Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Oncologia, Ematologia e Patologia

Ciclo XXXI

Settore Concorsuale: 06/A2

Settore Scientifico Disciplinare: MED04

Functional relationship of colorectal cancer-associated glycans with malignant phenotype and transcriptome changes

Presentata da: **Inês Isabel Gomes Ferreira**

Coordinatore Dottorato: Prof. Pier-Luigi Lollini

Supervisore: Prof. Fabio Dall'Olio

Esame finale anno 2019

Acknowledgements

To the Marie Sklodowska Curie European committee for awarded me as one of the PhD students of the project H2020 GlyCoCan;

To the Department of Experimental, Diagnostic and Specialty Medicine from *Università di Bologna*, for hosting me in this PhD project;

To Professor Fabio Dall'Olio for accepting me as his PhD student in the Department of Experimental, Diagnostic and Specialty Medicine from *Università di Bologna*; for all the guidance and all knowledge passed to me during my period in his laboratory;

To Professor Paula Videira from *New University of Lisbon* (Lisbon-Portugal) and Doctor Daniel Spencer from *Ludger* (Oxford-United Kingdom) for accepting me during my period aboard in their research laboratories;

To Doctor Mariella Chiricolo and Doctor Nadia Malagolini for all the friendship, technical support and guidance during this PhD;

To Doctor Francesco Vasuri from *S.Orsola-Malpighi Hospital* (Bologna-Italy) for all the work collaboration regarding the clinical samples;

To Professor Manuela Ferracin from *Università di Bologna* for all the work collaboration regarding microarray analysis/TCGA database and for all the wise advices;

To Professor Yoann Rombouts and Doctor Alexandre Stella from *Institut de Pharmacologie et de Biologie Structurale* (Toulouse-France) for the glycoproteomic analysis;

To Doctor Rad Kozak from *Ludger* (Oxford-United Kingdom) for the help in glycomic data analysis;

To LTTA microarray facility from *University of Ferrara* for all the technical support in the transcriptomic analysis;

To my laboratory colleagues (Mariangela Catera, Giulia Venturi and Michella Pucci). More than colleagues, Friends! Thank you for all the moments;

To all my department colleagues for the good environment, fun and sweet moments;

To all my housemates at Bologna, for the great environment and all the friendship, with a special thanks to Carmen Santomauro for all laughs, friendship and craziness;

To all my Portuguese friends specially my ever friends Susana Dias and Rita Martins;

To my GlyCoCan family, thank you for these three crazy years. A group of 13 young students so different but at the same time, with such a great team spirit. Once a GlyCoCan, always GlyCocan;

To all my family for all the personal support;

To Garcez for all the love, kindness, patience and support during these years apart.

Abstract

Tumor-associated glycosylation alterations are both tumor markers and engines of cancer progression. Glycosyltransferases ST6GAL1 and B4GALNT2 (and their cognate antigens Sia6LacNAc and Sd^a, respectively) are associated with colorectal cancer (CRC) but it is not fully clear their biological and clinical significance.

In this project, we explored the clinical relevance of both glycosyltransferases by interrogating The Cancer Genome Atlas (TCGA) database while the phenotypic/transcriptomic effects of ST6GAL1/B4GALNT2 overexpression were studied in genetically modified CRC cell lines. Transcriptomic data from CRC patients in TCGA database suggested a moderate impact of ST6GAL1 on CRC progression, although it was not possible to define a clear role as anti- or pro-tumoral effect. Transcriptomic analysis of ST6GAL1-transduced cell lines revealed a much deeper effect of ST6GAL1 on gene expression in SW948 than in SW48. Notably, no genes showed parallel modulation by ST6GAL1 in the two cell lines. The overexpression of ST6GAL1 induced opposite effects on growth in soft agar and wound healing in the two cell lines. These results indicate that the impact of a cancer-associated glycosyltransferase change on phenotype/transcriptome can be extremely variable, depending on the molecular context of the tumor cell. On the contrary, transcriptomic analysis of B4GALNT2-modified cell lines together with TCGA database survey demonstrated a strong impact of B4GALNT2 on the transcriptional activity of colon cancer cells, in particular its association with a better prognosis. We suggest an anti-tumoral role of this glycosyltransferase in CRC.

We also investigated the glycan changes related to ST6GAL1 and B4GALNT2 expression in a small cohort of tissues/plasma samples as well as the *N*-glycomic profile of CRC, normal and polyp tissues. We found an increase of ST6GAL1 activity in CRC and inflammatory bowel disease plasma samples comparing with plasma from healthy donors. A different Sd^a protein carrier pattern was observed between healthy donors and CRC plasma samples. β-arrestin 1 is a possible candidate as the Sd^a carrier protein around 55kDa in plasma samples. Future studies are needed to validate this candidate which can be a potential useful biomarker to help a better CRC diagnosis. The alterations found in the *N*-glycan pattern among normal, cancer and polyp tissues highlight the importance of *N*-glycome as a molecular signature in cancer.

Index

Index of Figures

Index of Tables

Abbreviations

A|

ACTB - β-actin

ADCC - Antibody dependent cellular cytotoxicity

APC - Adenomatous polyposis coli

Asn - Asparagine

B|

BRAF - Murine sarcoma viral oncogene homolog B1

BSA - Bovine serum albumin

B4GALNT2 - β1,4-N-acetylgalactosaminyltransferase II

C|

CD - Crohn disease

cDNA - Complementary DNA

CIMP - CpG island methylator phenotype

CIN - Chromosomal instability

CRC- Colorectal cancer

CSC - Cancer stem cell

CV - Coefficient of variation

C1GalT1 - Core 1 synthase

E|

EGFR - Epidermal growth factor receptor

ER - Endoplasmic reticulum

ERβ - Estrogen receptor-β

F|

FAK - Focal adhesion kinase

FTA - Phosphotungstic acid

Fuc - Fucose

FUT - Fucosyltransferase

G|

Gal - Galactose

GalNAc - N-acetylgalactosamine

GAPDH - Glyceraldehyde 3-phosphate dehydrogenase

Glc - Glucose

GlcNAc - N-acetylglucosamine

GlcNAcT-III - N-acetylglucosaminyltransferase 3

GlcNAcT-V - N-acetylglucosaminyltransferase 5

GT - Glycosyltransferase

H|

HNPCC - Hereditary nonpolyposis colon cancer

HPLC - High-performance liquid chromatography

I|

IBD - Inflammatory bowel disease

IGF2R - Insulin growth factor type 2 receptor

K|

KRAS - Kirsten rat sarcoma viral oncogene homolog

L|

LLO - Lipid-linked oligosaccharide

LOH - Loss of heterozygosity

M|

Man - Mannose

MMR - DNA mismatch repair

mRNA - Messenger RNA

MS - Mass spectrometry

MSI - Microsatellite instability

MSI-H - Microsatellite instability high

MSI-L - Microsatellite instability low

MSS - Microsatellite stable

N|

Neu5Ac -N-acetylneuraminic acid or sialic acid

O|

OST - Oligosaccharyltransferase

P|

PBS-T - Phosphate buffer saline with 0.1% Tween-20

PNGase F - Peptide *N*-glycosidase F

ppGalNAcT - Polypeptide-N-acetyl-galactosaminyltransferase ppGalNAcT

PSA - Polysialic acid

R|

RNA-seq - RNA sequencing

RT-PCR - Real Time-Polymerase Chain Reaction

S|

Sia6LacNAc - α2,6-sialylated lactosamine

sLe^a - Sialyl lewis a

sLe^x - Sialyl lewis x

SNA - *Sambucus nigra* agglutinin

ST - Sialyltransferase

sT - Sialyl T

sTn - Sialyl Tn

ST6GAL1 - β-galactoside α2,6-sialyltransferase

T|

T antigen - Thomsen–Friedenreich antigen

TCGA - The Cancer Genome Atlas

TGF - Transforming growth factor

TGFBR2 - Transforming growth factor β type 2 receptor

TNFR1 - Tumor necrosis factor receptor 1

TP53 - Tumor protein p53

U|

UC - Ulcerative colitis

UTR - Untranslated region

V|

VEGF - Vascular endothelial growth factor

VEGFR - Vascular endothelial growth factor receptor

DISSERTATION

Chapter I - Introduction

1. Colorectal cancer

1.1 Epidemiology

Colorectal cancer (CRC) is the second leading cause of cancer-related death and the third most common cancer throughout the world, accounting for over 10% of all cancers diagnosed in 2018[1]. By 2030, CRC is estimated to rise with more than 2.2 million new cases and 1.1 million cancer deaths[1].

CRC incidence and mortality rates vary widely across the world (**Figure I. 1**), occurring mainly in developed regions with a western culture[1]. However, the incidence is increasing rapidly in many low-income and middle-income countries, particularly those undergoing social and economic changes. In several high-income countries such as USA, CRC incidence and mortality have been diminishing or stabilising because of improvements in early detection, prevention, perioperative care as well as chemotherapy and radiotherapy techniques.

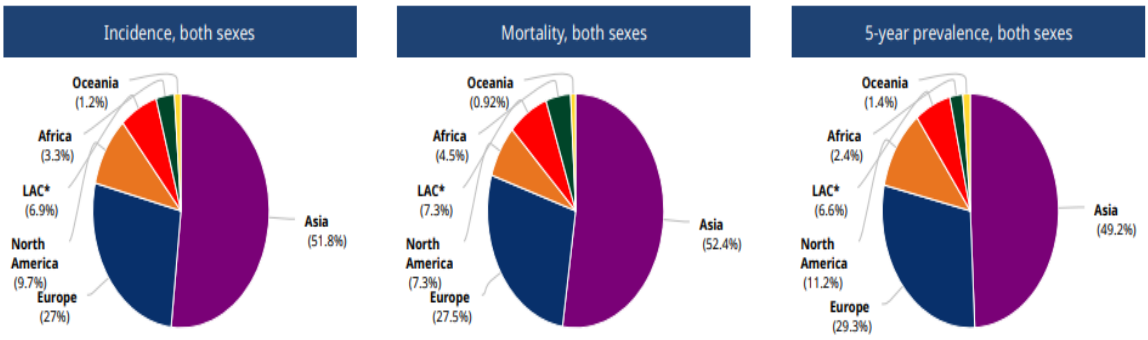


Figure I. 1 - Worldwide colorectal cancer estimated incidence, mortality and prevalence (5-years) for both genders in 2018. LAC: Latin America and the Caribbean. Data available from: http://globocan.iarc.fr, accessed on 1/10/2018.

CRC survival rates decrease with age and is highly dependent mainly with the stage of the disease at diagnosis: 90% 5-year survival for localized CRC (early stage), 70% 5-year survival when cancer has spread to surrounding tissues/organs (nearby lymph nodes) and 10% 5-year survival for patients diagnosed with metastatic CRC (advanced stage). In the cases of unknown stage, 5-year survival is about 35%[2].

1.2 Molecular pathogenesis

Colorectal cancer is a multifactorial disease that is developed through a multistep process with progressive accumulation of genetic and/or epigenetic changes. Two major carcinogenesis pathways have been identified in CRC: the most common one, the chromosomal instability (CIN) counting for 85% of CRC cases and microsatellite instability (MSI) corresponding to 15% of the cases[3].

Chromosomal instability model is characterized by the presence of structural and numerical chromosomal abnormalities in cancer cells. A key early mutation is present in the adenomatous polyposis coli (*APC*) tumour suppressor gene located on chromosome 5q21. This gene is involved in the Wnt pathway signalling: when *APC* is mutated, there is an accumulation of cytoplasmic β-catenin which allows its translocation into the nucleus and subsequently the binding to the T-cell factor (TCF)/lymphoid enhancer factor (LEF) family of transcription factors, altering the expression of several genes involved in proliferation, differentiation, migration, adhesion and apoptosis of colon cells[3,4]. APC plays also a role in the control of cell cycle progression and stabilization of microtubules thus contributing potentially to the initial chromosomal instability. These APC changes are presented mainly in adenomas, benign colorectal precursor lesions[5].

The adenoma-carcinoma sequence (**Figure I. 2**) is then followed by mutations in *KRAS* (Kirsten rat sarcoma viral oncogene homolog) proto-oncogene and lastly by the inactivation of *TP53* (tumour protein p53) tumour suppressor gene[3]. *KRAS* gene encodes for the GTPase protein involved in signal transduction pathways such as mitogen-activated protein kinase (MAPK). Mutation of this gene (normally at codon 12 and 13 of exon 1) leads to the permanently activated form of the protein affecting multiple cellular processes involved in cancer transformation[6]. *TP53* gene is located on chromosome 17p and the loss of heterozygosity (LOH) at this chromosome together with mutations in *TP53* is considered a late event in the process of CRC carcinogenesis[7]. *TP53* is a tumour suppressor gene that activates several genes involved in cell cycle, apoptosis, autophagy, cellular metabolism among others. A frequently co-occurring molecular alteration with the TP53 inactivation is the LOH at 18q21, a region containing some tumour suppressor genes including *SMAD2*, *SMAD4* and deleted in colorectal cancer (DCC)[3,8].

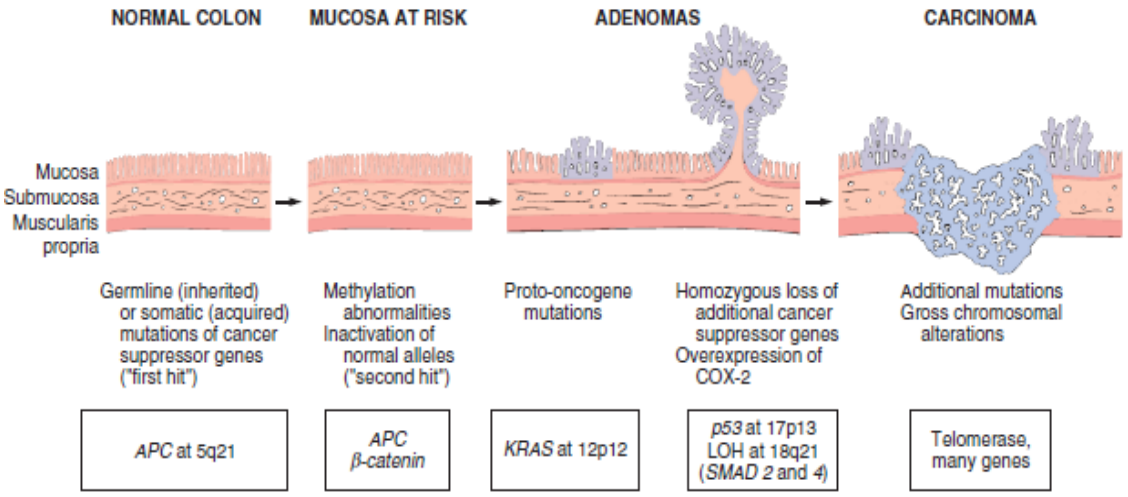


Figure I. 2 - Molecular and morphologic alterations in the adenoma-carcinoma sequence. Typically, involves an early event mutation in *APC* tumor suppressor gene. This is followed by additional mutations including *KRAS* oncogene and other tumor suppressor genes such as *TP53*, *SMAD2* and *SMAD4*. This molecular pathway is characterized also by aberrant chromosomal alterations, leading altogether to the formation of carcinoma. Even though there may be a preferred chronological order for these alterations, it is the combined effect of all these mutations rather than their order of occurrence that seems most critical in CRC. LOH - Loss of heterozygosity. Adapted from "Robbins Basic Pathology 9th Edition" (2012) by Kumar, V. *et al.*

The second main mechanism is the microsatellite instability pathway (**Figure I. 3**), characterized by the derangement of the DNA mismatch repair (MMR) genes. These genes including *MLH1* and *MSH2* can be mutated in the germline, originating the hereditary nonpolyposis colon cancer (HNPCC) or mutated sporadically (the most common form)[9]. In both cases, their inactivation leads to the instability of short repeat DNA sequences known as microsatellites. They can occur in intergenic regions as well as in genes controlling apoptosis, signal transduction and cell cycle. Normally, they have fewer mutations in *KRAS* and *TP53* genes but mutations in *BRAF* (Murine sarcoma viral oncogene homolog B1) oncogene are frequently observed in sporadic MSI cancers (not present in HNPCC form)[10]. Moreover, mutations affecting transforming growth factor β type 2 receptor (TGFBR2), pro-apoptotic tumor suppressor gene *BAX* and insulin growth factor type 2 receptor (IGF2R) are found in MSI tumors[3].

A panel comprising 5 microsatellites markers (D2S123, D5S346, D17S250, BAT26 and BAT25) was proposed as a standard test for MSI: microsatellite instability high (MSI-H) when at least two markers are positive and microsatellite instability low (MSI-L) when just one marker is positive[11]. Clinically, MSI cancers have particular clinical features: often localized in proximal colon, poorly differentiated with a mucinous phenotype, high infiltration of lymphocytes and a better overall

survival/prognosis comparing with patients with CIN positive CRC[3,8]. From a morphological point, MSI cancers derived from serrated adenomas or hyperplastic polyps, different from the classical adenoma-carcinoma sequence common in CIN cancers.

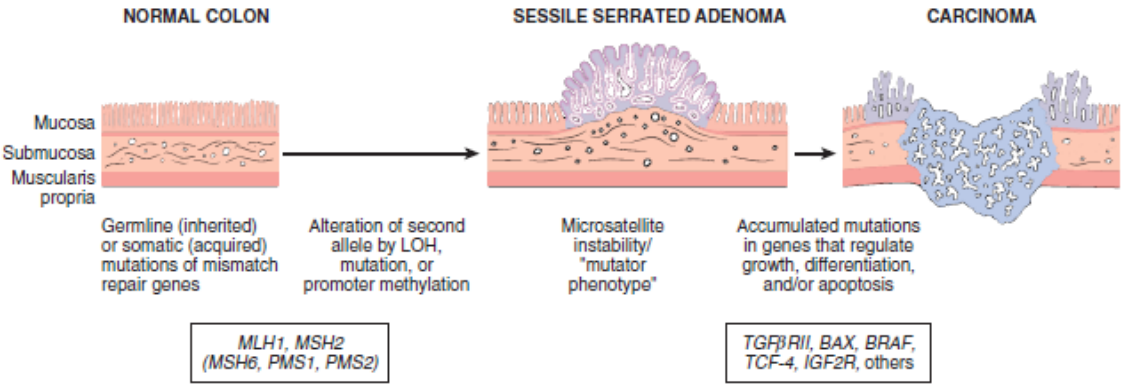


Figure I. 3 - Molecular and morphological alterations in microsatellite instability pathway of colorectal cancer transformation. It is characterized by the inactivation of DNA mismatch repair genes (mainly methylation of *MLH1*) leading to the accumulation of mutations in short repeat DNA sequences known as microsatellites. These mutations can affect genes involved in cell survival, regulation of cell growth and proliferation such as TGF-β type II receptor, pro-apoptotic protein *BAX* and *BRAF* oncogene. LOH - Loss of heterozygosity. Adapted from "Robbins Basic Pathology 9th Edition" (2012) by Kumar, V. *et al.*

A third particular pathway of CRC transformation is the CpG island methylator phenotype (CIMP), usually overlapping with sporadic MSI cancer features. It is characterized by the hypermethylation of the CpG island sites promoters resulting in the inactivation of several tumor related-genes[12].

1.3 Aetiology and risk factors

Several risk factors can influence the incidence and development of CRC. Inflammatory bowel disease (IBD), family history of CRC and/or adenomatous polyps are factors associated with a high risk of CRC[13].

IBD is characterized by a chronic inflammation of the gastrointestinal mucosa resulting from the oxidative stress-induced DNA damage that leads to the activation of oncogenes and silencing of tumour suppressor genes[14]. The two most common forms of IBD are ulcerative colitis (UC) and Crohn disease (CD). The former is typically localized in the large intestine with continuous inflammation of the colon and only affects the inner most lining of the colon. On the contrary, CD can affect any part of the gastrointestinal tract, occurs in all the layers of the bowel walls and is characterized by the presence of skip lesions. Regarding the risk of developing CRC, people with

ulcerative colitis are more prone to develop cancer comparing with those suffering from Crohn disease[15].

Other lifestyle and diet risk factors including smoking, excessive alcohol consumption, red meat intake, lack of physical activity, low vegetable/fruit consumption are also linked with an increased risk of CRC[16].

1.4 Diagnosis, staging classification and treatment

Diagnosis of CRC is made histologically from biopsy samples collected during colonoscopy. However, some patients can have an incomplete colonoscopy because of the invasiveness of this procedure or other technical difficulties. In those cases, further computed tomography colonography (CTC) can contribute to an appropriate CRC diagnosis. In 20% of CRC patients, metastasis is present at the time of first diagnosis[5]. The most common localization of CRC metastasis is the liver and for this reason, liver imaging should be realized in all patients[17].

CRC staging uses the TNM (tumour-node-metastasis) classification from the International Union Against Cancer (UICC)[18] (**Figure I. 4**). This classification is based on the local invasion depth (T stage), lymph node involvement (N stage) and presence of distant metastasis (M stage). The information of those categories is combined into an overall stage definition (stage I, II, III or IV).

	Definition
T stage	
Tx	No information about local tumour infiltration available
Tis	Tumour restricted to mucosa, no infiltration of lamina muscularis mucosae
T1	Infiltration through lamina muscularis mucosae into submucosa, no infiltration of lamina muscularis propria
T2	Infiltration into, but not beyond, lamina muscularis propria
T3	Infiltration into subserosa or non-peritonealised pericolic or perirectal tissue, or both; no infiltration of serosa or neighbouring organs
T4a	Infiltration of the serosa
T4b	Infiltration of neighbouring tissues or organs
N stage	
Nx	No information about lymph node involvement available
N0	No lymph node involvement
N1a	Cancer cells detectable in 1 regional lymph node
N1b	Cancer cells detectable in 2–3 regional lymph nodes
N1c	Tumour satellites in subserosa or pericolicor perirectal fat tissue, regional lymph nodes not involved
N2a	Cancer cells detectable in 4–6 regional lymph nodes
N2b	Cancer cells detectable in 7 or greater regional lymph nodes
M stage	
Mx	No information about distant metastases available
M0	No distant metastases detectable
M1a	Metastasis to 1 distant organ or distant lymph nodes
M1b	Metastasis to more than 1 distant organ or set of distant lymph nodes or peritoneal metastasis

	T	N	M
Stage 0	Tis	N0	M0
Stage I	T1/T2	N0	M0
Stage II	T3/T4	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
Stage III	Any	N+	M0
IIIA	T1–T2	N1	M0
	T1	N2a	M0
IIIB	T3–T4a	N1	M0
	T2–T3	N2a	M0
	T1–T2	N2b	M0
IIIC	T4a	N2a	M0
	T3–T4a	N2b	M0
	T4b	N1–N2	M0
Stage IV	Any	Any	M+
IVA	Any	Any	M1a
IVB	Any	Any	M1b

Figure I. 4 - Classification of colorectal cancers according to T stage (invasion depth), N stage (lymph node involvement) and M stage (presence of metastasis) - left side. Overall CRC staging from early stages (Stage I) to advanced stages (Stage IV) - right side. Adapted from "Seminar-Colorectal cancer" (2014) from Brenner, H. *et al.*

Histologically, CRC can be also classified based on the degree of preservation of normal glandular architecture and cytological features (well differentiated, moderately differentiated or poorly differentiated). About 20% of CRC cases are poorly differentiated, presenting a poor prognosis[19]. CRC can present also a mucinous phenotype, characterized by the presence of a prominent intracellular accumulation of mucins. This cancer characteristic turns out normally in a very aggressive cancer with a poor prognosis[20].

Regarding treatment options, surgery is the most common type of treatment, depending on several characteristics of the tumour such as its location, existence and extent of metastasis. If cancer is found only in a single polyp, it can be removed during colonoscopy. In early stages of cancer, treatment can include chemotherapy, radiotherapy, radiofrequency ablation, cryosurgery or targeted therapy. Radiofrequency ablation consists in the use of electrodes to kill cancer cells while in cryosurgery freezing techniques are used to destroy the tumour. The most common drug used in CRC chemotherapy is 5-fluorouracil[16].

2. Glycosylation

Glycosylation is defined as the process of the enzymatic addition of glycans to a non-carbohydrate moiety such as proteins, lipids or other organic compound. It is important not confuse the term glycosylation with glycation which is a non-enzymatic and irreversible process of adding glycans to proteins that is present in several diseases with particular emphasis in diabetes[21]. The glycome, analogous to genome, transcriptome and proteome, involves all glycans synthetized by an organism. It is estimated to be 10-10^4 times bigger than the proteome, depending on the species[22].

In mammals, the major glycans comprise 10 monosaccharide "building blocks": glucose (Glc), galactose (Gal), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), mannose (Man), fucose (Fuc), xylose (Xyl), glucuronic acid (GlcA), iduronic acid (IdoA) and 5-N-acetylneuraminic acid (Neu5Ac or sialic acid). The structural variety of glycans is enormous and arises from the available number of monosaccharides building blocks, glycosidic bond composition, anomeric configuration, presence of branching or linear structures and carbohydrates modifications such as sulfatation and phosphorylation[23,24]. Glycans are structures involved in processes such as cell adhesion, molecular trafficking and clearance,

receptor activation, cell growth and signal transduction[21,24]. Furthermore, the addition of glycans to proteins can affect the intrinsic properties of the modified protein including its solubility, proper folding, functional group orientation and protection from proteases[25].

Glycoconjugates can be classified into 5 main groups: glycoproteins, proteoglycans, glycosphingolipids (glycolipids), glycosylphosphatidylinositol (GPI)-linked proteins and *O*-GlcNAc glycoproteins (**Figure I. 5**). This PhD project is mainly focused on a particular group of glycoconjugates namely glycoproteins. Glycoproteins carry one or more glycans covalently linked to a polypeptide backbone. The major portion of the cell membrane is covered with a dense array of glycoproteins also known as glycocalyx. These glycoconjugates are also found as an integrant part of the extracellular matrix and moreover they can be secreted into biological fluids such as plasma[26].

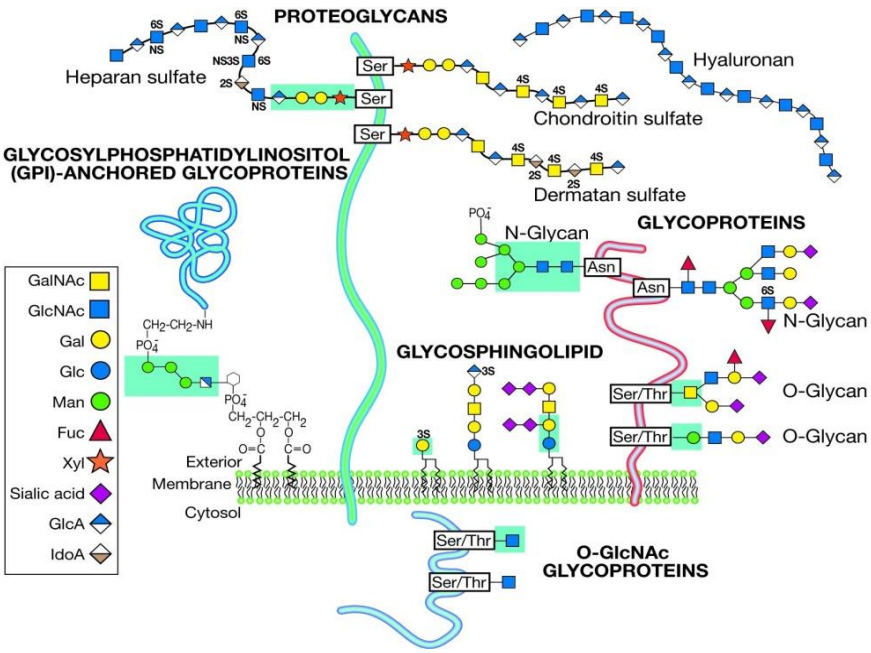


Figure I. 5 - Common classes of animal glycoconjugates. Proteins can be glycosylated by the covalently attachment of a saccharide to a polypeptide backbone, via *N*- linkage to asparagine (Asn) forming *N*-glycans or *O*-linkage to serine/threonine (Ser/Thr) forming *O*-glycans. Proteoglycans are glycoconjugates that present one or more glycosaminoglycans such as chondroitin sulfate, heparan sulfate and keratan sulfate. An exception is the hyaluronan, which is a glycosaminoglycan found as a free sugar chain. Glycosphingolipids are major components of cell plasma membrane, composed of glycans linked to a lipid ceramide. Glycosylphosphatidylinositol (GPI)-linked proteins are anchored in the outer leaflet of the plasma membrane by a glycan covalently linked to phosphatidylinositol. Various cytoplasmic and nuclear proteins contain *O*-linked N-acetylglucosamine (O-GlcNAc). Adapted from "Essentials of Glycobiology" (2009) by Varki, A. *et al.*

Glycosylation of proteins requires the specific action of glycosyltransferases (GTs), a large family of enzymes responsible for the transference of monosaccharide residues to an acceptor substrate using nucleotide-sugar donor as activated donor substrates[27]. GTs are specific for a nucleotide-sugar donor but may recognize more than one different acceptor. They are defined by the sugar that they transfer and the type of linkage that they catalyse, being located mainly in endoplasmic reticulum (ER) and Golgi complex. Besides GTs, glycosidases are also involved in the metabolism of glycoproteins, catalysing the hydrolysis of glycosidic linkages.

2.1 *N*-glycans

N-glycosylation is the most studied form of protein glycosylation in eukaryotic organisms (about 90% of eukaryotic proteins carry *N*-glycans)[28].

N-glycosylation biosynthesis (**Figure I. 6**) begins with the synthesis of a lipid-linked oligosaccharide (LLO) constituted by 2 GlcNAc, 9 Man and 3 Glc residues covalently attached to a lipid dolichol on the cytoplasmic face of ER. LLO biosynthesis is performed by a series of GTs that are encoded by asparagine linked glycosylation (*ALG*) genes. The formed glycan is then transferred *en bloc* to an asparagine (Asn) residue of a nascent polypeptide by a multi-enzyme complex named oligosaccharyltransferase (OST). There is a minimal consensus sequence, comprised of an Asn-X-Serine/Threonine tripeptide where X can be any amino acid except proline, that can accept a *N*-glycan. The next steps are a series of processing reactions trimming *N*-glycans: trimming of two Glc residues by α-glucosidases I and II originates $GlcMan_9GlcNAc_2$ structure that acts as a ligand for two chaperones, calnexin and calreticulin, molecules helping in the process of protein folding[28,29]. Misfolded proteins that do not pass quality control are translocated back to the cytosol for proteasomal degradation.

Properly folded proteins are transported to the Golgi complex where the first step is the demannosylation by the Golgi α-mannosidase I, forming the $Man_5GlcNAc_2$ structure which is the main substrate for the enzyme N-acetylglucosaminyltrasferase-I (GnT-I, product of the *MGAT1* gene). As the pathway progresses through the Golgi complex, the $GlcNAc_1Man_5GlcNAc_2$ structure can be further modified by the removal of 2 Man residues by α-mannosidase II and by the addition of a second GlcNAc residue, catalyzed by N-acetylglucosaminyltrasferase-I (GnT-II, product of the *MGAT2* gene).

N-glycans can be further decorated by the action of a number of GTs that add Gal, Fuc, sialic acid and sulphate to the antennae resulting in a heterogeneous collection of mature glycoconjugates[25,27,28,30].

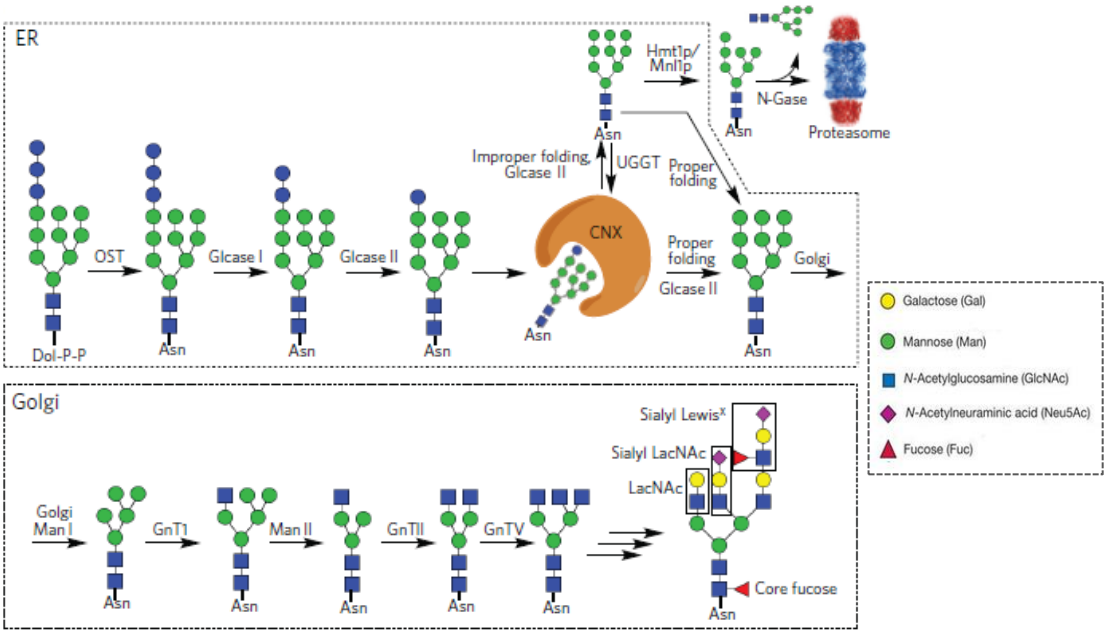


Figure I. 6 - Schematic representation of *N*-linked glycoproteins biosynthesis. The entire oligosaccharide $Glc_3Man_9GlcNAc_2$ is transferred to the asparagine residue of a nascent polypeptide chain by the complex oligosaccharyltransferase (OST). After the action of glucosidase I and II (Glcase I and II), protein glycosylation contributes to the quality control of protein biosynthesis, through the chaperones calnexin (CNX) and calreticulin. Glycoproteins with incorrect folding are transported to cytosol where a *N*-glycanase (*N*-Gase) removes the *N*-glycans and the resulting protein is delivered to proteasomal degradation. Properly folded proteins are transported to the Golgi, where an array of glycosyltransferases and glycosidases diversifies the various antennae of the glycans to give an array of complex structures. UGGT, UDP–glucose-glycoprotein glucosyltransferase; Hmt1p, HnRNP methyltransferase 1; Mnl1p, mannosidase-like protein 1; Man, α-mannosidase; GnT, N-acetylglucosaminyltransferase. Adapted from "Adaptive immune activation: glycosylation does matter" (2013) by Wolfert, MA. and Boons, GJ.

N-glycans share a common sugar core structure (Manα1,6(Manα1,3)Manβ1,4GlcNAcβ1,4GlcNAcβ1-Asn) and they are classified in three main types: oligomannose or high mannose, in which only mannose residues are attached to the core; complex, in which the core is extended by GlcNAc residues in both mannose arms; and hybrid, in which the Manα1,6 arm of the core contains only mannose residues whereas the Manα1,3 arm is extended by complex type structures[27]. They also include some hybrid and complex type glycan determinants (**Figure I. 7**): bisecting GlcNAc structure, where a third GlcNAc residue can be linked to the innermost mannose residue by the enzyme GlcNAc-transferase III; paucimannose structure, truncated structure from the N-glycan core; core fucosylation, where a fucose

residue is linked to the first GlcNAc of the chain by the fucosyltransferase VIII and further galactosylation, fucosylation and sialylation of the other antennas[27,31].

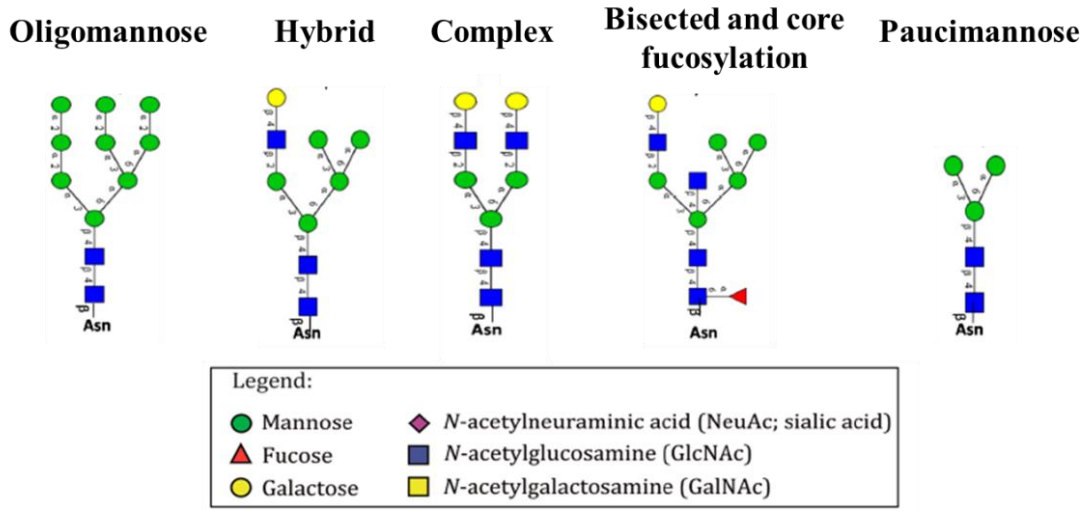


Figure I. 7 - Main types of *N*-glycans in vertebrates. There are three main types of *N*-glycans present in a mature glycoprotein: oligommanose, hybrid and complex. All types share a common core glycan structure that can be elongated by core fucosylation, bisecting GlcNAc and other glyco determinants. Paucimannose structures are characterized by truncated glycans. Adapted from "Mass spectrometry-based N-glycomics of colorectal cancer" (2015) by Sethi, MK. and Fanayan, S.

2.2 *O*-glycans

Several types of *O*-glycosylation have been identified but the most common form is the mucin-type *O*-glycosylation also denominated *O*-GalNAc glycans. Mucins are heavily *O*-glycosylated proteins (glycan moiety may comprise 80% of the molecule weight) that can be soluble, secreted or expressed in the cell membrane. They are present at many epithelial surfaces including respiratory, reproductive and gastro-intestinal tracts, playing an important role in the protection against pathogens[32].

It is characterized by the attachment of glycans to serine (Ser) or threonine (Thr) of a nascent protein through a GalNAc residue. This first step is initiated in Golgi complex and is controlled by a large family of up to 20 genes (*GALNT1-GALNT20*) encoding the polypeptide-N-acetyl-galactosaminyltransferase (ppGalNAcT). Subsequently, *O*-linked GalNAc residues can be further modified or extended by specific GTs, resulting in several heterogeneous structures. In literature, there are eight *O*-GalNAc glycan core structures described, being the most common ones those from core 1 to core 4[32,33] (**Figure I. 8**).

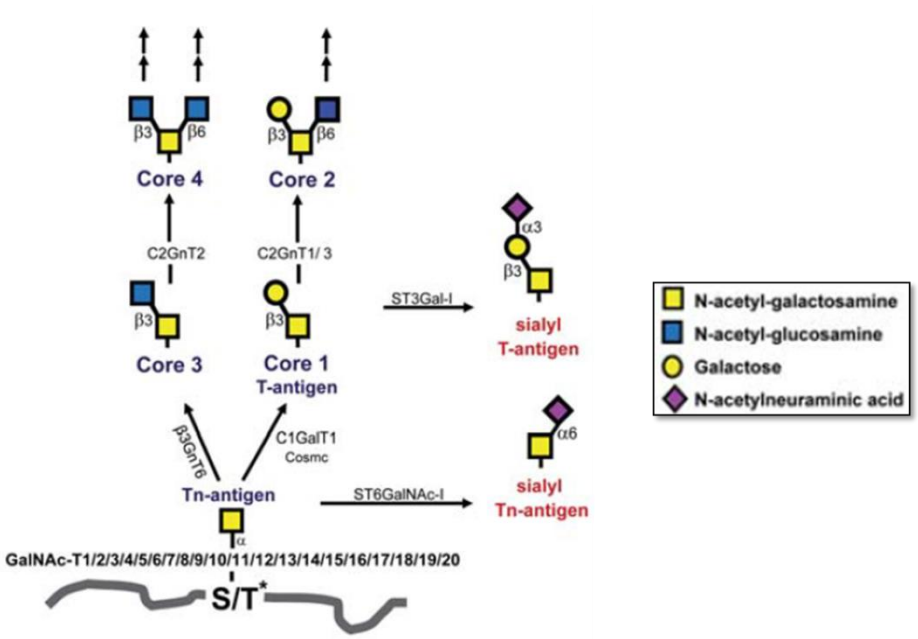


Figure I. 8 - Common *O*-GalNAc glycan cores structures and their biosynthetic pathway. It is initiated by up to 20 ppGalNAcTs forming the Tn antigen, which may be elongated by the core 1 synthase (C1GalT1) or core 3 synthase (β3GnT6), forming the T antigen and core 3 structures, respectively. The activity of C1GalT1 enzyme depends on the presence of the chaperone Cosmc. Both Tn and T antigens may be modified by sialic acid to form sialyl-Tn or sialyl-T antigens, correspondingly. Another common core structure contains a branching N-acetylglucosamine attached to core 1 and is named core 2. The different core structures can be further elongated and modified by several GTs originating several complex *O*-GalNAc glycans. C2GnT2, core 4 synthase; C2GnT1/3, core 3 synthase; ST3Gal-1, β-galactoside α-2,3-sialyltransferase 1; ST6GalNAc-1, GalNAc α-2,6-sialyltransferase 1. Adapted from "Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family" (2012) by Bennet, EP. *et al.*

2.3 Glycosylation in cancer

Altered glycosylation is a universal feature of cancer cells, being one of the main characteristics associated with malignant transformation and tumour progression.

In cancer cells, changes in glycans can take a variety of forms: loss or excessive expression of certain structure, incomplete or truncated glycans and appearance of novel antigens (less common way)[27]. One of the main mechanisms responsible for those aberrant glycan changes is the differential expression of glycosyltransferases involved in glycan biosynthesis. Some other factors can influence protein glycosylation in cancer cells such as availability and localization of nucleotide sugar donors or substrates, competition between GTs for similar glycan acceptors, levels of expression of glycosidases, inefficiency of enzymatic reactions and also localization of GTs[25]. Some of the most relevant cancer-associated glycans will be briefly described below.

Bisected and branched *N*-glycan structures

Altered glycosylation in *N*-glycans is mainly associated with an increase of β1–6 branching structures caused by the enhanced expression of N-acetylglucosaminyltransferase 5 (GlcNAcT-V) (**Figure I. 9**). The alteration in the expression of this enzyme seems to be a consequence of the increase on its gene transcription (*MGAT5*). This feature has been associated with an increase of polylactosaminic structures potentially recognized by galectins, which can be further sialylated and fucosylated, being recognized later also by selectins[27]. The increase of branching structures has been correlated with tumour growth and metastasis[24,34]. Indeed, cell lines with GlcNAcT-V overexpression displayed an increased frequency of metastasis in animal models while in another study, *MGAT5*-deficient mice showed a reduction in the growth and metastasis of breast tumours[24]. N-acetylglucosaminyltransferase 3 (GlcNAcT-III) is encoded by the gene *MGAT3* and catalyses the formation of glycans with a bisecting GlcNAc β1,4-linked to the innermost Man residue of the core. This modification supresses the processing and elongation of *N*-glycans and decreases *N*-glycan branching structures[35]. In fact, overexpression of GlcNAcT-III in several cancer types supresses the function of growth factor receptors, reducing cancer metastasis[36].

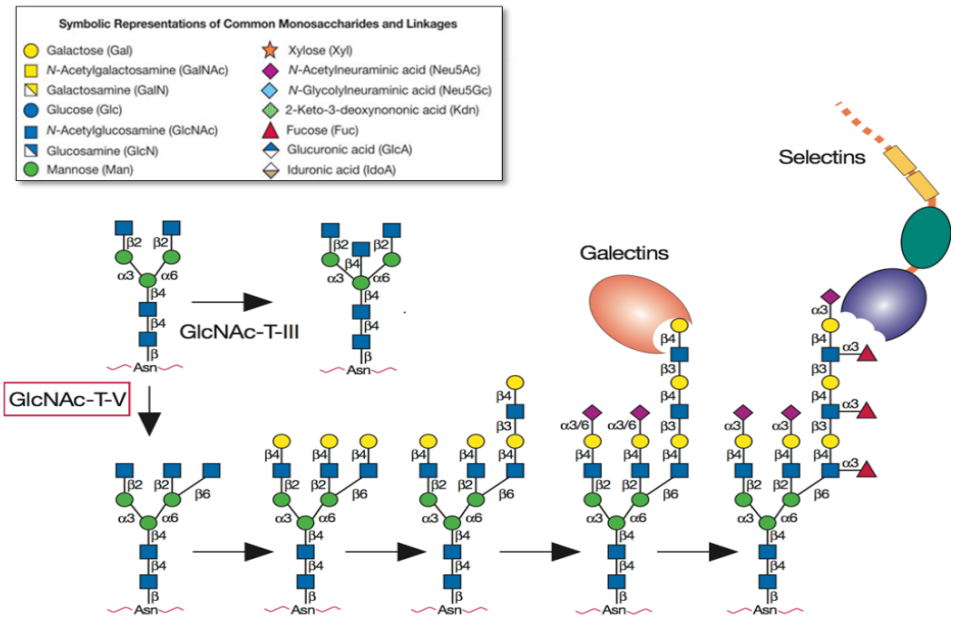


Figure I. 9 - During malignant transformation, there is an increase of GlcNAcT-V activity leading to the formation of branched *N*-glycans. Those structures can be further elongated, being recognized by galectins and/or selectins, lectins with an important role in cancer progression. The formation of bisected *N*-glycans by the expression of GlcNAcT-III is also shown. Adapted from "Essentials of Glycobiology" (2009) by Varki, A. *et al.*

Truncated and incomplete *O*-glycans

In cancers with epithelial origin, mucins appear to be the major carriers of changes in glycosylation. There is an overexpression of mucins observed in several carcinomas and there is also the presence of incomplete *O*-glycans[37]. This last point results in the abnormal expression of truncated glycans such as Thomsen–Friedenreich antigen (T antigen), the Tn antigen and their sialylated forms (sT and sTn antigens, respectively) (for the structures, see **Figure I. 8**)[25].

Changes in the expression of ppGalNAcTs, enzymes responsible for the initiation of *O*-mucin type glycosylation, are frequently observed in cancer[38,39]. Additionally, GTs competing for the same substrate can also induce the expression of truncated glycans in cancer cells that normally are not present in normal glycoproteins. An example is the sialyl Tn (sTn) antigen poorly expressed in normal cells but found highly expressed in most carcinomas including pancreas, stomach, colorectal and bladder cancer[40–42]. Its expression in cancer is highly correlated with poor prognosis and tumour malignancy. The abnormal expression of sTn in tumor cells arises from the overexpression of the GalNAc α-2,6-sialyltransferase I (ST6GalNAc-I) together with the fact of the loss ability to synthetize the core 1 *O*-glycan by C1GalT1 enzyme[43].

Only few glycoproteins are known to present truncated antigens in malignant tissues, being mucin-1 and CD44v6 the main carriers of sTn and sT antigens in colon, gastric and breast carcinomas[44–46]. High levels of sTn expression are associated with an overexpression of mucin-1 and CD44v6 in breast and gastric cancers[47].

Core fucosylation

Core fucose is the product of α1-6 fucosyltransferase VIII (FUT8) that adds a residue of fucose to the innermost GlcNAc residue of *N*-glycan structures. A well-known tumour marker for hepatocarcinomas is α-fetoprotein. However, this protein can be also increased in other diseases such as hepatitis and cirrhosis. Core fucosylated α-fetoprotein is an approved biomarker for the early detection of hepatocellular carcinoma which allows to distinguish it from other pathologies[48]. It has been reported an increase of core fucosylation in several cancers, including lung cancer and breast cancer[49,50]. The presence of core fucose in *N*-glycans is also important in the process of antibody dependent cellular cytotoxicity (ADCC). Indeed, deletion of core fucose from human immunoglobulin IgG1 enhances ADCC activity[51].

Core fucosylation can also modulate the activity of growth factor receptors, integrins and cadherins. In the absence of core fucosylation, TGF-β/Smad2/3 signalling is inactivated and epithelial-mesenchymal transition of renal epithelial cells is inhibited[52].

Polysialic acids

Polysialic acids (PSA) are comprised of long linear arrays of sialic acids α2,8 linked to *N*-linked glycans. The two enzymes responsible for the synthesis of PSA are ST8SiaII and ST8SiaIV. PSA are present frequently in neural cell adhesion molecule 1 (NCAM), being associated with aggressiveness, metastasis and poor prognosis in cancers such as lung cancer, neuroblastoma and glioma[53,54]. Several reports revealed an important role of PSA in the organisms, including nervous system development and maintenance, tissue regeneration and repair. Moreover, it is also implicated in psychiatric diseases such as schizophrenia[55].

