randomized, multicenter study to determine the safety and efficacy of the immunoconjugate SGN-15 plus docetaxel for the treatment of non-small cell lung carcinoma. *Lung Cancer* (2006). doi:10.1016/j.lungcan.2006.05.020
288. Scott, A. M. *et al.* Construction, production, and characterization of humanized anti-Lewis Y monoclonal antibody 3S193 for targeted immunotherapy of solid tumors. *Cancer Res.* (2000).
289. Shitara, K. *et al.* Application of anti-sialyl Le a monoclonal antibody, KM231, for immunotherapy of cancer. *Anticancer Res.* (1991).
290. Sawada, R. *et al.* Human monoclonal antibodies to sialyl-Lewis a (CA19.9) with potent CDC, ADCC, and antitumor activity. *Clin. Cancer Res.* (2011). doi:10.1158/1078-0432.CCR-10-2640
291. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–8 (2001).
292. Vankemmelbeke, M., Chua, J. X. & Durrant, L. G. Cancer cell associated glycans as targets for immunotherapy. *Oncoimmunology* **5**, e1061177 (2016).
293. Hatakeyama, K. *et al.* Novel protein isoforms of carcinoembryonic antigen are secreted from pancreatic, gastric and colorectal cancer cells. *BMC Res. Notes* **6**, 381 (2013).

294. Blanas, A., Sahasrabudhe, N. M., Rodríguez, E., van Kooyk, Y. & van Vliet, S. J. Fucosylated Antigens in Cancer: An Alliance toward Tumor Progression, Metastasis, and Resistance to Chemotherapy. *Front. Oncol.* **8**, 39 (2018).
295. Loureiro, L. R. *et al.* Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules* **5**, 1783–809 (2015).
296. Beatson, R. *et al.* The breast cancer-associated glycoforms of MUC1, MUC1-Tn and sialyl-Tn, are expressed in COSMC wild-type cells and bind the C-type lectin MGL. *PLoS One* (2015). doi:10.1371/journal.pone.0125994
297. Cutsem, E. Van *et al.* ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. 1386–1422 (2016). doi:10.1093/annonc/mdw235
298. Thapa, R. & Wilson, G. D. The Importance of CD44 as a Stem Cell Biomarker and Therapeutic Target in Cancer. *Stem Cells Int.* **2016**, 2087204 (2016).
299. Jakway, J. P. & Shevach, E. M. Mycoplasma contamination: a hazard of screening hybridoma supernatants for inhibition of [3H]thymidine incorporation. *J. Immunol. Methods* **67**, 337–345 (1984).
300. Carroll, K. & O'Kennedy, R. The elimination of mycoplasma from infected hybridomas by passaging in BALB/c mice. *J. Immunol. Methods* **108**, 189–193 (1988).
301. Spearman, M. & Butler, M. Glycosylation in Cell Culture. in 237–258 (Springer, Cham, 2015). doi:10.1007/978-3-319-10320-4_9
302. Trinh, A. *et al.* Practical and Robust Identification of Molecular Subtypes in Colorectal Cancer by Immunohistochemistry. *Clin. Cancer Res.* (2016).