Sialyl Lewis antigens

Sialyl Lewis structures are frequently overexpressed in carcinomas, being their overexpression correlated with tumour progression and poor prognosis in several types of cancers such as colorectal, lung and renal cancer[56–58]. These terminal end structures can be found on *N*-glycans, *O*-glycans and in glycosphingolipids. Sialyl Lewis x (sLe^x) and Sialyl Lewis a (sLe^a) result from the mono-fucosylation of the α2,3 sialylated type 2 or type 1 chains[59], respectively (**Figure I. 10**). The sLe^a tetrasaccharide is a tumour associated biomarker widely used in clinical practise (detected by the serological assay CA19.9), being mostly applied in patients detected with pancreatic, colorectal and gastric cancers[60].

The terminal steps for the Sialyl Lewis antigens biosynthesis involve the action of sialyltransferases (STs) and fucosyltransferases (FUTs). ST3GALs transfer a sialic acid residue in α2,3 linkage to a galactose. While α2,3 sialylation of type 1 chains can be mediated only by ST3GAL3, sialylation of type 2 chains can be mediated by ST3GAL3, ST3GAL4 and ST3GAL6[59]. FUTs catalyse the transference of a fucose residue to an acceptor, normally galactose or GlcNAc. There are 5 α-1,3-FUT (FUT3, FUT4, FUT5, FUT6 and FUT7) able to synthesize sLe^x, while sLe^a can be synthesized only by FUT3.

These structures are selectin ligands and when present at the surface of cancer cells, they interact with selectins expressed by the endothelial cells, regulating the metastatic cascade by favouring the arresting of tumour cells on endothelium[61]. Of all selectins, E-selectin is the major receptor that is involved in adhesion events during metastasis, although P- and L-selectin can also contribute to that process[62,63]. Indeed, studies in animal models showed a decrease in tumour metastasis after the inhibition of P-selectin-mediated interactions of platelets with sLe^x/sLe^a antigens present on the surface of cancer cells[64].

Studies in colon cancer elucidate a clear lack of relationship between sialyltransferases/fucosyltransferases modulation and sLe^x expression, suggesting that in this type of cancer, the expression of sialyl lewis antigens can be due to other mechanisms[65,66] (see chapter 2.4.2. B4GALNT2 enzyme and Sd^a antigen).

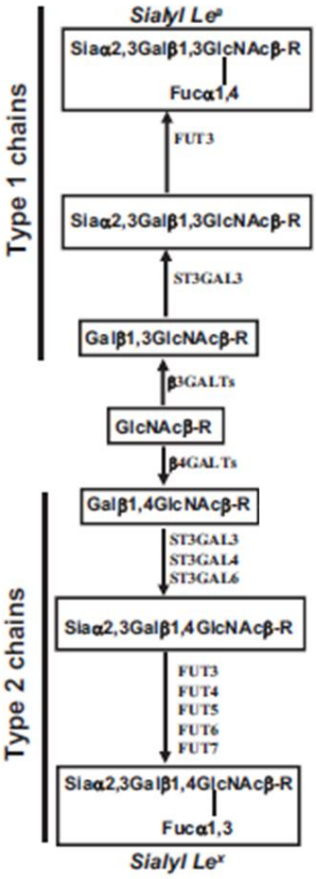


Figure I. 10 - Structures and glycosyltransferases involved in the biosynthesis of Sialyl Lewis antigens. Sialyl Lewis x (type 2 chain structure) and Sialyl Lewis a (type 1 chain structure) are synthetized by sequential enzyme reactions, finishing with the action of sialyltransferases and fucosyltransferases. β3GALTs, β-1,3-galactosyltransferase: β4GALTs, β-1,4-galactosyltransferase. Adapted from "Sialosignaling: sialyltransferases as engines of self-fueling loops in cancer progression" (2014) by Dall'Olio, F. *et al.*

Another common modification in the glycosylation *status* of cancer cells is the overall increase in sialylation (increase in cell-surface sialic acid content), especially in α2,6 and α2,3 linked structures[67]. In this project, we focus on two particular enzymes and respective carbohydrate antigens altered in colorectal cancer. The first one, a sialyltransferase responsible for α2,6 sialylation of glycan structures and the second one, an enzyme that synthesizes the Sd^a antigen and indirectly influences the expression of sLe^x antigen, affecting malignant transformation.

2.4 Colorectal cancer associated carbohydrate antigens and enzymes

2.4.1. ST6GAL1 enzyme and Sia6LacNAc antigen

β-galactoside α2,6-sialyltransferase 1 (ST6GAL1) is a type II membrane protein responsible for the transfer of α2,6-linked sialic acid to lactosaminic chains of *N*-glycans. The glycosidic structure formed is denominated α2,6-sialylated lactosamine (Sia6LacNAc). The sialic acid binding lectin from *Sambucus nigra* agglutinin (SNA) is the most used lectin to detect the expression of Sia6LacNAc in normal and tumour tissues[68].

ST6GAL1 enzyme is normally found in Golgi complex but can be processed to a soluble and secreted form, being also found in body fluids[69]. *ST6GAL1* gene is located in chromosome number 3, spanning at least 145000 base pairs of genomic DNA[70]. Three main ST6GAL1 mRNA isoforms have been identified in humans, only differing in the 5'-untranslated regions (UTRs) (**Figure I. 11**). The first one is considered the product of the basal expression of this gene: it was cloned from a placenta cDNA library; it is transcribed through the P_3 promoter and comprises the 5'-untranslated exons Y and Z (YZ form)[70]. The second form is characteristic of mature B-lymphocytes; it is transcribed through the P_2 promoter, lacks the exons Y and Z but contains the 5'-untranslated region X (X form)[71]. The third and last form (H form) was cloned from the hepatocarcinoma cell line HepG2; it lacks the exons Y, Z or X but contains a short specific sequence in front of exon I[72]. This short sequence is not separated from exon I by introns in genomic DNA therefore it cannot be considered as an exon. In all the three transcripts, translation begins within exon II and finishes within exon VI, originating identical polypeptides. The different transcripts of this gene have been shown to result from the regulation of diverse promoter regions that might be regulated in a tissue-specific manner[73]. YZ and H transcripts were detectable in normal

and colon cancer tissues, but H form has a tendency to accumulate in cancer[68]. In addition, the promoter activity of *ST6GAL1* is also regulated by epigenetic modifications[74].

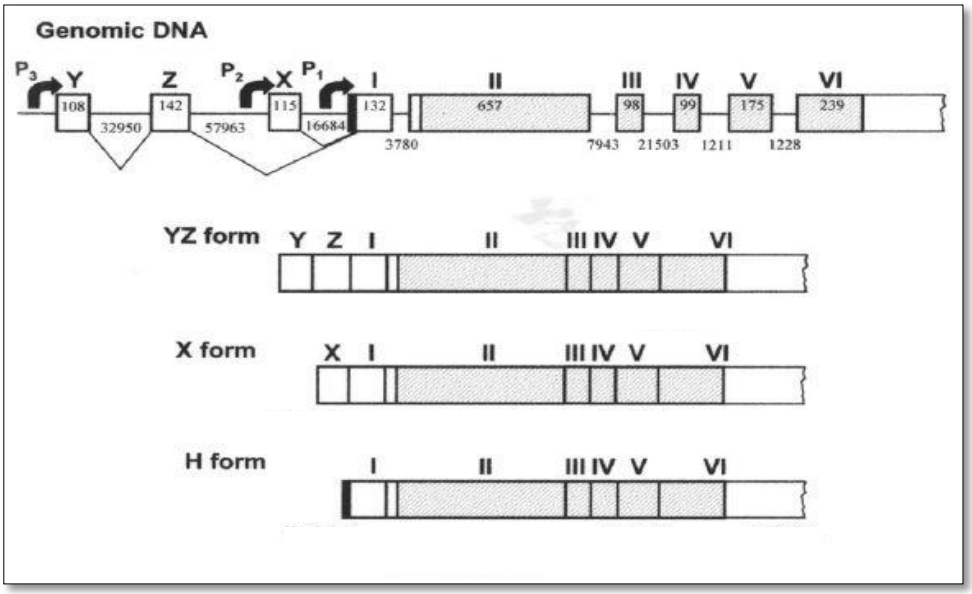


Figure I. 11 - Representation of *ST6GAL1* gene and the three main transcripts identified in humans. The numbers inside the exons and between exons indicate exon and intron length, respectively. The three transcripts share a common coding region (exons II, III, IV, V and IV) and differ in the 5'-untranslated region. Transcription can start by different promoters (from P1 to P3). YZ form is characterized by the presence of exons Y and Z, meanwhile X form is characterized by the exon X. In H transcript, there is a short specific sequence in front of exon I, absent in the other two forms. Adapted from "Expression of β-galactoside α2,6 sialyltransferase and of α2,6-sialylated glycoconjugates in normal human liver, hepatocarcinoma, and cirrhosis" (2004) by Dall'Olio, F. et al.

ST6GAL1 is expressed at different levels among tissues, being the liver the tissue with the highest level of expression. Several studies reported an overexpression of ST6GAL1 and consequently increase of Sia6LacNAc structure in several malignancies including colorectal[75], breast and liver cancers[76]. In CRC, it was observed an increase of SNA reactivity[77,78], ST6GAL1 activity[75] and also ST6GAL1 expression in colon cancer cell lines[79] and cancer tissues[73], being these alterations predictive of poor prognosis[80–82].

ST6GAL1 glycosyltransferase is directly controlled by the oncogene *KRAS*. In fact, the increase SNA reactivity in CRC tissues is mainly associated with a microsatellite stable phenotype most probably due to the dependence of ST6GAL1 on *KRAS* activation, commonly mutated in MSS CRC[83]. Indeed, it has been demonstrated that both *N-ras* and *H-ras* oncogenes stimulate ST6GAL1 expression[84]. In cancer cells with mutated *KRAS*, ST6GAL1 overexpression leads to the α2,6- sialylation of β-1 integrin chains, increasing the binding to extracellular matrix and downstream focal adhesion kinase (FAK) signaling[59,85] (**Figure I. 12**).

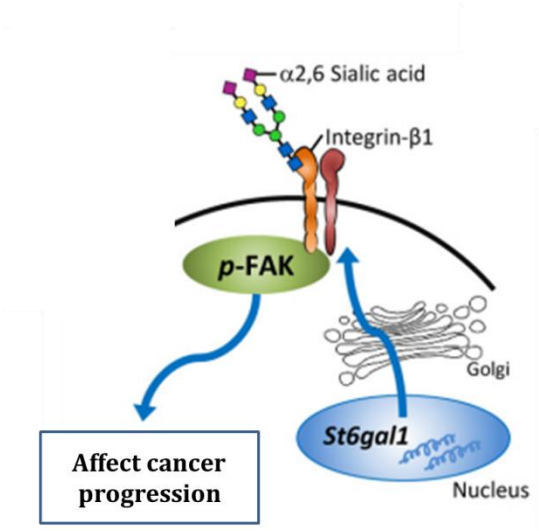


Figure I. 12 - The presence of ST6GAL1 leads to the α2,6-sialylation of integrin β1 chains and consequently activation of downstream signaling pathways, including FAK phosphorylation. Adapted from "The inhibitory role of α2,6-sialylation in adipogenesis" (2017) by Kaburagi, T. *et al.*

ST6GAL1 and its cognate Sia6LacNAc antigen can affect tumorigenesis through different features, being the most important ones resumed below.

Resistance to chemotherapy and radiotherapy

ST6GAL1 is involved in the resistance of tumors to chemotherapy and radiotherapy. Its expression confers drug resistance to ovarian cancer cells and leukemia cells whereas in CRC cells it diminishes the sensitivity to the epidermal growth factor receptor (EGFR)-kinase specific inhibitor *gefitinib*[86–88]. Animal tissues and cell lines exposure to ionizing radiation demonstrated an increase expression of ST6GAL1 while forced expression of ST6GAL1 cDNA in CRC cell lines induced resistance[89].

Cell survival

α2,6 sialylation plays an important role in cell survival. One well studied example is the effect of α2,6 sialic acids on the apoptosis mediated by galectins. Galectins are a family of lectins with affinity for β-galactosides, being capable of induce apoptosis though the interaction with cell-surface and extracellular matrix glycans. It has been reported that α2,6 sialylation of integrins prevents the binding to galectins (specially galectin-3), inhibiting cell apoptosis and promoting tumor cell survival[90]. Another well-studied pathway in tumor survival is the cell death pathway initiated by Fas and tumor necrosis factor receptor 1 (TNFR1). It has been shown that Fas pro-apoptotic activity can be masked by α2,6 sialylation by interfering with the formation of

death inducing signaling complex and limiting the internalization of Fas receptor[91]. Analogous to Fas, ST6GAL1 mediates the α2,6 sialylation of TNFR1 death receptor inhibiting apoptosis[92].

A recent study stated that ST6GAL1 can protect tumor cells from serum growth factor withdrawal. In ovarian and pancreatic cell lines, the overexpression of ST6GAL1 promoted the survival in serum-starved cells by increasing the activation of pro-survival signaling molecules including pNFκB, pAKT and p-p70S6K and the apoptosis inhibitor cIAP2[93].

Stem cell phenotype

ST6GAL1 has been associated with a stem cell phenotype in CRC cells. Cancer stem cells (CSCs) are a minority of cancer cells that are able of self-renewal and generation of differentiated progeny. Such cells have been found in several human carcinomas, being attractive targets for cancer therapy. In some reports, high levels of ST6GAL1 expression are correlated with human induced pluripotent stem cells and CSCs[68,94]. In fact, tumors induced in ST6GAL1 knockout mice were more differentiated compared with those in wild type background, suggesting an importance of this glycosyltransferase for an immature or undifferentiated cell phenotype[95]. In addition, as mentioned in the topic above, ST6GAL1 can confer resistance to apoptosis, thus possibly extending the cell lifespan of CSCs.

Growth factor receptors

ST6GAL1 plays an important role on receptor activation, exerting an inhibitory or activator effect. In CRC cells, α2,6 sialylation of EGFR reduces ligand binding and tyrosine phosphorylation, acting as an inhibitory signal in receptor activation[87]. Furthermore, in lung cancer cells ST6GAL1 overexpression supresses EGFR dimerization and phosphorylation, reducing EGFR-mediated invasion[50]. Other members of the ERBB family appear to be negatively regulated by ST6GAL1 expression, although through an indirect mechanism. Indeed, the reduction of ST6GAL1 expression leads to a reduced protein level of the nectin-like Molecule 2/Cell Adhesion Molecule 1, which acts as an inhibitor of the ERBB2/ERBB3 signalling[96].

Sialylation can also affect the vascular endothelial growth factor receptor (VEGFR) signalling pathway, supressing VEGF-independent angiogenesis in cancer growth by preventing galectin-1 binding[97]. Regarding hepatocyte growth factor receptor

(also known as c-Met), ST6GAL1 enzyme deficiency causes a reduction in α2,6-sialylation of this receptor and consequently abolishes cell motility in *ST6GAL1* knockdown HCT116 CRC cells[98].

Overall, the role of ST6GAL1 in cancer progression has been studied by several groups, with the pro- and anti-tumoral effect of this glycosyltransferase being discussed among them. In colon cancer, the expression of ST6GAL1 in the SW948 human colon cancer cell line (lacks the expression of endogenous ST6GAL1) increased the adhesion to integrin binding components of the extracellular matrix such as fibronectin, collagen and laminin[99]. In HT-29 CRC cells, suppression of ST6GAL1 decreases the invasiveness and anchorage-independent growth[100]. Breast cancer tumors developed by PyMT mice displayed increased differentiation when developed in a *ST6GAL1*-null mice, however no differences in terms of growth properties were found when comparing with *ST6GAL1*$^{+/+}$ mice[95]. On the contrary, it reduces the tumorigenic potential and the multilayer growth in colon cancer cell line SW948 and the invasive growth in glioma cells[99,101]. In bladder cancer, ST6GAL1 loss was also associated with increasing invasiveness[74].

Altogether, the relationship between α2,6 sialylation and malignancy is complex and controversial, being probably strongly tissue dependent.

2.4.2. B4GALNT2 enzyme and Sd^a antigen

The Sd^a glycan belongs to the "non ABO" histo-blood group antigens, being expressed on erythrocytes and found in secretions of nearly 95% of individuals with Caucasian origin[102]. It is formed by an α2,3 sialylated type 2 chain substituted in the O-4 position of a galactose residue with a GalNAc monosaccharide (**Figure I. 13**). The enzyme responsible for the addition of a GalNAc residue in the last step of the Sd^a antigen biosynthesis is the β1,4-N-acetylgalactosaminyltransferase II (B4GALNT2). The Sd^a antigen has been primarily identified in the gastrointestinal tract on the borders of epithelial cells and in the goblet cells of the large intestine, being expressed by *N*- or *O*-linked glycans chains of glycoproteins as well as by long gangliosides[66]. Several lectins have been used for the study of Sd^a antigen including *Vicia vilosa* B4 lectin and *Helix pomatia* lectin[103,104].

Figure I. 13 - Simple biosynthetic pathway of Sd^a biosynthesis. The addition of β1,4-linked GalNAc residue to Gal, mediated by B4GALNT2, leads to the formation of Sd^a antigen.

The enzyme B4GALNT2 was first described in guinea pig kidney[105] and successively identified in the colon of different species including human, rat and pig[66]. *B4GALNT2* gene maps in chromosome 17 and contains 11 exons. At least, two main transcripts have been identified in humans, only differing from the exon 1(**Figure I. 14**) [106]. The short form is comprised by a short exon 1 (exon 1S) with 38 nucleotides while the long form is comprised by a long exon 1 (exon 1L) with 253 nucleotides. These two transcripts contain a translational start site, thus originating at least two different transmembrane peptides: the long form with a very long cytoplasmic domain and a short form with a conventional length cytoplasmic domain. It was demonstrated experimentally that the short B4GALNT2 form presents a higher enzymatic activity comparing with the activity of the long form in the CRC cell line LS 174T[107]. Very recently, it has been shown that while the short form localizes exclusively in the Golgi apparatus, the long form localizes also on the plasma membrane and in post-Golgi vesicles[108].

Exons 1L and 1S are inserted in genomic sequences with characteristics of CpG islands, thus suggesting that DNA methylation can contribute for the regulation of *B4GALNT2* gene expression[109]. B4GALNT2 methylation was found in gastric cancer cases and in the majority of gastric and CRC cell lines analyzed in the study from Kawamura, Y. *et al.* [110]. Anti-DNA methylation treatment in cell lines induced a weak expression of B4GALNT2 and corresponding Sd^a antigen[110].

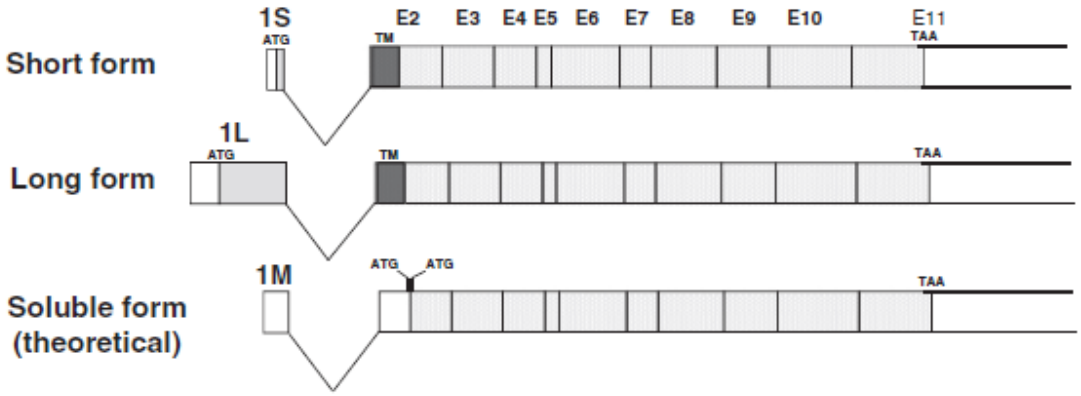


Figure I. 14 - Three main transcripts of the human *B4GALNT2* gene. The gene is constituted by 11 exons and the three different isoforms (short, long and a theoretical soluble form) differing mainly from exon 1. The theoretical soluble form is predicted by GenBank and has never been confirmed experimentally so far. The grey zones represent the coding regions, while the black zones represent predictive transmembrane domains. Adapted from "The expanding role of the Sd^a/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2" (2014) from Dall'Olio, F. *et al.*

In colorectal cancer, B4GALNT2 activity and Sd^a antigen were dramatically reduced comparing with normal colon mucosa[111,112]. As described before, the selectin ligand sLe^x antigen is highly overexpressed in colorectal cancer, contributing to cancer progression and metastasis. The overexpression of sLe^x antigen in CRC is not supported by a concomitant increase of the fucosyltransferases and sialyltransferases involved in its biosynthesis, which are expressed at comparable levels in normal colon mucosa and CRC. A role for B4GALNT2 and Sd^a in the regulation of sLe^x expression has been proposed. The similarity between the Sd^a and sLe^x antigen structures suggests that their biosynthesis might be mutually exclusive[107,113]. This hypothesis is supported by the observation that sLe^x cannot act as an acceptor for B4GALNT2. Both derive from the substitution of an α2,3-sialylated type 2 chain: by a GalNAc residue β1,4 linked to a galactose by B4GALNT2 for Sd^a antigen and fucose α1,3 linked to a GlcNAc residue for sLe^x. It was demonstrated *in vitro* that forced upregulation of B4GALNT2 in CRC cell lines resulted in the down regulation of sLe^x antigen and expression of Sd^a carbohydrate[107,113]. In colon specimens, mucins from normal colonic mucosa express high levels of Sd^a and low levels of sLe^x[114]. Since FUT6 is the main fucosyltransferase responsible for sLe^x biosynthesis in colonic tissues and is nearly unchanged in cancer[115], a model for the regulation of sLe^x expression by B4GALNT2 suggests that the low level of B4GALNT2 present in colon cancer tissues is responsible for the shift of the Sd^a/sLe^x equilibrium towards sLe^x (**Figure I. 15**)[66]. Accordingly, a significant linear relationship

between sLe^x and the FUT6/B4GALNT2 ratio was demonstrated in normal colon but not in cancer[65].

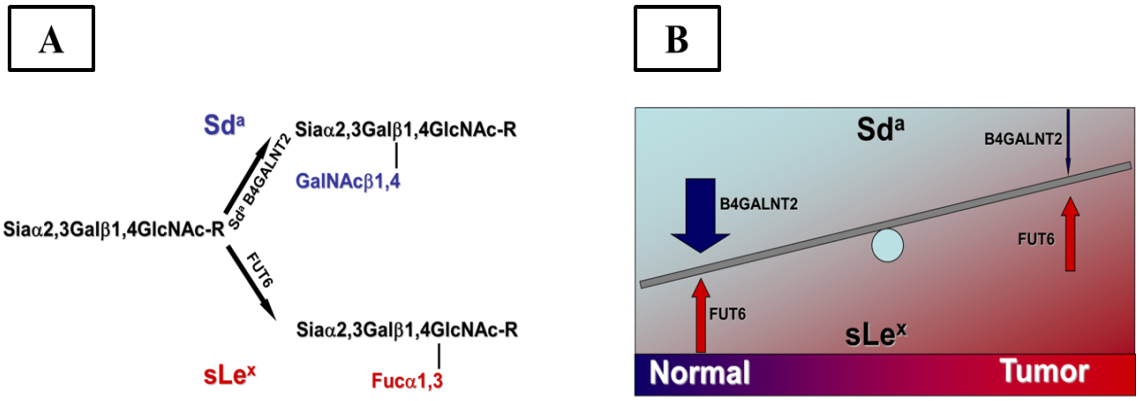


Figure I. 15 - A) Biosynthetic pathway of sLe^x and Sd^a antigens, showing the competition between B4GALNT2 and FUT6; B) In CRC, the increase of sLe^x expression is not correlated with an increase in fucosyltransferase expression (arrows with same shape) meanwhile the reduced B4GALNT2 expression in cancer tissues (slight arrow) is responsible to change the equilibrium towards sLe^x expression. Adapted from "The expanding role of the Sd^a/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2" (2014) from Dall'Olio, F. *et al.*

Although only partially understood, the role of B4GALNT2 and Sd^a antigen appears to be wide and different in different tissues and organs[66]. Examples are provided by the regulation of hemostasis in a murine model, by acting on the clearance of the Von Willebrand factor[116]; a role in embryo attachment[117] and a role in preventing muscle degeneration in a mouse model of Duchenne muscular dystrophy[118–120]. However, one of the most likely roles of the Sd^a antigen is to prevent the cell surface attachment of microorganisms expressing receptors for α2,3 sialylated glycans. In fact, a recent study has shown that B4GALNT2 is the major factor restricting the infectivity of influenza virus strains expressing receptors for α2,3 sialylated glycans[121].

3. Importance of "omics" in biomedical research

Nowadays, the revolution of "omics" technologies provides high system-level measurements for nearly all types of cellular components in a model organism, becoming powerful tools for disease studies in particular cancer. Omics are aimed primarily to the comprehensive detection of genes (genomics), mRNA (transcriptomics) and proteins (proteomics) in a specific biological sample (**Figure I. 16**). Other areas of biomedical sciences such as metabolomics (the study of metabolites produced during biochemical reactions), lipidomics (full characterization of lipid molecular species), epigenomics (study of epigenetic modifications on the genetic material of a cell) and

glycomics (study of all glycans and glycoconjugates present in an organism) are emerging in the field of omics[122].

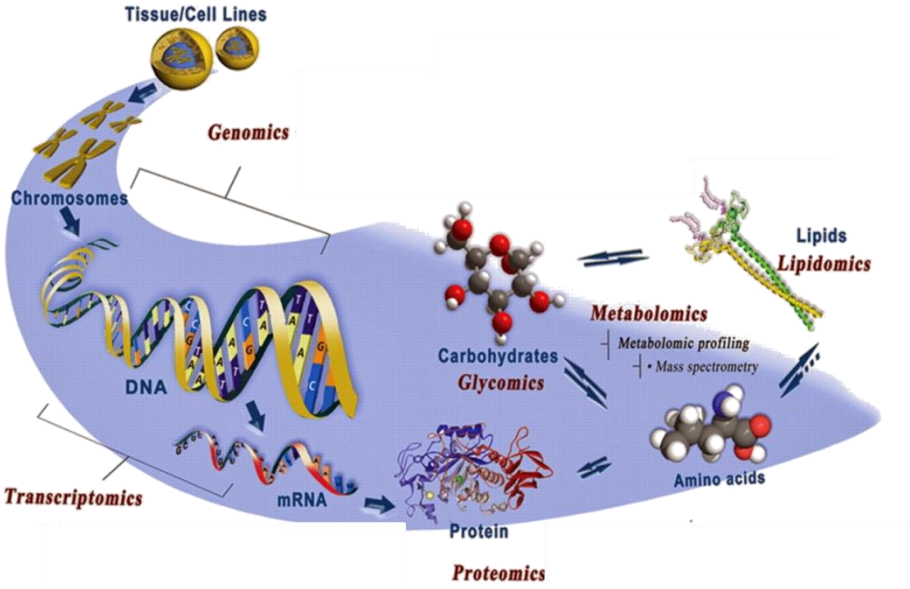


Figure I. 16 - Schematic representation of the most representative omics technologies and their corresponding analysis targets. Adapted from "Novel molecular events in oral carcinogenesis via integrative approaches" (2011) from Wu, R. *et al.*

Genomics is defined as the study of the whole genome sequence of an organism and the information contained therein, being one of the fist type of omics emerging in the biomedical research. Since 1995, nearly 3000 genome-sequencing projects derived from different types of species have been completed and hundreds more are ongoing[123]. The study of the genetic background of an individual is of great importance for the identification of individual mutations and/or variants underpinning pathways that discriminate health and disease. Current techniques to study the genome include Sanger sequencing, DNA-microarray (based on the hybridization of DNA samples with pre-defined oligonucleotide probes spread across the entire genome or enriched around regions of interest) and next-generating sequencing (NGS - based in the fragmentation of genomic DNA that are consequently sequenced and aligned to a reference sequence)[124].

The transcriptome is comprised of the total RNA transcripts in a cell/organism, consisting in coding (messenger RNA- mRNA) and non-coding (ribosomal, transfer, long-non coding, small nuclear, small interfering and micro) RNAs. The transcriptome is a "bridge" in the process of biological information transference from DNA to protein, being crucial to address the functions of genome, reveal the molecular constituents of cells and reflect the occurrence or development of a disease. Microarray and sequencing

of RNA (RNA-seq) represent the most well-used approaches in this field to study the transcriptome[125]. In both techniques, the general workflow to generate raw transcriptomic data involves purification of RNA samples, conversion of RNA to complementary DNA (cDNA), chemical labelling and hybridization of the cDNA with probes spotted on chips (microarray) or fragmentation of cDNA and building of a library to be sequenced (RNA-seq), run the microarray or sequence through a platform of choice and extensive data analysis (**Figure I. 17**)[124].

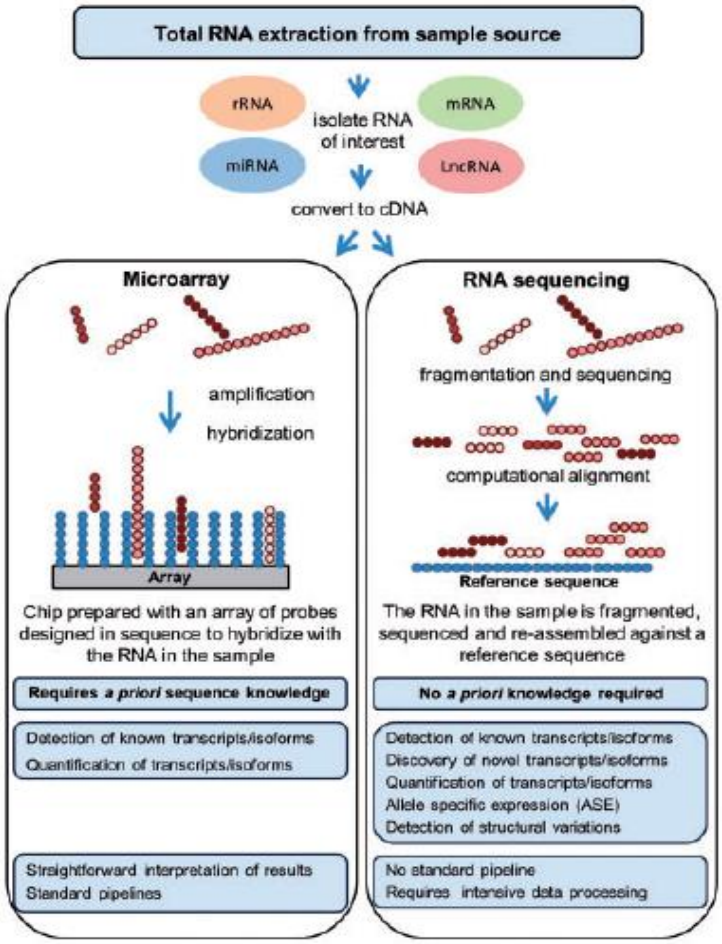


Figure I. 17 - Workflow and main features associated with RNA-microarray and RNA-sequencing. RNA-microarray approach is less costly comparing with RNA-seq, however has some limitations such as the knowledge *a priori* of the genome sequence. RNA-seq is a more comprehensive technique allowing to analyze any form of RNA at a much higher coverage level. Adapted from "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences" (2018) from Manzoni, Claudia. *et al.*

The proteome is defined as all the set of proteins expressed in a cell, tissue or organism. Methods based on 2-D gel electrophoresis and mass spectrometry are the most popular strategies in this field[126]. One important consideration is the high level of complexity to study the proteome. From mRNA, these molecules are translated into a much more complex code of possible 20 amino-acids, with primary sequence and

different levels of folding, conformations and chemical modifications, originating a huge heterogeneity of possible proteins.

Glycomic field is defined as the study of the complete set of glycans and glycoconjugates present by a cell or organism. Glycosylation is the most abundant post-translational modification in proteins and lipids, providing an enormous dynamic structural diversity in those molecules and consequently affecting cell phenotype, metabolic state and cell development. Presently, several high-resolution and high-sensitive methods are available to study the glycome including mass spectrometry (MS), capillary electrophoresis, high-performance liquid chromatography (HPLC) and lectin microarrays[127]. *N*-glycans from glycoconjugates are analyzed usually by HPLC after tagging the reducing end of the sugars with a fluorescent compound. Fractionated glycans from HPLC can be further analyzed by MS to get additional information about glycan structure and purity. Glycomic analysis allows a fast comparison of glycome within body fluids or tissues, which would allow the identification and application of a new type of biomarkers for cancer diagnosis and prognosis[128].

In summary, there is a clearly potential for the use of integrated multi-omics data for a better understanding of the molecular mechanisms, physiological processes and pathways involved in health and disease[122,124].

In 2005, the National Cancer Institute (from USA) funded The Cancer Genome Atlas (TCGA) project, aiming to study the key genomic changes that occurs in the major types of cancer. The TCGA dataset is publically available containing a large collection of clinical and molecular phenotypes of tumor tissues and matched normal tissues from more than 11 000 patients across 33 different tumor types (including 10 rare cancers)[129]. The structure of TCGA is well organized and involves several centers responsible for collection and sample processing, followed by high-throughput sequencing and sophisticated bioinformatics data analysis. This project uses several high-throughput techniques to analyze each patient sample including RNA-seq, MicroRNA sequencing, DNA sequencing, array-based detection of single nucleotide polymorphisms, reverse-phase protein array and array-based DNA methylation sequencing. In particular, for this thesis we focused mainly on data regarding RNA sequencing, high-throughput technology for transcriptome (total RNA) profiling. RNA-seq is able to identify and quantify rare and common transcripts, isoforms, novel transcripts and non-coding RNAs[125]. For transcriptome analysis, TCGA uses a platform based on the *Illumina* system.

4. Purpose of the work

The present work is based on the following premises:

1) CRC is a major cause of the cancer burden worldwide and early detection protocols are crucial in this type of disease. However, current screening techniques are either invasive or lack sensitivity and specificity.

2) Glycan biomarkers are novel and highly promising candidates for CRC diagnosis and prognosis.

3) ST6GAL1 and B4GALNT2 (and their associated antigens, Sia6LacNAc and Sd^a) are associated with cancer but it is not fully clear their role in tumour progression and development.

The aims of this project are the following:

1) Study of the relationship between ST6GAL1 and B4GALNT2 expression and transcriptomic profiles in colorectal cancer.´

This will be done through two different approaches:

- *In silico* study using data from the public database "The Cancer Genome Atlas" (TCGA) database. Using gene expression data, patients will be classified according to their level of expression of either glycosyltransferase. The clinical features of high or low expressers will be identified, as well as the genes showing up- or down-regulation in the high or low glycosyltransferase expressers. This study will identify clinical and transcriptomic features which are associated (but not necessarily causally related) with the high or low ST6GAL1/B4GALNT2 expression *status*.

- To establish a causal relationship between ST6GAL1 or B4GALNT2 expression and transcriptomic changes, microarray analysis will be performed on colon cancer cell lines forced to express either glycosyltransferase, by genetic manipulation.

2) Study of the phenotypic changes induced by the expression of either glycosyltransferase: the cell lines forced to express either glycosyltransferase, by genetic manipulation mentioned above will be phenotypically characterized with respect to the main properties of cancer cells.

3) Identification of new CRC-specific glycan changes in tissues and plasma samples:

- In a cohort of CRC/polyp tissues and also plasma samples, the enzymatic activity of ST6GAL1 and B4GALNT2 will be measured;

- Sd^a protein carriers in plasma samples will be explored;

- *N*-glycomic analysis will be performed on tissues of CRC, normal colonic mucosa and polyps.

Chapter II - Material and Methods

1. Cell lines

In this project, three main cell lines derived from colorectal cancer were used: SW948 (ATCC® Number: CCL-237™), SW48 (ATCC® Number: CCL-231™) and LS 174T (ATCC® Number: CL-188™). SW948 and SW48 cells were cultured in Leibovitz's L-15 medium while LS 174T cells in Dulbecco's Modified Eagle's Medium (DMEM), all from Microgem. The basal medium was supplemented with 10% fetal bovine serum (FBS), 2 mM of L-Glutamine and 100 µg/mL Penicillin/Streptomycin, all from Microgem. SW948 and SW48 cells were kept in culture in absence of CO_2 in a humidifier incubator at 37ºC while LS 174T cells were kept in an incubator with a humidified atmosphere of 5% CO_2 at 37ºC.

The cell line SW948 is derived from a 81 years old Caucasian female colorectal adenocarcinoma with grade III and stage III cancer. The cell line SW48 is derived from a 82 years old Caucasian female colorectal adenocarcinoma with grade IV and stage III cancer. The cell line LS 174T was established from a stage II colorectal adenocarcinoma in a 58 years old Caucasian female.

The lentivirus transduction with human ST6GAL1 cDNA (ST cells) or with an empty vector (NC cells) was reported previously by our group[130]. SW948 ST and SW48 ST were transduced with the cDNA of the whole coding sequence of human ST6GAL1 from Caco2 cells, using for the PCR amplification the forward primer (hST6Gal-IL.1), constituted by CACCATGATTCACACCAACCTGAAGAAAAAGTTCAGCTGCTGC nucleotides and the reverse primer (hST6Gal1.R1), constituted by TTAGCAGTGAATGGTCCGGAAGCCAGGCAGTGTG nucleotides. The underlined sequence in the forward primer is required for the cloning in TOPO vectors and is not gene specific. The PCR product was cloned in pLenti6/V5 Directional TOPO® Cloning Kit (Invitrogen, Carlsbad, CA) and used to transfect the 293FT cell line, following the instructions of the ViraPower Lentiviral Expression System (Invitrogen). The conditioned medium (containing virions) was centrifuged to eliminate cell debris, filtered on a 0.45 µM pore membrane, diluted with 9 volumes of fresh culture medium, added to SW948/SW48 cells and left for 48h. Cells were selected with 10 µg/mL of blasticidin. A mock transduction was performed in parallel with a retrovirus lacking the ST6GAL1 insert.

The construction of B4GALNT2 transfectants was reported previously by our group[107]. Briefly, the PCR amplification of the B4GALNT2 short form derived from the human colon cancer cell line Caco2 was performed using the forward primer L.19 (5'-CACCATGACTTCGGGCGGCTCG-3') and the reverse primer R.10 (5'-CCAGTAACTGAGCCATTTCCCTTTTCC-3'). The underlined sequence in the forward primer is required for the cloning in TOPO vectors and is not gene specific. The PCR product was cloned in pcDNA3.1 Directional TOPO® Expression vector (Invitrogen, Paisley, UK). LS 174T cells were transfected using the calcium phosphate method with either construct or empty vector and selected with 0.4 mg/mL G418. Resistant clones were isolated with cloning cylinders, expanded and screened for B4GALNT2 activity.

Cell lines were genotyped using highly-polymorphic short tandem repeat *loci* which were amplified by the PowerPlex® 16 HS System (Promega). Fragment analysis was performed on an ABI3730xl (Life Technologies) and data were analyzed with GeneMarker HID software (Softgenetics) by Microsynth (Switzerland).

1.1 Total RNA extraction and reverse transcription to cDNA

Total RNA extraction was performed according to Chomczynski & Sacchi method[131], being suspended at the end in 50 µl of DNase/RNase free water. RNA was quantified using the Nano Genius Photometer ONDA, measuring the absorbance at 260 nm with a 2.0 ±0.5 ratio of Abs_{260}/Abs_{280}. RNA integrity was assessed by running RNA samples on a 1% agarose gel. Conversion of RNA to cDNA was performed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following manufacturer's instructions.

1.2 Transcriptomic data analysis

Transcriptomic analysis of RNA extracted from SW948 (NC and ST cells), SW48 (NC and ST cells) and LS 174T (Neo population, S2 and S11 clones) was performed using Agilent whole human genome oligo microarray (G4851A) at LTTA Microarray Facility (Università di Ferrara). For ST6GAL1 transduced cells and respective negative controls, two independent replicates were analyzed for each cell line. For B4GALNT2 transfectants, two independent replicates of LS174T Neo were analyzed and for B4GALNT2-short form clones, one replicate of S2 and S11 were analyzed. Microarray results were examined using the GeneSpring GX 12 software

(Agilent Technologies, Palo Alto, CA). A filter on low gene expression was used to keep only the probes expressed in at least one sample and a fold change cut-off of 2 was applied to all panel of genes. Statistical analysis was done using moderated T-test and the false discovery rate was measured with the multiple testing correction Benjamini-Hochberg with Q=0.15 and Q=0.05, for ST6GAL1 transduced cells experiment and B4GALNT2 transfectants, respectively. Heat-maps were created by hierarchical clustering using the Manhattan correlation. Gene function was studied through an extensive literature search. Functional Ontology Enrichment analysis and Pathway map visualization was performed using MetaCore™ (Thomson Reuters v. 6.8) including GO (gene ontology) processes, process networks and canonical pathways maps.

1.3 Wound healing assay

The wound healing assay was performed using Culture-Insert 2 Well from Ibidi. Aliquots of 70000 (for SW48), 50000 (for SW948) and 50000 (for LS 174T) cells were seeded into each well of the Culture-Insert 2 Well. When cells reached total confluency, the insert was removed, creating a cell-free gap of approximately 500 μm. Then, the cell layer was washed with PBS to remove cell debris and non-attached cells and the well was filled with the corresponding complete medium for each cell line. The capacity to heal the wound was measured taking pictures with Nikon Eclipse TS100 inverted microscope at 10× magnification and a Digital c-Mount Camera Sony Color. The area of the wound was measured using the MRI wound healing tool from ImageJ. Statistical analysis was performed using two-way ANOVA and Tukey's multiple comparison test.

1.4 Soft agar assay

A bottom layer constituted of 0.5% agar solution in complete L-15 medium was dispensed in each well of a 6-multiwell plate and allowed to solidify. A top layer constituted by 0.3% agar solution in complete L-15 medium containing 10000 cells (SW48 or SW948 cells) was dispensed into each well above the bottom layer, in triplicate. The plates were incubated at 37ºC in a humidified incubator without CO_2 for two weeks. Then, the cells were fixed and colored for one hour with a solution containing 4% formaldehyde and 0.005% crystal violet in phosphate-buffered saline (PBS). Photographs of the stained colonies were taken using a Nikon Eclipse TS100 inverted microscopy at 5× magnification and a Digital c-Mount Camera Sony Color.

The colonies were counted and statistical analysis was performed using the non-parametric Kolmogorov-Smirnov test.

2. Biological samples

This study involves 19 CRC, 5 polyp and 27 IBD patients who underwent surgery due to suspicion of colorectal cancer at S.Orsola-Malpighi Hospital (Bologna-Italy). Among IBD cases, twenty-one patients were further diagnosed with Crohn disease and six with ulcerative colitis. The median age was 72 (Min= 43 years; Max= 84 years), 60 (Min= 50 years; Max= 84 years) and 44 (Min= 23 years; Max= 78 years) for CRC, polyp and IBD patients, respectively. For each CRC and polyp patients, fragments of colon tumor tissue and normal colonic tissue were collected. Plasma samples were collected from all patients. The determination of the histological type was performed by the Pathology Unit from S.Orsola-Malpighi Hospital (Bologna-Italy) under coordination of Dr. Francesco Vasuri. Tumor staging was classified according to TNM classification 7th edition system based on the International Union Against Cancer[18]. The study was approved by the ethic committee of S.Orsola-Malpighi Hospital and written informed consent was obtained from all patients.

2.1 Enzymatic activity assays

Tissues samples were homogenized in ice-cold water and the protein concentration of the homogenates was determined by the Lowry method[132].

ST6GAL1 enzyme activity was measured in whole tissue homogenates and plasma samples by incorporation of [^{3}H]-sialic acid on asialotransferrin as previously reported by our group[133,134]. The radioactivity measured in the absence of the acceptor was subtracted for each sample. Briefly for whole tissue homogenates, the assay mixture contained in a final volume of 50 µL: 80 mM Na-cacodylate buffer, pH 6.5; 10 mM $MnCl_2$; 0.5% Triton X-100; 300 µg asialotransferrin (prepared by the desialylation of transferrin in 50 mM H_2SO_4 at 80ºC for 2 hours, followed by dialysis); 30 µM unlabelled CMP-sialic acid, 1 µl CMP-[^{3}H] sialic acid with a specific activity of 1x10^5 dpm/µl and 100 µg of protein homogenate as enzyme source. For plasma samples, the assay mixture contained in a final volume of 50 µL: 80 mM Na-cacodylate buffer, pH 6.5; 500 µg asialotransferrin; 5 µM unlabelled CMP-sialic acid, 1 µl CMP-[^{3}H] sialic acid with a specific activity of 1x10^5 dpm/µl and 5 µl of human plasma as enzyme

source. After 3 hours incubation at 37ºC, the acid-insoluble radioactivity was precipitated with 1% phosphotungstic acid in 0.5 M HCl (FTA). Pellets were washed two times with FTA and once with methanol. Then, the samples were boiled for 20 minutes with 1 M HCl. At the end, the samples were resuspended and read with the Guardian 1414 Liquid Scintillation Counter (PerkinElmer) after the addition of 3.5 ml of scintillation liquid.

B4GALNT2 enzyme activity was measured only in whole tissue homogenates as the difference between the incorporation of [^{3}H]-GalNAc on fetuin and asialofetuin, as previously described by our group[135]. Briefly, the assay mixture contained in a final volume of 25 μL: 80 mM Tris/HCl buffer, pH 7.5; 10 mM $MnCl_2$; 0.5% Triton X-100; UDP-[^{3}H]GalNAc (ARC, St. Louis, MO) with a specific activity of 550 dpm/pmol, 2 mM ATP, 250 μg of either fetuin (Sigma) or asialofetuin (prepared by the desialylation of fetuin in 50 mM H_2SO_4 at 80ºC for 2 hours, followed by dialysis) as acceptors and 50-70 μg of protein homogenate as enzyme source. After 3 hours incubation at 37ºC, the acid-insoluble radioactivity was precipitated with 1% phosphotungstic acid in 0.5 M HCl (FTA). Pellets were washed two times with FTA and once with methanol. Then, the samples were boiled for 20 minutes with 1 M HCl. At the end, the samples were resuspended and read with the Guardian 1414 Liquid Scintillation Counter (PerkinElmer) after the addition of 3.5 ml of scintillation liquid.

Wilcoxon matched-pairs signed rank test was used to analyze the level of ST6GAL1/B4GALNT2 activity in normal and CRC/polyp tissues. Ordinary one-way Anova and Tukey's multiple comparisons test were used to compare ST6GAL1 activity in plasma samples from CRC, polyp, IBD patients and also healthy individuals.

2.2 Total RNA extraction and reverse transcription to cDNA

Total RNA extraction and reverse transcription to cDNA was performed as described before in chapter II - Material and Methods (section 1.1).

2.3 Real Time-PCR

Real Time-Polymerase Chain Reaction (RT-PCR) was performed using *Taqman* probes methodology, following the instructions of the manufacturer. For each primer/probe set, the Assay ID (Applied Biosystems) was the following: *ST6GAL1* (Hs00949382_m1), *B4GALNT2* (Hs00396440_m1), β-actin *(ACTB)* (Hs99999903_m1)

and glyceraldehyde 3-phosphate dehydrogenase (*GAPDH)* (Hs99999905_m1). The mRNA expression was normalized using the geometric mean of the expression of the endogenous controls, *ACTB* and *GAPDH* genes. Each reaction was performed in triplicate with TaqMan® Fast Universal PCR Master Mix (Applied Biosystems) supplemented with 100 ng cDNA and TaqMan probes (described above) in a final volume of 10 µL. The reaction was carried out with the following protocol: 95°C hold step for 20 seconds and 40 cycles of 95-60°C for 5 and 30 seconds, respectively. The relative mRNA level expression was calculated using the expression $2^{-\Delta Ct} \times 1000$, which infers the number of mRNA molecules of the gene of interest for each 1000 molecules of endogenous controls[136]. RT-PCR was performed in a CFX96 machine (Bio-Rad) and data were analysed using the Bio-Rad CFX Manager software version 3.1. Mann Whitney test was used to analyze the level of ST6GAL1/B4GALNT2 mRNA in normal and CRC tissues.

2.4 Western blotting

Plasma samples were diluted in water and sample buffer 4× containing 0.5 M Tris HCl pH 6.8, 10% sodium dodecyl sulphate (SDS), 30% glycerol, 6% β-mercaptoethanol and 0.012% bromophenol blue in a final volume of 20 µl. The prepared samples were boiled at 95ºC for 5 minutes and separated in a 8% acrylamide gel (0.75 mm width) in denaturing and reducing conditions. As molecular weight marker, it was used the PageRuler Prestained Protein Ladder from Thermo Scientific. Proteins were transferred to a 0.2 µm polyvinylidene difluoride membrane (Trans-Blot Turbo Mini PVDF Transfer Packs from Bio-Rad) for 7+7 minutes using the Trans-Blot® Turbo™ Transfer System from Bio-Rad. After complete transference, membranes were incubated on an orbital shaker for 1 hour at room temperature with 0.1% bovine serum albumin (BSA) in phosphate buffer saline with 0.1% tween-20 (PBS-T). After blocking, membranes were washed 3 times (5 minutes each) with PBS-T and incubated on an orbital shaker for 1 hour at room temperature with the primary antibody anti-Sd^a KM694 kindly provided by Kyowa Hakko Kogyo Co. Ltd. (Tokyo, Japan), diluted 1:1000 in 0.1% BSA in PBS-T. Then, the membranes were washed 3 times (5 minutes each) with PBS-T and incubated on an orbital shaker for 1 hour at room temperature with the secondary antibody anti-IgM conjugated with horseradish peroxidase from Sigma, diluted 1:10000 in 0.1% BSA in PBS-T. Membranes were washed three more times as described earlier.

The reaction was developed with Westar ηC 2.0 (Cyanagen) according to the manufacturer's instructions and detected with a photographic film (Kodak). Pictures of the films were taken using EDAS 290 camera (Kodak).

The purified band found bellow 55kDa together with mock-purification were isolated from the gel and sent to the Institute of Pharmacology and Structural Biology (Toulouse-France) for the glycoproteomic identification of the Sd^a carrier proteins using the mass spectrometry technique.

2.5 Purification of plasma samples by affinity column purification

A healthy human plasma sample (3 mL) was passed through a column resin made with anti-human IgM (μ-chain specific)-agarose antibody produced in goat (A9935-Sigma), to remove all human IgM molecules. Then, the flow-through was incubated at room temperature under agitation for 3 hours with 100 μl of anti-Sd^a KM694. The resulting mixture was passed four times through a second column made with anti-mouse IgM (μ-chain specific)-agarose antibody produced in goat (A4540-Sigma). Sd^a linked glycoproteins were then eluted from the column using 3M potassium thiocyanate (KSCN), followed by dialysis. In all the steps, part of total plasma, flow-through and eluate were kept at room temperature for further analysis by western blotting (procedure described previously in chapter II - Material and Methods (section 2.4). A mock purification without the addition of anti-Sd^a antibody was performed in parallel to exclude any unspecific bindings.

2.6 Glycomic analysis

Isolation of glycoproteins from tissues samples was performed after lipid extraction. After homogenization of tissues with water, methanol was added to the samples followed by vortexing. Then, chloroform was added resulting in a phase separation, followed by vortexing and centrifugation at 120000×g for 15 minutes. The upper phase was removed and replaced by methanol/water (50:50, v/v). Samples were vortexed and centrifuged followed by the removal of the upper phase. This extraction process was repeated twice. Finally, methanol was added to the lower phase and the interphase, containing most of the glycoproteins. The samples were vortexed and centrifuged at 120000×g for 15 minutes, resulting in the pelleting of the protein fraction. Pellet was resuspended in methanol, followed by centrifugation and removal of the

supernatant. This step was repeated three times and consequently the pellets were dried using a Savant SpeedVac™ Concentrator.

The *N*-glycans were released from glycoproteins of tissues samples using LudgerZyme™ PNGase F Release Kit (LZ-rPNGaseF-kit). Briefly, protein tissue pellets were homogenized for 45 min in a sonication bath followed by centrifugation at 500×g for 15 minutes. 9 µL of pure water and 1 µL of denaturing buffer were added to each sample and mixed. The samples were incubated for 10 minutes at 100°C to denature the proteins. After cooling the samples to room temperature, 2 µL of 10X reaction buffer, 2 µL of 10% NP-40 solution, 6 µL of pure water and 1 µL of PNGase F were added to each sample. Samples were vortexed and incubated overnight at 37°C. The released glycans were then converted to aldoses with 0.1% formic acid, filtered through a protein-binding membrane and dried.

Released *N*-glycans were fluorescently labelled by reductive amination with procainamide using LudgerTag™ Procainamide Glycan Labelling Kit (LT-KPROC-24). Briefly, samples in 10 µL of pure water were incubated for 60 minutes at 65°C with procainamide labelling solution. The procainamide labelled samples were cleaned-up using a Ludger Clean™ Procainamide Clean-up Plate (LC-PROC-96). The purified procainamide labelled *N*-glycans were eluted with pure water (200 µL). The samples were dried by vacuum centrifugation and resuspended in pure water (30 µL) for further analysis.

Procainamide labelled samples and system suitability standards were analyzed by HILIC-(U)HPLC-ESI-MS with fluorescence detection. Samples were injected in 24% pure water/76% acetonitrile (injection volume 25 µL) onto an ACQUITY UPLC® BEH Glycan 1.7 µm, 2.1 x 150 mm column at 40°C on a Ultimate 3000 UHPLC instrument (Thermo Scientific, Massachusetts, USA) with a fluorescence detector (λex = 310nm, λem = 370nm). The running conditions used were: Solvent A was 50 mM ammonium formate pH 4.4 made from Ludger Stock Buffer (Ludger), and solvent B was acetonitrile. Gradient conditions were: 0 to 53.5 min, 76 to 51% B; 53.5 to 55.5 min, 51 to 0% B at a flow rate of 0.4 ml/min; 55.5 to 57.5 min, 0% B at a flow rate of 0.25 ml/min; 57.5 to 59.5 min, 0 to 76% B at a flow rate of 0.2 ml/min; 59.5 to 65.5 min, 76% B at a flow rate of 0.2 ml/min; 65.5 to 66.5 min, 76% B at a flow rate of 0.4 ml/min; 66.5 to 70 min, 76% B at a flow rate of 0.4 ml/min. The UHPLC system was coupled on-line to an AmaZon Speed ETD electrospray mass spectrometer (Bruker Daltonics, Bremen, Germany) with the following settings: source temperature 250°C,

gas flow 10 L/min; Capillary voltage 4500 V; ICC target 200,000; maximum accumulation time 50 ms; rolling average 2; number of precursors ions selected 3, release after 0.2 min; Positive ion mode; Scan mode: enhanced resolution; mass range scanned, 180-1500; Target mass, 700. A glucose homopolymer ladder labelled with procainamide (Ludger Ltd) was used as a system suitability standard as well as an external calibration standard for GU allocation. ESI-MS and MS/MS data analysis was performed using Bruker Compass DataAnalysis v4.1 software and GlycoWorkbench software. Ordinary one-way Anova and Tukey's multiple comparisons test was used to compare glycosylation features in tissues samples from CRC, polyp and normal samples.

3. TCGA database analysis - *In silico* study

Gene expression data and clinical information for 623 colorectal adenocarcinoma samples and 51 normal tissues were retrieved from The Cancer Genome Atlas (TCGA) database using Firebrowse website (http://firebrowse.org). RSEM normalized data for colorectal adenocarcinoma (COADREAD) cohort were matched with clinical data from Clinical Pick Tier1 archive. Mutations for some of those patients were retrieved using cBioPortal web site (http://www.cbioportal.org). Genes with average expression lower than 5 (≈4000 genes) were removed from the list. The characteristics of this cohort are given in appendix section (**Table VI. 1**).

ST6GAL1 and B4GALNT2 mRNA expression was compared with stage, MSS/MSI *status*, response to treatment, histological type, survival, and KRAS, BRAF, APC and TP53 mutations. Since the samples did not present a normal distribution, non-parametric tests were used. Mann Whitney test was used to analyze the level of expression of ST6GAL1/B4GALNT2 in normal and tumor tissues and the association of its expression in cancer tissues with mucinous/non-mucinous histological type and the mutational *status* of KRAS, BRAF, APC or TP53. Kruskal-Wallis test was used to evaluate the association of ST6GAL1/B4GALNT2 mRNA expression with cancer stages and MSS/MSI *status*. The survival curve was calculated using the Mantel-Cox test for log-rank.

To identify genes showing modulation associated with high- or low ST6GAL1/B4GALNT2 expression, we selected patients falling in the 15% upper or lower percentile for the expression of either glycosyltransferase. Genes showing

association with either glycosyltransferase were selected according to the following criteria: 1) a level of expression higher than the arbitrary threshold of 500, either in the group of high or of low glycosyltransferase expressers or both. This was done to exclude genes with a negligible level of expression. 2) A percentage change ratio (PC) > 50 (arbitrary threshold) calculated according to the following formula: $PC = (Mean_{15\% high} - Mean_{15\% low})/Mean_{All} * 100$. $Mean_{15\% high}$ indicates the mean level of expression of the gene in the upper 15% ST6GAL1/B4GALNT2 expression cohort; $Mean_{15\% low}$ indicates the mean level of expression of the gene in the lower 15% ST6GAL1/B4GALNT2 expression cohort; $Mean_{All}$ indicates the mean level of expression of the gene among all patients. The statistical tests used were Two-way Anova and Bonferroni's multiple comparison test.

Chapter III - Results

1. Transcriptomic and phenotypic impact of ST6GAL1 expression in colorectal cancer

Note: Results presented in this section were taken from the manuscript "Venturi, G.*, Gomes Ferreira, I.*, Pucci, M., Ferracin, M., Malagolini, N., Chiricolo, M. & Dall'Olio, F. Cell type specific transcriptomic and phenotypic impact of ST6GAL1 overexpression in colon cancer."
* Co-first authorship. Manuscript submitted (under review).

1.1 Survey of TCGA database

Clinical information of 623 colorectal adenocarcinoma patients included in this study is summarized in **Table VI. 1** (see appendix section - chapter VI). Features such as age at initial diagnosis, gender (female/male), histological subtype of tumor (adenocarcinoma or mucinous adenocarcinoma), microsatellite *status* (microsatellite stable, high microsatellite instability or low microsatellite instability), stage (stage I, II, III or IV) and follow-up treatment success (complete remission/response, partial remission/response, stable disease or progressive disease) are the main characteristics collected for this TCGA survey. Some of those clinical data were not available for some CRC samples. In TCGA database, gene expression data from 51 cases of normal colonic mucosa (matching samples with those of CRC) were also accessible.

The analysis of transcriptomic data of CRC and normal specimens from TCGA database, allowed the identification of relationships between ST6GAL1 gene expression and clinical parameters. As shown in **Figure III. 1A and B**, ST6GAL1 mRNA expression is variable among CRC specimens but quite uniform in normal tissues (**Figure III. 1B**). No relationship existed between ST6GAL1 expression and clinical stage (**Figure III. 1C**), while a highly significant association was found between low ST6GAL1 expression and microsatellite instability (MSI) (**Figure III. 1D**). Low ST6GAL1 was associated also with a mucinous phenotype (**Figure III. 1E**). Regarding the follow-up treatment success, patients were divided in two groups: response (including complete remission/response and partial remission/response) and non-response (including progressive or stable disease). No relationship was found between ST6GAL1 expression and response to treatment (**Figure III. 1F**).

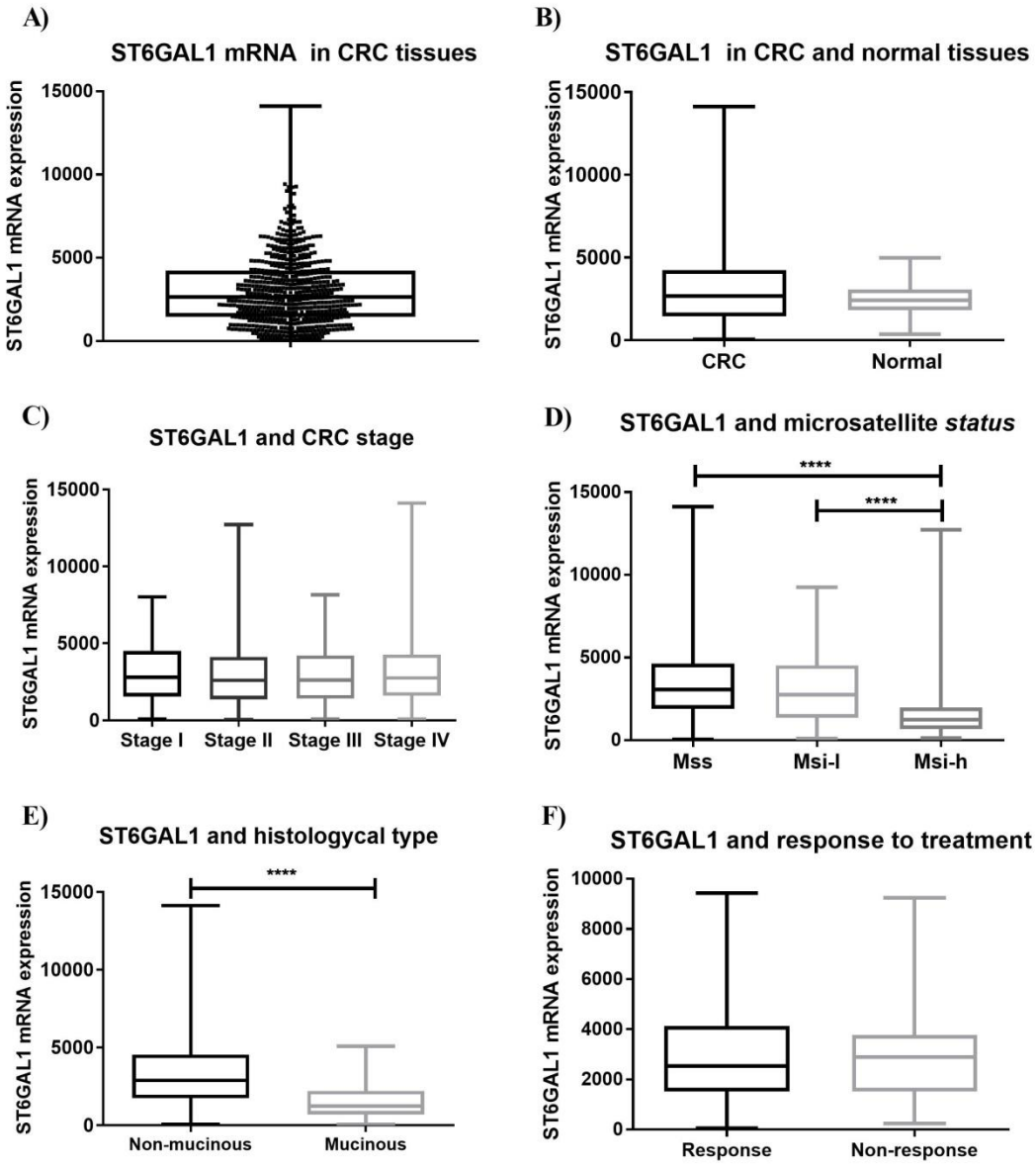


Figure III. 1 - Data from the TCGA database. A: ST6GAL1 mRNA expression in cancer tissues of CRC patients. Each black dot represents the value for one patient. B-F: box plot graphs showing median, Q1 quartile, Q3 quartile, minimal and maximal value of ST6GAL1 mRNA expression in CRC and normal tissues (B); in CRC tissue of stage I-IV patients (C); in CRC tissue of patients with microsatellite stability (MSS), low grade microsatellite instability (MSI-l) or high grade microsatellite instability (MSI-h) (D); with mucinous or non-mucinous phenotype (E) and showing response or non-response to treatment (F). Mann Whitney test was used in B, E and F. Kruskal-Wallis test was used in C and D. ****p≤0.0001

In TCGA database, data regarding DNA mutations were also available. We compared the level of ST6GAL1 mRNA expression with mutations in some genes relevant for CRC carcinogenesis such as oncogenes *KRAS* and *BRAF* and the tumor suppressor genes *TP53* and *APC*. We found a highly significant association between low ST6GAL1 expression and BRAF mutations (**Figure III. 2A**) in particular V600E mutation (substitution at position 600 in BRAF, from a valine to a glutamic acid) but we

did not observed any association between ST6GAL1 mRNA expression and KRAS (**Figure III. 2B**), TP53 (**Figure III. 2C**) and APC (**Figure III. 2D**) mutations. We also subjected TCGA data to Kaplan-Meier survival analysis to assess the relation between patient's overall survival and high or low level of ST6GAL1 mRNA expression. For this reason, patients were divided into two groups: high (upper 15% percentile values) and low (lower 15% percentile values) ST6GAL1 mRNA expression. No significant differences were found between the survival curves of the two groups (**Figure III. 2E**).

Collectively, these data suggest a moderate general impact of ST6GAL1 level on CRC progression, although this impact can be highly relevant in subgroups of patients.

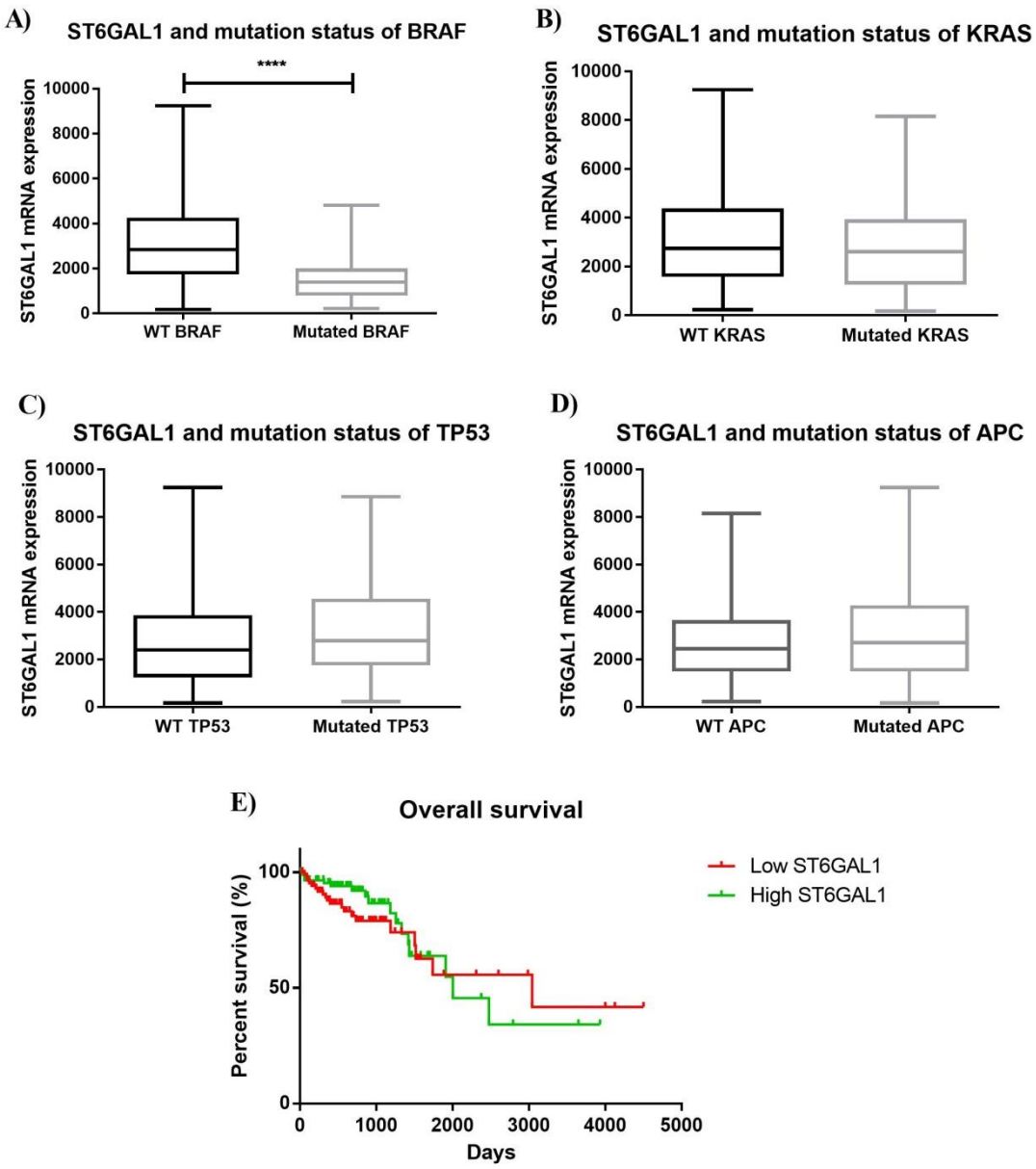


Figure III. 2 - Data from the TCGA database. A-D: box plot graphs showing median, Q1 quartile, Q3 quartile, minimal and maximal value of ST6GAL1 mRNA expression in CRC tissues of patients with wild type or mutated BRAF (A), KRAS (B), TP53 (C) and APC (D). E: Survival curves of patients falling in the upper or lower 15% percentile of ST6GAL1 mRNA expression. Mann Whitney test was used in A, B, C and D. The survival curve was calculated using the Mantel-Cox test for log-rank. ****p≤0.0001

In order to find a gene expression signature associated with high or low ST6GAL1 expression, two cohorts including the patients in the 15% upper and 15% lower percentiles of ST6GAL1 mRNA level were compared. **Table III. 1** shows the genes statistically modulated between the two cohorts.

Table III. 1 - Genes modulated in high/low ST6GAL1 cohorts. Genes showing a percentage change higher than 50, according to the formula: PC = ($Mean_{15\%\ high}$ - $Mean_{15\%\ low}$)/$Mean_{All}$ * 100 are shown. In red and green are reported genes with a PC higher that 50 or lower than -50, respectively. Bonferroni's multiple comparisons test was used for the analysis. *≤0.05, **≤0.01, ****p≤0.0001

Gene Symbol/Gene ID	Gene name (from Human Gene Database)	Adjusted P-value
CEACAM5|1048	Carcinoembryonic Antigen Related Cell Adhesion Molecule 5	****
CD24|100133941	CD24 molecule	****
PYGB|5834	Glycogen Phosphorylase B	****
PRDX5|25824	Peroxiredoxin 5	**
YWHAB|7529	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Beta	**
ERBB2|2064	Erb-B2 Receptor Tyrosine Kinase 2	*
SLC2A1|6513	Solute Carrier Family 2 Member 1	*
REG3A|5068	Regenerating Family Member 3 Alpha	**
LCN2|3934	Lipocalin 2	****
SPINK4|27290	Serine Peptidase Inhibitor, Kazal Type 4	****
ANXA2|302	Annexin A2	****
KRT19|3880	Keratin 19	****
SERPINA1|5265	Serpin Family A Member 1	****
CLCA1|1179	Chloride Channel Accessory 1	****
MUC5B|727897	Mucin 5B, Oligomeric Mucus/Gel-Forming	****
TFF3|7033	Trefoil Factor 3	****
PKM2|5315	Pyruvate Kinase M1/2	****
S100A6|6277	S100 Calcium Binding Protein A6	****
AGR2|10551	Anterior Gradient 2, Protein Disulphide Isomerase Family Member	****
LYZ|4069	Lysozyme	****
REG1A|5967	Regenerating Family Member 1 Alpha	****
REG4|83998	Regenerating Family Member 4	****
GAPDH|2597	Glyceraldehyde-3-Phosphate Dehydrogenase	****
FCGBP|8857	Fc Fragment Of IgG Binding Protein	****
MUC2|4583	Mucin 2, Oligomeric Mucus/Gel-Forming	****

The upper percentile of ST6GAL1 expressers showed an up-regulation of one CEA-related gene (CEACAM5) and CD24 molecule. The products of these two genes share a common involvement in cell adhesion. On the contrary, down-regulated genes are involved in several cellular functions including carbohydrate metabolism and transport, regeneration and repair, protein synthesis, cytoskeleton organization and mucin production (detail functions of modulated genes are presented in Table VI. **2** from appendix section- chapter VI). Down-regulation of genes related with mucin production (MUC2, MUC5B, AGR2, CLCA1 and FCGBP) in higher ST6GAL1 expresser patients is in line with the observation that non-mucinous tumor types display high ST6GAL1 expression (**Figure III. 1E**).

1.2 The impact of ST6GAL1 overexpression on the transcriptome is strongly cell-type specific

To establish whether and how the expression of ST6GAL1 affects the global transcriptional activity of colon cancer cells, we have used the colon cancer cell lines SW948 and SW48 previously permanently transduced with the cDNA of the human ST6GAL1[130] (SW948 ST and SW48 ST) and their respective negative controls, transduced with an empty vector (SW948 NC and SW48 NC). The negative controls of both cell lines lacked endogenous ST6GAL1 expression, like their wild type counterparts (**Figure VI. 1**, from appendix section - chapter VI)[99,137]. However, after ST6GAL1 transduction, both SW948 ST and SW48 ST expressed the ST6GAL1 mRNA and protein, as well as enzymatic activity and the presence of Sia6LacNAc on their cell membranes, as revealed by fluorescent SNA (**Figure VI. 1**, from appendix section-chapter VI). The two cell lines have been chosen because they are representative of the two main genetic pathways of transformation in CRC. SW48 cells exhibit microsatellite instability while SW948 cells display chromosomal instability.

To investigate whether and how the overexpression of ST6GAL1 could modify the gene expression profile of the colorectal cancer cell lines in study, we performed a RNA microarray analysis. **Figure III. 3** shows heat-maps representing the genes modulated in the two cell lines transduced with ST6GAL1 compared to their negative controls, revealing a clearly clustering in two different groups (NC and ST6GAL1 transduced cells) based on the similarity of gene expression pattern. The number of

genes modulated is higher in SW948 cells (**Figure III. 3B** - 137 genes modulated) than in SW48 cells (**Figure III. 3A** - 66 genes modulated).

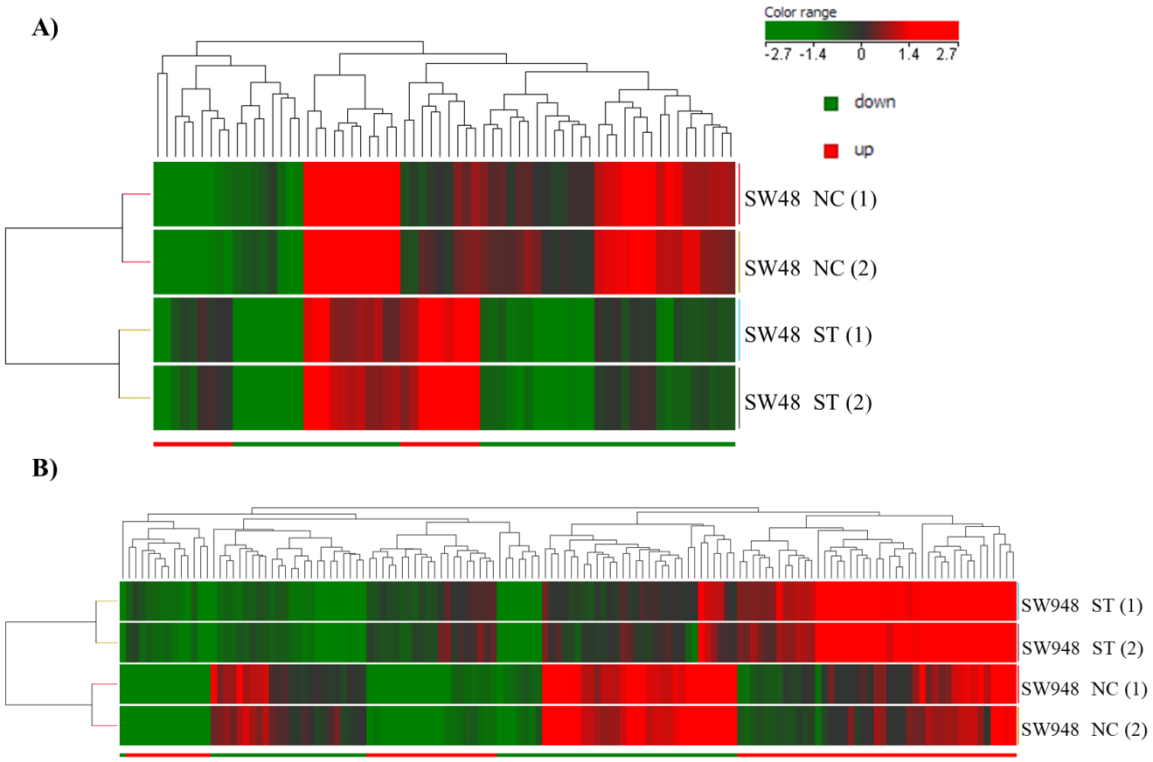


Figure III. 3 - Heat-maps using the list of genes that are differentially expressed between ST6GAL1-transduced SW48 cells and respective negative control (A) and ST6GAL1-transduced SW948 cells and respective negative control (B). Each sample was analyzed with two independent replicates (1 and 2). Genes (columns) and samples (rows) were grouped by hierarchical clustering (Manhattan correlation). High- and low- expression is normalized to the average expression across all samples. Test to analyze the differences between ST6GAL1-transduced cells and negative control: moderated t-test; corrected p-value cut-off: 0.15; multiple test correction used: Benjamini-Hochberg.

Table III. 2 and **Table III. 3** report the list of genes significantly modulated by ST6GAL1 expression with a fold change ≥2 in SW48 and SW948 cells, respectively (for complete gene name, see **Table VI. 3** and **Table VI. 4** from section Appendix-chapter VI). Unexpectedly, none of the genes modulated in either of the two cell lines showed parallel modulation in the other.

Table III. 2 - Genes modulated in SW48 ST cells compared to SW48 NC cells. Genes differentially expressed where selected by a fold change≥2 and a correct p-value cut-off<0.15. In red and green, are represented up-regulated and down-regulated genes, respectively. For correction of p-values, Benjamini-Hochberg multiple test comparison was applied. X represents a gene not identified or with unknown/unclear functions.

Gene Symbol	Fold change	P-value corrected	Gene Symbol	Fold change	P-value corrected
CALCA	-6.86	0.1046	UGT2B15	-2.40	0.1368

NUDT12	-5.77	0.1368	PDE4A	-2.38	0.1328
MAP1B	-5.49	0.1468	EPHA3	-2.37	0.1069
PARP11	-4.79	0.1368	HLA-DPB1	-2.37	0.1069
UGT2B7	-4.41	0.1368	INPP4B	-2.34	0.1069
NTRK2	-4.13	0.1368	TRAPPC2	-2.31	0.1069
X_1	-3.63	0.1368	X_5	-2.30	0.1368
KRT17	-3.61	0.1069	X_6	-2.28	0.1468
KRT14	-3.45	0.1368	INA	-2.27	0.1368
UGT2B10	-3.41	0.1368	ANK3	-2.16	0.1368
KRT16	-3.23	0.1069	SNORD83A	-2.13	0.1328
PADI1	-3.14	0.1368	HAS2	-2.06	0.1424
X_2	-3.13	0.1424	SERHL2	-2.05	0.1368
PDLIM5	-3.04	0.1490	INHBB	-2.04	0.1424
UGT2B11	-3.03	0.1328	TMEM41A	-2.03	0.1368
ADAM7	-3.02	0.1468	TLR1	2.01	0.1368
STRA6	-2.99	0.1368	DNAJB13	2.03	0.1368
X_3	-2.90	0.1368	PXDNL	2.06	0.1368
ANXA8L1	-2.84	0.1368	FHOD3	2.16	0.1069
SEMA3A	-2.84	0.1368	KDM5D	2.19	0.1468
C14orf37	-2.83	0.1328	X_7	2.24	0.1391
ASB4	-2.80	0.1399	X_8	2.30	0.1368
CNTLN	-2.77	0.1368	MUC2	2.42	0.1424
OBSCN	-2.76	0.1368	PCDH1	2.45	0.1490
SLC30A3	-2.74	0.1069	FABP6	2.49	0.1424
EPS8L1	-2.73	0.1328	CSAG1	2.61	0.1368
VWA5A	-2.72	0.1490	CCDC170	2.75	0.1424
COLCA1	-2.61	0.1368	ZP1	2.91	0.1448
CLEC2B	-2.60	0.1368	X_9	3.20	0.1328
ANO1	-2.60	0.1328	KLRC1	3.59	0.1424
MS4A15	-2.48	0.1368	LOC403323	3.80	0.1328
X_4	-2.44	0.1368	X_{10}	5.59	0.1368
ST8SIA1	-2.44	0.1069	LOC729224	5.88	0.1424

Table III. 3 - Genes modulated in SW948 ST cells compared to SW948 NC cells. Genes differentially expressed where selected by a fold change≥2 and a correct p-value cut-off<0.15. In red and green, are represented up-regulated and down-regulated genes, respectively. For correction of p-values, Benjamini-Hochberg multiple test comparison was applied. X represents a gene not identified or with unknown/unclear functions.

Gene Symbol	Fold change	P-value corrected	Gene Symbol	Fold change	P-value corrected
PEG3	-40.88	0.1162	S100A3	2.10	0.0533
EIF4G3	-21.08	0.0944	NRP2	2.11	0.1044
EML5	-18.79	0.1406	DEFB130	2.14	0.0944
X_{11}	-10.52	0.0914	C9orf84	2.14	0.0533
FAM230A	-10.04	0.1254	SCGB2A1	2.16	0.1181
LOC102725453	-8.37	0.0511	C3orf35	2.18	0.1254
LOC101060157	-6.72	0.0914	MUC17	2.18	0.1368
CXCL12	-6.18	0.1254	CAPN13	2.21	0.1368
X_{12}	-5.71	0.1254	DOC2GP	2.24	0.0914
SLC38A3	-5.38	0.0914	OR5F1	2.28	0.0944
WNT4	-5.22	0.1393	GGT6	2.33	0.1390
X_{13}	-5.19	0.0638	ERICH5	2.34	0.1254
PPP1R17	-4.72	0.0563	XIST	2.35	0.1284
X_{14}	-4.56	0.0511	FAM171B	2.35	0.1474
CTF1	-4.23	0.0914	ZC3H13	2.36	0.1390
POLH	-4.06	0.1115	PRAM1	2.39	0.1460
CDCA5	-4.03	0.0914	SLC22A13	2.40	0.1254
X_{15}	-3.87	0.0526	C6orf52	2.41	0.0980
TCF15	-3.84	0.0558	SLC15A1	2.44	0.1368
X_{16}	-3.83	0.0511	AHNAK	2.50	0.0533
X_{17}	-3.82	0.1254	KLF8	2.53	0.0862
X_{18}	-3.67	0.0526	X_{31}	2.54	0.1390
FLJ32154	-3.44	0.0511	ADRB2	2.57	0.1254
LOC100506458	-3.42	0.1016	CAPN13	2.61	0.1463
FANK1	-3.42	0.0914	ZNF543	2.66	0.0526
HOXC13	-3.40	0.0914	PPM1K	2.70	0.0526
DUSP26	-3.37	0.0914	CD33	2.71	0.0542
X_{19}	-3.28	0.1254	SH2D6	2.77	0.0809
LOC101929469	-3.28	0.1007	APOE	2.83	0.0914
ZNF71	-3.16	0.0914	RNF144A	2.90	0.1153
X_{20}	-3.10	0.1044	PHLDA3	2.91	0.0533
BEST2	-3.03	0.1368	UTRN	2.91	0.1291
X_{21}	-2.96	0.1390	CASC4	3.02	0.0914
X_{22}	-2.91	0.1367	PPP5D1	3.04	0.0526
ZNF483	-2.86	0.0944	P3H2	3.08	0.1442
X_{23}	-2.83	0.0526	UBL4B	3.09	0.0944
SYP	-2.82	0.0914	PRSS57	3.12	0.0533
CNIH2	-2.82	0.0914	X_{32}	3.17	0.0944

LOC100131860	-2.68	0.1254	X_{33}	3.21	0.0526
X_{24}	-2.67	0.1472	FAM171B	3.22	0.1461
X_{25}	-2.63	0.0526	LOC101927137	3.26	0.0914
ZNF727	-2.62	0.0809	HOXC5	3.28	0.1472
CHRFAM7A	-2.59	0.1463	NEK11	3.41	0.1254
FANCD2	-2.58	0.1153	MS4A8	3.41	0.0526
SPG7	-2.56	0.1070	KLF13	3.69	0.1254
AGR3	-2.54	0.0526	HSPB3	3.69	0.0511
CCDC121	-2.51	0.0526	ABCC8	3.73	0.0914
POMZP3	-2.50	0.0862	FMNL1	3.74	0.0533
X_{26}	-2.46	0.1390	TRIM9	3.75	0.0526
PDZD2	-2.41	0.0809	S1PR4	3.83	0.1472
FFAR2	-2.33	0.0533	FOLR1	3.88	0.0914
SCML2	-2.31	0.1254	IL1RN	3.88	0.0944
X_{27}	-2.27	0.1367	PRO1082	4.08	0.1027
SPDYE5	-2.21	0.1070	PTAFR	4.10	0.1368
SRRM3	-2.21	0.0914	SEMA7A	4.40	0.1368
X_{28}	-2.19	0.0944	ACE2	4.78	0.0914
LECT1	-2.16	0.0533	MEDAG	5.03	0.1144
X_{29}	-2.04	0.1254	PCDHGB1	5.07	0.1027
IRX2	-2.03	0.1144	SOHLH2	5.38	0.0526
X_{30}	-2.02	0.1254	TGFB2	5.40	0.1254
APBB1	-2.01	0.1390	DQX1	5.87	0.0740
KRIT1	-2.00	0.0833	CNTN5	5.97	0.0809
ZNF239	2.05	0.0820	H2AFY2	6.62	0.0685
EDN2	2.07	0.0845	BST2	6.80	0.1409
LOC388282	2.08	0.0862	ZNF439	6.97	0.0644
MX1	2.08	0.1144	SLC14A1	7.03	0.1472
STAT4	2.08	0.1254	ANKRD1	8.62	0.1368
KCNJ11	2.09	0.1410	MYEOV	9.68	0.0526
			X_{34}	17.24	0.1059

By functional annotation, it is possible to infer the association between the modulated genes and their biological functions. Pathway analysis (**Figure III. 4**) showed that ST6GAL1 affected several pathways in SW48 cells, being the most significant ones the estradiol metabolism and inhibition of ephrin receptors in colorectal cancer. On contrary, SW948-modulated genes are involved in totally different pathways including upregulation of interleukin 8 (IL-8) expression in colorectal cancer (for more detailed data, including p-value and false discovery rate see **Table VI. 5** and **Table VI. 6** in section Appendix-chapter VI).

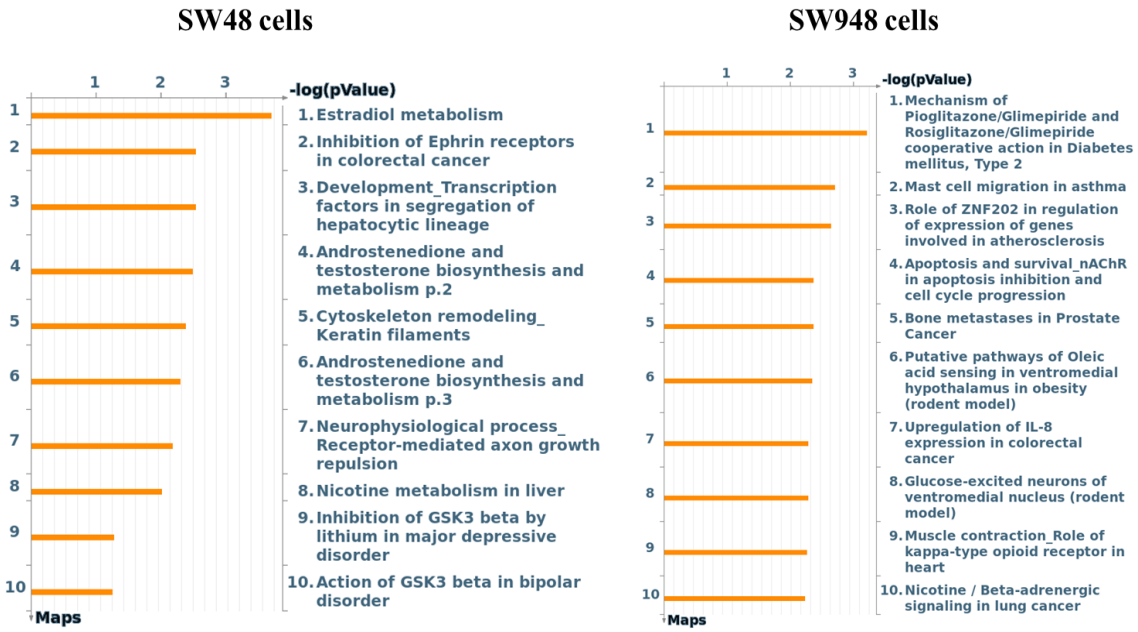


Figure III. 4 - Top 10 significant pathways where the genes modulated by ST6GAL1 in SW48 cells (left side) and SW948 cells (right side) are involved. Pathway map visualization was performed using MetaCore pathway analysis by GeneGo.

Regarding GeneGo process networks (**Figure III. 5**) in SW48 cells, ST6GAL1 gene affected several process networks related with carcinogenesis including cell adhesion-glycoconjugates, cytoskeleton-intermediate filaments, immune response-antigen presentation and development-regulation of angiogenesis. By other way of the top 10 significant process networks in SW948 cells, five of them appeared to be relevant for colorectal cancer: signal transduction-WNT signaling, cell adhesion-cadherins, lymphocyte proliferation, potassium transport and translation-regulation of initiation (for more detailed data, including p-value and false discovery rate, see **Table VI. 5** and **Table VI. 6** in section Appendix-Chapter VI).

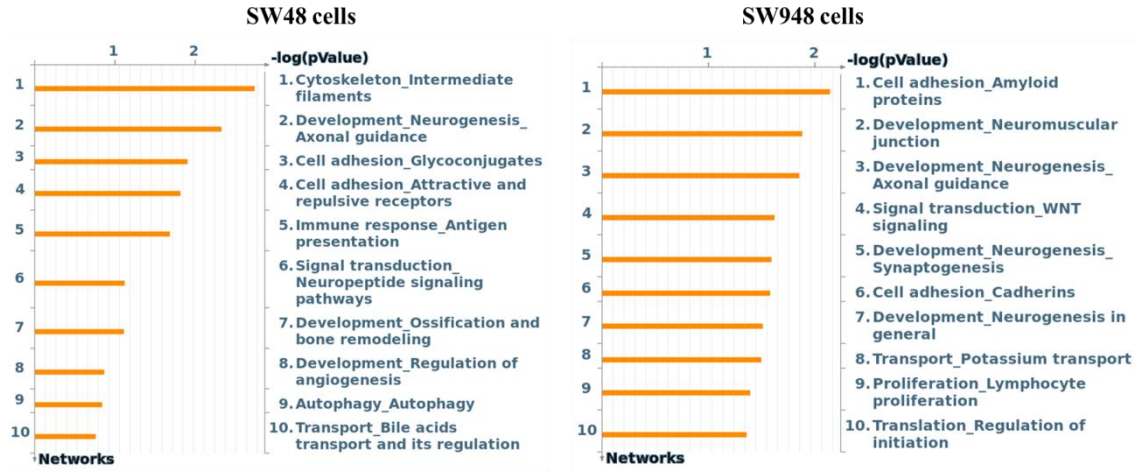


Figure III. 5 - Top 10 significant process networks where the genes modulated by ST6GAL1 in SW48 cells (left side) and SW948 cells (right side) are involved. Pathway map visualization was performed using MetaCore pathway analysis by GeneGo.

Altogether, these data indicate that the impact of ST6GAL1 overexpression on the transcriptional activity of colon cancer cells is strongly cell line specific.

1.3 The phenotype of colon cancer cells is differentially modulated by ST6GAL1 overexpression

To investigate the effect of ST6GAL1 expression on the malignant phenotype, SW48 and SW948 cells transduced with the ST6GAL1 cDNA and their corresponding negative controls were assayed for two key phenotypic features associated with malignant transformation.

To assess if ST6GAL1 gene could modify the ability of cells to proliferate and migrate, wound healing assay was performed using Ibidi culture-insert. The time required to heal a scratch wound in a cell culture (**Figure III. 6**) was shorter for SW948 ST than for SW948 NC. This was mainly due to the fact that SW948 ST cells tend to proliferate first in monolayer and only successively as a multilayer, while SW948 NC cells show a stronger tendency to a multi-layer growth. The expression of ST6GAL1 produced a contrary effect in SW48 cells. Indeed, SW48 NC cells closed faster the wound comparing with SW48 ST cells.

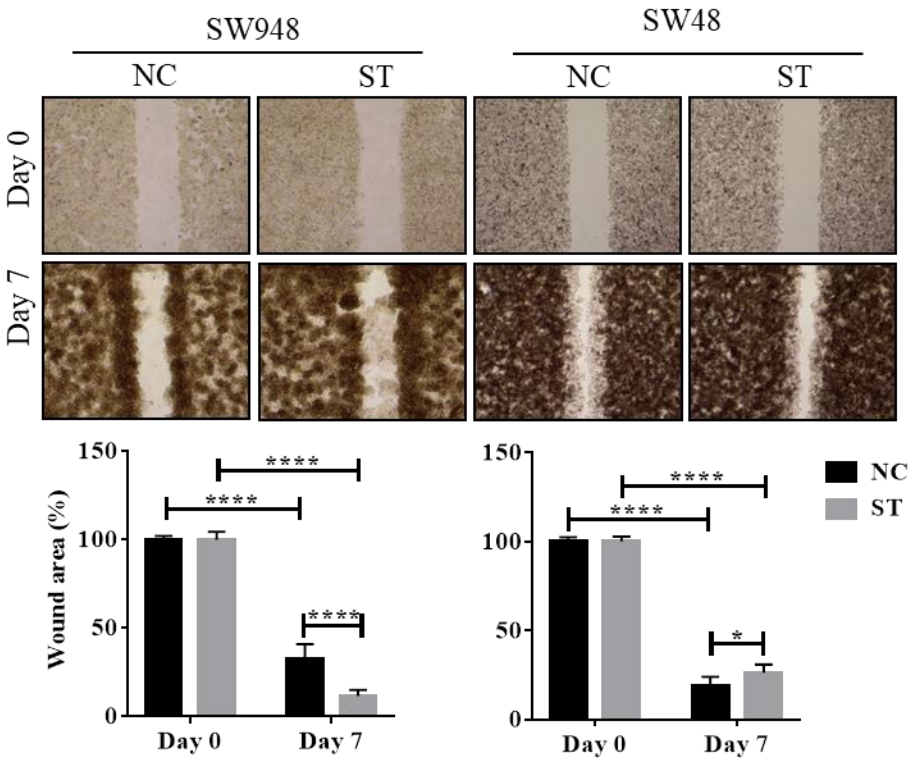


Figure III. 6 - Scratch wound test. A wound was made in a confluent cell layer and the healing of the wound was monitored after 7 days. Representative fields are shown at the top (original magnification 10×), while the percentage of free wound measured after seven days of healing is reported at the bottom. Statistical analysis was performed using two-way ANOVA and Tukey's multiple comparisons test. *≤0.05, ****p≤0.0001

Anchorage-independent growth is an ability of cells to grow independently on a solid surface and is considered as a hallmark of carcinogenesis. This feature was estimated by a soft-agar colony formation assay. The ability to grow in a semi-solid medium, such as the soft agar (**Figure III. 7**) was differentially modulated by ST6GAL1 in SW948 and SW48 cells. In fact, it resulted in the formation of fewer clones in the SW948 ST cells compared with SW948 NC cells while in SW48 cells it was the opposite effect.

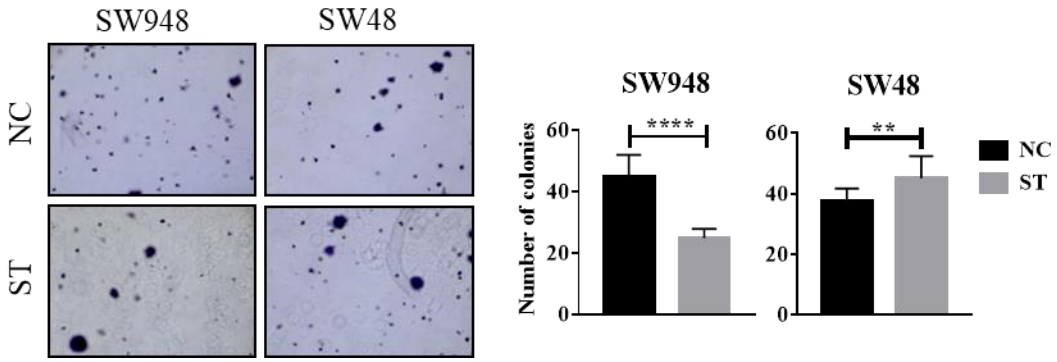


Figure III. 7 - Soft agar growth. Mock- or ST6GAL1-transduced cells were seeded in soft agar and the number of the colonies was determined as described in Materials and Methods section. Representative fields are shown on the left (original magnification 5×).The total number of colonies is reported on the right. Statistical analysis was performed using non-parametric Kolmogorov-Smirnov test. **≤0.01, ****p≤0.0001

2. Transcriptomic and phenotypic impact of B4GALNT2 expression in colorectal cancer

2.1 Survey of TCGA database

Data retrieved from TCGA database included also the gene expression of B4GALNT2 in the same cohort of 623 colorectal adenocarcinoma patients used for the ST6GAL1 study.

The analysis of transcriptomic data of CRC and normal specimens from TCGA database allowed the identification of relationships between B4GALNT2 gene expression and clinical features. As shown in **Figure III. 8A**, B4GALNT2 mRNA expression is variable among CRC specimens, being present at residual levels in the majority of the patients. Because of CRC residual levels, all data was presented in histograms instead of box-plots graphics. In **Figure III. 8B**, it is possible to observe a clear down-regulation of B4GALNT2 mRNA in CRC comparing with normal colonic

mucosa. No relationship was found between CRC stage and microsatellite *status* with B4GALNT2 mRNA expression in CRC tissues (**Figure III. 8C and D**, respectively). High B4GALNT2 mRNA levels were associated with mucinous phenotype and a better response to treatment (**Figure III. 8E and F**, respectively).

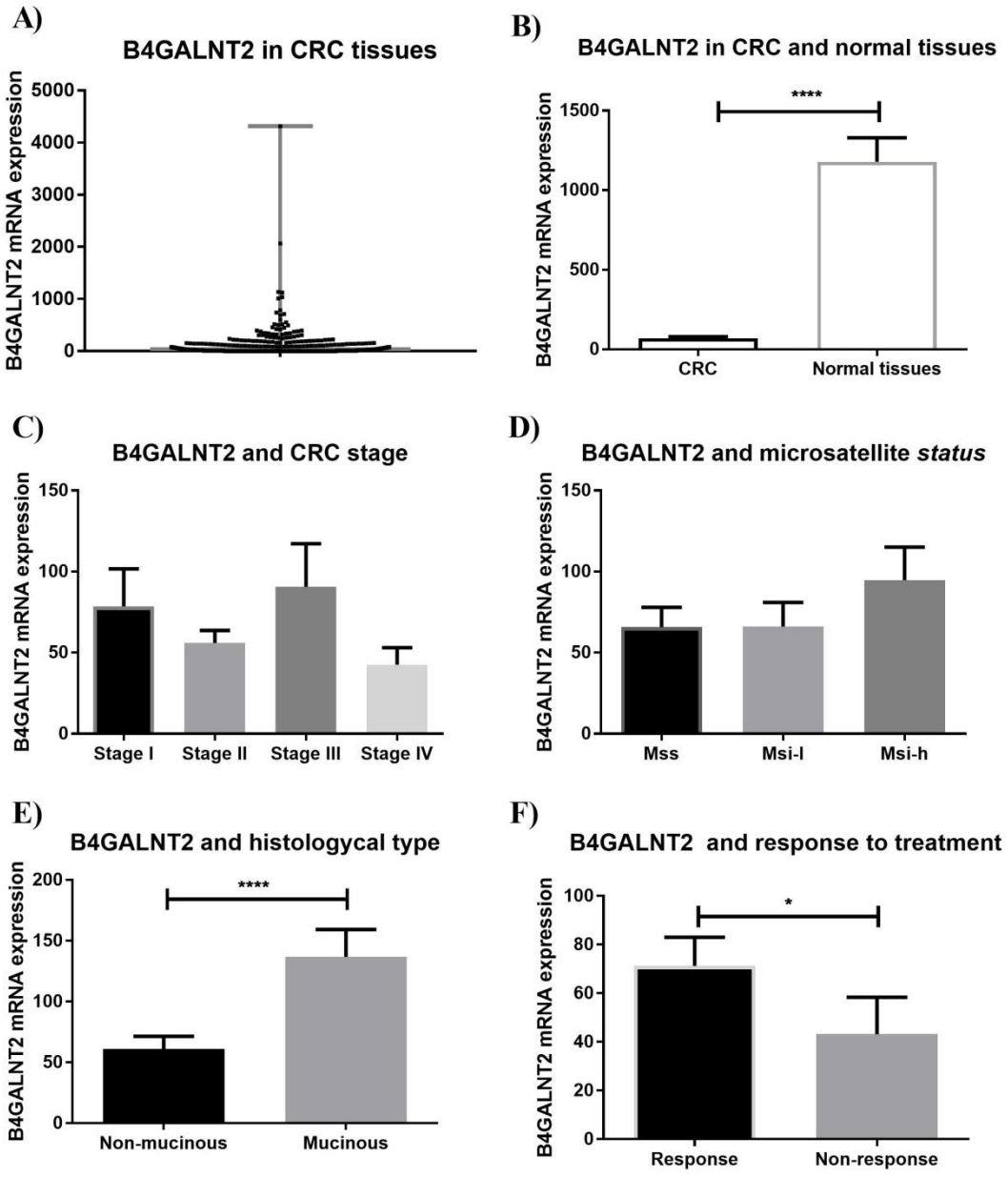


Figure III. 8 - Data from the TCGA database. A: B4GALNT2 mRNA expression in cancer tissues of CRC patients. Each black dot represents the value for one patient. B-F: Histograms showing mean+ standard error of mean (SEM) of B4GALNT2 mRNA expression in CRC and normal tissues (B); in CRC tissue of stage I-IV patients (C); in CRC tissue of patients with microsatellite stability (MSS), low grade microsatellite instability (MSI-l), high grade microsatellite instability (MSI-h) (D); with mucinous or non-mucinous phenotype (E) and showing response or non-response to treatment (F). Mann Whitney test was used in B, E and F. Kruskal-Wallis test was used in C and D. *≤0.05, ****p≤0.0001

We compared the level of B4GALNT2 mRNA in patients with or without mutations in genes relevant for CRC carcinogenesis such as oncogenes *KRAS* and *BRAF* and the tumor suppressor genes *TP53* and *APC*. No association was observed between B4GALNT2 expression and APC (**Figure III. 9D**) and BRAF (**Figure III. 9A**). Surprisingly, we found a significant association between high B4GALNT2 expression and KRAS-mutated (**Figure III. 9B**) and also between high B4GALNT2 expression and wild-type TP53 (**Figure III. 9C**).

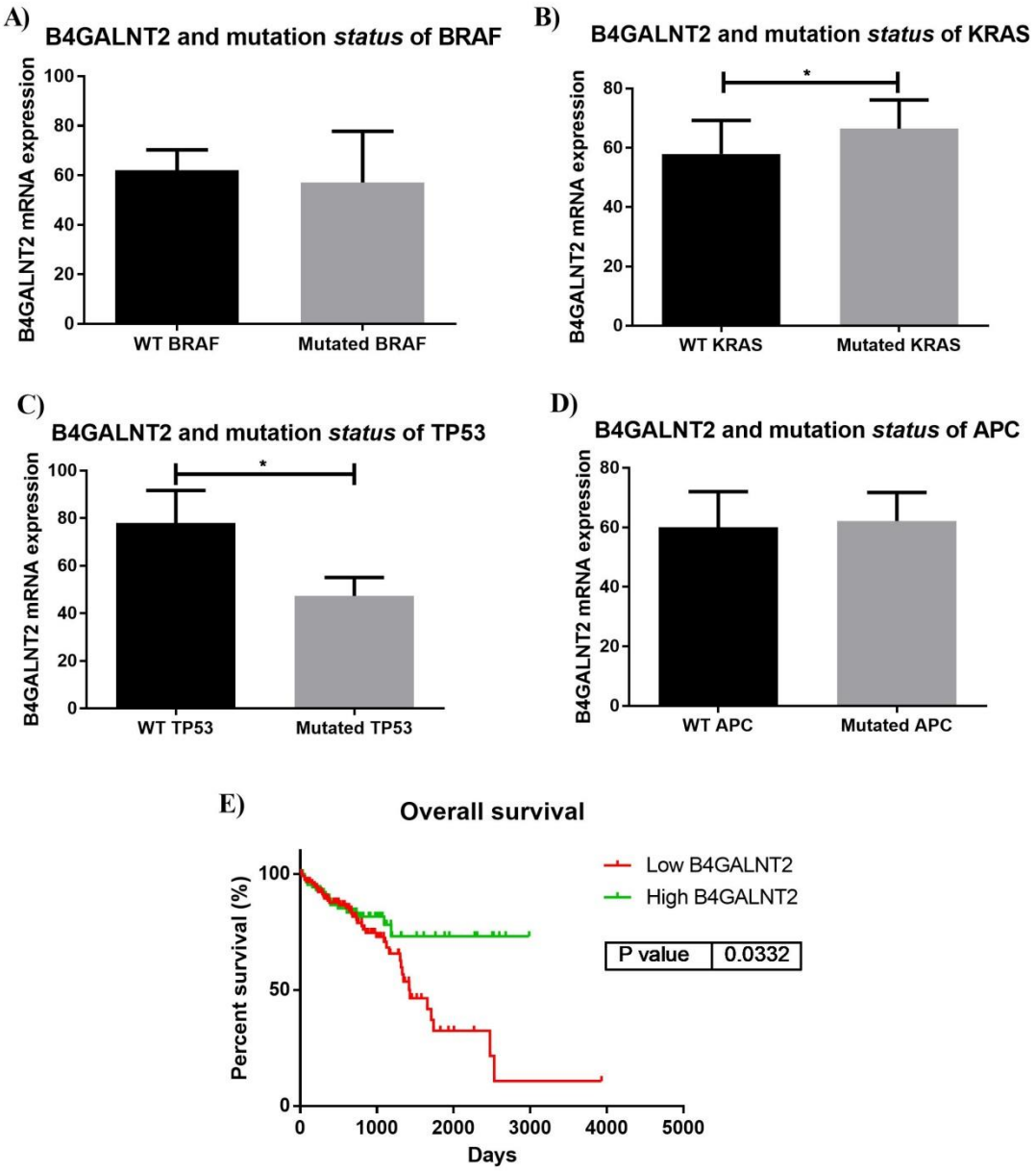


Figure III. 9 - Data from the TCGA database. A-D: Histograms mean+ standard error of mean (SEM) of B4GALNT2 mRNA expression in CRC tissues of patients with wild type or mutated BRAF (A), KRAS (B), TP53 (C) and APC (D). E: Survival curves of patients falling in the upper or lower 15% percentile of B4GALNT2 mRNA expression. Mann Whitney test was used in A, B, C and D. The survival curve was calculated using the Mantel-Cox test for log-rank. *p≤0.05

We also subjected TCGA data to Kaplan-Meier survival analysis to assess the relation between patient's overall survival and high or low level of B4GALNT2 mRNA expression. Once more, two groups were formed: high (upper 15% percentile values) and low (lower 15% percentile values) B4GALNT2 mRNA expression. High B4GALNT2 expressers showed a better overall survival compared with those with low B4GALNT2 expression (**Figure III. 9E**). Collectively, these data suggest a strong impact of B4GALNT2 level in CRC patients, in particular its relation with a better prognosis and a better response to treatment.

In search of gene expression signatures associated with high or low B4GALNT2 expression, two cohorts including the patients in the 15% upper and 15% lower percentiles of B4GALNT2 mRNA level were compared. **Table III. 4** represents the genes statistically modulated between high and low level of B4GALNT2 mRNA expression in CRC patients.

Table III. 4 - Genes modulated in high/low B4GALNT2 cohorts. Genes showing a percentage change higher than 50, according to the formula: PC = $(Mean_{15\%\ high} - Mean_{15\%\ low})/Mean_{All} * 100$ are shown. In red and green are reported genes with a PC higher that 50 or lower than -50, respectively. Bonferroni's multiple comparisons test was used for the comparison between the two cohorts. *≤0.05, **≤0.01, *** p≤0.001, ****p≤0.0001

Gene symbol/Gene ID	Gene name (from Human Gene Database)	Adjusted P-value
FCGBP|8857	Fc Fragment Of IgG Binding Protein	****
PIGR|5284	Polymeric Immunoglobulin Receptor	****
MUC2|4583	Mucin 2, Oligomeric Mucus/Gel-Forming	****
ADAM6|8755	A Disintegrin And Metalloproteinase Domain 6	****
REG4|83998	Regenerating Family Member 4	****
CLCA1|1179	Chloride Channel Accessory 1	****
TFF3|7033	Trefoil Factor 3	****
LGALS4|3960	Galectin 4	****
AGR2|10551	Anterior Gradient 2, Protein Disulphide Isomerase Family Member	****
IGJ|3512	Joining Chain Of Multimeric IgA And IgM	****
REG1A|5967	Regenerating Family Member 1 Alpha	****
LOC96610|96610	Ribosome assembly protein (yeast) pseudogene (LOC96610)	****
SERPINA1|5265	Serpin Family A Member 1	***
ITM2C|81618	Integral Membrane Protein 2C	**
MUC5B|727897	Mucin 5B, Oligomeric Mucus/Gel-Forming	**
SPINK4|27290	Serine Peptidase Inhibitor, Kazal Type 4	*

LCN2|3934	Lipocalin 2	*
IGF2|3481	Insulin Like Growth Factor 2	****

Data reported in the table above indicate that the number of up-regulated genes in high B4GALNT2 expressers is much higher than the number of down-regulated genes. Functions such as regeneration and repair, mucin production, immune regulation and protein synthesis are some of the biological functions related with the up-regulated genes (detail functions of modulated genes are presented in **Table VI. 7** from appendix section- chapter VI). Interestingly, the genes up-regulated here are mainly the same genes downregulated in high ST6GAL1 mRNA expressers. Up-regulation of genes related with mucin production (MUC2, MUC5B, AGR2, CLCA1 and FCGBP) in higher B4GALNT2 expresser patients is in line with the observation that mucinous tumor types display high B4GALNT2 expression (**Figure III. 8E**). Only one gene appeared down-regulated, insulin like growth factor 2 (IGF2), acting mainly with a growth-promoting function.

2.2 The anti-tumoral effect of B4GALNT2 overexpression on the colorectal cancer transcriptome

To establish whether and how the expression of B4GALNT2 affects the global transcriptional activity of colon cancer cells, we have used the colon cancer cell line LS 174T transfected with the cDNA of the short form of B4GALNT2 (originating subsequently two clones S2 and S11) and respective negative control (transfection with an empty vector, originating a polyclonal Neo population)[107]. As shown in **Figure VI. 2** (see chapter VI-appendix section), Neo clones lacked B4GALNT2 expression. Previous experiences with the polyclonal Neo population also demonstrated the lack of B4GALNT2 expression. After B4GALNT2 short-form transfection, selected clones S2 and S11 expressed the B4GALNT2 mRNA and activity as well as the presence of its cognate antigen Sd^a, as revealed by dot-blot analysis.

To investigate whether and how the overexpression of B4GALNT2 could modify the gene expression profile of the colorectal cancer cell clones in study, we performed a RNA microarray analysis. RNA preparation of the Neo population was performed independently twice and analysed by microarray as two replicates for the negative control. RNA preparation of short-form B4GALNT2 transfectants was prepared once for S2 and once for S11 and analysed by microarray as two replicates for

the B4GALNT2 overexpression transfectants. **Figure III. 10** illustrates a heat-map with the modulated genes between short-form B4GALNT2 transfectants (clone S2 and S11) and mock-transfectant (Neo population), revealing a clearly clustering in two different groups (Neo population and B4GALNT2 transfectant clones) based on the similarity of gene expression pattern.

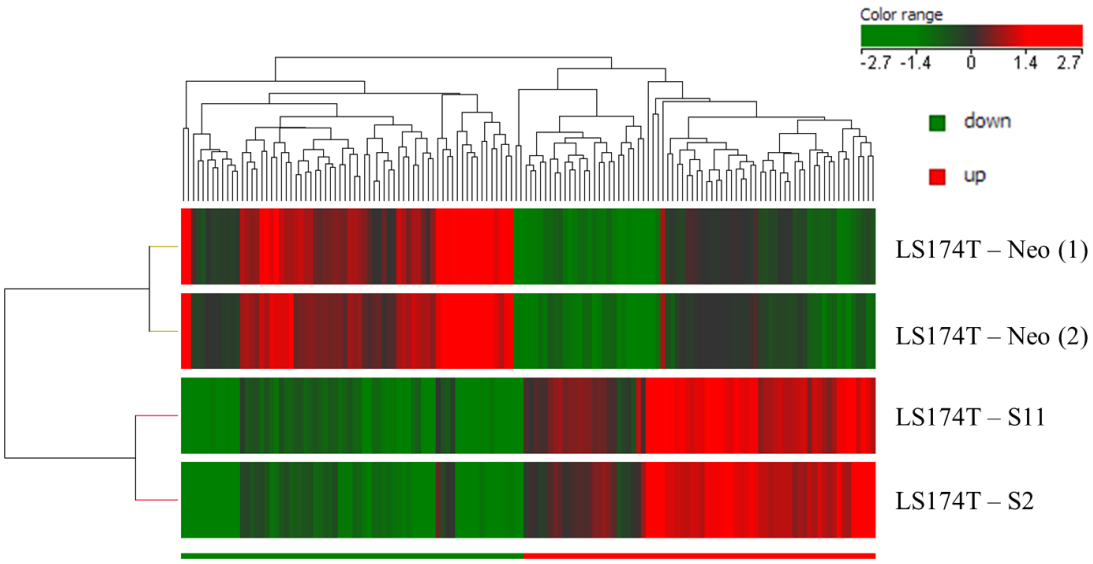


Figure III. 10 - Heat-map using the list of genes that are differentially expressed between LS 174T B4GALNT2-short form transfectants (S11 and S2) and mock-transfectant Neo population (two independent replicates, Neo 1 and Neo 2). Genes (columns) and samples (rows) were grouped by hierarchical clustering (Manhattan correlation). High- and low- expression is normalized to the average expression across all samples. Test to analyze the differences between short-form and mock-B4GALNT2 transfected cells: moderated t-test; corrected p-value cut-off: 0.05; multiple test correction used: Benjamini-Hochberg.

Table III. 5 reports a list of 141 genes significantly modulated by B4GALNT2 expression with a fold change ≥2 in LS 174T cell model (for complete gene name, see **Table VI. 8** from section Appendix - chapter VI).

Table III. 5 - Genes modulated in short-form B4GALNT2 transfectant clones compared to Neo population. Genes differentially expressed where selected by a fold change≥2 and a correct p-value cut-off<0.05. In red and green, are represented up-regulated and down-regulated genes, respectively. For correction of p-values, Benjamini-Hochberg multiple test comparison was applied. X represents a gene not identified or with unknown/unclear functions.

Gene Symbol	Fold change	P-value corrected	Gene Symbol	Fold change	P-value corrected
NID1	-89.36	0.0306	PNPLA1	2.04	0.0383
ZNF22	-44.57	0.0077	HNMT	2.04	0.0489
STARD3NL	-27.41	0.0346	CLDN22	2.04	0.0483
GALC	-24.90	0.0077	MYLPF	2.05	0.0386
NPTX1	-17.33	0.0123	KIT	2.06	0.0479
LGALS2	-16.82	0.0142	CALHM1	2.06	0.0479

SOX2	-16.51	0.0163	SUPT3H	2.06	0.0238
MID2	-15.01	0.0152	CHD5	2.07	0.0463
ARMC4	-8.70	0.0196	CD274	2.11	0.0317
NINL	-8.59	0.0339	X_{40}	2.12	0.0489
PEG10	-8.38	0.0233	RNF216	2.12	0.0387
FMO3	-7.70	0.0253	PRDM8	2.13	0.0383
RAI14	-6.64	0.0190	IGF1	2.14	0.0379
ROR1	-6.44	0.0163	SGOL2	2.14	0.0274
MBOAT2	-5.42	0.0383	PHACTR3	2.17	0.0489
MYH3	-5.02	0.0167	DPF3	2.18	0.0339
INMT	-4.72	0.0142	C14orf119	2.24	0.0385
F5	-4.72	0.0331	X_{41}	2.25	0.0331
NINL	-4.65	0.0339	MARCH10	2.26	0.0238
ALX1	-4.57	0.0167	X_{42}	2.27	0.0383
FAM110B	-4.45	0.0142	LOC101930346	2.28	0.0238
X_{35}	-4.39	0.0236	X_{43}	2.29	0.0386
X_{36}	-4.29	0.0331	LOC84843	2.30	0.0430
FAM26F	-4.15	0.0195	SLC7A8	2.32	0.0219
HEY2	-3.90	0.0411	X_{44}	2.35	0.0203
TBATA	-3.66	0.0379	BAI2	2.37	0.0254
RNF152	-3.62	0.0335	F2R	2.39	0.0238
ITGAX	-3.53	0.0167	DGKK	2.43	0.0238
TMEM255B	-3.53	0.0238	C17orf67	2.44	0.0352
EPB41L3	-3.43	0.0339	BTN2A1	2.45	0.0317
KRTAP5-4	-3.31	0.0451	BTN1A1	2.46	0.0331
SOHLH1	-3.19	0.0411	X_{45}	2.47	0.0231
CALCA	-3.17	0.0436	ISX	2.51	0.0302
ANK1	-3.14	0.0167	PALLD	2.52	0.0430
INHBB	-3.10	0.0203	X_{46}	2.53	0.0196
CHSY3	-3.08	0.0167	X_{47}	2.54	0.0411
OBSL1	-2.98	0.0195	SPANXC	2.58	0.0233
TREX2	-2.98	0.0195	UBD	2.62	0.0233
CRIP2	-2.98	0.0383	X_{48}	2.69	0.0253
TRAM1L1	-2.94	0.0167	X_{49}	2.70	0.0331
ADAM8	-2.90	0.0197	LOC100128325	2.75	0.0238
SPRR3	-2.85	0.0238	RHBDL2	2.78	0.0203
WFDC2	-2.68	0.0343	F2R	2.78	0.0317
LIPF	-2.66	0.0190	PHLDB2	2.83	0.0254
FBXW2	-2.56	0.0203	LINC00883	2.85	0.0253
PDZD2	-2.55	0.0411	LBH	2.87	0.0396
VWA5B2	-2.55	0.0331	RHBDL2	2.88	0.0311
PAGE2	-2.53	0.0343	BARX2	2.92	0.0377
ARHGEF4	-2.48	0.0343	X_{50}	2.96	0.0381
FGFR3	-2.46	0.0383	IL26	2.97	0.0317
SFTPB	-2.45	0.0253	X_{51}	3.00	0.0339

ANO1	-2.43	0.0317	C2orf78	3.05	0.0383
X_{37}	-2.41	0.0231	CT45A5	3.06	0.0460
AK4	-2.37	0.0331	X_{52}	3.12	0.0167
AKR1C4	-2.36	0.0474	SLC26A3	3.24	0.0203
HBE1	-2.30	0.0356	SHISA6	3.27	0.0451
CACNA2D1	-2.29	0.0238	LOC101927820	3.45	0.0254
TRIM54	-2.28	0.0343	FLJ36777	3.48	0.0451
NDUFA4L2	-2.26	0.0479	TTC6	3.53	0.0167
X_{38}	-2.24	0.0411	MEF2C	3.57	0.0253
OPRL1	-2.23	0.0231	MUC12	3.88	0.0167
MRC2	-2.22	0.0253	LUM	3.90	0.0451
X_{39}	-2.22	0.0411	LOC727799	4.11	0.0389
SUV420H2	-2.21	0.0383	X_{53}	4.52	0.0233
MUC19	-2.14	0.0356	X_{54}	4.83	0.0254
CPQ	-2.10	0.0236	GNAS-AS1	5.00	0.0317
STAC2	-2.07	0.0360	SLC14A1	5.23	0.0360
CLEC19A	-2.07	0.0331	X_{55}	5.82	0.0137
ALOX15B	-2.04	0.0349	SKAP1	6.59	0.0231
CALY	-2.04	0.0253	NGFRAP1	11.57	0.0383
			CD200	16.84	0.0411

The most up-regulated gene in B4GALNT2 transfectant clones is CD200 (Fold change = 16.84), a cell surface glycoprotein that plays an important role in immunosuppression and regulation of anti-tumor activity. By other way, the most down-regulated genes in B4GALNT2 transfectant clones include nidogen 1 (NOD1), galectin-2 (LGALS2) and transcription factor SOX2, all genes with pro-tumoral properties.

By functional annotation, it is possible to infer the association between the modulated genes and their biological functions. Pathway analysis showed 50 statistical significant pathways where B4GALNT2-modulated genes seem to be involved. The most relevant ones to colorectal cancer are listed in **Table III. 6**. Regarding GeneGo process networks, B4GALNT2 gene affected several process networks such as Cytoskeleton-Actin filaments, Cytoskeleton-Regulation of cytoskeleton rearrangement and Cardiac development-Wnt-beta-catenin, Notch, VEGF, IP3 and integrin signaling. (for more detailed data, including false discovery rate values see **Table VI. 9** in section Appendix-chapter VI).

Overall, these data together with TCGA database survey indicate a strong impact of B4GALNT2 overexpression in the transcriptional activity of cancer cells. In fact, B4GALNT2 overexpression seems to affect genes with crucial roles in malignant

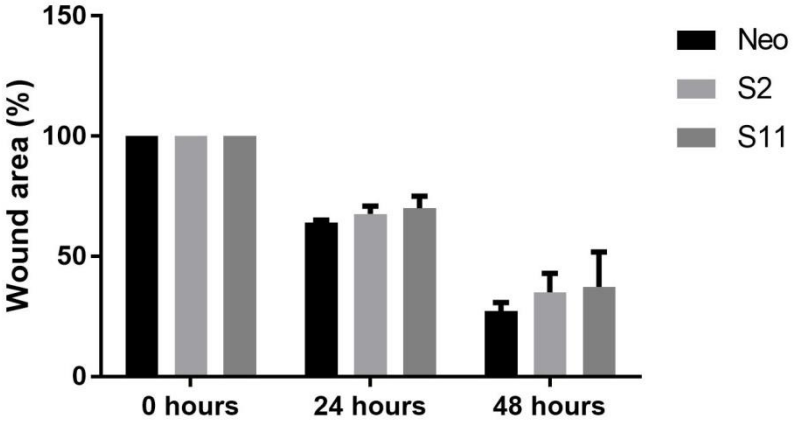
transformation, acting mainly as an anti-tumoral glycosyltransferase in colorectal cancer.

Table III. 6 - Ten most relevant pathway maps related with carcinogenesis transformation altered by B4GALNT2-modulated genes. Pathway map visualization was performed using MetaCore pathway analysis by GeneGo.

GeneGo - Pathway maps	P-value
Cell adhesion_Integrin-mediated cell adhesion and migration	1.697E-02
Cell adhesion_ECM remodeling	2.193E-02
Immune response_Function of MEF2 in T lymphocytes	1.903E-02
Macrophage-induced immunosuppression in the tumor microenvironment	6.051E-02
Transcription_HIF-1 targets	5.940E-02
Immune response _CCR3 signaling in eosinophils	4.079E-02
Immune response_Gastrin in inflammatory response	3.340E-02
Development_Regulation of endothelial progenitor cell differentiation from adult stem cells	2.580E-02
Stem cells_Role of growth factors in the maintenance of embryonic stem cell pluripotency	2.046E-02
Cytoskeleton remodeling_Regulation of actin cytoskeleton organization by the kinase effectors of Rho GTPases	2.422E-02

2.3 Cell migration is not affected by B4GALNT2 overexpression in colorectal cancer cell lines

To evaluate if B4GALNT2 gene could modify the ability of cells to proliferate and migrate (an important feature associated with malignant transformation), wound healing assay was performed with LS 174T Neo population and S2/S11 clones using Ibidi culture-inserts. The time required to heal a scratch wound in a cell culture was shorter for Neo population comparing with S2 and S11 clones, although not reaching statistical significance (**Figure III. 11**).



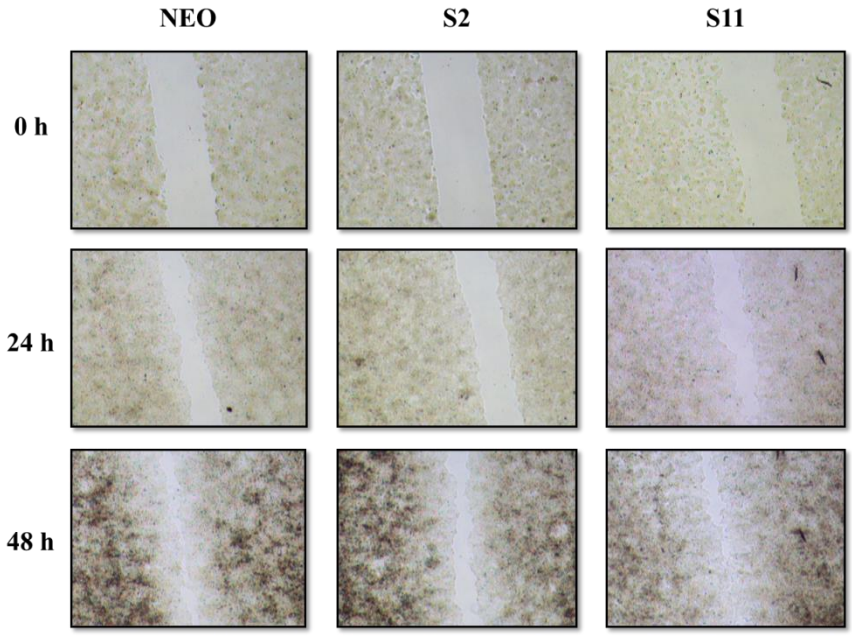


Figure III. 11 - Scratch wound test. A wound was made in a confluent cell layer and the healing of the wound was monitored after 24 hours and 48 hours. The percentage of free wound measured after 24h/48h of healing is reported at the top, while representative fields are shown at the bottom (original magnification 10×). Statistical analysis was performed using two-way ANOVA and Tukey's multiple comparisons test.

3. CRC-specific glycan changes in tissues and plasma samples

3.1 ST6GAL1 activity is up-regulated in CRC meanwhile B4GALNT2 is down-regulated. ST6GAL1 is increased in CRC and IBD compared with healthy individuals plasma

This study involves 19 CRC, 5 polyp and 27 IBD patients who underwent surgery due to suspicion of colorectal cancer at S.Orsola-Malpighi Hospital (Bologna-Italy). A summary of the clinical data is available in **Table VI. 10**. Features such as age at initial diagnosis, gender (female/male), histological subtype of tumor, grade, cancer stage (stage II or III) and IBD subtype (Crohn disease or ulcerative colitis) are the main characteristics collected for this cohort.

We performed enzymatic activity assays in plasma and tissues samples from patients diagnosed with CRC, polyp and IBD. In tissues samples, we analyzed the enzymatic activity of ST6GAL1 and B4GALNT2, while in plasma samples it was only possible to analyze ST6GAL1 activity. ST6GAL1 activity of plasma samples from healthy individuals were taken as reference from a recent work published by our group[134].

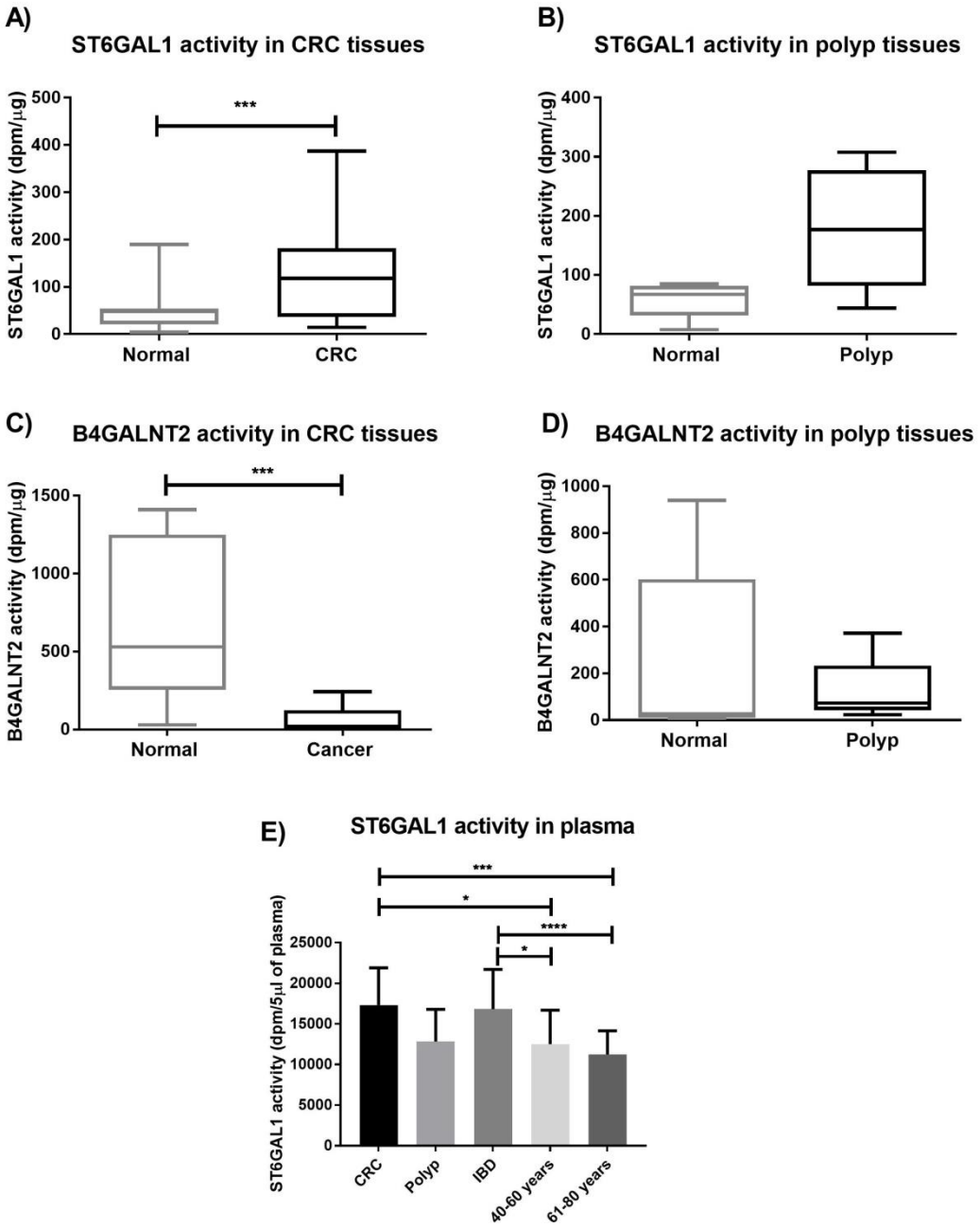


Figure III. 12 - ST6GAL1 and B4GALNT2 enzymatic activity in tissues and plasma samples. A-B: box plot graphs showing median, Q1 quartile, Q3 quartile, minimal and maximal value of ST6GAL1 activity in CRC and corresponding normal tissues (A); in polyp and corresponding normal tissues (B). C-D: box plot graphs showing median, Q1 quartile, Q3 quartile, minimal and maximal value of B4GALNT2 activity in CRC and corresponding normal tissues (C); in polyp and corresponding normal tissues (D). E: ST6GAL1 activity in plasma samples from CRC, polyp, IBD patients and healthy individuals (range 40-60 years and 61-80 years). Values are presented in histograms as mean±SD. Wilcoxon test was used in A-D graphics. Ordinary one-way Anova and Tukey's multiple comparisons test was used in graphic E. *≤0.05, ***p≤0.001, ****p≤0.0001

In **Figure III. 12A** and **Figure III. 12C**, we observe a high ST6GAL1 activity in CRC tissues compared with normal tissues and on the contrary, a high B4GALNT2 activity in normal colonic mucosa compared with cancer tissues. Regarding ST6GAL1

activity in polyp samples (**Figure III. 12B**), it is possible to observe a high activity of ST6GAL1 in polyp tissues compared with normal tissues but, owing to the small number of cases examined, statistical significance was obtained only at a confidence interval of 94% (p-value=0.0625). For the enzymatic activity of B4GALNT2 in polyps (**Figure III. 12D**), they do not follow the same tendency of CRC samples due to the great variability observed among the 5 analyzed polyp samples. In plasma samples (**Figure III. 12E**), it is possible to observe a different level of ST6GAL1 activity between CRC and IBD plasma samples with healthy individuals plasma: ST6GAL1 activity is higher in CRC and IBD plasma samples compared with plasma from healthy individuals. No enzymatic activity differences were detected between the two subtypes of IBD samples, ulcerative colitis and Crohn disease.

Data from TCGA database (section 1.1 from chapter III - Results) showed no differences between the median level of ST6GAL1 mRNA expression between normal and CRC tissues meanwhile B4GALNT2 was down-regulated in CRC samples. For that reason, we decided to evaluate in our small cohort the level of mRNA expression of ST6GAL1 and B4GALNT2 and compare them with the values of enzymatic activity. It was not possible to analyze all samples due to poor RNA quality. In line with the results of enzymatic activity, we observed a low level of B4GALNT2 mRNA expression in CRC comparing with normal colonic mucosa samples (**Figure III. 13B**). On the contrary, no significant differences were detected between ST6GAL1 mRNA level in normal and CRC tissues (**Figure III. 13A**), supporting the data found in TCGA database.

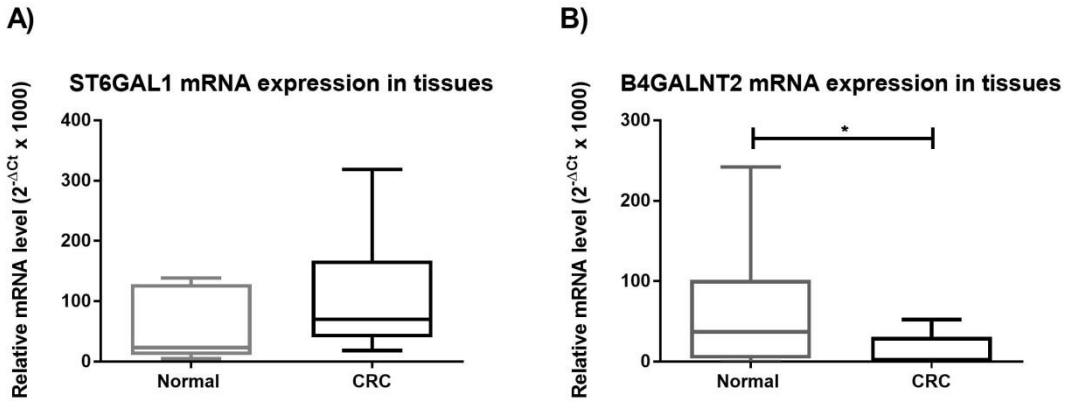


Figure III. 13 - ST6GAL1 and B4GALNT2 mRNA levels in tissues samples. A-B: box plot graphs showing median, Q1 quartile, Q3 quartile, minimal and maximal value of ST6GAL1 mRNA expression in CRC and normal tissues (A); of B4GALNT2 mRNA expression in CRC and normal tissues (B). The relative mRNA levels were analyzed by RT-PCR as described in section Material and Methods. Mann Whitney test was used in A and B. *≤0.05

3.2 Sd^a antigen in plasma samples

The TCGA database analysis revealed an association of the B4GALNT2 glycosyltransferase with patient overall survival, thus we decided to evaluate by western blotting the presence of its cognate antigen (Sd^a antigen) in plasma samples from CRC patients and healthy individuals, trying to observe any differences in the protein pattern. In **Figure III. 14**, it is represented a western blotting picture with some healthy individual plasma samples (H, each H represents a different sample) and CRC plasma samples (K, each K represents a different sample) blotted for the Sd^a antigen. It is possible to observe the presence of a reactive band around 130kDa in almost all the samples but with different levels of intensity among them. Moreover, it was interesting to observe the presence of another band below 55kDa that appears to be more related with a healthy status. The intense band observed bellow 100kDa corresponds to the IgM molecules present in plasma samples due to the cross-reactivity with the secondary antibody used (anti-mouse IgM).

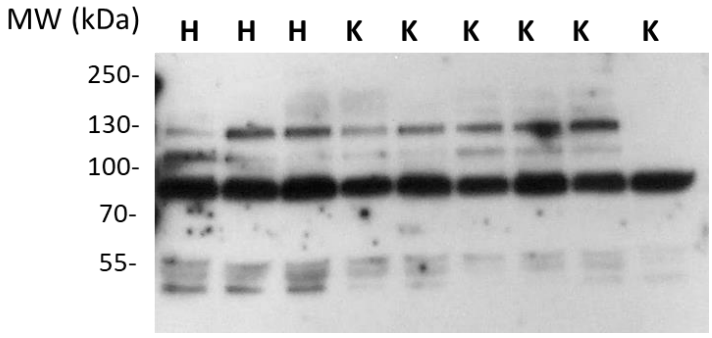


Figure III. 14 - Western blot analysis of Sd^a antigen present in plasma samples from CRC patients (K) and healthy individuals (H). Each K and H represents different samples from cancer and healthy individual plasma, respectively. Western blotting procedure was performed as described in section "Material and Methods- chapter II". Band sizes are indicated in kilodaltons (kDa) at the left side.

As regards to protein carriers of the Sd^a antigen in plasma samples, not a lot of information has been reported in literature. For that reason, we developed an affinity method for the purification of the Sd^a positive bands from plasma samples. Briefly, we performed a first step where a plasma sample passed on an anti-human IgM column, obtaining a flow-through without human IgM molecules. Then, we incubated the resulting sample with anti-Sd^a antibody (mouse IgM) and subsequently, we passed the sample on an anti-mouse IgM column, which retains the added mouse antibody with the bound Sd^a-proteins. After elution and dialysis, the final aim was to obtain a partially purified preparation of Sd^a positive bands. A mock purification without the addition of

anti-Sd^a antibody was performed in parallel to exclude any unspecific bindings to the column. To confirm that our process functioned correctly, all samples and flow-through fractions obtained during the procedure were run by western blotting against Sd^a antigen, confirming that our purification process was successful. The purified band bellow 55kDa together with mock-purification were isolated from the gel and sent to the Institute of Pharmacology and Structural Biology (Toulouse-France) for the glycoproteomic identification of the Sd^a carrier protein by mass spectrometry. A list of several hints were obtained but we selected the proteins with the highest ratio between presence of that particular protein in eluate versus mock-eluate and a p-value<0.05 (**Table III. 7**).

Table III. 7- Glycoproteins identified by mass spectrometry as possible Sd^a carrier proteins in plasma samples regarding the band below 55kDa found by western blotting. Protein name/gene symbol, molecular weight in kilodaltons (kDa), ratio between the protein presence in eluate versus. mock eluate and p-value are represented in this table.

Protein name (gene symbol)	**MW (kDa)**	**Ratio eluate/mock eluate**	**P-value**
Glutathione reductase, mitochondrial (GPHR)	56,3	104.12	0.011
Fetuin A (FETUA)	39,3	103.93	0.012
β-arrestin-1 (ARBB1)	47,1	100.59	0.012
Immunoglobulin heavy constant gamma 3 (IGHG3)	41,3	72.72	0.023
Complement C4-A (CO4A)	192,8	62.98	0.030
Arrestin-C (ARRC)	42,8	57.55	0.035
Dihydrolipoamide S-Succinyltransferase (DLST)	48,8	49.51	0.046
Ceruloplasmin (CERU)	122	47.43	0.049

The tissue expression of identified proteins was checked in NCBI - section Gene (https://www.ncbi.nlm.nih.gov/gene/). Owing to the strict tissue expression of B4GALNT2 in kidney and colon (and its absence in liver), it is unlikely that plasma proteins produced exclusively by the liver can be decorated by the Sd^a antigen. Fetuin A was excluded because there is no expression of this protein in kidney or colon. Complement C4-A, immunoglobulin heavy constant gamma 3, arrestin-C and ceruloplasmin were also excluded as possible Sd^a protein carriers due to their molecular weight. Dihydrolipoamide S-succinyltransferase is a mitochondrial enzyme thus

unlikely to be conventionally glycosylated. At the end, β-arrestin-1 is the main possible candidate as Sd^a carrier protein in plasma samples with molecular weight around 50kDa.

3.3 *N*-glycomic profile of CRC, polyp and normal tissues

Lately, the interest in glycomic profiles has increased in the biomedical field due to the potential utility of glycans in clinical aspects. Here, we describe the *N*-glycosylation profile of 7 CRC, 7 normal and 3 polyp samples. *N*-glycans structures were released by PNGase-F and fluorescent labelled with procainamide, followed by HILIC-UHPLC-FLR-ESI-MS analysis. In **Figure III. 15**, a representative chromatogram of each type of sample (normal, CRC and polyp) is shown. Although the three types of samples contain peaks with similar retention times, their relative abundance is dramatically different.

HILIC profiles obtained were very complex with a broad range of *N*-glycan structures detected in all the samples including co-elution of more than one structure in one single peak. Structures composition, peak id, GU values, coefficient of variation (CV) and average of relative peak area for all the analyzed samples are summarized in **Table VI. 11**, see chapter VI - Appendix. A detailed example is demonstrated below for a CRC tissue sample (**Table III. 8**). Possible glycan structures were allocated to UPLC peaks by: (i) comparison of the GU values to glycan standards; (ii) mass composition data and (iii) from matching m/z values from MS/MS with database of glycans and diagnostic ions. The MS data for some of these peaks was poor and structural identification was therefore not possible. In some cases, the m/z values and MS/MS data for these peaks did not match with m/z values consistent with *N*-glycan structures. In order to gain more detailed information on the glycan structures (e.g. monosaccharide type, linkage and sequence), different analytical techniques such as exoglycosidase sequencing are required.

All samples were processed in triplicate and at the end it was calculated the coefficient of variation for all the identified peaks. Overall, CVs values obtained were less than <15%, suggesting a good reproducibility.

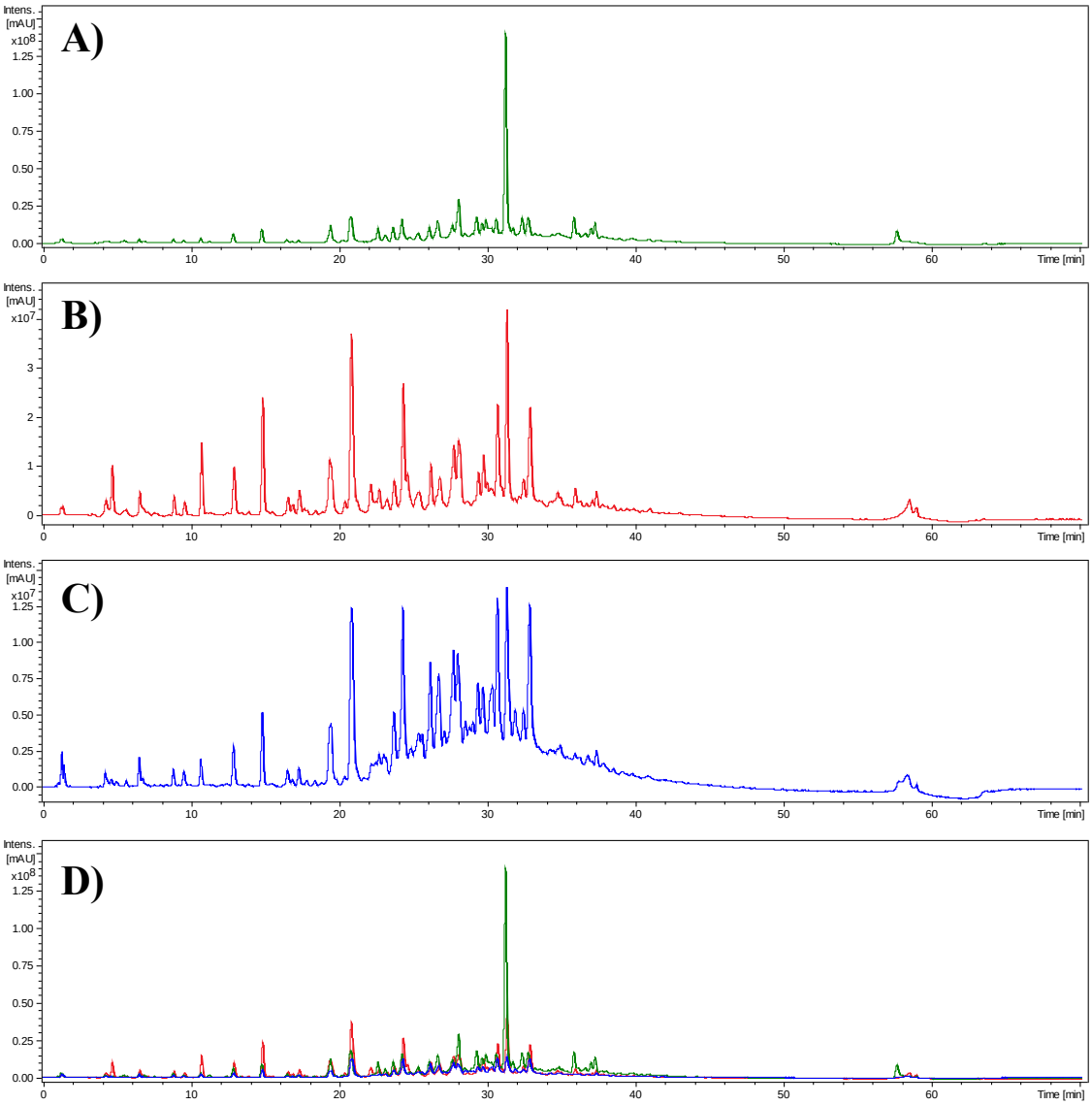


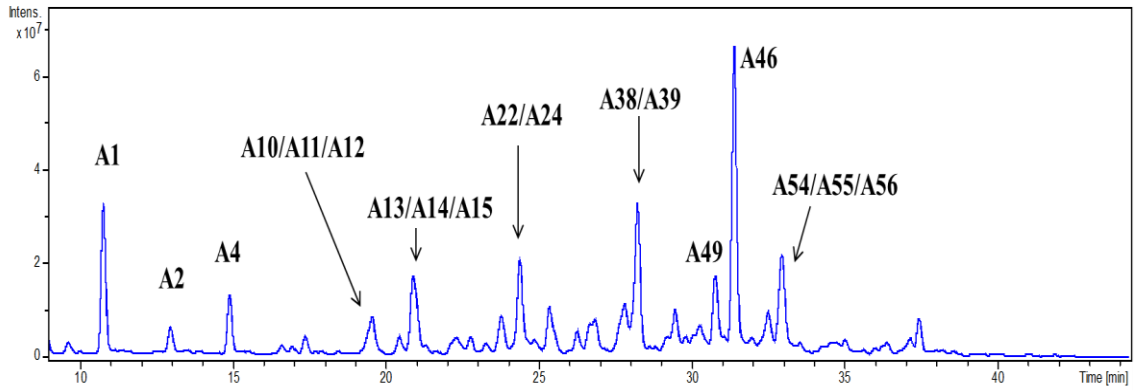
Figure III. 15 - HILIC-UHPLC profile of *N*-glycans present in normal colonic mucosa (A), colorectal cancer tissue (B) and polyp tissue (C). Picture D) represents the visualization of all the three representative profiles together in one spectrum. *N*-glycans were released in triplicate for each sample by PNGase-F followed by procainamide labelling and HILIC-UHPLC analysis with fluorescent detection.

Table III. 8 - *N*-glycans identified in a representative CRC tissue sample. A) Top: Peak identification (peak id), GU values, monosaccharides composition, possible structure, average relative peak area (in percentage) and coefficient of variance (CV) are represented in this table. Regarding monosaccharides composition, H represents a hexose (Hex) normally glucose or galactose; N represents a HexNAc normally GlcNAc or GalNAc; F represents a fucose and S represents a sialic acid. The average of relative peak area was calculated with three replicates. Bottom: Chromatograph with some of the most abundant structures assigned. B) Possible *N*-glycan structures identified in a representative CRC tissue. For glycan symbols, see figure I.7 from introduction section.

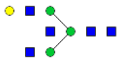
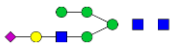
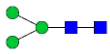
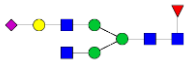
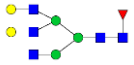
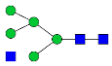
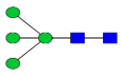
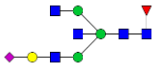
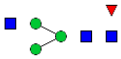
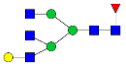
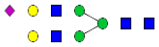
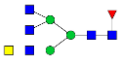
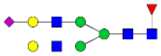
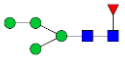
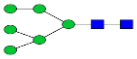
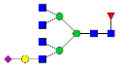
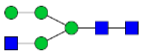
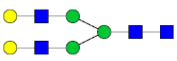
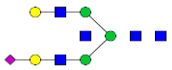
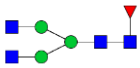
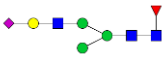
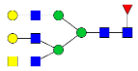
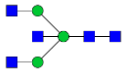
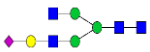
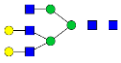
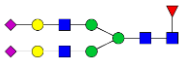
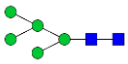
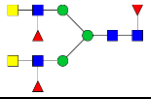
A)

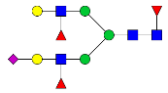
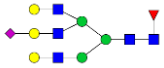
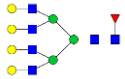
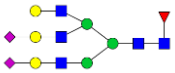
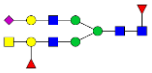
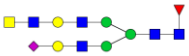
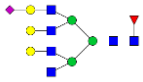
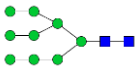
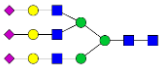
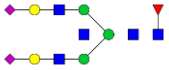
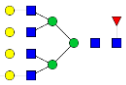
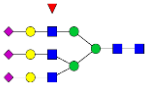
Cancer tissue sample					
UHPLC-FLR-ESI-MS/MS					
Peak id	**GU**	**Composition**	**Possible structure**	**Average % Area**	**CV**
1	**3.69**	H2N2F1	A1	5.85	10.75
5	**4.19**	H3N2	A2	1.12	14.60
6	**4.30**	H2N3F1	A3	0.14	27.83
8	**4.63**	H3N2F1	A4	2.38	9.42
11	**5.01**	H4N2	A5	0.53	14.37
12	**5.08**	H4N2	A5	0.37	1.61
13	**5.18**	H3N3F1	A6	0.85	12.63
14	**5.30**	H3N4	A7	0.12	8.10
15	**5.43**	H4N2F1	A8	0.10	3.91
16	**5.64**	H4N3/H3N4F1	A9/A10	0.65	6.62
17	**5.68**	H3N5/H3N4F1/H3N3F1S1	A11/A10/A12	2.39	22.67
18	**5.89**	H5N2	A13	0.86	3.58
19	**6.00**	H5N2/H4N3F1/H3N5F1	A13/A14/A15	4.99	2.19
20	**6.10**	H4N4	A16	0.31	35.81
22	**6.34**	H3N5F1/H4N5/H5N2F1	A15/A17/A18	1.33	21.44
23	**6.46**	H4N4F1/H5N3	A19/A20	1.13	32.42
24	**6.58**	H4N4F1/H4N3S1	A20/A21	0.63	27.54
25	**6.71**	H4N5F1/H4N3S1/H3N6F1	A22/A21/A23	1.96	13.78
26	**6.87**	H6N2/H4N5F1	A24/A22	5.93	2.79
27	**6.99**	H5N4	A25	0.67	65.30
28	**7.12**	H4N3F1S1/H4N4S1/H5N5	A26/A27/A28	3.37	3.95
29	**7.29**	H5N4F1/H4N4F2	A29/A30	0.23	10.91
30	**7.36**	H5N4F1	A29	1.12	21.15
31	**7.47**	H4N6F1/H4N5S1/H5N3S1	A31/A32/A33	1.29	14.67
32	**7.52**	H4N4F1S1/H4N5S1/H5N5F1	A34/A32/A35	1.97	13.14
34	**7.74**	H7N2	A36	1.26	15.29
35	**7.79**	H7N2/H4N5F1S1/H5N4S1	A36/A37/A38	2.55	4.30

36	**7.91**	H5N4F1S1/H5N4S1	A39/A38	8.51	7.49
37	**8.02**	H5N4S1/H4N6F1S1	A38/A40	0.42	38.61
39	**8.19**	H5N5S1/H5N6F1	A41/A42	0.93	13.19
40	**8.27**	H6N3S1/H5N4F1S1	A43/A39	2.62	2.77
41	**8.37**	H5N4F1S2/H3N6F2	A44/A45	0.99	2.82
42	**8.45**	H5N4S2	A46	0.80	5.08
43	**8.51**	H8N2/H5N5F1S1	A47/A48	2.22	1.64
44	**8.67**	H8N2	A47	4.27	4.38
45	**8.75**	H5N4F3S1	A49	0.71	10.90
46	**8.86**	H5N4S2	A46	15.88	8.98
47	**9.04**	H5N4F2S1/H6N5S1/H5N5S2	A50/A51/A46	1.18	0.80
48	**9.21**	H5N4F3S1/H5N5F2S1/H5N6F1S1	A49/A52/A53	1.70	14.87
49	**9.36**	H9N2/H5N5F1S2/H6N5F1S1	A54/A55/A56	8.02	2.43
51	**9.56**	H6N5F1S1	A56	0.87	12.60
52	**9.81**	H7N6F1/H6N5F1S2	A57/A58	0.64	14.06
53	**9.90**	H6N5S2/H6N6F1S1	A59/A60	1.21	15.26
56	**10.41**	H6N5S3	A61	0.39	2.67
58	**10.55**	H7N6F1S1	A62	0.84	10.11
59	**10.84**	H6N5S3	A61	1.39	4.96
60	**10.95**	H6N5F1S3	A63	1.95	8.02
61	**11.23**	H6N5F1S3	A63	0.72	7.71



B)

Number	Possible N-glycan structure	Number	Possible N-glycan structure	Number	Possible N-glycan structure
A1		A17		A33	
A2		A18		A34	
A3		A19		A35	
A4		A20		A36	
A5		A21		A37	
A6		A22		A38	
A7		A23		A39	
A8		A24		A40	
A9		A25		A41	
A10		A26		A42	
A11		A27		A43	
A12		A28		A44	
A13		A29		A45	
A14		A30		A46	
A15		A31		A47	
A16		A32		A48	

Number	Possible N-glycan structure	Number	Possible N-glycan structure
A49		A56	
A50		A57	
A51		A58	
A52		A59	
A53		A60	
A54		A61	
A55		A62	
		A63	

In all tissues, it was possible to identify the three main classes of *N*-glycans described in the literature (high mannose, complex and hybrid type) and also further modifications that can occur in a glycan structure including fucosylation and sialylation. Truncated structures denominated paucimmanosidic glycans were also detected.

Glycosylation features such as high-mannose, paucimmanose, fucosylation, di- and/or tri-fucosylation, sialylation and di- and/or tri- sialylation were calculated for all the samples and compared among them. Firstly, we compared the levels of high mannose structures (**Figure III. 16A**). A pronounced increase in high mannose *N*-glycans was observed in cancer and polyp samples comparing with normal colonic mucosa tissues. Regarding to paucimmanosidic structures, a high level was found in CRC tissues but not polyp tissues comparing to normal tissues (**Figure III. 16B**).

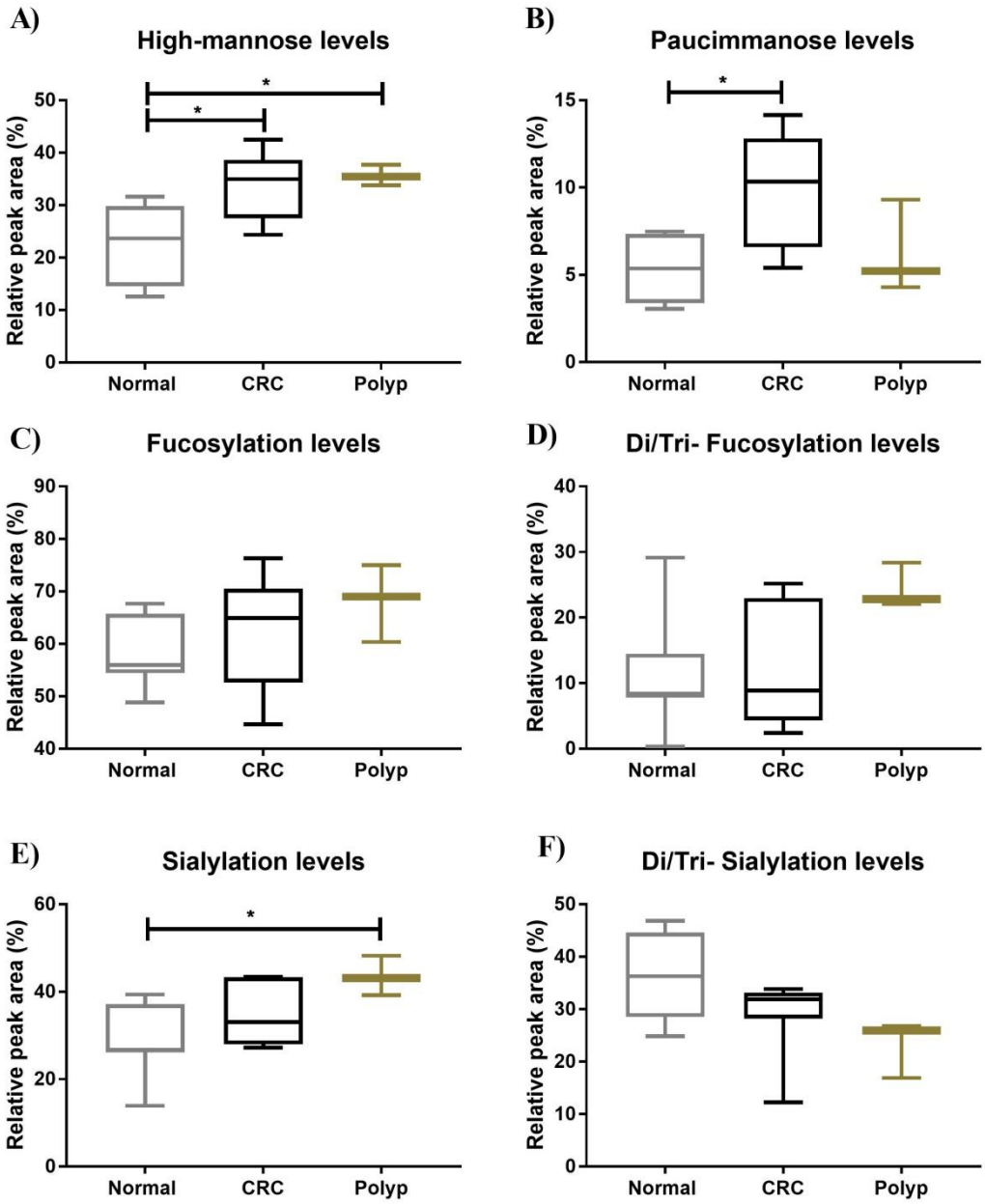


Figure III. 16 - *N*-glycosylation features comparison between normal colonic mucosa, cancer and polyp tissues: high-mannose structures (A); paucimmanose structures (B); overall fucosylation levels (C); di- and/or tri-fucosylated structures (D); overall sialylation levels (E) and di- and/or tri-sialylated structures (F). Box plot graphs show the median, Q1 quartile, Q3 quartile, minimal and maximal value of relative peak area (in percentage) in normal, CRC and polyp tissues. Ordinary one-way Anova and Tukey's multiple comparisons test was used in all graphics. *≤0.05.

Next, we looked at changes in the fucosylation levels. We couldn't observe differences between the level of fucosylation of normal, CRC and polyp samples (**Figure III. 16C** and **Figure III. 16D**). Regarding overall fucosylation and di- and/or tri-fucosylation levels, it was possible to observe a huge variability among normal and CRC samples. The last feature evaluated was the total level of sialylation. An increase level of overall sialylation was observed in cancer and polyp samples, only reaching

statistical significance in the last one (**Figure III. 16E**). By other way, di- and/or tri-sialylation levels appeared to be increased in normal tissues comparing with CRC and polyp tissues (**Figure III. 16F**). Regarding overall sialylation and di- and/or tri- sialylation levels, it was also possible to observe a huge variability among normal and CRC samples.

Chapter IV - Discussion and conclusions

Glycosylation is one of the most common and essential post-translational modifications in cell surface and secreted proteins. Glycans can impact a wide spectrum of cellular functions including cell adhesion, migration, cell signaling and immune recognition. Therefore, alterations in glycan structures contribute significantly to changes in properties of the cells, playing an important role in cancer development and progression.

Colorectal cancer is the second most common cancer in Europe. Unlike many other malignancies, CRC is preventable and potentially curable when detected in its early stages, which however, often remain undiagnosed due to their unspecific symptoms. Current screening techniques are invasive or lack either sensitivity or specificity.

Our main goal in this project is to improve the understanding of the structure-function relationship of glycosylation in CRC to discover in the future improved diagnostic and prognostic biomarkers and pave the way for novel therapeutic targets. More specifically, we performed a glyco(proteo)mic and transcriptomic analysis of CRC cells, plasma and tissues samples, trying to establish a relationship with malignant phenotype and transcriptome changes.

The first and second parts of this project aimed at clarifying the role of ST6GAL1 and B4GALNT2 glycosyltransferases and their respective associated antigens SiaLacNAc and Sd^a, in colorectal cancer progression. These glycosyltransferases and antigens have been associated with cancer but it is not fully clear their role in tumor progression and development. The last part regards the identification of new CRC-specific glycan changes in tissues and plasma samples using a small cohort of clinical samples collected in Sant'Orsola Hospital (Bologna-Italy).

4.1. Impact of ST6GAL1 expression in CRC progression

Several clinical and experimental studies reported a positive association between high level of ST6GAL1 and/or its cognate antigen Sia6LacNAc in cancer progression[138–141] meanwhile others studies reported the opposite effect of this glycosyltransferase[74,99,101,142,143]. To clarify the role of ST6GAL1 in clinics, we used the TCGA data to correlate the ST6GAL1 mRNA expression in CRC with patients'

features. ST6GAL1 mRNA level presented a huge variability among CRC tissues (cohort of 623 patients) but with similar median levels between normal and cancer colonic mucosa. This last point is contrary to what was observed in small cohorts of clinical samples[75,81,144]. Different mechanisms could be responsible for the discrepancy between mRNA and enzymatic activity including different translational efficiency, protein degradation and post-translational modifications. However in our small cohort of CRC patients, we confirm an increased level of ST6GAL1 activity in CRC and report for the first time increased ST6GAL1 activity in polyp tissues, although no significant differences were detected at the mRNA level.

Low ST6GAL1 mRNA expression was observed in patients with a high microsatellite instability (Msi-h) phenotype, being in line with a previous study that reported a correlation between high SNA reactivity and a microsatellite stable CRC phenotype[83]. Another study described a decrease in α2,6 sialylation levels upon transforming growth factor β receptor 2 reconstitution (one of the most interesting Msi target genes) in a MSI CRC cell line model[145]. Low ST6GAL1 level was also associated with a mucinous adenocarcinoma phenotype. A study comparing the profile of mucinous and non-mucinous adenocarcinomas demonstrated that mucinous type presents more frequently a high microsatellite instability *status*[146]. We found also an association between low ST6GAL1 mRNA level and *BRAF* mutated *status*. This observation is consistent with the fact that *BRAF* mutation is strongly associated with microsatellite instability in CRC[147]. Usually, Msi cancers display a slightly better prognosis comparing with Mss cancers[9], so in this CRC cohort ST6GAL1 seems to play a malignant role.

Correlations between ST6GAL1 expression and KRAS mutation[84], drug resistance[148–150], metastasis formation[151–153] and poor prognosis[81] have been reported in previous studies. However, TCGA data did not support any association between ST6GAL1 mRNA expression and KRAS mutation, response to treatment, CRC stage and overall patient survival.

The level of ST6GAL1 allows the classification of CRC patients in two distinct groups displaying different gene expression profiles. High ST6GAL1 expressers displayed mainly high levels of CEACAM5 and CD24, known molecules as promoters of invasion and metastasis *in vitro*[154,155]. Also glycogen phosphorylase B (PYGB) was up-regulated in high ST6GAL1 expressers. PYGB is an enzyme that metabolizes glycogen whose function is to provide energy for an organism in an emergency state. A

recent publication showed that *PYGB* silencing suppressed cell growth and promoted the apoptosis of prostate cancer cells[156]. Peroxiredoxin 5 is an antioxidant enzyme with peroxidase activity that was also up-regulated in high ST6GAL1 cohort. This enzyme showed elevated levels of expression in CRC tissues comparing with normal colonic mucosa[157] and has been associated with tumorigenesis[158] in particular with epithelial-mesenchymal transition in colon cancer[159]. On the other hand, downregulated genes in high ST6GAL1 expressers are involved in a variety of key features of malignancy including carbohydrate metabolism, tissue repair, protein synthesis, cytoskeleton organization and mucin secretion. The most down-regulated gene was mucin 2, an abundant secreted gastrointestinal mucin. It has been reported that its silencing can alter cell migration and also promote CRC metastasis[160–162]. In this group of down-regulated genes, it is not clear if they affect CRC in a pro- or anti-tumoral way: some of them have been described as anti-cancer promoters (such as lipocalin 2[163], keratin 19[164] and CLCA1[165]) and others as pro-tumoral promoters (such as annexin A2[166], REG family members[167,168], human anterior gradient-2[169] and S100A6[170]). It is important to refer that these associations do not necessarily imply a causal relationship between ST6GAL1 expression and that of the other genes.

Collectively, these data suggest a moderate general impact of ST6GAL1 level on CRC progression, although it is not possible to define a clear role as anti- or pro-tumoral effect of this glycosyltransferase.

4.2. The transcriptomic and phenotypical impact of ST6GAL1 overexpression in CRC is strongly cell-type specific

We studied the causal effect of ST6GAL1 on transcriptome modulation in two CRC cell lines, SW48 and SW948, representative of the two main types of genomic instability pathways. Surprisingly, the impact of ST6GAL1 on the transcriptome of the two cell lines was different. Indeed, the number of genes modulated by ST6GAL1 expression in SW948 cells was much larger comparing with those modulated in SW48 cells. Notably, no genes displayed parallel and consistent ST6GAL1-induced modulation in the two cell lines, indicating that the influence of a glycosyltransferase on the transcriptional activity depends on the cellular context.

Analyzing the molecular pathways affected by the ST6GAL1-modulated genes in SW48 cells, estradiol metabolism and inhibition of ephrin receptors were two of the

most significant pathways. Estradiol is the most abundant and potent estrogen in humans, being a member of the steroid hormone family (estrogens). Traditionally, they are associated with female reproduction, however, although is not the central pathway in CRC, some studies reported an important role of estrogens in the initiation and progression of CRC. In fact, loss of expression of estrogen receptor-β (ERβ) is considered part of the CRC tumorigenesis process[171,172]. A study in an APC mouse model reported that estradiol treatment increased the ratio of ERβ and protection against colon cancer[173]. In this transcriptomic analysis, UGT2B11, UGT2B7, UGT2B10 and UGT2B15 were down-regulated in SW48 ST cells. These enzymes play a major role in glucuronidation, an intermediate step in the metabolism of steroids, being able of regulate the levels and activity of estrogen metabolites. Changes in estrogen molecular pathways can activate the transcription of several genes involved in angiogenesis, cellular adhesion, proliferation and apoptosis[174], being in line with the process networks modulated by ST6GAL1 in this cell line (development and regulation of angiogenesis, cell adhesion and signal transduction). It is also important to note that pathways involving other steroid hormones (androstenedione and testosterone) appeared to be modulated in SW48 ST, supporting the importance of steroid hormones metabolism in CRC. Ephrin receptors are one of the largest family of receptor tyrosine kinases (RTKs), being divided into two subclasses, EphA and EphB. In this microarray analysis, a particular receptor EphA3 was down-regulated. Some studies showed an increased level of EphA receptors in early stages of CRC but a low level in advanced stages, suggesting a link between decreased EphA receptors and invasiveness[175,176]. It was suggested that those reduced levels of EphA receptors, in particular EphA1 and EphA2 can alter the adhesion and motility of cancer cells[177]. Other interesting genes in CRC pathogenesis were the down-regulated microtubule-associated protein 1B (MAP1B), a potential oncogene in MSI CRC[178] and neurotrophic tyrosine kinase receptor type 2 (NTRK2), involved in early-stage colon adenocarcinoma[179].

Overall, the few changes induced by ST6GAL1 in SW48 were predominantly towards reduced malignancy.

In SW948 cell line, the extent of ST6GAL1 modulation was much higher and ST6GAL1-modulated genes were completely different from the ones observed in SW48 cells. Overall, the changes induced by ST6GAL1 expression in SW948 cells appear to be equally oriented towards both increased and decreased malignancy. Paternally-expressed 3 (PEG3) was the most down-regulated gene in SW948 ST cells, with a fold

change around 40. It plays a role in cell proliferation and p53-mediated apoptosis, being considered a tumor suppressor gene in ovarian and glioma cells[180,181]. Stromal cell-derived factor 1 (SDF-1 also known as CXCL12) was other interesting downregulated gene in SW948 ST cells. This chemokine and its cell surface receptor (CXCR4) have been well studied in CRC and its presence in the microenvironment is involved in tumor growth, migration and invasion[182]. According to GeneGo pathways analysis, it is possible to observe the influence of this gene in some molecular pathways such as mast cell migration and bone metastasis even not strictly correlated with CRC but other type of diseases.

Some of the most up-regulated genes in SW948 ST cells with known functions in cancer were the myeloma overexpressed (MYEOV), ankyrin repeat domain 1 (ANKRD1), TGFB2, angiotensin-converting enzyme 2 (ACE2) and semaphorin 7A (SEMA7A). MYEOV is involved in malignancy in several cancers including CRC, gastric, lung and breast cancer[183–186]. ANKRD1 is associated with poor prognosis and platinum resistance in ovarian cancer[187] but a recent study suggested a possible role as a tumor suppressor gene with a clearly dependence on the presence of *TP53*[188]. Indeed, ectopic expression of this gene reduced colony formation in prostate, lung and colon cancer cell lines[188]. Other interesting gene up-regulated in SW948 ST, which plays a double edge role in CRC, is TGFB2, that inhibits the proliferation of normal cells and early stage cancer cells while it promotes tumor and stroma growth in late stages of tumorigenesis[189]. Considering that SW948 cells are from grade III and stage III adenocarcinoma type, we interpret TGFB2 up-regulation as a tumor-promoting change. ACE2 is a key enzyme of the renin-angiotensin system with an anti-tumoral role in cancer in particular lung[190] and breast cancer[191]. SEMA7A plays an essential role in tumoral growth and metastasis of several types of cancer[192–194]

One of the genes down-regulated in SW948 ST cells was WNT4 gene. This gene is involved in the up-regulation of IL-8 in CRC, one of the significant molecular pathways modulated in this cell line. WNT4 (Wingless-Type MMTV Integration Site Family, Member 4) is a secreted molecule that binds to members of the frizzled family receptors, resulting in transcriptional regulation of target genes. It acts by a non-canonical WNT signaling, where pathways activated by WNT proteins do not lead to β-catenin stabilization or β-catenin-mediated gene transactivation[195]. The up-regulation of WNT4 is observed in gastric, breast and pulmonary cancer, being involved in the growth and development of a tumorigenic phenotype[196–198]. By other way, IL-8 is one of

the most significant up-regulated chemokines in CRC, enhancing the proliferation and survival of cancer cells through an autocrine activation and promoting angiogenesis and neutrophil infiltration into the tumor[199]. The association between IL-8 cytokine and WNT4 pathway is not well studied in CRC. In a model of chronic obstructive pulmonary disease, WNT4 protein activates the expression of IL-8 gene via the non-canonical pathway, leading to neutrophil infiltration and inflammation[198]. In our study, WNT4 is downregulated and its involved in the up-regulation of IL-8. Our hypothesis is that perhaps other members of WNT family are altered in these cells (although not in a prominent way), influencing together with WNT4 the IL-8 molecular pathway. A study in human squamous cell carcinoma reported the down-regulation of WNT4 and up-regulation of WNT5a as possible markers of the malignant phenotype is this type of disease[200], showing the mutual influence of two different Wnt proteins in cancer progression.

Current data showing the impact of sialyltransferase ST6GAL1 on gene expression are consistent with recent data showing the impact of sialyltransferases ST3GAL1[201] and ST6GALNAC1[202] on the transcriptome of bladder cancer cells.

In this project, we also studied two key phenotypic features of malignancy in CRC and evaluated if they were affected or not by the presence of ST6GAL1. The phenotypic changes in ST6GAL1-expressing cells can be explained by two partially overlapping mechanisms: a direct modulation of cell membrane receptors and an indirect mechanism, through modulation of gene expression which influences multiple functions such as apoptosis, cell migration and cell growth. In SW948 and SW48 cells, ST6GAL1 induced not overlapping and sometimes divergent phenotypic changes, which could be related to some gene expression changes. We observed that SW948 ST cells formed fewer colonies when grown in adhesion-independent conditions, indicating that only a minority of the population had the potential to grow in these conditions. This could suggest the presence of a restricted portion of cancer cells with a stem cell phenotype in SW948 ST cells comparing with NC cells. The increased ability to heal a scratch wound displayed by SW948 ST, but not by SW48 ST cells, may depend both on their reduced ability to multilayer growth and from increased migration.

Analysis of genes modulated by ST6GAL1 transduction indicated that some phenotypic aspects of transduced cells could be explained by the transcriptional changes induced by ST6GAL1. For example, ST6GAL1-modulated genes involved in cell adhesion and WNT signaling in SW948 cells and ST6GAL1-modulated genes involved

in cell adhesion and cytoskeleton remodeling in SW48 cells. In particular, increased migration of SW948 ST cells can be partially explained by *MYEOV* overexpression, as reported in another CRC cell line[183].

Current results showing mild phenotypic changes in ST6GAL1 transduced cells are consistent with TCGA data that show no association between ST6GAL1 expression and patients overall survival, response to treatment and CRC stage. Although transcriptomic and phenotypic analysis of ST6GAL1 expressing cells does not allow to clearly establish whether this modification has a general effect of promoting or inhibiting malignancy, there is no doubt that the effect is strongly cell-type specific.

4.3 Strong impact of B4GALNT2 in CRC patients, in particular its relation with a better prognosis and response to treatment

So far, the role of B4GALNT2 glycosyltransferase in cancer is not well understood. In CRC, B4GALNT2 is dramatically down-regulated comparing with normal colonic mucosa and is responsible for the increased level of sLe^x selectin-ligand, the well-known tumor-associated antigen involved in metastasis[65,107]. Although this association has been reported in CRC, the direct role of B4GALNT2 in CRC progression is not clear. To clarify the role of this enzyme in clinics, we used the TCGA data to correlate the B4GALNT2 mRNA expression in CRC with patients' features. As reported before in terms of protein level[65,107], B4GALNT2 mRNA was markedly down-regulated in CRC tissues comparing with normal tissues, presenting in general residual levels in CRC patients. The same tendency was observed in our small cohort of analyzed tissues samples: B4GALNT2 mRNA and enzymatic activity level was reduced in CRC tissues. For polyp tissues, this was not observed however our limitation in the dimension of polyps cohort could explain this point.

High B4GALNT2 was associated with a better prognosis and a better response to therapy, suggesting an anti-tumoral role of this enzyme in CRC. TCGA data did not show any association between B4GALNT2 mRNA expression and CRC stage or microsatellite instability *status*. However, high B4GALNT2 level was associated with a mucinous adenocarcinoma phenotype, mutations in *KRAS* and wild-type *TP53*. Although occasionally associated with a poor prognosis mainly at advanced stages[203,204], mucinous adenocarcinoma phenotype was recently considered an independent factor of CRC patients survival[146,205]. Regarding the mutational *status*

analysis, it is important to refer that mutations in both *KRAS* and *TP53* are not common in CRC[206]. Several studies showed that *KRAS* mutation is essential for CRC progression and metastasis[207]. By other way, *TP53* is a tumor suppressor gene encoding a DNA-binding transcription factor that regulates multiple cellular functions involved in DNA repair, metabolism, apoptosis, cell cycle and senescence[208]. In high B4GALNT2 expressers, the tumor suppressor role of *TP53* seems to overlap the oncogenic potential of *KRAS*, contributing for the better overall survival and better response to therapy observed in those patients and supporting the new anti-tumoral role of B4GALNT2 in this type of cancer.

The level of B4GALNT2 allows the classification of CRC patients in two distinct groups displaying different gene expression profiles. Only one gene appeared down-regulated in high B4GALNT2 expressers, insulin like growth factor 2 (IGF2). This gene is a member of the IGF/insulin signaling pathway involved in cell development and growth. It possesses a growth-promoting activity and its overexpression has been associated with resistance to chemotherapy and worse prognosis in several malignancies[209]. Several genes involved in mucin production (MUC2, MUC5B, AGR2, CLCA1 and FCGBP) were up-regulated, being in line with the previous association obtained in TCGA database between mucinous adenocarcinoma phenotype and high B4GALNT2 mRNA level. As mentioned before, mucin 2 plays a role in CRC progression. Its silencing promotes CRC metastasis through interleukin-6 signaling[160]. In this group of up-regulated genes, the majority of them have been described with tumor-suppressor functions including lipocalin 2[163], CLCA1[165] and galectin-4[210] (LGALS4), although some of them presented a pro-tumoral function in CRC (serpina-1[211] and human anterior gradient-2[169]). Trefoil factor 3 (TFF3), polymeric immunoglobulin receptor (PIGR) and Fc fragment of the IgG binding protein (FCGBP) present a dual role in cancer, being associated with a protective role against carcinogenesis[212–215] and contrary, with metastasis[216–218].

Altogether, TCGA data analysis suggests that B4GALNT2 glycosyltransferase may play an important anti-tumoral role in CRC.

4.4 The anti-tumoral effect of B4GALNT2 overexpression on the colorectal cancer transcriptome

We studied the causal effect of B4GALNT2 on colorectal cancer transcriptome modulation in LS 174T cells transfected with the cDNA of the short form of B4GALNT2 (originating subsequently two clones S2 and S11) and respective negative control (Neo population). This cell line was chosen because it is an excellent model to also study the expression of sLe^x tumor-associated antigen in CRC. 141 genes were significantly modulated by B4GALNT2 in these cells.

The most down-regulated genes in B4GALNT2-expressing clones with known functions in cancer included nidogen-1 (NID1)[219,220], neuronal pentraxin I (NPTX1)[221], galectin-2 (LGALS2)[222], paternally expressed gene 10 (PEG10)[223], retinoic acid induced 14 (RAI14)[224] and receptor-tyrosine-kinase-like orphan receptor 1 (ROR1)[225]; all of them are described in literature as cancer promoting genes.

One of the most down-regulated genes in B4GALNT2-expressing clones was also the cancer stem cell (CSC) related gene, transcription factor SOX2. SOX2 belongs to the SOX (Sry related HMG Box) family of proteins and is crucial for the maintenance of self-renewal and pluripotency in embryonic stem cells as well as adult tissue progenitor cells. SOX2 overexpression is associated with tumorigenicity and CSC phenotype[226–228]. This gene is involved in pathways related with stemness, pathways that were altered by B4GALNT2-modulated genes as displayed in GeneGO enrichment analysis. Surprisingly, other gene related with cancer stemness (CD200)[229,230] was found modulated in S2/S11 clones but in an opposite direction: CD200 was the most up-regulated gene in B4GALNT2 transfectant clones. However, lately this gene has been described as a suppressor of tumor formation and metastasis[231,232]. Taking in consideration the TCGA data analysis and preliminary data regarding the formation of spheroids in Neo population comparing with B4GALNT2-transfectant clones, our hypothesis is that in this particular context, CD200 cell surface glycoprotein is not involved in the formation of a cancer stem cell phenotype, playing possibly an anti-tumoral role.

The second gene most up-regulated was the nerve growth factor receptor-associated protein 1 (NGFRAP1, most known as BEX3). BEX3, the first identified BEX (Brain-Expressed X-linked) family gene, is involved in the control of mitogenic

signaling, acting mainly as an inducer of apoptosis thus in a tumor suppressor manner[233].

From the most relevant pathway maps, several modulated genes were related with immune response. In particular up-regulated myocyte enhancer factor 2C (MEF2C), up-regulated PD-L1 (also known as CD274) and down-regulated 15-lipoxygenase 2 (ALOX15B) seemed to have an important role in cancer. In colon and breast cancer, MEF2C was found up-regulated during disease progression meanwhile in hepatocarcinoma displayed both oncogenic and tumor suppressive properties[234]. On one hand, it facilitates migration, invasion and VEGF-dependent angiogenesis, while on the other hand it negatively affects cell proliferation, by blocking β-catenin nuclear translocation and WNT signaling[235].

PD-L1, one of the ligands for programmed cell death 1 (PD-1), is a key immune-regulatory molecule that, upon interaction with its receptor, suppresses the CD8 cytotoxic immune response[236]. Recently, PD-L1/PD-1 axis has come under close scrutiny as potential targets for cancer therapy, being involved in the current Nobel Prize in Physiology or Medicine. The role of PD-L1 expression in CRC is not clear, with conflicting results as to whether PD-L1 expression points out to a better or worse prognosis[237,238]. A recent study associated PD-L1 expression with BRAF mutations, microsatellite instability, poor differentiation and frequent tumor-infiltrating lymphocytes[239]. In our TCGA study, B4GALNT2 expression was not associated with any microsatellite *status* not even with BRAF mutations, so remains unclear the increase of PD-L1 in B4GALNT2-expressing clones and its main role in colon carcinogenesis.

Several studies have shown the downregulation of ALOX15B in epithelial tumors, suggesting that ALOX15B has an anti-proliferative role in cancer[240]. This gene belongs to the lipoxygenases family which catalyzes the oxygenation of polyunsaturated fatty acids. Subsequently, the oxygenated lipids can activate cellular signaling mechanisms through specific cell surface receptors or can be metabolized into lipid mediators[241].

We also evaluated the capability of closing the wound (scratch wound assay) between B4GALNT2 transfectant clones and Neo population, to evaluate the effect of this glycosyltransferase in the migration process *in vitro*. The time required to heal a scratch wound was slightly shorter for Neo population comparing with S2 and S11 clones. Although not reaching statistical significance, we can consider that the absence

of B4GALNT2 influence marginally cell migration towards a more aggressive phenotype. Alterations related with myosin heavy chain 3 (MYH3 or MYHC-EMB), a gene related with cell migration[242], can explain the changes observed in wound healing assay since MYH3 was found down-regulated in B4GALNT2 transfectant clones.

Overall, these data together with TCGA database survey indicate a strong impact of B4GALNT2 overexpression in the transcriptional activity of cancer cells. In fact, B4GALNT2 overexpression seems to affect genes with crucial roles in malignant transformation, acting mainly as an anti-tumoral glycosyltransferase in colorectal cancer.

4.5 Increased ST6GAL1 activity in CRC and IBD compared with healthy individuals plasma

We found an increased level of ST6GAL1 activity in CRC and inflammatory bowel disease (IBD) plasma samples comparing with plasma from healthy donors. This observation supports the overall serum sialylation changes reported in several cancers[243]. IBD is an inflammatory condition of the intestinal mucosa, being one of the most important high-risk factors for CRC development[244]. α2,6-sialic acid content in CRC plasma samples appears to be useful as a potential diagnostic and/or prognostic tumor marker. However, ST6GAL1 activity was also elevated in IBD samples without differences between IBD and CRC plasma, limiting their possible use for cancer screening. Moreover, we did not observe differences between ulcerative colitis and Crohn's disease plasma (the two main subtypes of IBD) being not the appropriate plasma biomarker to help distinguishing between these two inflammatory conditions, sometimes difficult in a clinical setting[245].

4.6 Different Sd^a linked-glycoprotein pattern between healthy donors and CRC plasma samples

Since data from TCGA database showed a positive association between B4GALNT2 and patient overall survival, we decided to evaluate the presence of its Sd^a cognate antigen in plasma samples from CRC patients and healthy donors. It was possible to observe the presence of a reactive band around 130kDa in almost all the samples but with different levels of intensity among them. Furthermore and even more interesting, it was possible to detect the presence of another band below 55kDa that

appears to be more related with a healthy status. So far and to our knowledge, it is not described in the literature any paper showing the identification of protein carriers of Sd^a antigen in plasma samples. Protein band bellow 55kDa was sent to the Institute of Pharmacology and Structural Biology in Toulouse (France) for the glyco-proteomic analysis. After exclusions due to protein molecular weight and their expression by kidney or colon tissues, one main protein hint was found: β-arrestin 1. This protein belongs to a family of four cytosolic adaptor proteins, known to mediate the trafficking of a variety of cell-surface receptors including seven-transmembrane receptors (7TMRs). They can also recruit cytoplasmic proteins and modulate downstream signaling pathways, being involved in cancer invasion and metastasis[246–248]. β-arrestin 1 is a cytosolic protein that translocates to the plasma membrane or to the nucleus; it is not described as a secreted protein and consequently its detection in the plasma was not expected. However, this protein was previously identified in extracellular vesicles derived from various cancer cell lines[249], suggesting that β-arrestin 1 may be shed by the cells and released into biological fluids (in this particular case, plasma) via exosomes or other extracellular vesicles. Our hypothesis is that in healthy people and in high B4GALNT2 expressing CRC patients, β-arrestin 1 is decorated with Sd^a antigen which prevents its cancer-promoting activity. On the contrary in CRC patients, the lack of Sd^a antigen probably unmasks the cancer promoting activity of β-arrestin 1.

Future studies are needed to validate this candidate as the Sd^a carrier protein in plasma samples. This carrier could be a possible useful biomarker to help a better CRC diagnosis.

4.7 *N*-glycan differences between CRC, normal and polyp tissues

In this study, we characterized the *N*-glycomic profiles of CRC, polyp and normal tissues by HILIC-UHPLC-FLR-ESI-MS analysis.

Detailed *N*-glycan characterization and relative quantitation identified an extensive structural heterogeneity of *N*-glycans. We found increased levels of high mannose, paucimmanosidic structures and overall sialylation in CRC tissues compared with normal tissues, supporting previous data[250]. In polyp tissues, the same tendency was observed for high-mannose structures but not for paucimmanosidic structures: no differences were detected between polyp and normal tissues. We could not identify significant differences between adenoma (polyp samples) and adenocarcinoma (CRC

samples) profiles, on the contrary to what is already reported for rectal adenomas and carcinomas[251]. However, in this study, we had the limitation regarding the number of polyp samples collected from the hospital. No differences were observed in terms of overall fucosylation, differing from the previously observation that fucosylation levels are increased in tumoral tissue and serum from CRC patients[252].

The alterations found in the *N*-glycan pattern between the analyzed samples highlight the importance of *N*-glycome as a molecular signature in cancer. Unfortunately, in this type of analysis it was not possible to go in details in terms of linkage, conformation and other features of glycan structures. Therefore, the use of different analytical techniques like exoglycosidases sequencing coupled with powerful bioinformatics tools may provide valuable information, to allow a better understanding of the alterations associated with *N*-glycosylation in cancer, in particular CRC.

4.8 General conclusion

Malignant transformation is associated with abnormal glycosylation and one of the main mechanisms that leads to changes in glycosylation is the deregulation of glycosyltransferases expression.

ST6GAL1 and B4GALNT2 glycosyltransferases play an important role in CRC. In this work, we used an *in silico* approach, genetically modified cell lines and patient samples to study the influence of both glycosyltransferases in CRC transcriptome, phenotypic features of cancer cells and also relationship of those enzymes with clinical characteristics. Current results showed a moderate impact of ST6GAL1 on CRC progression, though it was not possible to define a clear role as anti- or pro-tumoral. Additionally, the impact of ST6GAL1 overexpression on the transcriptome and phenotypical features of colon cancer cells is strongly cell-type specific. On the contrary, our work indicated a strong impact of B4GALNT2 overexpression on CRC transcriptome and clinical patient's features, acting mainly as an anti-tumoral glycosyltransferase in CRC. A different Sd^a protein carrier pattern was observed between healthy donors and CRC plasma samples. β-arrestin 1 is a possible candidate as the Sd^a carrier protein in plasma samples. Future studies are needed to validate this candidate which can be a potential useful biomarker to help a better CRC diagnosis. Furthermore, the alterations found in the *N*-glycan pattern between normal, cancer and polyp tissues highlight the importance of *N*-glycome as a molecular signature in cancer.

In summary, this project contributed to a better understanding of the impact of B4GALNT2/ST6GAL1 glycosyltransferases (and their cognate antigens) in CRC, allowing identifying in the future new diagnosis/prognosis biomarkers and potential therapeutic targets for CRC.

Chapter V - References

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.* **0,** 1–31 (2018).
2. Howlader, N. *et al.* Cancer Statistics Review, 1975-2014 - SEER Statistics, National Cancer Institute. *SEER Cancer Statistics Review, 1975-2014* http://seer.cancer.gov/csr/1975_2014/ (2016).
3. Colussi, D., Brandi, G., Bazzoli, F. & Ricciardiello, L. Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *International journal of molecular sciences* **14,** 16365–16385 (2013).
4. Behrens, J. The role of the Wnt signalling pathway in colorectal tumorigenesis. *Biochem. Soc. Trans.* **33,** 672–5 (2005).
5. Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *Lancet* **383,** 1490–1502 (2014).
6. Malumbres, M. & Barbacid, M. RAS oncogenes: The first 30 years. *Nat. Rev. Cancer* **3,** 459–65 (2003).
7. Fearon, E. R. Molecular Genetics of Colorectal Cancer. *Annu. Rev. Pathol. Mech. Dis.* **6,** 479–507 (2011).
8. Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M. & Warusavitarne, J. Molecular pathways in colorectal cancer. *J. Gastroenterol. Hepatol.* **27,** 1423–1431 (2012).
9. Boland, C. R. & Goel, A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* **138,** 2073–2087 (2010).
10. Deng, G. *et al.* BRAF Mutation Is Frequently Present in Sporadic Colorectal Cancer with Methylated hMLH1, but Not in Hereditary Nonpolyposis Colorectal Cancer. *Clin. Cancer Res.* **10,** 191–195 (2004).
11. De La Chapelle, A. & Hampel, H. Clinical relevance of microsatellite instability in colorectal cancer. *J. Clin. Oncol.* **28,** 3380–3387 (2010).
12. Mojarad, E. N., Kuppen, P. J. K., Aghdaei, H. A. & Zali, M. R. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol. Hepatol. from Bed to Bench* **6,** 120–128 (2013).
13. Johnson, C. M. *et al.* Meta-analyses of colorectal cancer risk factors. *Cancer Causes Control* **24,** 1207–22 (2013).
14. Beaugerie, L. & Itzkowitz, S. H. Cancers Complicating Inflammatory Bowel Disease. *N. Engl. J. Med.* **372,** 1441–1452 (2015).
15. Andersen, N. N. & Jess, T. Has the risk of colorectal cancer in inflammatory bowel disease decreased? *World J. Gastroenterol.* **19,** 7561–7568 (2013).
16. Marley, A. R. & Nan, H. Epidemiology of colorectal cancer. *International Journal of Molecular Epidemiology and Genetics* **7,** 105–114 (2016).
17. De Rosa, M. *et al.* Genetics, diagnosis and management of colorectal cancer (Review). *Oncol. Rep.* **34,** 1087–1096 (2015).

18. Sobin, L., Gospodarowicz, M. & Wittekind, C. *TNM classification of malignant tumours, 7th edition. Wiley* (2009).
19. Hassan, C., Zullo, A., Risio, M., Rossini, F. P. & Morini, S. Histologic risk factors and clinical outcome in colorectal malignant polyp: A pooled-data analysis. *Dis. Colon Rectum* **48,** 1588–96 (2005).
20. Kang, H., O'Connell, J. B., Maggard, M. A., Sack, J. & Ko, C. Y. A 10-year outcomes evaluation of mucinous and signet-ring cell carcinoma of the colon and rectum. *Dis. Colon Rectum* **48,** 1161–8 (2005).
21. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126,** 855–867 (2006).
22. Freeze, H. H. Genetic defects in the human glycome. *Nat. Rev. Genet.* **7,** 537–551 (2006).
23. Ghazarian, H., Idoni, B. & Oppenheimer, S. B. A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **113,** 236–247 (2011).
24. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* **15,** 540–555 (2015).
25. Stowell, S. R., Ju, T. & Cummings, R. D. Protein Glycosylation in Cancer. *Annu. Rev. Pathol. Mech. Dis.* **10,** 473–510 (2015).
26. Taylor, M.E. & Drickamer, K. *Introduction to Glycobiology*. *Oxford University Press* (2006).
27. Varki, A. *et al. Essentials of Glycobiology*. *Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Print* (1999). doi:10.1016/S0962-8924(00)01855-9
28. Helenius, A. & Aebi, M. Roles of N-Linked Glycans in the Endoplasmic Reticulum. *Annu. Rev. Biochem.* **73,** 1019–49 (2004).
29. Vasconcelos-dos-Santos, A. *et al.* Biosynthetic Machinery Involved in Aberrant Glycosylation: Promising Targets for Developing of Drugs Against Cancer. *Front. Oncol.* **5,** 138 (2015).
30. Aebi, M. N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 2430–2437 (2013).
31. Sethi, M. K. & Fanayan, S. Mass spectrometry-based N-glycomics of colorectal cancer. *Int. J. Mol. Sci.* **16,** 29278–29304 (2015).
32. Bennett, E. P. *et al.* Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22,** 736–756 (2012).
33. Tian, E. & Ten Hagen, K. G. Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconj. J.* **26,** 325–334 (2009).
34. Dennis, J. W., Laferté, S., Waghorne, C., Breitman, M. L. & Kerbel, R. S. β1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science (80-.).* **236,** 582–5 (1987).
35. Yoshimura, M., Nishikawa, A., Ihara, Y., Taniguchi, S. & Taniguchi, N. Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 8754–8 (1995).
36. Takahashi, M., Kuroki, Y., Ohtsubo, K. & Taniguchi, N. Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydr. Res.* **344,** 1387–90 (2009).

37. Kudelka, M. R., Ju, T., Heimburg-Molinaro, J. & Cummings, R. D. Simple sugars to complex disease-mucin-type O-glycans in cancer. *Adv. Cancer Res.* **126,** 53–135 (2015).
38. Gill, D. J. *et al.* Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. *Proc. Natl. Acad. Sci.* **110,** E3152-61 (2013).
39. Gomes, J. *et al.* Expression of UDP-N-acetyl-D-galactosamine: Polypeptide N-acetylgalactosaminyltransferase-6 in gastric mucosa, intestinal metaplasia, and gastric carcinoma. *J. Histochem. Cytochem.* **57,** 79–86 (2009).
40. Munkley, J. The role of sialyl-Tn in cancer. *Int. J. Mol. Sci.* **17,** 275 (2016).
41. Ferreira, J. A. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol. Oncol.* **7,** 719–31 (2013).
42. Pinho, S. *et al.* Biological significance of cancer-associated sialyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. *Cancer Lett.* **249,** 157–70 (2007).
43. Ju, T. *et al.* Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res.* **68,** 1636–46 (2008).
44. Burdick, M. D., Harris, A., Reid, C. J., Iwamura, T. & Hollingsworth, M. A. Oligosaccharides expressed on MUC1 produced by pancreatic and colon tumor cell lines. *J. Biol. Chem.* **272,** 24198–202. (1997).
45. Singh, R. *et al.* Cell surface-expressed Thomsen-Friedenreich antigen in colon cancer is predominantly carried on high molecular weight splice variants of CD44. *Glycobiology* **11,** 587–92 (2001).
46. Storr, S. J. *et al.* The O-linked glycosylation of secretory/shed MUC1 from an advanced breast cancer patient's serum. *Glycobiology* **18,** 456–62 (2008).
47. Häuselmann, I. & Borsig, L. Altered Tumor-Cell Glycosylation Promotes Metastasis. *Front. Oncol.* **4,** (2014).
48. Miyoshi, E. *et al.* Fucosylation Is a Promising Target for Cancer Diagnosis and Therapy. *Biomolecules* **2,** 34–45 (2012).
49. Tu, C. F., Wu, M. Y., Lin, Y. C., Kannagi, R. & Yang, R. B. FUT8 promotes breast cancer cell invasiveness by remodeling TGF-β receptor core fucosylation. *Breast Cancer Res.* **19,** (2017).
50. Liu, Y.-C. *et al.* Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proc. Natl. Acad. Sci.* **108,** 11332–11337 (2011).
51. Shields, R. L. *et al.* Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* **277,** 26733–26740 (2002).
52. Lin, H. *et al.* Blocking core fucosylation of TGF-β1 receptors downregulates their functions and attenuates the epithelial-mesenchymal transition of renal tubular cells. *Am. J. Physiol. Renal Physiol.* **300,** F1017-25 (2011).
53. A. Falconer, R., .J. Errington, R., .D. Shnyder, S., .J. Smith, P. & .H. Patterson, L. Polysialyltransferase: A New Target in Metastatic Cancer. *Curr. Cancer Drug Targets* **12,** 925–939 (2012).
54. Tanaka, F. *et al.* Prognostic significance of polysialic acid expression in resected non-small cell

lung cancer. *Cancer Res.* **61,** 1666–1670 (2001).

55. Colley, K. J., Kitajima, K. & Sato, C. Polysialic acid: Biosynthesis, novel functions and applications. *Crit. Rev. Biochem. Mol. Biol.* **49,** 498–532 (2014).
56. Tozawa, K. *et al.* Positive correlation between sialyl Lewis X expression and pathologic findings in renal cell carcinoma. *Kidney Int.* **67,** 1391–1396 (2005).
57. Nakamori, S. *et al.* Increased expression of sialyl Lewisx antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study. *Cancer Res.* **53,** 3632–3637 (1993).
58. Mizuguchi, S. *et al.* Clinical value of serum cytokeratin 19 fragment and sialyl-Lewis x in non-small cell lung cancer. *Ann. Thorac. Surg.* **83,** 216–221 (2007).
59. Dall'Olio, F., Malagolini, N., Trinchera, M. & Chiricolo, M. Sialosignaling: Sialyltransferases as engines of self-fueling loops in cancer progression. *Biochim. Biophys. Acta - Gen. Subj.* **1840,** 2752–2764 (2014).
60. Ugorski, M. & Laskowska, A. Sialyl Lewisa: A tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells. *Acta Biochim. Pol.* **49,** 303–311 (2002).
61. Läubli, H. & Borsig, L. Selectins promote tumor metastasis. *Semin. Cancer Biol.* **20,** 169–177 (2010).
62. Kannagi, R., Izawa, M., Koike, T., Miyazaki, K. & Kimura, N. Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis. *Cancer Sci.* **95,** 377–84 (2004).
63. Gout, S., Tremblay, P. L. & Huot, J. Selectins and selectin ligands in extravasation of cancer cells and organ selectivity of metastasis. *Clin. Exp. Metastasis* **25,** 335–344 (2008).
64. Borsig, L. *et al.* Heparin and cancer revisited: Mechanistic connections involving platelets, P-selectin, carcinoma mucins, and tumor metastasis. *Proc. Natl. Acad. Sci.* **98,** 3352–3357 (2001).
65. Groux-Degroote, S. *et al.* B4GALNT2 gene expression controls the biosynthesis of Sda and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract. *Int. J. Biochem. Cell Biol.* **53,** 442–449 (2014).
66. Dall'Olio, F., Malagolini, N., Chiricolo, M., Trinchera, M. & Harduin-Lepers, A. The expanding roles of the Sda/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2. *Biochim. Biophys. Acta - Gen. Subj.* **1840,** 443–453 (2014).
67. Vajaria, B. N., Patel, K. R., Begum, R. & Patel, P. S. Sialylation: an Avenue to Target Cancer Cells. *Pathol. Oncol. Res.* **22,** 443–7 (2016).
68. Lu, J. & Gu, J. Significance of β-galactoside α2,6 sialyltranferase 1 in cancers. *Molecules* **20,** 7509–7527 (2015).
69. Kitazume-Kawaguchi, S., Dohmae, N., Takio, K., Tsuji, S. & Colley, K. J. The relationship between ST6Gal I Golgi retention and its cleavage-secretion. *Glycobiology* **9,** 1397–1406 (1999).
70. Weinstein, J., Lee, E. U., McEntee, K., Lai, P. H. & Paulson, J. C. Primary structure of β-galactoside α2,6-sialyltransferase. Conversion of membrane-bound enzyme to soluble forms by cleavage of the NH2-terminal signal anchor. *J. Biol. Chem.* **262,** 17735–17743 (1987).
71. Keppler, O. T. *et al.* Human Golgi beta-galactoside alpha-2,6-sialyltransferase generates a group of sialylated B lymphocyte differentiation antigens. *Eur. J. Immunol.* **22,** 2777–2781 (1992).

72. Aas-Eng, D. A., Asheim, H. C., Deggerdal, A., Smeland, E. & Funderud, S. Characterization of a promoter region supporting transcription of a novel human beta-galactoside alpha-2,6-sialyltransferase transcript in HepG2 cells. *Biochim. Biophys. Acta* **1261,** 166–169 (1995).
73. Dall'Olio, F. *et al.* Beta-Galactoside alpha 2,6 sialyltransferase in human colon cancer: Contribution of multiple transcripts to regulation of enzyme activity and reactivity with Sambucus nigra agglutinin. *Int. J. Cancer* **88,** 58–65 (2000).
74. Antony, P. *et al.* Epigenetic inactivation of ST6GAL1 in human bladder cancer. *BMC Cancer* **14,** 901 (2014).
75. Dall'Olio, F. *et al.* Increased CMP-NeuAc:Gal beta 1,4GlcNAc-R alpha 2,6 sialyltransferase activity in human colorectal cancer tissues. *Int. J. Cancer* **44,** 434–439 (1989).
76. Dall'Olio, F. & Chiricolo, M. Sialyltransferases in cancer. *Glycoconj. J.* **18,** 841–850 (2001).
77. Dall'Olio, F. & Trerè, D. Expression of alpha 2,6-sialylated sugar chains in normal and neoplastic colon tissues. Detection by digoxigenin-conjugated Sambucus nigra agglutinin. *Eur. J. Histochem.* **37,** 257–65 (1993).
78. Sata, T., Roth, J., Zuber, C., Stamm, B. & Heitz, P. U. Expression of alpha 2,6-linked sialic acid residues in neoplastic but not in normal human colonic mucosa. A lectin-gold cytochemical study with Sambucus nigra and Maackia amurensis lectins. *Am. J. Pathol.* **139,** 1435–48 (1991).
79. Dall'Olio, F., Chiricolo, M. & Lau, J. T. Differential expression of the hepatic transcript of beta-galactoside alpha2,6-sialyltransferase in human colon cancer cell lines. *Int J Cancer* **81,** 243–7 (1999).
80. Inagaki, Y. *et al.* Clinicopathological utility of sialoglycoconjugates in diagnosing and treating colorectal cancer. *World J. Gastroenterol.* **20,** 6123–6132 (2014).
81. Lise, M. *et al.* Clinical Correlations of α2,6-Sialyltransferase Expression in Colorectal Cancer Patients. *Hybridoma* **19,** 281–286 (2000).
82. Vierbuchen, M. J., Fruechtnicht, W., Brackrock, S., Krause, K. T. & Zienkiewicz, T. J. Quantitative lectinhistochemical and immunohistochemical studies on the occurrence of alpha(2,3)- and alpha(2,6)-linked sialic acid residues in colorectal carcinomas. Relation to clinicopathologic features. *Cancer* **76,** 727–735 (1995).
83. Gebert, J. *et al.* Colonic carcinogenesis along different genetic routes: Glycophenotyping of tumor cases separated by microsatellite instability/stability. *Histochem. Cell Biol.* **138,** 339–350 (2012).
84. Dalziel, M., Dall'Olio, F., Mungul, A., Piller, V. & Piller, F. Ras oncogene induces β-galactoside α2,6-sialyltransferase (ST6Gal I) via a RalGEF-mediated signal to its housekeeping promoter. *Eur. J. Biochem.* **271,** 3623–3634 (2004).
85. Seales, E. C., Jurado, G. A., Singhal, A. & Bellis, S. L. Ras oncogene directs expression of a differentially sialylated, functionally altered β1 integrin. *Oncogene* **22,** 7137–7145 (2003).
86. Schultz, M. J. *et al.* ST6Gal-I sialyltransferase confers cisplatin resistance in ovarian tumor cells. *J. Ovarian Res.* **6,** 25 (2013).
87. Park, J.-J. *et al.* Sialylation of epidermal growth factor receptor regulates receptor activity and chemosensitivity to gefitinib in colon cancer cells. *Biochem. Pharmacol.* **83,** 849–857 (2012).

88. Ma, H. *et al.* Reversal effect of ST6GAL 1 on multidrug resistance in human leukemia by regulating the PI3K/Akt pathway and the expression of P-gp and MRP1. *PLoS One* **9,** e85113 (2014).
89. Lee, M., Lee, H.-J., Bae, S. & Lee, Y.-S. Protein Sialylation by Sialyltransferase Involves Radiation Resistance. *Mol. Cancer Res.* **6,** 1316–1325 (2008).
90. Zhuo, Y. & Bellis, S. L. Emerging role of alpha2,6-sialic acid as a negative regulator of galectin binding and function. *J. Biol. Chem.* **25,** 5935–41 (2011).
91. 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.* **286,** 22982–90 (2011).
92. Liu, Z. *et al.* ST6Gal-I regulates macrophage apoptosis via α2-6 sialylation of the TNFR1 death receptor. *J. Biol. Chem.* **286,** 39654–62 (2011).
93. Britain, C. M., Dorsett, K. A. & Bellis, S. L. The glycosyltransferase ST6Gal-I protects tumor cells against serum growth factor withdrawal by enhancing survival signaling and proliferative potential. *J. Biol. Chem.* **292,** 4663–4673 (2017).
94. Swindall, A. F. *et al.* ST6Gal-I protein expression is upregulated in human epithelial tumors and correlates with stem cell markers in normal tissues and colon cancer cell lines. *Cancer Res.* **73,** 2368–2378 (2013).
95. Hedlund, M., Ng, E., Varki, A. & Varki, N. M. α2-6-linked sialic acids on N-glycans modulate carcinoma differentiation in vivo. *Cancer Res.* **68,** 388–394 (2008).
96. Minami, A. *et al.* Reduction of the ST6 β-Galactosamide α-2,6-sialyltransferase 1 (ST6GAL1)-catalyzed sialylation of nectin-like molecule 2/cell adhesion molecule 1 and enhancement of ErbB2/ErbB3 signaling by microRNA-199a. *J. Biol. Chem.* **288,** 11845–11853 (2013).
97. Croci, D. O. *et al.* Glycosylation-dependent lectin-receptor interactions preserve angiogenesis in anti-VEGF refractory tumors. *Cell* **156,** 744–758 (2014).
98. Qian, J. *et al.* α2,6-hyposialylation of c-Met abolishes cell motility of ST6Gal-I-knockdown HCT116 cells. *Acta Pharmacol. Sin.* **30,** 1039–1045 (2009).
99. Chiricolo, M., Malagolini, N., Bonfiglioli, S. & Dall'Olio, F. Phenotypic changes induced by expression of beta-galactoside alpha2,6 sialyltransferase I in the human colon cancer cell line SW948. *Glycobiology* **16,** 146–154 (2006).
100. Zhu, Y. *et al.* Suppression of a sialyltransferase by antisense DNA reduces invasiveness of human colon cancer cells in vitro. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1536,** 148–60 (2001).
101. Yamamoto, H., Oviedo, A., Sweeley, C., Saito, T. & Moskal, J. R. Alpha 2, 6-sialylation of cell-surface N-glycans inhibits glioma formation in vivo. *Cancer Res.* **61,** 6822–9 (2001).
102. Renton, P. H., Howell, P., Ikin, E. W., Giles, C. M. & Goldsmith, K. L. G. Anti Sda, a New Blood Group Antibody. *Vox Sang.* **13,** 493–501 (1967).
103. Tollefsen, S. E. & Kornfeld, R. The B4 lectin from Vicia villosa seeds interacts with N-acetylgalactosamine residues alpha-linked to serine or threonine residues in cell surface glycoproteins. *J. Biol. Chem.* **123,** 1099–1106 (1983).
104. Blanchard, D., Piller, F., Gillard, B., Marcus, D. & Cartron, J. P. Identification of a novel

ganglioside on erythrocytes with blood group Cad specificty. *J. Biol. Chem.* **260,** 7813–6 (1985).

105. Serafini-Cessi, F. & Dall'Olio, F. Guinea-pig kidney beta-N-acetylgalactosaminyltransferase towards Tamm-Horsfall glycoprotein. Requirement of sialic acid in the acceptor for transferase activity. *Biochem. J.* **215,** 483–489 (1983).
106. Lo Presti, L., Cabuy, E., Chiricolo, M. & Dall'Olio, F. Molecular Cloning of the Human β1,4 N-Acetylgalactosaminyltransferase Responsible for the Biosynthesis of the Sda Histo-Blood Group Antigen: The Sequence Predicts a Very Long Cytoplasmic Domain. *J. Biochem.* **134,** 675–682 (2003).
107. Malagolini, N., Santini, D., Chiricolo, M. & Dall'Olio, F. Biosynthesis and expression of the Sda and sialyl Lewis x antigens in normal and cancer colon. *Glycobiology* **17,** 688–697 (2007).
108. Groux-Degroote, S. *et al.* The extended cytoplasmic tail of the human B4GALNT2 is critical for its Golgi targeting and post-Golgi sorting. *FEBS J.* **285,** 3442–3463 (2018).
109. Wang, H.-R., Hsieh, C.-Y., Twu, Y.-C. & Yu, L.-C. Expression of the human Sd(a) beta-1,4-N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. *Glycobiology* **18,** 104–13 (2008).
110. Kawamura, Y. I. *et al.* DNA Hypermethylation Contributes to Incomplete Synthesis of Carbohydrate Determinants in Gastrointestinal Cancer. *Gastroenterology* **135,** 142–151 (2008).
111. Malagolini, N. *et al.* Expression of UDP-GalNAc:NeuAc alpha 2,3Gal beta-R beta 1,4(GalNAc to Gal) N-acetylgalactosaminyltransferase involved in the synthesis of Sda antigen in human large intestine and colorectal carcinomas. *Cancer Res.* **49,** 6466–6470 (1989).
112. Robbe-Masselot, C. *et al.* Expression of a core 3 disialyl-Lex hexasaccharide in human colorectal cancers: A potential marker of malignant transformation in colon. *J. Proteome Res.* **8,** 702–711 (2009).
113. Kawamura, Y. I. *et al.* Introduction of Sda carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. *Cancer Res.* **65,** 6220–7 (2005).
114. Capon, C., Maes, E., Michalski, J. C., Leffler, H. & Kim, Y. S. Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *Biochem. J.* **358,** 657–664 (2001).
115. 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,** 130–9 (2011).
116. Mohlke, K. L. *et al.* Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. *Cell* **96,** 111–120 (1999).
117. Li, P. T., Liao, C. J., Yu, L. C., Wu, W. G. & Chu, S. T. Localization of B4GALNT2 and its role in mouse embryo attachment. *Fertil. Steril.* **97,** 1206–1212 (2012).
118. Jayasinha, V., Hoyte, K., Xia, B. & Martin, P. T. Overexpression of the CT GalNAc transferase inhibits muscular dystrophy in a cleavage-resistant dystroglycan mutant mouse. *Biochem. Biophys. Res. Commun.* **302,** 831–836 (2003).
119. Xu, R., Devries, S., Camboni, M. & Martin, P. T. Overexpression of Galgt2 reduces dystrophic pathology in the skeletal muscles of alpha sarcoglycan-deficient mice. *Am. J. Pathol.* **175,** 235–

247 (2009).

120. Thomas, P. J., Xu, R. & Martin, P. T. B4GALNT2 (GALGT2) Gene Therapy Reduces Skeletal Muscle Pathology in the FKRP P448L Mouse Model of Limb Girdle Muscular Dystrophy 2I. *Am. J. Pathol.* **186,** 2429–2448 (2016).
121. Heaton, B. E. *et al.* A CRISPR Activation Screen Identifies a Pan-avian Influenza Virus Inhibitory Host Factor. *Cell Rep.* **20,** 1503–1512 (2017).
122. Joyce, A. R. & Palsson, B. O. The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell Biol.* **7,** 198–2010 (2006).
123. Pagani, I. *et al.* The Genomes OnLine Database (GOLD) v.4: Status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* **40,** D571-9 (2012).
124. Manzoni, C. *et al.* Genome , transcriptome and proteome : the rise of omics data and their integration in biomedical sciences. *Brief. Bioinform.* **19,** 286–302 (2018).
125. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10,** 57–63 (2009).
126. Patterson, S. D. & Aebersold, R. H. Proteomics: The first decade and beyond. *Nat. Genet.* **33,** 311–323 (2003).
127. Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143,** 672–676 (2010).
128. Kam, R. K. T. & Poon, T. C. W. The potentials of glycomics in biomarker discovery. *Clin. Proteomics* **4,** 67–79 (2008).
129. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Wspolczesna Onkol.* **1A,** A68–A77 (2015).
130. Malagolini, N., Chiricolo, M., Marini, M. & Dall'Olio, F. Exposure of α2,6-sialylated lactosaminic chains marks apoptotic and necrotic death in different cell types. *Glycobiology* **19,** 172–181 (2009).
131. Chomczyński, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinum thiocyanate-phenol-chlotoform extraction. *Anal. Biochem.* **162,** 156–159 (1987).
132. LOWRY, O. H., ROSEBROUGH, N. J., FARR, A. L. & RANDALL, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193,** 265–275 (1951).
133. Dall'Olio, F., Chiricolo, M., Mariani, E. & Facchini, A. Biosynthesis of the cancer-related sialyl-alpha 2,6-lactosaminyl epitope in colon cancer cell lines expressing beta-galactoside alpha 2,6-sialyltransferase under a constitutive promoter. *Eur. J. Biochem.* **268,** 5876–5884 (2001).
134. Catera, M. *et al.* Identification of novel plasma glycosylation-associated markers of aging. *Oncotarget* **7,** 7455–7468 (2016).
135. Malagolini, N. *et al.* Expression of UDP-GalNAc:NeuAc alpha 2,3Gal beta-R beta 1,4(GalNAc to Gal) N-acetylgalactosaminyltransferase involved in the synthesis of Sda antigen in human large intestine and colorectal carcinomas. *Cancer Res.* **49,** 6466–6470 (1989).
136. 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).
137. Dall'Olio, F., Malagolini, N. & Serafini-Cessi, F. Enhanced CMP-NeuAc:Gal beta 1,4GlcNAc-R alpha 2,6 sialyltransferase activity of human colon cancer xenografts in athymic nude mice and of

xenograft-derived cell lines. *Int. J. Cancer* **50,** 325–330 (1992).

138. Meng, Q., Ren, C., Wang, L., Zhao, Y. & Wang, S. Knockdown of ST6Gal-I inhibits the growth and invasion of osteosarcoma MG-63 cells. *Biomed. Pharmacother.* **72,** 172–178 (2015).
139. Wei, A. *et al.* ST6Gal-I overexpression facilitates prostate cancer progression via the PI3K/Akt/GSK-3beta/beta-catenin signaling pathway. *Oncotarget* **7,** :65374-65388 (2016).
140. Yuan, Q. *et al.* Modification of α2,6-sialylation mediates the invasiveness and tumorigenicity of non-small cell lung cancer cells in vitro and in vivo via Notch1/Hes1/MMPs pathway. *Int. J. Cancer* **143,** 2319–2330 (2018).
141. Lu, J. *et al.* β-Galactoside α2,6-sialyltranferase 1 promotes transforming growth factor-β-mediated epithelial-mesenchymal transition. *J. Biol. Chem.* **289,** 34627–34641 (2014).
142. Dawson, G., Moskal, J. R. & Dawson, S. A. Transfection of 2,6 and 2,3-sialyltransferase genes and GlcNAc-transferase genes into human glioma cell line U-373 MG affects glycoconjugate expression and enhances cell death. *J. Neurochem.* **89,** 1436–1444 (2004).
143. Yen, H.-Y. *et al.* Effect of sialylation on EGFR phosphorylation and resistance to tyrosine kinase inhibition. *Proc. Natl. Acad. Sci.* **112,** 6955–60 (2015).
144. Dall'Olio, F. *et al.* B-Galactoside alpha 2,6 sialyltransferase in human colon cancer: Contribution of multiple transcripts to regulation of enzyme activity and reactivity with Sambucus nigra agglutinin. *Int. J. Cancer* **88,** 58–65 (2000).
145. Lee, J. *et al.* Transforming Growth Factor Beta Receptor 2 (TGFBR2) Changes Sialylation in the Microsatellite Unstable (MSI) Colorectal Cancer Cell Line HCT116. *PLoS One* **8,** e57074 (2013).
146. Park, J. S. *et al.* Prognostic comparison between mucinous and nonmucinous adenocarcinoma in colorectal cancer. *Med. (United States)* **94,** e658 (2015).
147. Kloor, M., Staffa, L., Ahadova, A. & Von Knebel Doeberitz, M. Clinical significance of microsatellite instability in colorectal cancer. *Langenbeck's Arch. Surg.* **399,** 23–31 (2014).
148. Chakraborty, A. *et al.* ST6Gal-I sialyltransferase promotes chemoresistance in pancreatic ductal adenocarcinoma by abrogating gemcitabine-mediated DNA damage. *J. Biol. Chem.* **293,** 984–994 (2018).
149. Britain, C. M., Holdbrooks, A. T., Anderson, J. C., Willey, C. D. & Bellis, S. L. Sialylation of EGFR by the ST6Gal-I sialyltransferase promotes EGFR activation and resistance to gefitinib-mediated cell death. *J. Ovarian Res.* **11,** 12 (2018).
150. Chen, X. *et al.* ST6GAL1 modulates docetaxel sensitivity in human hepatocarcinoma cells via the p38/MAPK caspase pathway. *Oncotarget* **7,** 51955–51964 (2016).
151. Wichert, B., Milde-Langosch, K., Galatenko, V., Schmalfeldt, B. & Oliveira-Ferrer, L. Prognostic role of the sialyltransferase ST6GAL1 in ovarian cancer. *Glycobiology* **28,** 898–903 (2018).
152. Hsieh, C.-C. *et al.* Elevation of β-galactoside α2,6-sialyltransferase 1 in a fructoseresponsive manner promotes pancreatic cancer metastasis. *Oncotarget* **8,** 7691–7709 (2017).
153. 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.* **75,** 143–149 (1993).
154. Blumenthal, R. D., Hansen, H. J. & Goldenberg, D. M. Inhibition of adhesion, invasion, and

metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (carcinoembryonic antigen). *Cancer Res.* **65,** 8809–8817 (2005).

155. Jinesh, G. G., Manyam, G. C., Mmeje, C. O., Baggerly, K. A. & Kamat, A. M. Surface PD-L1, E-cadherin, CD24, and VEGFR2 as markers of epithelial cancer stem cells associated with rapid tumorigenesis. *Sci. Rep.* **7,** 9602 (2017).
156. Wang, Z., Han, G., Liu, Q., Zhang, W. & Wang, J. Silencing of PYGB suppresses growth and promotes the apoptosis of prostate cancer cells via the NF-κB/Nrf2 signaling pathway. *Mol. Med. Rep.* **18,** 3800–3808 (2018).
157. Knoops, B., Goemaere, J., Van der Eecken, V. & Declercq, J.-P. Peroxiredoxin 5: Structure, Mechanism, and Function of the Mammalian Atypical 2-Cys Peroxiredoxin. *Antioxid. Redox Signal.* **15,** 817–829 (2011).
158. Kim, B. *et al.* Peroxiredoxin 5 overexpression enhances tumorigenicity and correlates with poor prognosis in gastric cancer. *Int. J. Oncol.* **51,** 298–306 (2017).
159. Ahn, H. M. *et al.* Peroxiredoxin 5 promotes the epithelial-mesenchymal transition in colon cancer. *Biochem. Biophys. Res. Commun.* **487,** 580–586 (2017).
160. Hsu, H. P. *et al.* Mucin 2 silencing promotes colon cancer metastasis through interleukin-6 signaling. *Sci. Rep.* **7,** 5823 (2017).
161. Velcich, A. *et al.* Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science (80-.).* **295,** 1726–1729 (2002).
162. Wang, H. *et al.* Expression of survivin, MUC2 and MUC5 in colorectal cancer and their association with clinicopathological characteristics. *Oncol. Lett.* **14,** 1011–1016 (2017).
163. Toyonaga, T. *et al.* Lipocalin 2 prevents intestinal inflammation by enhancing phagocytic bacterial clearance in macrophages. *Sci. Rep.* **6,** 35014 (2016).
164. Saha, S. K. *et al.* KRT19 directly interacts with β-catenin/RAC1 complex to regulate NUMB-dependent NOTCH signaling pathway and breast cancer properties. *Oncogene* **36,** 332–349 (2017).
165. Li, X. *et al.* CLCA1 suppresses colorectal cancer aggressiveness via inhibition of the Wnt/beta-catenin signaling pathway. *Cell Commun. Signal.* **15,** 38 (2017).
166. Rocha, M. R. *et al.* Annexin A2 overexpression associates with colorectal cancer invasiveness and TGF-ß induced epithelial mesenchymal transition via Src/ANXA2/STAT3. *Sci. Rep.* **8,** 11285 (2018).
167. Rafa, L. *et al.* REG4 acts as a mitogenic, motility and pro-invasive factor for colon cancer cells. *Int. J. Oncol.* **36,** 689–698 (2010).
168. Astrosini, C. *et al.* REG1A expression is a prognostic marker in colorectal cancer and associated with peritoneal carcinomatosis. *Int. J. Cancer* **123,** 409–413 (2008).
169. Tian, S. *et al.* Secreted AGR2 promotes invasion of colorectal cancer cells via Wnt11-mediated non-canonical Wnt signaling. *Exp. Cell Res.* **364,** 198–207 (2018).
170. Duan, L. *et al.* S100A6 stimulates proliferation and migration of colorectal carcinoma cells through activation of the MAPK pathways. *Int. J. Oncol.* **44,** 781–90 (2014).
171. Hartman, J. *et al.* Tumor repressive functions of estrogen receptor beta in SW480 colon cancer

cells. *Cancer Res.* **69,** 6100–6 (2009).

172. Giroux, V., Lemay, F., Bernatchez, G., Robitaille, Y. & Carrier, J. C. Estrogen receptor β deficiency enhances small intestinal tumorigenesis in ApcMin/+mice. *Int. J. Cancer* **23,** 303–11 (2008).
173. Weyant, M. J. *et al.* Reciprocal expression of ERα and ERβ is associated with estrogen-mediated modulation of intestinal tumorigenesis. *Cancer Res.* **61,** 2547–51 (2001).
174. Barzi, A., Lenz, A. M., Labonte, M. J. & Lenz, H. J. Molecular pathways: Estrogen pathway in colorectal cancer. *Clin. Cancer Res.* **19,** 5842–5848 (2013).
175. Dong, Y. *et al.* Downregulation of EphA1 in colorectal carcinomas correlates with invasion and metastasis. *Mod. Pathol.* **22,** 151–160 (2009).
176. Dunne, P. D. *et al.* EphA2 expression is a key driver of migration and invasion and a poor prognostic marker in colorectal cancer. *Clin. Cancer Res.* **22,** 230–242 (2016).
177. Wu, B. O., Jiang, W. E. N. G., Zhou, D. & Cui, Y. Knockdown of EPHA1 by CRISPR / CAS9 Promotes Adhesion and Motility of HRT18 Colorectal Carcinoma Cells. *Anticancer Res.* **36,** 1211–1220 (2016).
178. Gylfe, A. E. *et al.* Identification of candidate oncogenes in human colorectal cancers with microsatellite instability. *Gastroenterology* **145,** 540–3.e22 (2013).
179. Liu, J. *et al.* Screening key genes and miRNAs in early-stage colon adenocarcinoma by RNA-sequencing. *Tumor Biol.* **39,** 1010428317714899 (2017).
180. Kohda, T. *et al.* Tumour suppressor activity of human imprinted gene PEG3 in a glioma cell line. *Genes to Cells* **6,** 237–47 (2001).
181. Feng, W. *et al.* Imprinted tumor suppressor genes ARHI and PEG3 are the most frequently down-regulated in human ovarian cancers by loss of heterozygosity and promoter methylation. *Cancer* **112,** 1489–502 (2008).
182. Domanska, U. M. *et al.* A review on CXCR4/CXCL12 axis in oncology: No place to hide. *Eur. J. Cancer* **49,** 219–230 (2013).
183. Lawlor, G., Doran, P. P., MacMathuna, P. & Murray, D. W. MYEOV (myeloma overexpressed gene) drives colon cancer cell migration and is regulated by PGE2. *J. Exp. Clin. Cancer Res.* **29,** (2010).
184. Leyden, J. *et al.* Net1 and Myeov: Computationally identified mediators of gastric cancer. *Br. J. Cancer* **94,** 1204–1212 (2006).
185. Fang, L. *et al.* MYEOV functions as an amplified competing endogenous RNA in promoting metastasis by activating TGF-β pathway in NSCLC. *Oncogene* (2018). doi:10.1038/s41388-018-0484-9
186. Janssen, J. W. G. *et al.* MYEOV: A candidate gene for DNA amplification events occurring centromeric to CCND1 in breast cancer. *Int. J. Cancer* **102,** 608–14 (2002).
187. Lei, Y., Henderson, B. R., Emmanuel, C., Harnett, P. R. & Defazio, A. Inhibition of ANKRD1 sensitizes human ovarian cancer cells to endoplasmic reticulum stress-induced apoptosis. *Oncogene* **34,** 485–95 (2015).
188. Jiménez, A. P. *et al.* The tumor suppressor RASSF1A induces the YAP1 target gene ANKRD1

that is epigenetically inactivated in human cancers and inhibits tumor growth. *Oncotarget* **8,** 88437–88452 (2017).

189. Villalba, M., Evans, S. R., Vidal-Vanaclocha, F. & Calvo, A. Role of TGF-β in metastatic colon cancer: it is finally time for targeted therapy. *Cell Tissue Res.* **370,** 29–39 (2017).
190. Qian, Y. R. *et al.* Angiotensin-converting enzyme 2 attenuates the metastasis of non-small cell lung cancer through inhibition of epithelial-mesenchymal transition. *Oncol. Rep.* **29,** 2408–2414 (2013).
191. Yu, C. *et al.* Downregulation of ACE2/Ang-(1-7)/Mas axis promotes breast cancer metastasis by enhancing store-operated calcium entry. *Cancer Lett.* **376,** 268–77 (2016).
192. Garcia-Areas, R. *et al.* Suppression of tumor-derived Semaphorin 7A and genetic ablation of host-derived Semaphorin 7A impairs tumor progression in a murine model of advanced breast carcinoma. *Int. J. Oncol.* **51,** 1395–1404 (2017).
193. Saito, T. *et al.* Semaphorin7A promotion of tumoral growth and metastasis in human oral cancer by regulation of g1 cell cycle and matrix metalloproteases: Possible contribution to tumoral angiogenesis. *PLoS One* **10,** e0137923 (2015).
194. Ma, B. *et al.* Role of chitinase 3-like-1 and semaphorin 7a in pulmonary melanoma metastasis. *Cancer Res.* **75,** 487–96 (2015).
195. Veeman, M. T., Axelrod, J. D. & Moon, R. T. A second canon: Functions and mechanisms of β-catenin-independent Wnt signaling. *Dev. Cell* **5,** 367–77 (2003).
196. Zhu, Y. *et al.* Anti-cancer drug 3,3'-diindolylmethane activates Wnt4 signaling to enhance gastric cancer cell stemness and tumorigenesis. *Oncotarget* **7,** 16311–16324 (2016).
197. Vouyovitch, C. M. *et al.* WNT4 mediates the autocrine effects of growth hormone in mammary carcinoma cells. *Endocr. Relat. Cancer* **23,** 571–585 (2016).
198. Durham, A. L. *et al.* Regulation of Wnt4 in chronic obstructive pulmonary disease. *FASEB J.* **27,** 2367–2381 (2013).
199. Ning, Y. & Lenz, H.-J. Targeting IL-8 in colorectal cancer. *Expert Opin. Ther. Targets* **16,** 491–497 (2012).
200. Taki, M. *et al.* Down-regulation of Wnt-4 and up-regulation of Wnt-5a expression by epithelial-mesenchymal transition in human squamous carcinoma cells. *Cancer Sci.* **94,** 593–597 (2003).
201. Severino, P. F. *et al.* Oxidative damage and response to Bacillus Calmette-Guérin in bladder cancer cells expressing sialyltransferase ST3GAL1. *BMC Cancer* **18,** 198 (2018).
202. Severino, P. F. *et al.* Expression of sialyl-Tn sugar antigen in bladder cancer cells affects response to Bacillus Calmette-Guérin (BCG) and to oxidative damage. *Oncotarget* **8,** 54506–54517 (2017).
203. Mekenkamp, L. J. M. *et al.* Mucinous adenocarcinomas: Poor prognosis in metastatic colorectal cancer. *Eur. J. Cancer* **48,** 501–9 (2012).
204. Catalano, V. *et al.* Mucinous histology predicts for poor response rate and overall survival of patients with colorectal cancer and treated with first-line oxaliplatin- and/or irinotecan-based chemotherapy. *Br. J. Cancer* **100,** pages 881–887 (2009).
205. Warschkow, R. *et al.* Predictive value of mucinous histology in colon cancer: A population-based, propensity score matched analysis. *Br. J. Cancer* **114,** 1027–1032 (2016).

206. Smith, G. *et al.* Mutations in APC, Kirsten-ras, and p53--alternative genetic pathways to colorectal cancer. *Proc. Natl. Acad. Sci.* **99,** 9433–9438 (2002).
207. Haigis, K. M. KRAS Alleles: The Devil Is in the Detail. *Trends in Cancer* **3,** 686–697 (2017).
208. Aubrey, B. J., Strasser, A. & Kelly, G. L. Tumor-suppressor functions of the TP53 pathway. *Cold Spring Harb. Perspect. Med.* **6,** a026062 (2016).
209. Brouwer-Visser, J. & Huang, G. S. IGF2 signaling and regulation in cancer. *Cytokine Growth Factor Rev.* **26,** 371–377 (2015).
210. Satelli, A., Rao, P. S., Thirumala, S. & Rao, U. S. Galectin-4 functions as a tumor suppressor of human colorectal cancer. *Int. J. Cancer* **129,** 799–809 (2011).
211. Kwon, C. H., Park, H. J., Choi, J. H., Lee, J. R. & Kim, H. K. Snail and serpinA1 promote tumor progression and predict prognosis in colorectal cancer. *Oncotarget* **6,** 20312–26 (2014).
212. John, R., El-Rouby, N. M., Tomasetto, C., Rio, M. C. & Karam, S. M. Expression of TFF3 during multistep colon carcinogenesis. *Histol. Histopathol.* **22,** 743–51 (2007).
213. Teng, X., Xu, L. F., Zhou, P., Sun, H. W. & Sun, M. Effects of trefoil peptide 3 on expression of TNF-α, TLR4, and NF-κB in trinitrobenzene sulphonic acid induced colitis mice. *Inflammation* **32,** 120–9 (2009).
214. Fristedt, R. *et al.* Expression and prognostic significance of the polymeric immunoglobulin receptor in esophageal and gastric adenocarcinoma. *J. Transl. Med.* **12,** (2014).
215. Lee, S., Bang, S., Song, K. & Lee, I. Differential expression in normal-adenoma-carcinoma sequence suggests complex molecular carcinogenesis in colon. *Oncol. Rep.* **16,** 747–754 (2006).
216. Babyatsky, M. *et al.* Trefoil factor-3 expression in human colon cancer liver metastasis. *Clin. Exp. Metastasis* **26,** 143–51 (2009).
217. Liu, F. *et al.* COLORECTAL Polymeric Immunoglobulin Receptor Expression is Correlated with Hepatic Metastasis and Poor Prognosis in Colon Carcinoma Patients with Hepatic Metastasis. *Hepatogastroenterology.* **61,** 652–9 (2014).
218. Qi, C., Hong, L., Cheng, Z. & Yin, Q. Identification of metastasis-associated genes in colorectal cancer using metaDE and survival analysis. *Oncol. Lett.* **11,** 568–574 (2016).
219. Zhou, Y. *et al.* NID1, a new regulator of EMT required for metastasis and chemoresistance of ovarian cancer cells. *Oncotarget* **8,** 33110–33121 (2017).
220. Alečković, M. *et al.* Identification of Nidogen 1 as a lung metastasis protein through secretome analysis. *Genes Dev.* **31,** 1439–1455 (2017).
221. Mori, Y. *et al.* Novel candidate colorectal cancer biomarkers identified by methylation microarray-based scanning. *Endocr. Relat. Cancer* **18,** 465–78 (2011).
222. Chen, C. *et al.* Circulating galectins-2,-4 and-8 in cancer patients make important contributions to the increased circulation of several cytokines and chemokines that promote angiogenesis and metastasis. *Br. J. Cancer* **110,** 741–752 (2014).
223. Xie, T. *et al.* PEG10 as an oncogene: Expression regulatory mechanisms and role in tumor progression. *Cancer Cell Int.* **18,** 1–10 (2018).
224. Yuan, C. *et al.* Super enhancer associated RAI14 is a new potential biomarker in lung adenocarcinoma. *Oncotarget* **8,** 105251–105261 (2017).

225. Zhou, J.-K. *et al.* ROR1 expression as a biomarker for predicting prognosis in patients with colorectal cancer. *Oncotarget* **8,** 32864–32872 (2017).
226. Maurizi, G., Verma, N., Gadi, A., Mansukhani, A. & Basilico, C. Sox2 is required for tumor development and cancer cell proliferation in osteosarcoma. *Oncogene* **37,** 4626–4632 (2018).
227. Zheng, J. *et al.* Sox2 modulates motility and enhances progression of colorectal cancer via the Rho-ROCK signaling pathway. *Oncotarget* **8,** 98635–98645 (2017).
228. Lundberg, I. V. *et al.* SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. *BMC Cancer* **16,** (2016).
229. Zhang, S. S., Huang, Z. W., Li, L. X., Fu, J. J. & Xiao, B. Identification of CD200+colorectal cancer stem cells and their gene expression profile. *Oncol. Rep.* **36,** 2252–60 (2016).
230. Kawasaki, B. T., Mistree, T., Hurt, E. M., Kalathur, M. & Farrar, W. L. Co-expression of the toleragenic glycoprotein, CD200, with markers for cancer stem cells. *Biochem. Biophys. Res. Commun.* **364,** 778–82 (2007).
231. Erin, N. *et al.* Bidirectional effect of CD200 on breast cancer development and metastasis, with ultimate outcome determined by tumor aggressiveness and a cancer-induced inflammatory response. *Oncogene* **34,** 3860–3870 (2015).
232. Talebian, F. & Bai, X. F. The role of tumor expression of CD200 in tumor formation, metastasis and susceptibility to T lymphocyte adoptive transfer therapy. *Oncoimmunology* **1,** 971–973 (2012).
233. Kazi, J. U., Kabir, N. N. & Rönnstrand, L. Brain-Expressed X-linked (BEX) proteins in human cancers. *Biochim. Biophys. Acta - Rev. Cancer* **1856,** 226–233 (2015).
234. Di Giorgio, E., Hancock, W. W. & Brancolini, C. MEF2 and the tumorigenic process, hic sunt leones. *Biochim. Biophys. Acta - Rev. Cancer* In press (2018). doi:10.1016/j.bbcan.2018.05.007
235. Bai, X. L. *et al.* Myocyte enhancer factor 2C regulation of hepatocellular carcinoma via vascular endothelial growth factor and Wnt/β-catenin signaling. *Oncogene* **34,** 4089–4097 (2015).
236. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12,** 252–264 (2012).
237. Droeser, R. A. *et al.* Clinical impact of programmed cell death ligand 1 expression in colorectal cancer. *Eur. J. Cancer* **49,** 2233–2242 (2013).
238. Song, M. *et al.* PTEN Loss Increases PD-L1 Protein Expression and Affects the Correlation between PD-L1 Expression and Clinical Parameters in Colorectal Cancer. *PLoS One* **8,** e65821 (2013).
239. Rosenbaum, M. W., Bledsoe, J. R., Morales-Oyarvide, V., Huynh, T. G. & Mino-Kenudson, M. PD-L1 expression in colorectal cancer is associated with microsatellite instability, BRAF mutation, medullary morphology and cytotoxic tumor-infiltrating lymphocytes. *Mod. Pathol.* **29,** 1104–1112 (2016).
240. Mashima, R. & Okuyama, T. The role of lipoxygenases in pathophysiology; new insights and future perspectives. *Redox Biol.* **6,** 297–310 open access (2015).
241. Tian, R. *et al.* ALOX15 as a suppressor of inflammation and cancer: Lost in the link. *Prostaglandins Other Lipid Mediat.* **132,** 77–83 (2017).

242. Li, Y.-R. & Yang, W.-X. Myosins as fundamental components during tumorigenesis: diverse and indispensable. *Oncotarget* **7,** 46785–46812 (2016).
243. Zhang, Z., Wuhrer, M. & Holst, S. Serum sialylation changes in cancer. *Glycoconj. J.* **35,** 139–160 (2018).
244. Stidham, R. W. & Higgins, P. D. R. Colorectal Cancer in Inflammatory Bowel Disease. *Clin. Colon Rectal Surg.* **31,** 168–178 (2018).
245. Feakins, R. M. Ulcerative colitis or crohn's disease? Pitfalls and problems. *Histopathology* **64,** 317–35 (2014).
246. Song, Q., Ji, Q. & Li, Q. The role and mechanism of β-arrestins in cancer invasion and metastasis (Review). *Int. J. Mol. Med.* **41,** 631–639 (2018).
247. Shenoy, S. K. & Lefkowitz, R. J. β-arrestin-mediated receptor trafficking and signal transduction. *Trends Pharmacol. Sci.* **32,** 521–533 (2011).
248. Buchanan, F. G. *et al.* Role of beta-arrestin 1 in the metastatic progression of colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 1492–1497 (2006).
249. Hurwitz, S. N. *et al.* Proteomic profiling of NCI-60 extracellular vesicles uncovers common protein cargo and cancer type-specific biomarkers. *Oncotarget* **7,** 86999–87015 (2016).
250. Sethi, M. K. *et al.* In-depth N-glycome profiling of paired colorectal cancer and non-tumorigenic tissues reveals cancer-, stage- and EGFR-specific protein N-glycosylation. *Glycobiology* **25,** 1064–1078 (2015).
251. Kaprio, T. *et al.* N-glycomic Profiling as a Tool to Separate Rectal Adenomas from Carcinomas. *Mol. Cell. Proteomics* **14,** 277–288 (2015).
252. Fernández-Rodríguez, J., De La Cadena, M. P., Martínez-Zorzano, V. S. & Rodríguez-Berrocal, F. J. Fucose levels in sera and in tumours of colorectal adenocarcinoma patients. *Cancer Lett.* **121,** 147–53 (1997).

Chapter VI - Appendix

Table VI. 1 - Clinical information of patients included in current study and their tumor characteristics. Data was retrieved from The Cancer Genome Atlas database.

Clinical features	N=623	Percentage (%)
Age at initial diagnosis (years)		
Median age	68 years	
Range	31-90 years	
Gender		
Male	331	53.1%
Female	290	46.5%
Non available data	2	0.4%
Histological subtype		
Adenocarcinoma	535	85.9%
Mucinous adenocarcinoma	75	12%
Non available data	13	2.1%
Microsatellite status		
MSI-H	87	14%
MSI-L	101	16.2%
MSS	430	69%
Non available or indeterminate	5	0.8%
Stage		
Stage I	105	16.9%
Stage II	229	36.8%
Stage III	179	28.7%
Stage IV	88	14.1%
Non available data	22	3.5%
Follow-up treatment success		
Complete remission/response	247	39.6%
Partial remission/response	16	2.6%
Stable disease	7	1.1%
Progressive disease	53	8.5%
Non available data	300	48.2%

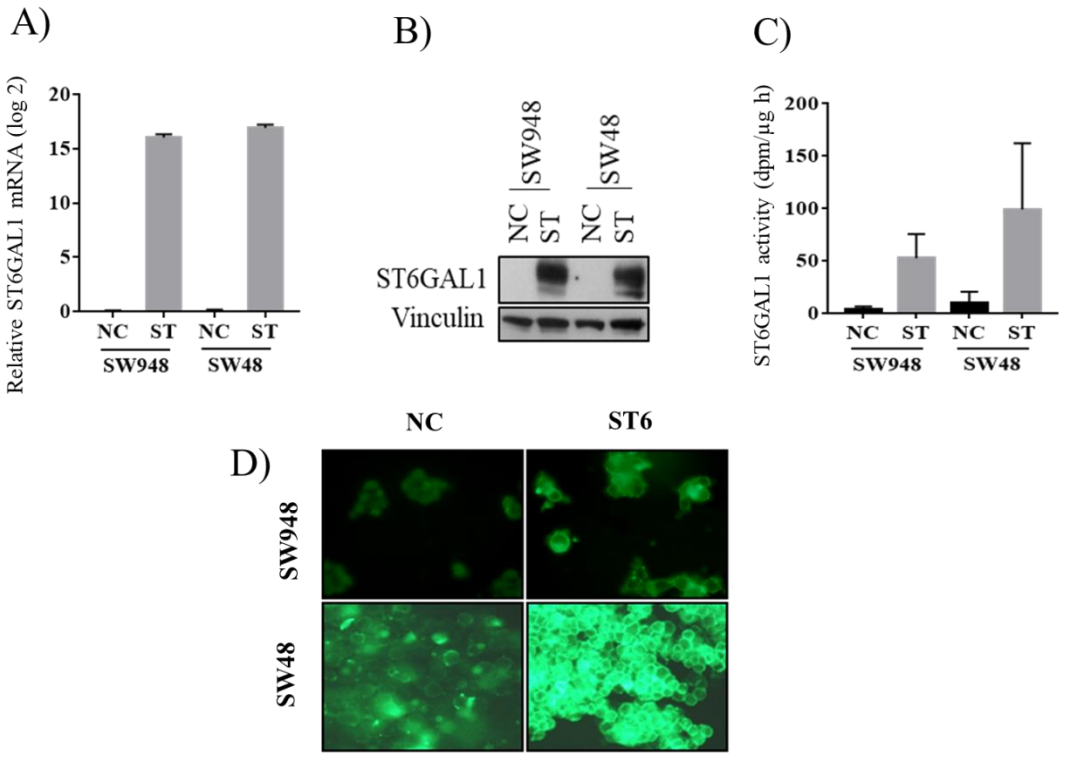


Figure VI. 1 - Biochemical characterization of negative control (NC) and ST6GAL1 (ST)-transduced cells. A: ST6GAL1 mRNA expression measured by real-time PCR. B: ST6GAL1 enzyme protein detected in cell homogenates by Western blotting. C: ST6GAL1 enzyme activity measured by incorporation of radioactive sialic acid on asialotransferrin. D: Expression of Sia6LacNAc structures detected with FITC-labelled *Sambucus nigra* agglutinin (SNA). From manuscript "Venturi, G., Gomes Ferreira, I., Pucci, M., Ferracin, M., Malagolini, N., Chiricolo, M. & Dall'Olio, F. Cell type specific transcriptomic and phenotypic impact of ST6GAL1 overexpression in colon cancer."

Table VI. 2 - Genes modulated and respective biological functions between high/low ST6GAL1 mRNA expression in CRC patients. Red color indicates up-regulation and green color indicates down-regulation.

Gene symbol/Gene ID	Fuctions (specifications)	General function
CEACAM5|1048	Founding member of the carcinoembryonic antigen family. Clinical biomarker for gastrointestinal cancers. Promotes tumor development as a cell adhesion molecule and regulates differentiation, apoptosis, and cell polarity.	**Adhesion**
CD24|100133941	Sialo-glycoprotein cell adhesion molecule expressed mainly on granulocytes and B lymphocytes.	
PRDX5|25824	Member of the peroxiredoxin family of antioxidant enzymes, which reduce hydrogen peroxide and alkyl hydroperoxides. Plays an antioxidant protective role in different tissues under normal conditions and during inflammatory processes.	**Antioxidative properties**
PYGB|5834	Glycogen phosphorylase, which catalyzes the rate-determining step in glycogen degradation.	**Carbohydrate metabolism and transport**
PKM2|5315	Glycolytic enzyme that plays a general role in caspase independent cell death of tumor cells. Contributes to the control of glycolysis and is important for tumor cell proliferation and survival.	
GAPDH|2597	Multiple functions, including carbohydrate metabolism.	
SLC2A1|6513	This gene encodes a major glucose transporter in the mammalian blood-brain barrier.	
REG3A|5068	Regenerating proteins are acute phase reactants, lectins, anti-apoptotic or growth factors. This gene encodes a pancreatic secretory protein that may be involved in cell proliferation or differentiation.	**Regeneration and repair**
TFF3|7033	Involved in the maintenance and repair of the intestinal mucosa. Promotes the mobility of epithelial cells in healing processes.	
REG1A|5967	Regenerating proteins are acute phase reactants, lectins, anti-apoptotic or growth factors. REG1A is secreted by the exocrine pancreas associated with islet cell regeneration.	
REG4|83998	Regenerating proteins are acute phase reactants, lectins, anti-apoptotic or growth factors. REG4 is a calcium-independent mannose-binding lectin, involved in inflammatory and metaplastic responses of the gastrointestinal epithelium.	
SERPINA1|5265	Serine protease inhibitor for elastase, plasmin, thrombin, trypsin, chymotrypsin, and plasminogen activator.	**Protease Inhibitor**
SPINK4|27290	Serine-type endopeptidase inhibitor activity.	
LYZ|4069	Lysozymes are bacteriolytic and enhance the activity of immunoagents	**Antimicrobial and immune function**
LCN2|3934	Transport small hydrophobic molecules. This protein is a neutrophil gelatinase-associated lipocalin and plays a role in innate immunity by limiting bacterial growth as a result of sequestering iron-containing siderophores.	
ANXA2|302	Involved in cell motility, linkage of membrane-associated protein complexes to the actin cytoskeleton and organization of exocytosis.	**Cytoskeleton organization**
KRT19|3880	Member of the intermediate filament family.	

S100A6|6277	Calcium sensor and modulator. Indirectly plays a role in the reorganization of the actin cytoskeleton and in cell motility.	
FCGBP|8857	May be involved in the maintenance of the mucosal structure as a gel-like component of the mucosa.	**Mucin-related**
MUC2|4583	Coats the epithelia of the intestines, airways and other mucus membrane-containing organs. Downregulation of this gene has been observed in patients with Crohn disease and ulcerative colitis.	
AGR2|10551	Required for MUC2 post-transcriptional synthesis and secretion. Proto-oncogene that may play a role in cell migration, cell differentiation and cell growth.	
CLCA1|1179	Involved in chloride conductance. May be involved in the regulation of mucus production and/or secretion by goblet cells and inflammation in the innate immune response. May play a role as a tumor suppressor. Induces MUC5AC.	
MUC5B|727897	Gel-forming mucin that contributes to the lubricating and viscoelastic properties of mucus.	
YWHAB|7529	Mediate signal transduction by binding to phosphoserine-containing proteins. May play a role in linking mitogenic signaling and the cell cycle machinery.	**Signal transduction**
ERBB2|2064	Member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases. Enhances kinase-mediated activation of downstream signalling pathways.	

Table VI. 3 - Genes modulated in SW48 ST cells compared to SW48 NC cells. In red and green, are represented up-regulated and down-regulated genes, respectively. For correction of p-values, Benjamini-Hochberg multiple test comparison was applied. X represents a gene not identified or with unknown/unclear functions.

Gene Symbol	P-value corrected	Gene Name
CALCA	0.1046	Calcitonin-related polypeptide alpha
NUDT12	0.1368	Nudix Hydrolase 12
MAP1B	0.1468	Microtubule-associated protein 1B
PARP11	0.1368	Poly (ADP-ribose) polymerase family, member 11
UGT2B7	0.1368	UDP glucuronosyltransferase 2 family, polypeptide B7
NTRK2	0.1368	Neurotrophic tyrosine kinase, receptor, type 2
X_1	0.1368	BX114010 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGp998G154069
KRT17	0.1069	Keratin 17, type I
KRT14	0.1368	Keratin 14, type I
UGT2B10	0.1368	UDP glucuronosyltransferase 2 family, polypeptide B10
KRT16	0.1069	Keratin 16, type I
PADI1	0.1368	Peptidyl arginine deiminase, type I
X_2	0.1424	Keratin 16 pseudogene 6
PDLIM5	0.1490	PDZ and LIM domain 5
UGT2B11	0.1328	UDP glucuronosyltransferase 2 family, polypeptide B11
ADAM7	0.1468	ADAM metallopeptidase domain 7
STRA6	0.1368	Stimulated by retinoic acid 6
X_3	0.1368	Homo sapiens cDNA FLJ26766 fis, clone PRS02790
ANXA8L1	0.1368	Annexin A8-like 1
SEMA3A	0.1368	Sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A
C14orf37	0.1328	Chromosome 14 open reading frame 37
ASB4	0.1399	Ankyrin repeat and SOCS box containing 4
CNTLN	0.1368	Centlein, centrosomal protein
OBSCN	0.1368	Obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF
SLC30A3	0.1069	Solute carrier family 30 (zinc transporter), member 3
EPS8L1	0.1328	EPS8-like 1
VWA5A	0.1490	Von Willebrand factor A domain containing 5A
COLCA1	0.1368	Colorectal cancer associated 1
CLEC2B	0.1368	C-type lectin domain family 2, member B
ANO1	0.1328	Anoctamin 1, calcium activated chloride channel
MS4A15	0.1368	Membrane-spanning 4-domains, subfamily A, member 15
X_4	0.1368	No information available

ST8SIA1	0.1069	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 1
UGT2B15	0.1368	UDP glucuronosyltransferase 2 family, polypeptide B15
PDE4A	0.1328	Phosphodiesterase 4A, cAMP-specific
EPHA3	0.1069	EPH receptor A3
HLA-DPB1	0.1069	Major histocompatibility complex, class II, DP beta 1
INPP4B	0.1069	Inositol polyphosphate-4-phosphatase, type II, 105kDa
TRAPPC2	0.1069	Trafficking protein particle complex 2
X_5	0.1368	No information available
X_6	0.1468	No information available
INA	0.1368	Internexin neuronal intermediate filament protein, alpha
ANK3	0.1368	Ankyrin 3, node of Ranvier (ankyrin G)
SNORD83A	0.1328	Small nucleolar RNA, C/D box 83A
HAS2	0.1424	Hyaluronan synthase 2
SERHL2	0.1368	Serine hydrolase-like 2
INHBB	0.1424	Inhibin, beta B
TMEM41A	0.1368	Transmembrane protein 41A
TLR1	0.1368	Toll-like receptor 1
DNAJB13	0.1368	DnaJ (Hsp40) homolog, subfamily B, member 13
PXDNL	0.1368	Peroxidasin-like
FHOD3	0.1069	Formin homology 2 domain containing 3
KDM5D	0.1468	Lysine (K)-specific demethylase 5D
X_7	0.1391	No information available
X_8	0.1368	Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens cDNA 5'
MUC2	0.1424	Mucin 2, oligomeric mucus/gel-forming
PCDH1	0.1490	Protocadherin 1
FABP6	0.1424	Fatty acid binding protein 6, ileal
CSAG1	0.1368	Chondrosarcoma associated gene 1
CCDC170	0.1424	Coiled-coil domain containing 170
ZP1	0.1448	Zona pellucida glycoprotein 1 (sperm receptor)
X_9	0.1328	No information available
KLRC1	0.1424	Killer cell lectin-like receptor subfamily C, member 1
LOC403323	0.1328	Uncharacterized LOC403323
X_{10}	0.1368	No information available
LOC729224	0.1424	Uncharacterized LOC729224

Table VI. 4 - Genes modulated in SW948 ST cells compared to SW948 NC cells. In red and green, are represented up-regulated and down-regulated genes, respectively. For correction of p-values, Benjamini-Hochberg multiple test comparison was applied. X represents a gene not identified or with unknown functions.

Gene Symbol	P-value corrected	Gene Name
PEG3	0.1162	paternally expressed 3
EIF4G3	0.0944	eukaryotic translation initiation factor 4 gamma, 3
EML5	0.1406	echinoderm microtubule associated protein like 5
X_{11}	0.0914	No information available
FAM230A	0.1254	family with sequence similarity 230, member A
LOC102725453	0.0511	uncharacterized LOC102725453
LOC101060157	0.0914	uncharacterized LOC101060157
CXCL12	0.1254	chemokine (C-X-C motif) ligand 12
X_{12}	0.1254	No information available
SLC38A3	0.0914	solute carrier family 38, member 3
WNT4	0.1393	wingless-type MMTV integration site family, member 4
X_{13}	0.0638	CLPP_DESPS (Q6AK59) ATP-dependent Clp protease proteolytic subunit (Endopeptidase Clp), partial (7%)
PPP1R17	0.0563	protein phosphatase 1, regulatory subunit 17
X_{14}	0.0511	No information available
CTF1	0.0914	cardiotrophin 1
POLH	0.1115	polymerase (DNA directed), eta
CDCA5	0.0914	cell division cycle associated 5
X_{15}	0.0526	No information available
TCF15	0.0558	transcription factor 15 (basic helix-loop-helix)
X_{16}	0.0511	MORC family CW-type zinc finger 3
X_{17}	0.1254	Q5VT28_HUMAN (Q5VT28) Family with sequence similarity 27, member B , partial (85%)
X_{18}	0.0526	No information available
FLJ32154	0.0511	uncharacterized protein FLJ32154
LOC100506458	0.1016	putative uncharacterized protein LOC65996-like
FANK1	0.0914	fibronectin type III and ankyrin repeat domains 1
HOXC13	0.0914	homeobox C13
DUSP26	0.0914	dual specificity phosphatase 26 (putative)
X_{19}	0.1254	cs47h05.y2 Human Retinal pigment epithelium [CA394149]
LOC101929469	0.1007	uncharacterized LOC101929469
ZNF71	0.0914	zinc finger protein 71
X_{20}	0.1044	Q6IMN1_HUMAN (Q6IMN1) Growth arrest-specific 6, partial (13%)
BEST2	0.1368	bestrophin 2
X_{21}	0.1390	No information available
X_{22}	0.1367	No information available
ZNF483	0.0944	zinc finger protein 483

X_{23}	0.0526	No information available
SYP	0.0914	synaptophysin
CNIH2	0.0914	cornichon family AMPA receptor auxiliary protein 2
LOC100131860	0.1254	uncharacterized LOC100131860
X_{24}	0.1472	No information available
X_{25}	0.0526	Q2J851_FRASC (Q2J851) ATPases involved in chromosome partitioning-like, partial (5%)
ZNF727	0.0809	zinc finger protein 727
CHRFAM7A	0.1463	CHRNA7 (cholinergic receptor, nicotinic, alpha 7, exons 5-10) and FAM7A (family with sequence similarity 7A, exons A-E) fusion
FANCD2	0.1153	Fanconi anemia, complementation group D2
SPG7	0.1070	spastic paraplegia 7 (pure and complicated autosomal recessive)
AGR3	0.0526	anterior gradient 3
CCDC121	0.0526	coiled-coil domain containing 121
POMZP3	0.0862	POM121 and ZP3 fusion
X_{26}	0.1390	immunoglobulin lambda variable 5-45
PDZD2	0.0809	PDZ domain containing 2
FFAR2	0.0533	free fatty acid receptor 2
SCML2	0.1254	sex comb on midleg-like 2 (Drosophila)
X_{27}	0.1367	No information available
SPDYE5	0.1070	speedy/RINGO cell cycle regulator family member E5
SRRM3	0.0914	serine/arginine repetitive matrix 3
X_{28}	0.0944	No information available
LECT1	0.0533	leukocyte cell derived chemotaxin 1
X_{29}	0.1254	No information available
IRX2	0.1144	iroquois homeobox 2
X_{30}	0.1254	No information available
APBB1	0.1390	amyloid beta (A4) precursor protein-binding, family B, member 1 (Fe65)
KRIT1	0.0833	KRIT1, ankyrin repeat containing
ZNF239	0.0820	zinc finger protein 239
EDN2	0.0845	endothelin 2
LOC388282	0.0862	uncharacterized LOC388282
MX1	0.1144	MX dynamin-like GTPase 1
STAT4	0.1254	signal transducer and activator of transcription 4
KCNJ11	0.1410	potassium channel, inwardly rectifying subfamily J, member 11
S100A3	0.0533	S100 calcium binding protein A3
NRP2	0.1044	neuropilin 2
DEFB130	0.0944	defensin, beta 130
C9orf84	0.0533	chromosome 9 open reading frame 84
SCGB2A1	0.1181	secretoglobin, family 2A, member 1
C3orf35	0.1254	chromosome 3 open reading frame 35
MUC17	0.1368	mucin 17, cell surface associated

CAPN13	0.1368	calpain 13
DOC2GP	0.0914	double C2-like domains, gamma, pseudogene
OR5F1	0.0944	olfactory receptor, family 5, subfamily F, member 1
GGT6	0.1390	gamma-glutamyltransferase 6
ERICH5	0.1254	glutamate-rich 5
XIST	0.1284	X inactive specific transcript (non-protein coding)
FAM171B	0.1474	family with sequence similarity 171, member B
ZC3H13	0.1390	zinc finger CCCH-type containing 13
PRAM1	0.1460	PML-RARA regulated adaptor molecule 1
SLC22A13	0.1254	solute carrier family 22 (organic anion/urate transporter), member 13
C6orf52	0.0980	chromosome 6 open reading frame 52
SLC15A1	0.1368	solute carrier family 15 (oligopeptide transporter), member 1
AHNAK	0.0533	AHNAK nucleoprotein
KLF8	0.0862	Kruppel-like factor 8
X_{31}	0.1390	No information available
ADRB2	0.1254	adrenoceptor beta 2, surface
CAPN13	0.1463	calpain 13
ZNF543	0.0526	zinc finger protein 543
PPM1K	0.0526	protein phosphatase, Mg2+/Mn2+ dependent, 1K
CD33	0.0542	CD33 molecule
SH2D6	0.0809	SH2 domain containing 6
APOE	0.0914	apolipoprotein E
RNF144A	0.1153	ring finger protein 144A
PHLDA3	0.0533	pleckstrin homology-like domain, family A, member 3
UTRN	0.1291	utrophin
CASC4	0.0914	cancer susceptibility candidate 4
PPP5D1	0.0526	PPP5 tetratricopeptide repeat domain containing 1
P3H2	0.1442	prolyl 3-hydroxylase 2
UBL4B	0.0944	ubiquitin-like 4B
PRSS57	0.0533	protease, serine, 57
X_{32}	0.0944	NFKB activating protein pseudogene 1
X_{33}	0.0526	No information available
FAM171B	0.1461	family with sequence similarity 171, member B
LOC101927137	0.0914	uncharacterized LOC101927137
HOXC5	0.1472	homeobox C5
NEK11	0.1254	NIMA-related kinase 11
MS4A8	0.0526	membrane-spanning 4-domains, subfamily A, member 8
KLF13	0.1254	Kruppel-like factor 13
HSPB3	0.0511	heat shock 27kDa protein 3
ABCC8	0.0914	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
FMNL1	0.0533	formin-like 1

TRIM9	0.0526	tripartite motif containing 9
S1PR4	0.1472	sphingosine-1-phosphate receptor 4
FOLR1	0.0914	folate receptor 1 (adult)
IL1RN	0.0944	interleukin 1 receptor antagonist
PRO1082	0.1027	uncharacterized protein PRO1082
PTAFR	0.1368	platelet-activating factor receptor
SEMA7A	0.1368	semaphorin 7A, GPI membrane anchor (John Milton Hagen blood group)
ACE2	0.0914	angiotensin I converting enzyme 2
MEDAG	0.1144	mesenteric estrogen-dependent adipogenesis
PCDHGB1	0.1027	protocadherin gamma subfamily B, 1
SOHLH2	0.0526	spermatogenesis and oogenesis specific basic helix-loop-helix 2
TGFB2	0.1254	transforming growth factor, beta 2
DQX1	0.0740	DEAQ box RNA-dependent ATPase 1
CNTN5	0.0809	contactin 5
H2AFY2	0.0685	H2A histone family, member Y2
BST2	0.1409	bone marrow stromal cell antigen 2
ZNF439	0.0644	zinc finger protein 439
SLC14A1	0.1472	solute carrier family 14 (urea transporter), member 1 (Kidd blood group)
ANKRD1	0.1368	ankyrin repeat domain 1 (cardiac muscle)
MYEOV	0.0526	myeloma overexpressed
X_{34}	0.1059	No information available

Table VI. 5 - 10 top pathway maps (first part), cellular processes (second part) and process networks (third part) affected by ST6GAL1 in SW48 cells. It includes the p-value, false discovery rate (FDR), name and number of genes modulated involved in certain pathway or process.

GeneGO - pathway maps	P-value	FDR	Number of genes	Modulated genes
Estradiol metabolism	2.075E-04	1.909E-02	3	UGT2B11, UGT2B7, UGT2B28
Inhibition of Ephrin receptors in colorectal cancer	3.096E-03	8.092E-02	2	Ephrin-A receptors, Ephrin-A receptor 3
Development_Transcription factors in segregation of hepatocytic lineage	3.096E-03	8.092E-02	2	Activin B, Activin
Androstenedione and testosterone biosynthesis and metabolism p.2	3.518E-03	8.092E-02	2	UGT2B15, UGT2B28
Cytoskeleton remodeling_Keratin filaments	4.438E-03	8.166E-02	2	Keratin 16, Keratin 17
Androstenedione and testosterone biosynthesis and metabolism p.3	5.457E-03	8.367E-02	2	UGT2B7, UGT2B28
Neurophysiological process_Receptor-mediated axon growth repulsion	7.166E-03	9.418E-02	2	Semaphorin 3A, Ephrin-A receptors
Nicotine metabolism in liver	1.048E-02	1.205E-01	2	UGT2B10, UGT2B7
Inhibition of GSK3 beta by lithium in major depressive disorder	5.940E-02	2.040E-01	1	TrkB
Action of GSK3 beta in bipolar disorder	6.202E-02	2.040E-01	1	MAP-1B

Gene ontology - cellular processes	p-value	FDR	Number of genes	Modulated genes
Cellular glucuronidation	5.78E-09	1.40E-05	5	UGT, UGT2B11, UGT2B15, UGT2B7, UGT2B28
Uronic acid metabolic process	1.68E-08	1.40E-05	5	UGT, UGT2B11, UGT2B15, UGT2B7, UGT2B28
Glucuronate metabolic process	1.68E-08	1.40E-05	5	UGT, UGT2B11, UGT2B15, UGT2B7, UGT2B28
Negative regulation of hepatocyte growth factor biosynthetic process	1.40E-07	4.98E-05	3	Activin B, Activin, Activin beta B
Negative regulation of hepatocyte growth factor production	1.40E-07	4.98E-05	3	Activin B, Activin, Activin beta B
Regulation of hepatocyte growth factor biosynthetic process	1.40E-07	4.98E-05	3	Activin B, Activin, Activin beta B
Regulation of hepatocyte growth factor production	1.40E-07	4.98E-05	3	Activin B, Activin, Activin beta B
Regulation of biological quality	3.57E-07	1.11E-04	31	Semaphorin 3A, Keratin 16, ANXA8L1, Ephrin-A receptors, Ephrin-A receptor 3, UGT, TMEM16A, Mucin 2, FHOD3, DRF, Activin B, Activin, Activin beta B, INPP4B, MHC class II beta chain, PDE4, PDE, UGT2B11, TrkB, CGRP, Calcitonin, MAP-1B, HAS, HAS2, Annexin VIII, UGT2B7, STRA6, ENH, EPS8L1, Ankyrin-G, ZnT3
Reproductive senescence	4.87E-07	1.35E-04	3	Activin B, Activin, Activin beta B
Positive regulation of multicellular organismal process	8.31E-07	2.08E-04	20	Semaphorin 3A, Ephrin-A receptors, Ephrin-A receptor 3, Activin B, Activin, Activin beta B, ASB4, HLA-DPB, MHC class II beta chain, HLA-DPB1, Keratin 17, PDE4, PDE, TLR1, TrkB, CGRP, Calcitonin, MAP-1B, HAS, HAS2

GeneGo - process networks	p-value	FDR	Number of genes	Modulated genes
Cytoskeleton_Intermediate filaments	1.869E-03	6.168E-02	3	Alpha-internexin, Keratin 16, Keratin 17
Development_Neurogenesis_Axonal guidance	4.906E-03	8.095E-02	4	Semaphorin 3A, Ephrin-A receptors, Ephrin-A receptor 3, TrkB
Cell adhesion_Glycoconjugates	1.296E-02	1.315E-01	3	TrkB, HAS, HAS2
Cell adhesion_Attractive and repulsive receptors	1.594E-02	1.315E-01	3	Semaphorin 3A, Ephrin-A receptors, Ephrin-A receptor 3
Immune response_Antigen presentation	2.181E-02	1.439E-01	3	MHC class II beta chain, HLA-DPB1, NKG2A
Signal transduction_Neuropeptide signaling pathways	8.190E-02	3.947E-01	2	CGRP, Calcitonin
Development_Ossification and bone remodeling	8.373E-02	3.947E-01	2	Activin, Calcitonin
Development_Regulation of angiogenesis	1.490E-01	5.347E-01	2	Ephrin-A receptors, TrkB
Autophagy_Autophagy	1.571E-01	5.347E-01	1	TLR1
Transport_Bile acids transport and its regulation	1.906E-01	5.347E-01	1	Gastrotropin

Table VI. 6 - 10 top pathway maps (first part), cellular processes (second part) and process networks (third part) affected by ST6GAL1 in SW948 cells. It includes the p-value, false discovery rate (FDR), name and number of genes modulated involved in certain pathway or process.

GeneGO - pathway maps	P-value	FDR	Number of genes	Modulated genes
Mechanism of Pioglitazone/Glimepiride and Rosiglitazone/Glimepiride cooperative action in Diabetes mellitus, Type 2	6.336E-04	1.301E-01	2	SUR1, Kir6.2
Mast cell migration in asthma	2.037E-03	1.301E-01	3	PTAFR, TGF-beta 2, SDF-1
Role of ZNF202 in regulation of expression of genes involved in atherosclerosis	2.367E-03	1.301E-01	2	HDL proteins, APOE
Apoptosis and survival_nAChR in apoptosis inhibition and cell cycle progression	4.496E-03	1.301E-01	2	nAChR alpha, nAChR alpha-7
Bone metastases in Prostate Cancer	4.496E-03	1.301E-01	2	SDF-1, WNT
Putative pathways of Oleic acid sensing in ventromedial hypothalamus in obesity (rodent model)	4.806E-03	1.301E-01	2	SUR1, Kir6.2
Upregulation of IL-8 expression in colorectal cancer	5.457E-03	1.301E-01	2	WNT4, WNT
Glucose-excited neurons of ventromedial nucleus (rodent model)	5.457E-03	1.301E-01	2	SUR1, Kir6.2
Muscle contraction_Role of kappa-type opioid receptor in heart	5.796E-03	1.301E-01	2	Beta-2 adrenergic receptor, Kir6.2
Nicotine / Beta-adrenergic signaling in lung cancer	6.145E-03	1.301E-01	2	Beta-adrenergic receptor, Beta-2 adrenergic receptor

Gene ontology - cellular processes	p-value	FDR	Number of genes	Modulated genes
Regulation of ion transport	2.865E-13	1.220E-09	26	Utrophin, PEPT1, CNIH2, PTAFR, TGF-beta, TGF-beta 2, SDF-1, HDL proteins, APOE, AHNAK, DRF, Olfactory receptor, IL1RN, Galpha(i)-specific EDG GPCRs, Galpha(q)-specific EDG GPCRs, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, nAChR alpha, S100, SUR, SUR1, SN1, K(+) channel, subfamily J, Kir6.2
Positive regulation of heart contraction	7.156E-13	1.220E-09	10	ACE2, TGF-beta, TGF-beta 2, Endothelin-2, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, nAChR alpha, nAChR alpha-7
Negative regulation of blood vessel morphogenesis	1.091E-12	1.220E-09	13	Krit1, Chondromodulin-I, TGF-beta, TGF-beta 2, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, WNT4, WNT, SUR, SUR1, STAT4
Positive regulation of blood circulation	1.095E-12	1.220E-09	12	ACE2, PTAFR, TGF-beta, TGF-beta 2, Endothelin-2, Olfactory receptor, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, nAChR alpha, nAChR alpha-7
Negative regulation of vasculature development	2.104E-12	1.875E-09	13	Krit1, Chondromodulin-I, TGF-beta, TGF-beta 2, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, WNT4, WNT, SUR, SUR1, STAT4
Positive regulation of ion transport	2.282E-11	1.695E-08	17	PTAFR, SDF-1, HDL proteins, APOE, Galpha(i)-specific EDG GPCRs, Galpha(q)-specific EDG GPCRs, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, nAChR alpha, S100, SUR, SUR1, SN1, K(+) channel, subfamily J, Kir6.2
Regulation of transporter activity	3.048E-11	1.804E-08	16	Utrophin, CNIH2, PTAFR, HDL proteins, AHNAK, Olfactory receptor, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, nAChR alpha, S100, SUR, SUR1, K(+) channel, subfamily J, Kir6.2
Negative regulation of wound healing	3.239E-11	1.804E-08	10	HDL proteins, APOE, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, WNT4, WNT, SUR, SUR1
Regulation of cation transmembrane transport	6.268E-11	3.103E-08	17	Utrophin, CNIH2, TGF-beta, TGF-beta 2, AHNAK, DRF, Olfactory receptor, Beta-adrenergic receptor, Beta-2 adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, nAChR alpha, S100, SUR, SUR1, K(+) channel, subfamily J, Kir6.2
Positive regulation of cell growth	8.204E-11	3.656E-08	14	TGF-beta, TGF-beta 2, SDF-1, HDL proteins, APOE, Semaphorin 7A, Galpha(i)-specific EDG GPCRs, Galpha(q)-specific EDG GPCRs, Beta-adrenergic receptor, Galpha(i)-specific amine GPCRs, Galpha(s)-specific amine GPCRs, WNT, S100, eIF4G1/3

GeneGo - process networks	p-value	FDR	Number of genes	Modulated genes
Cell adhesion_Amyloid proteins	7.352E-03	3.213E-01	5	PDZK3, TGF-beta 2, Fe65, WNT4, WNT
Development_Neuromuscular junction	1.349E-02	3.213E-01	4	SYP, Utrophin, nAChR alpha, nAChR alpha-7
Development_Neurogenesis_Axonal guidance	1.437E-02	3.213E-01	5	Neuropilin-2, APOE, AHNAK, Semaphorin 7A, Fe65
Signal transduction_WNT signaling	2.491E-02	3.213E-01	4	TGF-beta, TGF-beta 2, WNT4, WNT
Development_Neurogenesis_Synaptogenesis	2.630E-02	3.213E-01	4	SYP, APOE, WNT, nAChR alpha
Cell adhesion_Cadherins	2.725E-02	3.213E-01	4	Protocadherin gamma B1, PDZK3, WNT4, WNT
Development_Neurogenesis in general	3.231E-02	3.213E-01	4	Galpha(i)-specific amine GPCRs, WNT4, WNT, Cardiotrophin-1
Transport_Potassium transport	3.338E-02	3.213E-01	4	SUR, SUR1, K(+) channel, subfamily J, Kir6.2
Proliferation_Lymphocyte proliferation	4.211E-02	3.519E-01	4	BST2, SDF-1, CD33, STAT4
Translation_Regulation of initiation	4.570E-02	3.519E-01	3	nAChR alpha-7, eIF4G1/3, eIF4G3

Table VI. 7 - Genes modulated and respective biological functions between high/low B4GALNT2 mRNA expression in CRC patients. Red color indicates up-regulation and green color indicates down-regulation.

Gene symbol/Gene ID	Fuctions (specifications)	General function
MUC5B|727897	Gel-forming mucin that contributes to the lubricating and viscoelastic properties of mucus.	**Mucin-related**
MUC2|4583	Coats the epithelia of the intestines, airways and other mucus membrane-containing organs. Downregulation of this gene has been observed in patients with Crohn disease and ulcerative colitis.	
AGR2|10551	Required for MUC2 post-transcriptional synthesis and secretion. Proto-oncogene that may play a role in cell migration, cell differentiation and cell growth.	
FCGBP|8857	May be involved in the maintenance of the mucosal structure as a gel-like component of the mucosa.	
CLCA1|1179	Involved in chloride conductance. May be involved in the regulation of mucus production and/or secretion by goblet cells and inflammation in the innate immune response. May play a role as a tumor suppressor. Induces MUC5AC.	
REG4|83998	Regenerating proteins are acute phase reactants, lectins, anti-apoptotic or growth factors. REG4 is a calcium-independent mannose-binding lectin, involved in inflammatory and metaplastic responses of the gastrointestinal epithelium.	**Regeneration and repair**
TFF3|7033	Involved in the maintenance and repair of the intestinal mucosa. Promotes the mobility of epithelial cells in healing processes.	
REG1A|5967	Regenerating proteins are acute phase reactants, lectins, anti-apoptotic or growth factors. REG1A is secreted by the exocrine pancreas associated with islet cell regeneration.	
SERPINA1|5265	Serine protease inhibitor for elastase, plasmin, thrombin, trypsin, chymotrypsin, and plasminogen activator.	**Protease Inhibitor**
SPINK4|27290	Serine-type endopeptidase inhibitor activity.	
PIGR|5284	This receptor binds polymeric IgA and IgM at the basolateral surface of epithelial cells.	**Immune function**
LCN2|3934	Transport small hydrophobic molecules. This protein is a neutrophil gelatinase-associated lipocalin and plays a role in innate immunity by limiting bacterial growth as a result of sequestering iron-containing siderophores.	
LGALS4|3960	Implicated in modulating cell-cell and cell-matrix interactions. The expression of this gene is restricted to small intestine, colon, and rectum, and it is underexpressed in colorectal cancer.	**Cell adhesion/surface interactions**
IGJ|3512	Serves to link two monomer units of either IgM or IgA. It also helps to bind these immunoglobulins to secretory component.	
ITM2C	Negative regulator of amyloid-beta peptide production.	**Amyloid beta-binding**
ADAM6|8755	Pseudogene. No functions known.	
LOC96610|96610	Pseudogene. No functions known.	

IGF2|3481	The insulin-like growth factors possess growth-promoting activity. *In vitro*, they are potent mitogens for cultured cells.	**Growth factor activity**

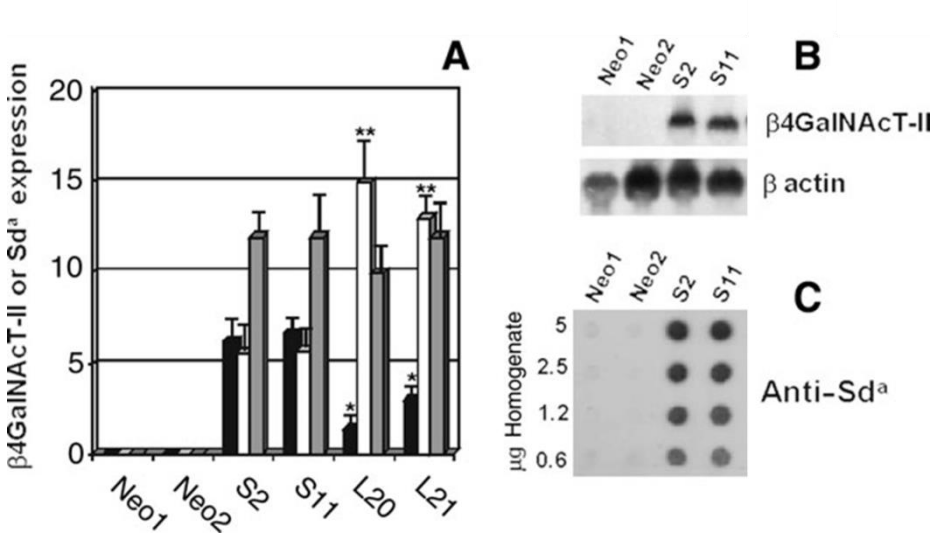


Figure VI. 2 - Biochemical characterization of mock transfectants (LS 174T clone Neo1 and Neo2) and short-form B4GALNT2 transfectants (LS 174T S2 and S11 clones) A: B4GALNT2 enzyme activity. B: B4GALNT2 mRNA expression by northern blot. C: Sd^a antigen expression detected in cell homogenates by dot-blot analysis. In picture A, its possible to observe other two clones, L20 and L21, long-form B4GALNT2 transfectants, which are not the scope of this phD project. In (A) is reported the B4GALNT2 activity (nmol/mg protein h; black bars), the quantification of the mRNA signal, normalized for the b-actin signal (white bars), and the quantification of the dot blot signals (gray bars), both expressed in arbitrary units. Adapted from "Biosynthesis and expression of the Sd^a and sialyl Lewis x antigens in normal and cancer colon" (2007) from Malagolini, Nadia. *et al.*

Table VI. 8 - Genes modulated in short form-B4GALNT2 clones compared to Neo clone. In red and green, are represented up-regulated and down-regulated genes, respectively. For correction of p-values, Benjamini-Hochberg multiple test comparison was applied. X represents a gene not identified or with unknown/unclear functions.

Gene Symbol	P-value corrected	Gene name
NID1	0.0306	nidogen 1
ZNF22	0.0077	zinc finger protein 22
STARD3NL	0.0346	STARD3 N-terminal like
GALC	0.0077	galactosylceramidase
NPTX1	0.0123	neuronal pentraxin I
LGALS2	0.0142	lectin, galactoside-binding, soluble, 2
SOX2	0.0163	SRY (sex determining region Y)-box 2
MID2	0.0152	midline 2
ARMC4	0.0196	armadillo repeat containing 4
NINL	0.0339	ninein-like
PEG10	0.0233	paternally expressed 10
FMO3	0.0253	flavin containing monooxygenase 3
RAI14	0.0190	retinoic acid induced 14
ROR1	0.0163	receptor tyrosine kinase-like orphan receptor 1
MBOAT2	0.0383	membrane bound O-acyltransferase domain containing 2
MYH3	0.0167	myosin, heavy chain 3, skeletal muscle, embryonic
INMT	0.0142	indolethylamine N-methyltransferase
F5	0.0331	coagulation factor V (proaccelerin, labile factor)
NINL	0.0339	ninein-like
ALX1	0.0167	ALX homeobox 1
FAM110B	0.0142	family with sequence similarity 110, member B
X_{35}	0.0236	No information available
X_{36}	0.0331	PREDICTED: Homo sapiens ZFP92 zinc finger protein (ZFP92), transcript variant X1, mRNA [XM_005274652]
FAM26F	0.0195	family with sequence similarity 26, member F
HEY2	0.0411	hes-related family bHLH transcription factor with YRPW motif 2
TBATA	0.0379	thymus, brain and testes associated
RNF152	0.0335	ring finger protein 152
ITGAX	0.0167	integrin, alpha X (complement component 3 receptor 4 subunit)
TMEM255B	0.0238	transmembrane protein 255B
EPB41L3	0.0339	erythrocyte membrane protein band 4.1-like 3
KRTAP5-4	0.0451	keratin associated protein 5-4
SOHLH1	0.0411	spermatogenesis and oogenesis specific basic helix-loop-helix 1
CALCA	0.0436	calcitonin-related polypeptide alpha
ANK1	0.0167	ankyrin 1, erythrocytic
INHBB	0.0203	inhibin, beta B
CHSY3	0.0167	chondroitin sulfate synthase 3

OBSL1	0.0195	obscurin-like 1
TREX2	0.0195	three prime repair exonuclease 2
CRIP2	0.0383	cysteine-rich protein 2
TRAM1L1	0.0167	translocation associated membrane protein 1-like 1
ADAM8	0.0197	ADAM metallopeptidase domain 8
SPRR3	0.0238	small proline-rich protein 3
WFDC2	0.0343	WAP four-disulfide core domain 2
LIPF	0.0190	lipase, gastric
FBXW2	0.0203	F-box and WD repeat domain containing 2
PDZD2	0.0411	PDZ domain containing 2
VWA5B2	0.0331	von Willebrand factor A domain containing 5B2
PAGE2	0.0343	P antigen family, member 2 (prostate associated)
ARHGEF4	0.0343	Rho guanine nucleotide exchange factor (GEF) 4
FGFR3	0.0383	fibroblast growth factor receptor 3
SFTPB	0.0253	surfactant protein B
ANO1	0.0317	anoctamin 1, calcium activated chloride channel
X_{37}	0.0231	No information available
AK4	0.0331	adenylate kinase 4
AKR1C4	0.0474	aldo-keto reductase family 1, member C4
HBE1	0.0356	hemoglobin, epsilon 1
CACNA2D1	0.0238	calcium channel, voltage-dependent, alpha 2/delta subunit 1
TRIM54	0.0343	tripartite motif containing 54
NDUFA4L2	0.0479	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2
X_{38}	0.0411	No information available
OPRL1	0.0231	opiate receptor-like 1
MRC2	0.0253	mannose receptor, C type 2
X_{39}	0.0411	Q2GWG0_CHAGB (Q2GWG0) Predicted protein, partial (5%)
SUV420H2	0.0383	suppressor of variegation 4-20 homolog 2 (Drosophila)
MUC19	0.0356	mucin 19, oligomeric
CPQ	0.0236	carboxypeptidase Q
STAC2	0.0360	SH3 and cysteine rich domain 2
CLEC19A	0.0331	C-type lectin domain family 19, member A
ALOX15B	0.0349	arachidonate 15-lipoxygenase, type B
CALY	0.0253	calcyon neuron-specific vesicular protein
PNPLA1	0.0383	patatin-like phospholipase domain containing 1
HNMT	0.0489	histamine N-methyltransferase
CLDN22	0.0483	claudin 22
MYLPF	0.0386	myosin light chain, phosphorylatable, fast skeletal muscle
KIT	0.0479	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
CALHM1	0.0479	calcium homeostasis modulator 1
SUPT3H	0.0238	suppressor of Ty 3 homolog (S. cerevisiae)

CHD5	0.0463	chromodomain helicase DNA binding protein 5
CD274	0.0317	CD274 molecule
X_{40}	0.0489	No information available
RNF216	0.0387	ring finger protein 216
PRDM8	0.0383	PR domain containing 8
IGF1	0.0379	insulin-like growth factor 1 (somatomedin C)
SGOL2	0.0274	shugoshin-like 2 (S. pombe)
PHACTR3	0.0489	phosphatase and actin regulator 3
DPF3	0.0339	D4, zinc and double PHD fingers, family 3
C14orf119	0.0385	chromosome 14 open reading frame 119
X_{41}	0.0331	No information available
MARCH10	0.0238	membrane-associated ring finger (C3HC4) 10, E3 ubiquitin protein ligase
X_{42}	0.0383	peroxiredoxin 2 pseudogene 3
LOC101930346	0.0238	uncharacterized LOC101930346
X_{43}	0.0386	No information available
LOC84843	0.0430	uncharacterized LOC84843
SLC7A8	0.0219	solute carrier family 7 (amino acid transporter light chain, L system), member 8
X_{44}	0.0203	No information available
BAI2	0.0254	brain-specific angiogenesis inhibitor 2
F2R	0.0238	coagulation factor II (thrombin) receptor
DGKK	0.0238	diacylglycerol kinase, kappa
C17orf67	0.0352	chromosome 17 open reading frame 67
BTN2A1	0.0317	butyrophilin, subfamily 2, member A1
BTN1A1	0.0331	butyrophilin, subfamily 1, member A1
X_{45}	0.0231	myosin, heavy chain 16 pseudogene
ISX	0.0302	intestine-specific homeobox
PALLD	0.0430	palladin, cytoskeletal associated protein
X_{46}	0.0196	No information available
X_{47}	0.0411	PREDICTED: Homo sapiens monofunctional C1-tetrahydrofolate synthase, mitochondrial-like (LOC100996643), transcript variant X5
SPANXC	0.0233	SPANX family, member C
UBD	0.0233	ubiquitin D
X_{48}	0.0253	No information available
X_{49}	0.0331	No information available
LOC100128325	0.0238	uncharacterized LOC100128325
RHBDL2	0.0203	rhomboid, veinlet-like 2 (Drosophila)
F2R	0.0317	coagulation factor II (thrombin) receptor
PHLDB2	0.0254	pleckstrin homology-like domain, family B, member 2
LINC00883	0.0253	long intergenic non-protein coding RNA 883
LBH	0.0396	limb bud and heart development
RHBDL2	0.0311	rhomboid, veinlet-like 2 (Drosophila)
BARX2	0.0377	BARX homeobox 2
X_{50}	0.0381	No information available

IL26	0.0317	interleukin 26
X_{51}	0.0339	No information available
C2orf78	0.0383	chromosome 2 open reading frame 78
CT45A5	0.0460	cancer/testis antigen family 45, member A5
X_{52}	0.0167	Homo sapiens cDNA FLJ25969 fis, clone CBR02250
SLC26A3	0.0203	solute carrier family 26 (anion exchanger), member 3
SHISA6	0.0451	shisa family member 6
LOC101927820	0.0254	uncharacterized LOC101927820
FLJ36777	0.0451	uncharacterized LOC730971
TTC6	0.0167	tetratricopeptide repeat domain 6
MEF2C	0.0253	myocyte enhancer factor 2C
MUC12	0.0167	mucin 12, cell surface associated
LUM	0.0451	lumican
LOC727799	0.0389	uncharacterized LOC727799
X_{53}	0.0233	No information available
X_{54}	0.0254	No information available
GNAS-AS1	0.0317	GNAS antisense RNA 1
SLC14A1	0.0360	solute carrier family 14 (urea transporter), member 1 (Kidd blood group)
X_{55}	0.0137	T cell receptor delta constant [Source:HGNC Symbol;Acc:HGNC:12253]
SKAP1	0.0231	src kinase associated phosphoprotein 1
NGFRAP1	0.0383	nerve growth factor receptor (TNFRSF16) associated protein 1
CD200	0.0411	CD200 molecule
B4GALNT2	0.0297	beta-1,4-N-acetyl-galactosaminyl transferase 2

Table VI. 9 - Most relevant pathway maps (first part), cellular processes (second part) and process networks (third part) affected by B4GALNT2 in LS 174T cells. It includes the p-value, false discovery rate (FDR) and name of the modulated genes involved in a certain pathway or process.

GeneGo - Pathway maps	P-value	FDR	Modulated genes
Cell adhesion_Integrin-mediated cell adhesion and migration	1.697E-02	1.918E-01	MyHC, MRLC
Cell adhesion_ECM remodeling	2.193E-02	1.964E-01	Nidogen, IGF-1
Immune response_Function of MEF2 in T lymphocytes	1.903E-02	1.918E-01	MEF2, MEF2C
Macrophage-induced immunosuppression in the tumor microenvironment	6.051E-02	2.517E-01	PD-L1, ALOX15B
Transcription_HIF-1 targets	5.940E-02	2.517E-01	AK3, SOX2
Immune response _CCR3 signaling in eosinophils	4.079E-02	2.375E-01	MyHC, MRLC
Immune response_Gastrin in inflammatory response	3.340E-02	2.134E-01	MEF2, MEF2C
Development_Regulation of endothelial progenitor cell differentiation from adult stem cells	2.580E-02	1.988E-01	c-Kit, PAR1
Stem cells_Role of growth factors in the maintenance of embryonic stem cell pluripotency	2.046E-02	1.964E-01	SOX2, IGF-1
Cytoskeleton remodeling_Regulation of actin cytoskeleton organization by the kinase effectors of Rho GTPases	2.422E-02	1.964E-01	MyHC, MRLC
Gene Ontology - cellular processes	**P-value**	**FDR**	**Modulated genes**
muscle hypertrophy	1.648E-09	6.553E-06	Galpha(i)-specific peptide GPCRs, MyHC, Galpha(q)-specific peptide GPCRs, MRLC, IGF-1, MEF2, MEF2C, HEY2
cardiac muscle hypertrophy	1.769E-08	2.939E-05	Galpha(i)-specific peptide GPCRs, MyHC, Galpha(q)-specific peptide GPCRs, MRLC, MEF2, MEF2C, HEY2
striated muscle hypertrophy	2.217E-08	2.939E-05	Galpha(i)-specific peptide GPCRs, MyHC, Galpha(q)-specific peptide GPCRs, MRLC, MEF2, MEF2C, HEY2
negative regulation of gonadotropin secretion	3.207E-08	3.188E-05	Galpha(i)-specific peptide GPCRs, Galpha(q)-specific peptide GPCRs, Activin B, Activin, Activin beta B
cell development	5.185E-08	3.832E-05	Galpha(i)-specific peptide GPCRs, MyHC, c-Kit, ROR1, Ankyrin 1, Integrin, CACNA2D, Galpha(q)-specific peptide GPCRs, SOHLH1, SPANXB, Lumican, FGFR3, TRIM54, SLC26A3 (DRA), Activin B, Activin, Activin beta B, MRLC, OBSL1, DAL1, MYH3, NPX1, Prdm8, Cart1, IGF-1, Palladin, CHD5, CGRP, Calcitonin, MEF2, MEF2C, HEY2
regulation of system process	6.514E-08	3.832E-05	Nociceptin receptor, Galpha(i)-specific peptide GPCRs, MyHC, ISX, c-Kit, CACNA2D1, CACNA2D, PAR1, Galpha(q)-specific peptide GPCRs, Activin B, Activin, Activin beta B, MRLC, IGF-1, CGRP, Calcitonin, MEF2, MEF2C, HEY2
muscle cell development	6.746E-08	3.832E-05	Galpha(i)-specific peptide GPCRs, MyHC, CACNA2D, Galpha(q)-specific peptide GPCRs, TRIM54, MRLC, OBSL1, MYH3, IGF-1, MEF2, HEY2
cellular process involved in reproduction in multicellular organism	9.162E-08	4.553E-05	Galpha(i)-specific peptide GPCRs, MyHC, c-Kit, Galpha(q)-specific peptide GPCRs, SOHLH1, SPANXB, SLC26A3 (DRA), Activin B, Activin, Activin beta B, SGOL2, Spatial, CHD5, CGRP, Calcitonin
regulation of gonadotropin secretion	1.079E-07	4.765E-05	Galpha(i)-specific peptide GPCRs, Galpha(q)-specific peptide GPCRs, Activin B, Activin, Activin beta B
striated muscle cell differentiation	1.874E-07	7.196E-05	Galpha(i)-specific peptide GPCRs, MyHC, CACNA2D, Galpha(q)-specific peptide GPCRs, MRLC, OBSL1, MYH3, BARX2, IGF-1, MEF2, MEF2C, HEY2
GeneGO - process networks	**P-value**	**FDR**	**Modulated genes**
Signal transduction_Neuropeptide signaling pathways	2.651E-04	2.015E-02	Galpha(i)-specific peptide GPCRs, Galpha(q)-specific peptide GPCRs, CT, CGRP, Calcitonin
Reproduction_Feeding and Neurohormone signaling	1.324E-03	3.447E-02	Galpha(i)-specific peptide GPCRs, CT, CGRP, Calcitonin
Development_Skeletal muscle development	1.470E-03	3.447E-02	Galpha(i)-specific peptide GPCRs, MyHC, SP-B, Galpha(q)-specific peptide GPCRs, GBR1, IGF-1, CT, CGRP, Calcitonin
Cardiac development_Wnt_beta-catenin, Notch, VEGF, IP3 and integrin signaling	1.814E-03	3.447E-02	Galpha(q)-specific peptide GPCRs, CT, CGRP, Calcitonin
Cardiac development_BMP_TGF_beta_signaling	4.790E-03	7.281E-02	CT, CGRP, Calcitonin
Cardiac development_FGF_ErbB signaling	5.884E-03	7.454E-02	MyHC, Coagulation factor V, Integrin, ITGAX, Galpha(q)-specific peptide GPCRs, Activin B, Activin, Activin beta B, IGF-1, CT, CGRP, Calcitonin
Development_Ossification and bone remodeling	1.329E-02	1.443E-01	Galpha(i)-specific peptide GPCRs, Coagulation factor V, CT, CGRP, Calcitonin
Muscle contraction	1.837E-02	1.642E-01	Galpha(i)-specific peptide GPCRs, HNMT, Integrin, Galpha(q)-specific peptide GPCRs, CT, CGRP, Calcitonin
Cytoskeleton_Actin filaments	1.945E-02	1.642E-01	Galpha(q)-specific peptide GPCRs, Activin B, Activin, Activin beta B, IGF-1
Cytoskeleton_Regulation of cytoskeleton rearrangement	2.210E-02	1.679E-01	Galpha(i)-specific peptide GPCRs, Coagulation factor V, CT, CGRP, Calcitonin

Table VI. 10 - Clinical information of patients included in current study (CRC, polyp and inflammatory bowel disease samples) and their characteristics. Tissues and plasma samples were collected at S.Orsola-Malpighi Hospital (Bologna-Italy).

Clinical features (Polyp)	**N=5**	**Percentage (%)**
Age at initial diagnosis (years)		
Median age	60 years	
Range	50-84 years	
Gender		
Male	1	20%
Female	4	80%
Clinical features (CRC)	**N=19**	**Percentage (%)**
Age at initial diagnosis (years)		
Median age	72 years	
Range	43-84 years	
Gender		
Male	8	42.1%
Female	11	57.9%
Stage		
II	10	52.6%
III	9	47.4%
Grade		
2	16	84.2%
3	3	15.8%
Histological subtype		
Adenocarcinoma	18	94.7%
Mucinous adenocarcinoma	1	5.3%
Clinical features (IBD)	**N=27**	**Percentage (%)**
Age at initial diagnosis (years)		
Median age	44 years	
Range	23-78 years	
Gender		
Male	18	66.7%
Female	9	33.3%
Subtype		
Crohn disease	21	77.8%
Ulcerative colitis	6	22.2%

Table VI. 11 - *N*-glycan composition for each analyzed sample including CRC, normal and polyp tissues. Peak id, GU values, structures composition, average of relative peak area and coefficient of variation (CV) are represented in these tables. In the column of CV values, green and red colors indicate CVs less than 5% and between 5-15%, respectively. Samples with same number (*e.g.* sample 1- Normal and sample 1- Cancer) mean that cancer or polyp tissue and normal corresponding tissue were analyzed for the same patient sample.

Sample 1 - Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
12	3.69	H2N2F1	0.90	4.75
13	4.19	H3N2	1.84	9.69
15	4.62	H3N2F1	2.16	5.33
16	5.01	H4N2	1.28	32.51
17	5.17	H3N3F1	0.39	3.11
18	5.68	H3N4F1	4.64	11.34
19	5.89	H5N2/ H3N3S1	0.53	27.96
20	6.00	H5N2/ H3N5F1	9.95	6.51
21	6.37	H4N5/ H5N2F1	0.73	9.96
22	6.45	H4N4F1	2.53	13.32
23	6.57	H4N4F1	1.42	9.14
24	6.70	H4N5F1	4.43	2.90
25	6.86	H6N2	6.28	11.05
26	7.14	H5N4F1/ H5N5/ H4N3F1S1	1.93	11.15
27	7.35	H5N4F1	1.88	3.72
28	7.44	H4N5F2	1.77	4.52
29	7.50	H5N5F1/ H4N5F2	6.63	1.69
30	7.77	H7N2/ H4N5F2	3.59	2.53
31	7.89	H5N4S1/ H5N4F1S1	4.24	2.69
34	8.24	H5N5F2/ H5N4F1S1	5.97	18.54
35	8.36	H3N6F3/ H5N4F1S2	2.07	18.97
36	8.43	H5N4S2	1.08	15.82
37	8.55	H4N5F3	3.23	9.71
38	8.65	H8N2	3.29	4.67
39	8.85	H5N4S2	14.08	2.50
40	9.01	H5N5F3	4.62	14.32
41	9.21	H5N4F3S1/ H5N5F2S1	1.26	8.51
42	9.35	H5N5F1S2	6.09	9.15
45	10.40	H6N5S3	0.95	7.81
47	10.83	H6N5S3	0.59	13.35

Sample 1 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.69	H2N2F1	1.14	24.59
3	4.19	H3N2	2.00	20.57
6	4.63	H3N2F1	3.58	12.99
7	4.76	H3N3	0.31	48.46
8	5.01	H4N2	1.33	10.67
9	5.08	H4N2	0.56	13.10
10	5.18	H3N3F1/H3N4	1.12	11.74
11	5.26	H3N3F1/H3N4	0.63	14.89
12	5.43	H4N2F1	0.19	13.77
13	5.54	H4N2F1/H4N3	0.19	6.26
14	5.65	H3N4F1/H3N5	3.39	18.98
15	5.69	H3N4F1	2.01	17.23
16	5.90	H5N2/ H3N3S1	0.59	4.88
17	6.00	H5N2/ H3N5F1/H4N3F1	13.75	17.34
19	6.11	H5N2/ H3N4F1/H4N3F1	0.77	6.18
20	6.40	H5N2F1/H5N4F1	1.70	6.09
21	6.46	H4N4F1/H5N3	0.84	8.32
22	6.57	H3N4F2/H4N4F1	1.26	6.08
23	6.71	H4N5F1/H4N3S1	2.32	3.71
24	6.87	H6N2/H4N3F2	8.81	3.26
25	7.00	H5N4	0.50	16.05
26	7.12	H4N4S1/ H4N3F1S1	2.71	6.51
28	7.36	H5N4F1	0.48	9.94
29	7.46	H5N3S1/H4N5S1	0.15	74.64
30	7.53	H4N4F1S1/H5N5F1/H4N5S1	2.41	53.12
31	7.62	H7N2	0.20	63.73
32	7.74	H7N2	1.24	11.25
33	7.79	H7N2	4.52	4.96
34	7.86	H5N3F1S1/ H5N4F1S1/H4N5F1S1	1.64	13.86
35	7.89	H5N4S1/ H5N4F1S1	1.81	12.09
37	8.17	H5N5S1/H5N4F2	0.48	52.52
38	8.27	H5N4F1S1/H6N3S1	2.95	22.68
39	8.37	H3N6F3/ H5N4F1S2/H6N3S1	1.88	55.60
40	8.57	H4N5F2S1/H4N5F3	2.82	14.42
41	8.66	H8N2/H5N4F1S1	5.90	6.64
42	8.86	H5N4S2	8.75	10.50
43	9.05	H5N4F2S1	3.26	9.36
44	9.21	H5N4F3S1/ H5N5F2S1	2.08	9.88
45	9.36	H9N2	6.70	2.06
46	9.50	H5N4F2S2/H6N5S2	0.31	11.78
47	9.73	H5N4F3S2/H5N4F3S1	0.71	14.00
51	10.40	H6N5S3	0.61	8.61
53	10.79	H6N5F3S1	1.36	13.79

Sample 2 -Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.69	H2N2F1	0.59	10.12
3	4.18	H3N2	0.80	8.71
5	4.62	H3N2F1	1.11	3.24
6	4.76	H3N3	0.13	2.92
7	5.01	H4N2	0.49	24.72
8	5.08	H4N2	0.26	9.30
9	5.17	H3N3F1	0.26	4.30
10	5.30	H3N4	0.16	10.89
11	5.64	H3N5	1.32	1.45
12	5.68	H3N4F1	2.21	0.55
13	5.75	H3N3S1	0.14	14.32
14	5.88	H5N2	0.33	11.47
15	6.01	H5N2/H3N5F1	5.51	2.26
16	6.10	H4N4	0.15	7.20
18	6.36	H4N5/H5N2F1	0.45	6.55
19	6.45	H4N4F1/H5N3	1.25	3.16
20	6.57	H4N4F1/H4N3S1	1.03	2.65
21	6.70	H4N5F1/H4N3S1	1.46	4.34
22	6.86	H6N2/H4N5F1	3.58	2.33
23	6.98	H5N4	0.93	5.90
24	7.15	H4N3F1S1/H4N4S1/H5N5	1.17	5.41
25	7.35	H5N4F1	1.87	3.01
26	7.51	H4N5S1/H5N5F1/H4N5F1S1	2.56	1.80
27	7.60	H7N2	0.53	5.06
28	7.78	H7N2/H5N4F1S1	2.54	1.45
29	7.90	H5N4S1/H5N4F1S1	7.00	1.80
30	8.02	H5N4S2	1.13	11.70
31	8.17	H5N5S1	1.11	7.26
32	8.26	H5N4S1F1	2.69	0.56
33	8.37	H3N6F3/H5N4F1S1	2.44	1.52
34	8.44	H5N4S2	0.85	3.44
35	8.50	H5N5F1S1	1.90	12.03
37	8.66	H8N2/H5N4F1S1	3.12	0.89
38	8.74	H5N4F3S1/H5N4S2	0.70	6.08
39	8.85	H5N4S2	26.73	2.99
40	9.00	H6N5S1	0.93	4.20
41	9.14	H5N4F3S1	1.29	2.28
42	9.20	H5N4F3S1	3.37	2.36
43	9.33	H9N2/H6N5F1S1/H5N5F1S2	4.39	1.01
44	9.49	H6N5S2	0.54	9.33
47	9.90	H6N5S2	0.83	10.34
48	10.06	H6N5F1S2	1.85	9.04
50	10.40	H6N5S3	0.79	11.18
51	10.49	H6N5F1S2	1.37	8.37
53	10.83	H6N5S3	2.01	4.45
54	10.94	H6N5S3F1	4.24	2.54

Sample 2 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.69	H2N2F1	5.85	10.75
5	4.19	H3N2	1.12	14.60
6	4.30	H2N3F1	0.14	27.83
8	4.63	H3N2F1	2.38	9.42
11	5.01	H4N2	0.53	14.37
12	5.08	H4N2	0.37	1.61
13	5.18	H3N3F1	0.85	12.63
14	5.30	H3N4	0.12	8.10
15	5.43	H4N2F1	0.10	3.91
16	5.64	H4N3/H3N4F1	0.65	6.62
17	5.68	H3N5/H3N4F1/H3N3F1S1	2.39	22.67
18	5.89	H5N2	0.86	3.58
19	6.00	H5N2/H4N3F1/H3N5F1	4.99	2.19
20	6.10	H4N4	0.31	35.81
22	6.34	H3N5F1/H4N5/H5N2F1	1.33	21.44
23	6.46	H4N4F1/H5N3	1.13	32.42
24	6.58	H4N4F1/H4N3S1	0.63	27.54
25	6.71	H4N5F1/H4N3S1/H3N6F1	1.96	13.78
26	6.87	H6N2/H4N5F1	5.93	2.79
27	6.99	H5N4	0.67	65.30
28	7.12	H4N3F1S1/H4N4S1/H5N5	3.37	3.95
29	7.29	H5N4F1/H4N4F2	0.23	10.91
30	7.36	H5N4F1	1.12	21.15
31	7.47	H4N6F1/H4N5S1/H5N3S1	1.29	14.67
32	7.52	H4N4F1S1/H4N5S1/H5N5F1	1.97	13.14
34	7.74	H7N2	1.26	15.29
35	7.79	H7N2/H4N5F1S1/H5N4S1	2.55	4.30
36	7.91	H5N4F1S1/H5N4S1	8.51	7.49
37	8.02	H5N4S1/H4N6F1S1	0.42	38.61
39	8.19	H5N5S1/H5N6F1	0.93	13.19
40	8.27	H6N3S1/H5N4F1S1	2.62	2.77
41	8.37	H5N4F1S2/H3N6F2	0.99	2.82
42	8.45	H5N4S2	0.80	5.08
43	8.51	H8N2/H5N5F1S1	2.22	1.64
44	8.67	H8N2	4.27	4.38
45	8.75	H5N4F3S1	0.71	10.90
46	8.86	H5N4S2	15.88	8.98
47	9.04	H5N4F2S1/H6N5S1/H5N5S2	1.18	0.80
48	9.21	H5N4F3S1/H5N5F2S1/H5N6F1S1	1.70	14.87
49	9.36	H9N2/H5N5F1S2/H6N5F1S1	8.02	2.43
51	9.56	H6N5F1S1	0.87	12.60
52	9.81	H7N6F1/H6N5F1S2	0.64	14.06
53	9.90	H6N5S2/H6N6F1S1	1.21	15.26
56	10.41	H6N5S3	0.39	2.67
58	10.55	H7N6F1S1	0.84	10.11
59	10.84	H6N5S3	1.39	4.96
60	10.95	H6N5F1S3	1.95	8.02
61	11.23	H6N5F1S3	0.72	7.71

Sample 3 - Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
5	3.69	H2N2F1	0.45	11.80
6	4.19	H3N2	1.78	25.38
8	4.62	H3N2F1	2.43	29.75
9	5.02	H4N2	2.69	13.64
11	5.17	H3N3F1	0.50	23.98
13	5.68	H3N4F1	5.12	10.60
15	5.99	H5N2/ H3N5F1	9.18	5.79
16	6.44	H4N4F1	3.14	11.26
17	6.57	H4N4F1	0.90	6.47
18	6.70	H4N5F1/H4N3S1	1.56	4.16
19	6.85	H6N2	6.34	5.12
20	7.13	H4N4S1/H5N5	1.41	4.89
21	7.34	H5N4F1	2.02	13.34
22	7.50	H4N5S1/H5N5F1/H5N3S1	3.61	0.59
23	7.77	H7N2/H5N4F1S1	3.35	45.98
24	7.88	H5N4S1/ H5N4F1S1	7.32	13.94
27	8.24	H5N5F2/ H5N4F1S1	3.70	10.65
28	8.35	H3N6F3/ H5N4F1S2	2.26	10.92
29	8.42	H5N4S2	2.11	34.68
30	8.51	H5N5F1S1/H5N4S2	1.81	18.82
31	8.64	H8N2	4.67	13.24
32	8.84	H5N4S2	19.71	1.68
33	9.00	H5N5F1S1	1.46	10.04
34	9.19	H5N4F3S1/H5N5F2S1	2.43	7.88
35	9.34	H9N2	8.06	13.56
36	10.39	H6N5S3	2.03	13.20
39	10.82	H6N5S3	0.88	14.27

Sample 3 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
12	3.70	H2N2F1	1.81	11.19
13	4.09	H3N2	0.25	10.09
14	4.22	H3N2	7.97	5.69
17	4.64	H3N2F1	2.41	5.88
20	5.03	H4N2	1.56	19.62
21	5.10	H4N2/H3N3F1	0.16	80.58
22	5.19	H3N3F1	0.48	3.27
23	5.31	H3N4	0.16	11.02
24	5.45	H4N2F1	0.11	0.99
25	5.70	H3N5/H3N4F1	6.30	3.53
26	5.91	H5N2	0.47	2.61
27	6.01	H5N2/H3N5F1	5.60	16.59
29	6.12	H4N4/H4N3F1	0.27	8.49
30	6.36	H3N5F1/H3N3F1S1	0.38	2.62
31	6.44	H4N4F1/H5N2F1	2.82	4.77
32	6.59	H4N4F1/H3N4F2	0.64	3.30
33	6.73	H4N5F1/H4N3S1	1.00	3.36
34	6.88	H6N2	8.28	9.05
35	7.00	H5N4	0.51	36.52
37	7.17	H4N4S1	1.57	26.63
38	7.30	H4N4F2	0.20	74.93
39	7.37	H5N4F1	2.92	1.87
40	7.48	H5N3S1/ H5N4S1/H4N5S1	0.56	15.95
41	7.54	H4N5F1S1/H4N4F1S1/H5N4S1	1.57	15.42
43	7.75	H7N2	0.49	135.97
44	7.80	H7N2/H5N4F1S1/H4N5F1S1	4.42	15.31
45	7.90	H5N4F1S1/H5N4S1	5.81	30.89
47	8.09	H5N4S1	0.56	20.38
48	8.19	H5N5S1	0.42	15.44
49	8.29	H5N4F1S1	2.07	3.14
50	8.39	H5N4F1S2/H3N6F1S1	4.95	8.95
51	8.47	H5N4S2	0.05	30.50
52	8.53	H8N2/H5N5F1S1/H6N5F1	0.41	8.31
53	8.68	H8N2	7.87	2.94
54	8.78	H5N4F1S2/H6N5F1	0.05	25.62
55	8.88	H5N4S2	8.61	9.22
56	9.05	H5N4F2S1/H5N5F1S1	1.09	5.56
57	9.12	H5N4F1S2/H6N5F1S1	0.72	10.99
58	9.24	H5N5F2S1/H5N4F3S1	1.29	35.37
59	9.38	H9N2/H5N5F1S2	10.07	2.75
61	9.66	H5N5F1S2/H5N5F2S1	0.70	7.46
63	9.93	H6N5S2/H5N5F2S1	0.45	3.05
64	10.02	H6N5F1S3	0.71	10.06
67	10.44	H6N5S3	0.95	31.57
69	10.73	H6N5F1S3	0.54	3.08
70	10.87	H6N5S3	0.43	3.90

Sample 4 - Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
13	3.70	H2N2F1	0.36	3.00
16	4.20	H3N2	0.86	3.04
17	4.31	H2N3F1	0.06	18.93
18	4.63	H3N2F1	1.24	3.40
19	5.02	H4N2	0.38	9.11
20	5.09	H4N2/H3N3F1	0.15	4.18
21	5.19	H3N3F1	0.22	1.23
22	5.31	H3N4	0.11	10.54
23	5.44	H4N2F1	0.06	31.75
24	5.69	HN4F1/H3N5	2.32	6.21
25	5.76	H3N3S1	0.12	5.85
26	5.89	H5N2/H3N3S1	0.29	2.54
27	6.02	H5N2/H3N5F1	4.25	5.08
28	6.20	H4N4	0.05	8.82
29	6.37	H5N2F1/H4N5	0.36	25.91
30	6.46	H4N4F1	1.59	5.36
31	6.58	H4N4F1/H4N3S1	1.05	7.31
32	6.71	H4N5F1/H4N3S1	1.71	4.59
33	6.87	H6N2/H4N5F1	3.28	1.79
34	7.00	H5N4	0.78	7.28
35	7.16	H4N4S1/H5N5	1.67	6.57
36	7.36	H5N4F1	2.19	4.40
38	7.51	H4N4F1S1/H5N5F1	3.21	9.15
39	7.61	H5N5F1	0.68	3.68
40	7.79	H7N2/H4N5F2/H4N5F1S1	2.96	0.67
41	7.91	H5N4S1/H5N4F1S1	6.05	1.72
42	8.03	H5N4S1	1.46	2.12
43	8.18	H5N5S1	1.18	0.17
44	8.26	H5N5F2/H5N4F1S1	3.45	1.17
45	8.38	H5N4F1S2/H3N6F1S1	2.09	0.27
46	8.45	H5N4S2/H5N5F1S1	2.45	0.54
47	8.50	H5N5F1S1	0.84	3.98
48	8.56	H4N5F1S1	1.62	3.66
49	8.66	H8N2	2.96	2.00
50	8.75	H5N4F1S2/H5N4S2	0.86	1.85
51	8.86	H5N4S2	21.48	2.67
52	9.02	H5N5F3/H6N5S1	1.79	2.71
53	9.21	H5N4F3S1	4.07	1.92
54	9.35	H9N2/H5N5F1S2/H6N5F1S1	3.65	3.26
55	9.50	H6N5S2	1.22	3.74
57	9.81	H6N5F1S2	0.67	11.77
58	9.90	H6N5S2	1.27	29.24
59	10.05	H6N5S3/H6N5F1S2	1.78	1.84
62	10.42	H6N5S2	2.73	4.02
63	10.52	H6N5S2/H6N5F1S2	1.05	0.30
65	10.71	H6N5F1S2	1.31	14.55
66	10.85	H6N5S2	1.52	4.94
67	10.95	H6N5F1S3	2.07	4.79
68	11.12	H6N5F1S2	0.80	2.11
71	11.60	H7N6S3	0.70	10.48
72	11.94	H7N6S3	0.83	7.09
75	12.43	H7N6S3	0.38	27.83

Sample 4 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.69	H2N2F1	2.45	4.68
4	4.08	H3N2	0.13	9.78
5	4.19	H3N2	1.91	3.69
6	4.31	H3N2F1	0.12	16.72
8	4.63	H3N2F1	4.25	4.63
10	4.77	H3N3/H3N2F1	0.22	3.07
12	5.01	H4N2	0.78	1.90
13	5.08	H4N2/H3N3F1	0.49	3.71
14	5.19	H3N3F1/H3N4	1.09	5.13
15	5.26	H3N4/H3N3F1	0.25	4.74
16	5.30	H3N4	0.21	1.94
17	5.43	H4N2F1	0.26	8.67
18	5.54	H4N3/H4N2F1/H5N2	0.21	3.72
19	5.65	H3N4F1/H3N5	2.16	4.17
20	5.69	H3N4F1	1.89	4.20
21	5.90	H5N2/H3N3S1/H3N4F1	0.66	5.06
22	6.00	H5N2/H3N5F1	8.93	14.47
24	6.10	H4N3F1/H3N5F1/H5N2	0.73	8.70
25	6.31	H3N6F1/H3N5F1	1.69	6.22
26	6.39	H5N2F1/H4N4F1/H4N5	0.43	9.77
27	6.46	H4N4F1/H5N3	1.27	4.22
29	6.58	H4N4F1/H3N4F2	1.04	9.61
30	6.71	H4N5F1/H3N6F1/H4N3S1	1.15	0.42
31	6.87	H6N2/H4N5F1	7.43	13.88
32	6.94	H3N7F1/H6N2/H5N4	2.33	75.88
33	7.15	H4N4S1/H4N6F1/H4N3F1S1	1.49	38.92
34	7.29	H5N4F1/H4N4F2/H3N6F2	0.28	4.01
35	7.36	H5N4F1	2.33	1.30
36	7.46	H4N6F1/H5N3S1	0.60	1.99
37	7.53	H5N5F1/H4N5F1S1/H4N5S1/H4N4F1S1	2.11	1.40
38	7.62	H7N2/H4N5F2	0.43	2.56
39	7.74	H7N2/H4N7F1	0.01	30.44
40	7.79	H7N2/H5N4F1S1/H4N5F1S1	4.52	1.80
41	7.88	H5N4S1/H5N4F1S1	5.09	2.78
42	8.06	H5N4S2	1.01	4.65
43	8.19	H5N5S1	0.72	7.68
44	8.27	H5N4F1S1/H5N5F2	1.67	39.59
45	8.38	H5N4F1S2/H5N4S2/H3N6F3	5.47	12.05
46	8.45	H5N4S2	0.17	13.56
47	8.54	H5N5F1S1/H8N2/H6N5F1	0.21	9.55
48	8.67	H8N2	5.99	1.51
49	8.76	H5N4F1S2	0.06	12.77
50	8.86	H5N4S2	8.85	6.66
51	9.02	H5N5F1S1/H6N5S1	0.76	4.34
52	9.12	H6N5F1S1/H5N4F1S2	0.82	3.73
53	9.21	H5N4F1S2/H5N5S2	0.64	9.10
54	9.36	H9N2/H5N5F1S2/H6N5F1S1	7.42	2.86
55	9.55	H6N5S2	1.36	4.33
56	9.82	H6N5F1S2	0.74	6.62
57	9.91	H6N5S2	0.50	5.25
58	9.99	H6N5F1S3	0.97	8.36
60	10.21	H6N5F1S2	0.50	41.20
62	10.42	H6N5S3	1.16	9.28
65	10.84	H6N5S3	0.62	8.62
66	10.95	H6N5F1S3	0.94	9.90
67	11.11	H6N5F1S3	0.66	27.69

Sample 5 - Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
11	3.69	H2N2F1	0.96	13.70
14	4.18	H3N2	1.89	5.55
16	4.62	H3N2F1	2.60	4.60
17	4.77	H3N3	0.27	21.62
19	5.00	H4N2	0.77	4.10
20	5.08	H4N2/H3N3F1	0.36	1.95
21	5.18	H3N3F1/H3N4/H4N2	0.63	4.28
22	5.29	H3N4	0.36	3.08
23	5.43	H4N2F1	0.11	0.46
24	5.64	H3N5/H3N4F1	1.77	10.18
25	5.68	H3N4F1	4.54	5.17
27	5.88	H5N2/H3N3S1	0.41	13.97
28	5.98	H5N2/ H3N5F1/H4N3F1	3.12	9.15
29	6.02	H3N5F1/H5N2	9.55	10.69
30	6.36	H5N2F1/H4N4F1/H4N5	0.50	63.41
31	6.45	H4N4F1/H4N3F1	2.59	13.51
32	6.56	H4N4F1/H4N3S1	1.79	1.19
33	6.70	H4N5F1	2.06	5.04
34	6.85	H6N2/H4N5F1	5.43	2.64
35	6.99	H5N4/H4N4S1	0.60	6.36
36	7.14	H4N4S1	1.73	6.79
37	7.27	H4N4F2	0.21	20.69
38	7.34	H5N4F1	1.48	3.36
39	7.44	H5N3S1/H4N5S1/H5N5F1	0.86	1.55
40	7.51	H5N5F1/H4N4F1S1	2.74	0.86
41	7.60	H7N2	0.60	9.67
42	7.76	H7N2/H4N5F2/H4N5F1S1	1.45	6.65
43	7.89	H5N4F1/H5N4F1S1	6.67	2.21
44	8.01	H4N5F2S1	1.37	8.95
45	8.16	H5N5S1	0.90	10.13
46	8.24	H5N4F1S1/H5N5F2	2.86	0.33
47	8.35	H3N6F3/H5N4F1S2	2.34	0.16
48	8.43	H5N4S2	1.64	1.42
49	8.49	H5N5F1S1/H5N4S2	2.87	0.11
50	8.64	H8N2	3.90	10.03
51	8.73	H5N4F1S2	0.47	27.98
52	8.84	H5N4S2	14.74	4.27
53	8.99	H5N5F1S1/H5N5S2	1.79	10.84
54	9.19	H5N4F1S2	1.46	51.34
55	9.33	H9N2/H5N5F1S2	5.36	15.82
56	9.48	H6N5S2	0.83	37.25
58	9.87	H6N5S2	0.21	82.10
59	10.04	H6N5S3	1.39	28.62
60	10.39	H6N5S3	1.31	17.24
63	10.82	H6N5S3	0.58	14.84
64	10.93	H6N5F1S3	1.26	3.83

Sample 5 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.68	H2N2F1	1.17	15.50
3	4.07	H3N2	0.10	32.68
4	4.18	H3N2	1.00	12.49
6	4.62	H3N2F1	1.47	17.64
7	4.76	H3N3/H3N2	0.21	10.18
9	5.00	H4N2	0.49	16.58
10	5.07	H4N2/H3N3F1	0.27	15.83
11	5.17	H3N3F1/H3N4/H4N2	0.69	17.28
12	5.29	H3N4	0.30	9.77
14	5.53	H4N3/H4N2F1	0.12	7.00
15	5.64	H3N5/H3N4F1	1.48	11.12
16	5.68	H3N4F1/H3N5/H3N3F1	2.60	17.58
17	5.89	H5N2/H3N4F1/H3N3S1	0.40	12.68
18	6.01	H5N2/H3N5F1/H4N3F1	9.06	13.27
19	6.10	H4N3F1/H3N4F1/H5N2	0.42	20.43
20	6.18	H4N4/H5N2	0.07	0.76
21	6.33	H4N5/H5N2F1	0.37	13.28
22	6.40	H4N3F1/H5N2F1/H4N4F1/H3N2F1	0.49	11.83
23	6.45	H4N4F1/H4N3F1/H5N3	1.27	14.37
24	6.52	H3N4F2/H4N4F1	0.23	2.43
25	6.57	H4N4F1/H4N3S1	1.05	12.19
26	6.70	H4N4F1/H4N3S1/H3N6F1	1.36	13.65
27	6.86	H6N2/H4N3S1/H4N4F1	5.95	7.57
29	7.14	H4N3F1S1/H4N5F1	1.94	0.91
30	7.29	H4N4F2	0.59	7.30
31	7.35	H5N4F1/H4N3F1	3.23	5.60
32	7.45	H5N3S1/H5N4S1	0.50	14.36
33	7.52	H4N4F1S1/H5N5F1/H4N5F1S1/H4N5F2	2.07	3.40
34	7.61	H7N2/H5N3F2	0.59	20.81
36	7.78	H7N2/H5N4F1S1/H4N5F2/H5N3F1S1	4.01	1.57
37	7.87	H5N4S1/H5N4F1S1/H4N5F1S1	5.65	3.26
41	8.16	H5N4F2/H5N3F2S1/H5N5S1/H4N5F2S1	1.32	23.20
42	8.26	H5N4F1S1/H6N3S1	2.23	2.10
43	8.36	H3N6F3/ H5N4F1S2	4.51	1.74
44	8.44	H5N4S2/H5N5F1S1	0.03	15.47
45	8.54	H4N5F1S1/H5N4F2S1/H4N5S2/H5N5F1S1	3.60	3.02
46	8.65	H8N2/H5N4F1S1	5.88	8.39
47	8.74	H5N4F1S2	0.21	163.26
48	8.85	H5N4S2	8.86	6.90
49	8.92	H5N4S2/H4N5F3S1/H5N4F1S1	4.09	12.43
50	9.04	H5N4F2S1	0.97	10.44
51	9.20	H5N4F1S2/H5N5F1S1	1.08	76.19
52	9.34	H9N2	8.61	9.03
53	9.49	H5N5S2	0.85	49.13
54	9.72	H5N4F3S2	3.21	22.88
55	9.87	H6N5S2	0.47	66.02
56	9.99	H6N5S3	0.78	39.83
60	10.40	H6N5S3	1.32	39.86
63	10.82	H6N5S3	1.01	51.94
64	10.94	H6N5F1S3	1.89	36.18

Sample 6 - Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
2	3.68	H2N2F1	1.24	1.55
4	4.07	H3N2	0.03	2.41
5	4.18	H3N2	0.49	2.58
6	4.29	H2N3F1/H3N2	0.23	3.99
7	4.33	H2N3F1	0.08	13.03
9	4.61	H3N2F1/H2N2F1	0.64	5.81
10	4.75	H3N3	0.10	27.30
12	5.00	H4N2	0.21	30.71
13	5.07	H4N2/H3N3F1	0.14	42.23
14	5.17	H3N3F1/H4N2	0.83	5.49
15	5.28	H3N4/H3N3	0.49	14.61
16	5.41	H3N4F1/H4N2F1	0.11	58.31
17	5.48	H3N4F1/H4N2F1/H4N3	0.14	56.12
18	5.67	H3N4F1/H3N3F1	18.03	3.18
19	5.87	H5N2/H3N3S1	0.44	20.30
20	5.97	H5N2/H4N3F1	0.66	13.49
21	6.01	H3N5F1/H5N2/H3N4F1	3.65	4.03
22	6.08	H4N4/ H3N5F1	0.35	12.83
23	6.17	H4N4/H3N3F1S1	0.23	41.37
24	6.28	H3N3F1S1	0.14	68.01
25	6.35	H4N5/H4N4/H4N3S1/H5N2F1	0.34	22.45
26	6.44	H4N4F1/H4N3F1	4.18	3.42
27	6.56	H4N4F1/H4N3S1	1.92	2.83
28	6.69	H4N5F1/H4N3S1	2.04	1.07
29	6.80	H4N5F1/H6N2	0.39	6.78
30	6.84	H6N2/H5N2/H4N5F1	1.68	1.36
31	6.97	H5N4	0.65	11.35
32	7.09	H4N3F1S1	0.36	16.64
33	7.13	H4N4S1/H5N5/H4N3F1S1	0.94	10.72
34	7.34	H5N4F1/H4N4F1S1	1.52	4.61
35	7.50	H4N4F1S1/H4N5F1S1/H4N5S1/H5N5F1	3.28	2.33
36	7.72	H7N2	0.42	3.94
37	7.76	H7N2/H5N4F1S1	0.71	26.71
38	7.81	H4N5F1S1/H5N4S1/H7N2	0.97	4.50
39	7.89	H5N4S1	5.20	7.03
41	8.15	H5N5S1	0.69	9.55
42	8.25	H5N4F1S1/H5N5F1S1	3.20	2.53
43	8.35	H5N4F1S2/H5N4S2/H5N4F1S1	0.91	3.93
44	8.43	H5N4S2/H5N5F1S1	1.39	1.10
45	8.48	H5N5F1S1	2.37	1.16
46	8.64	H8N2/H5N4S2	1.60	1.05
47	8.73	H5N4F1S2/H5N4S2	0.61	6.49
48	8.83	H5N4S2	19.84	4.93
50	9.02	H5N4F2S1/H6N5S1	0.36	63.91
51	9.18	H5N4F1S2	6.76	0.67
52	9.30	H5N5F1S2/H9N2	2.06	8.93
53	9.47	H6N5S2	0.51	18.06
54	9.58	H6N5S2	0.27	43.23
55	9.77	H6N5F1S2	0.10	154.81
56	9.87	H6N5S2/H6N5F1S2	0.39	22.03
57	10.03	H6N5F1S2/H6N5S2	0.69	21.71
58	10.18	H6N5F1S2	0.15	41.87
60	10.38	H6N5S3	2.02	7.45
61	10.51	H6N5S3	0.21	99.95
62	10.67	H6N5F1S3	0.26	49.38
63	10.81	H6N5S3	1.11	7.37
64	10.91	H6N5F1S3/H6N5S3	1.64	6.51
65	11.08	H6N5F1S3	0.28	39.86

Sample 6 - Polyp				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.68	H2N2F1	1.92	7.35
2	3.76	H2N2F1	0.07	13.87
4	4.07	H3N2	0.12	17.83
5	4.18	H3N2	1.36	7.47
6	4.29	H2N3F1	0.09	12.47
8	4.61	H3N2F1	3.42	2.60
9	4.69	H3N2F1	0.07	31.83
10	4.75	H3N3/H3N2	0.20	10.97
11	5.00	H4N2/H3N2	0.56	2.06
12	5.07	H4N2/H3N2	0.33	15.46
13	5.17	H3N3F1/H4N2	1.19	6.73
14	5.25	H3N4/H3N3F1	0.12	25.35
15	5.29	H3N4	0.30	13.43
16	5.42	H4N2F1	0.16	24.64
17	5.53	H4N3/H4N2F1	0.19	20.55
18	5.63	H3N4F1/H3N5/H4N3	0.01	110.30
19	5.67	H3N4F1/H3N5/H3N3F1	4.49	12.56
20	5.88	H5N2/H3N3S1	0.51	22.68
21	5.99	H5N2/H4N3F1/H3N5F1	6.97	0.66
22	6.08	H5N2/H4N4	0.44	17.38
23	6.17	H4N4/H5N2F1/H5N2	0.15	35.04
24	6.30	H3N3F1S1/H3N5F1	1.19	2.01
26	6.44	H4N4F1/H4N3F1	1.30	12.25
27	6.52	H3N4F2	0.37	10.91
28	6.56	H4N4F1/H3N4F2/H4N3S1	0.87	9.59
29	6.69	H4N5F1/H4N3S1	1.15	2.81
30	6.85	H6N2/H5N2	6.43	3.97
31	6.98	H5N4/H3N5F2/H6N2	0.90	5.99
32	7.11	H4N3F1S1/H4N5F1	2.45	2.77
33	7.28	H4N4F2/H3N6F2/H5N4F1	1.11	5.29
34	7.34	H5N4F1	1.50	1.00
35	7.44	H5N3S1/H4N5S1	0.50	110.42
36	7.51	H4N5F2/H4N4F1S1/H5N5F1	3.44	14.94
37	7.60	H7N2/H5N3S1	0.47	6.75
38	7.72	H7N2/H5N4F1	0.01	4.91
39	7.77	H7N2/H4N5F1S1	4.35	2.37
40	7.89	H5N4S1/H5N4F1S1	3.96	5.91
42	8.15	H5N5F1/H4N5S1/H5N4F2	1.51	3.85
43	8.25	H5N4F1S1/H5N5F2	2.35	2.42
44	8.34	H3N6F3/H5N4F1S2/H3N6F2	0.47	12.95
45	8.43	H5N4S2	0.12	6.11
46	8.49	H5N5F1S1/H8N2	0.03	14.26
47	8.54	H4N5F1S1/H4N5F2/H5N5F1S1	5.27	2.21
48	8.64	H8N2	6.25	1.68
49	8.84	H5N4S2/H5N4F1S1	10.95	6.75
50	9.00	H5N5F1S1	2.08	47.80
51	9.10	H5N4S2/H5N5F1S1/H6N5F1S1	0.58	3.31
52	9.19	H5N4F1S2/H5N5F2S1	2.74	5.35
53	9.33	H9N2/H5N5F1S2	7.44	2.05
54	9.61	H5N5F2S1/H5N4F2S1	1.78	6.43
56	9.88	H6N5S2	0.65	7.91
57	10.05	H4N7/H5N5F3S1	2.02	4.86
59	10.39	H6N5S3	1.56	7.61
62	10.81	H6N5S3	0.90	8.12
63	10.92	H6N5F1S3	1.42	8.89

Sample 7 - Normal				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
6	3.71	H2N2F1	0.67	21.60
7	4.21	H3N2	1.41	21.04
8	4.64	H3N2F1	2.10	16.96
9	5.03	H4N2	0.87	38.32
10	5.10	H4N2	0.31	39.66
11	5.20	H3N3F1	0.63	12.51
12	5.32	H3N4	0.23	31.81
14	5.70	H3N4F1/H3N5	3.01	41.61
16	6.02	H5N2/H3N5F1	8.97	19.47
17	6.40	H5N2F1/H4N5/H4N4F1	0.75	14.41
18	6.48	H4N4F1/H5N3	1.48	26.29
20	6.60	H4N4F1	1.12	22.71
21	6.73	H4N5F1	2.73	12.34
22	6.89	H6N2	6.71	17.72
23	7.02	H5N4	0.40	84.66
25	7.17	H5N5/H5N4F1	2.76	10.17
26	7.38	H5N4F1	3.07	1.69
27	7.53	H5N5F1/H4N5S1	6.83	8.23
28	7.65	H7N2	0.71	46.84
29	7.81	H7N2/H4N5F2	5.29	1.62
30	7.91	H5N4S1/H5N5F1/H5N4F1S1	5.63	7.00
32	8.18	H5N5S1/H5N4F2	1.11	39.06
33	8.28	H5N5F2/H5N4F1S1	6.15	30.34
34	8.40	H5N4F1S2/H3N6F1S1	3.40	8.15
35	8.47	H5N4S2/H5N5F1S1	1.52	27.49
36	8.59	H4N5F1S1	3.50	15.11
37	8.69	H8N2	4.40	21.78
38	8.89	H5N4S2	10.41	2.20
39	9.04	H5N5F1S1	3.27	23.56
41	9.24	H5N4F3S1	1.92	33.88
42	9.38	H9N2/H5N5F1S2	5.48	17.88
48	10.45	H6N5S3	1.38	21.69
51	10.89	H6N5S3	0.93	8.99
52	10.99	H6N5F1S3	1.02	7.73

Sample 8 - Polyp				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.68	H2N2F1	1.01	37.42
3	4.08	H3N2	0.12	31.74
6	4.40	H3N2	0.10	14.91
7	4.62	H3N2F1	2.42	30.14
8	5.00	H4N2	0.71	35.92
9	5.07	H4N2/H3N3F1	0.36	38.73
10	5.17	H3N3F1/H4N2	0.50	27.47
11	5.29	H3N4	0.10	12.14
12	5.42	H4N2F1	0.13	16.89
13	5.68	H3N4F1/H3N5	1.97	20.61
14	5.89	H5N2/H3N3S1	0.37	11.24
15	5.99	H5N2/H3N5F1	6.86	15.72
17	6.39	H5N2F1/H4N3F1/H4N4F1/H4N5	1.00	9.33
18	6.45	H4N4F1/H5N3	0.88	8.40
19	6.53	H3N4F2/H4N4F1	0.95	4.89
20	6.70	H4N5F1/H5N4/H4N3S1	1.42	25.12
21	6.86	H6N2	8.13	7.85
22	6.98	H5N4	0.82	3.85
23	7.14	H4N3F1S1/H4N4F2/H5N4F1/H5N5	2.40	0.93
24	7.28	H4N4F2/H6N2F1/H5N4F1	0.58	10.57
25	7.35	H5N4F1	2.53	2.88
26	7.50	H5N5F1/H4N4F1/H4N5S1	4.22	1.88
27	7.61	H7N2	0.64	5.81
28	7.78	H7N2/H4N5F2	6.14	0.81
29	7.89	H5N4S1/H5N4F1S1	4.74	2.34
30	8.02	H5N5F1S1	0.46	6.92
31	8.15	H5N5S1	0.20	8.39
32	8.25	H5N4F1S1/H5N5F2	6.09	6.80
33	8.35	H3N6F3/H5N4F1S2/H3N6F2	3.33	24.16
34	8.43	H5N4S2/H5N5F1S1	0.11	16.97
35	8.49	H5N5F1S1/H4N5F1S1	1.12	95.15
36	8.55	H4N5F1S1/H5N5F1S1/H4N5F2	3.31	30.98
37	8.65	H8N2	7.92	2.94
38	8.85	H5N4S2	8.71	2.86
39	9.02	H5N5F1S1/H5N4F2S1	4.79	5.51
40	9.10	H5N5F1S1/H6N5F1S1	0.02	48.33
41	9.20	H5N4F3S1/H5N5F2S1	2.31	10.50
42	9.34	H9N2/H6N5F1S1	7.65	4.21
44	9.70	H6N5F3/H6N5F1S2	0.76	34.01
45	9.81	H6N5F1S2	0.75	25.78
46	9.89	H6N5S2	0.96	54.89
49	10.40	H6N5S3	1.18	29.47
52	10.83	H6N5S3	0.50	49.40
53	10.92	H6N5S3	0.58	54.23

Sample 9 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
1	3.68	H2N2F1	1.81	1.85
4	4.07	H3N2	0.04	6.55
5	4.18	H3N2	0.74	4.99
6	4.29	H2N3F1/H3N2	0.09	10.86
7	4.62	H3N2F1	1.77	0.87
8	4.75	H3N3	0.11	8.43
9	4.96	H4N2	0.38	6.71
10	4.99	H4N2/H3N2	0.36	9.41
11	5.07	H4N2/H3N3F1/H3N2	0.25	4.48
12	5.17	H3N3F1/H4N2	1.02	5.05
13	5.29	H3N4	0.22	5.83
14	5.44	H4N2F1/H3N3F1	0.70	7.54
15	5.52	H4N2F1/H4N3	0.11	13.70
16	5.67	H3N4F1/H3N5	3.90	3.39
17	5.88	H5N2/H3N3S1	0.55	1.63
18	5.98	H5N2/H4N3F1/H3N5F1	6.06	2.27
19	6.08	H4N3F1/H4N4	0.46	3.38
20	6.17	H4N4	0.26	2.46
21	6.32	H3N5F1/H3N3F1S1	0.52	2.81
22	6.39	H5N2F1	0.46	1.58
23	6.45	H4N4F1/H4N3F1/H3N5F1S1	1.45	4.38
24	6.56	H4N4F1/H4N3S1	1.07	4.34
25	6.70	H4N5F1/H4N3S1	1.26	1.85
26	6.85	H6N2/H4N3F2/H5N2	4.39	0.46
27	6.98	H5N4	1.25	4.42
28	7.11	H4N3F1S1/H4N4F2	1.66	1.42
29	7.29	H4N4F2/H4N3F1S1	1.62	4.25
30	7.34	H5N4F1	1.61	1.69
31	7.44	H5N3S1/H4N5S1	0.93	2.38
32	7.51	H4N4F1S1/H4N5S1	2.28	2.94
33	7.73	H7N2/H4N4F1S1	2.22	1.18
34	7.77	H7N2/H5N4F1S1/H4N5F1S1	1.79	1.26
36	7.90	H5N4S1	6.11	2.12
37	8.01	H5N4S1	0.73	16.40
38	8.07	H4N5F1S1/H5N4S2	0.35	31.80
39	8.16	H5N5S1/H4N5F1S1	0.80	0.27
40	8.25	H5N4F1S1/H6N3S1	2.49	1.03
41	8.36	H5N4F1S2/H3N6F1S1	1.36	2.67
42	8.43	H5N4S2/H5N5F1S1	1.29	2.79
43	8.49	H5N5F1S1/H8N2/H5N4S2/H5N4F1S1	1.07	2.18
44	8.55	H4N5F3/H4N6F1S1	0.81	1.14
45	8.65	H8N2	3.88	0.78
46	8.74	H5N4F3S1/H5N4S2	0.72	1.59
47	8.84	H5N4S2	18.06	2.31
48	9.02	H5N4F2S1/H6N5S1/H5N5F3/H4N5F2S1	1.53	6.37
49	9.19	H5N4F3S1/H5N5F2S1	2.61	2.56
50	9.33	H9N2/H5N4F2S1	4.39	1.01
51	9.43	H5N4F1S2/H6N5S2	1.10	1.77
52	9.61	H5N5F2S1	1.45	2.24
53	9.76	H5N4F3S1/H6N5F1S2	0.82	2.12
54	9.88	H6N5S2	0.97	3.00
55	10.05	H6N5F1S2/H5N5F3S1/H6N5S3	1.45	1.72
56	10.14	H6N5S3	0.54	13.49
58	10.40	H6N5S3	1.43	0.88
59	10.48	H6N5S3/H6N5F1S2	0.98	2.97
61	10.82	H6N5S3	1.46	0.62
62	10.93	H6N5F1S3	2.28	3.64

Sample 10 - Cancer				
UHPLC-FLR-ESI-MS/MS				
Peak id	GU	Composition	Average % Area	CV
2	3.68	H2N2F1	2.33	13.02
3	4.07	H3N2	0.09	7.98
4	4.18	H3N2	1.97	3.31
6	4.61	H3N2F1/H2N2F1	2.60	9.09
9	5.01	H4N2	5.82	26.37
10	5.16	H3N3F1	0.56	4.90
11	5.42	H4N2F1	0.10	8.06
12	5.49	H4N2F1/H4N3	0.13	50.22
13	5.67	H3N4F1/H3N5	3.92	30.09
14	5.71	H3N4F1	4.00	33.75
15	5.88	H5N2	0.69	12.79
16	5.98	H5N2/H4N3F1/H3N5F1	5.31	11.38
17	6.08	H4N3F1/H5N2	0.27	18.30
18	6.17	H5N2F1	0.10	15.09
19	6.28	H3N3F1S1	0.15	21.64
20	6.44	H5N2F1/H4N4F1	2.94	5.73
21	6.56	H4N4F1	1.01	16.65
22	6.70	H4N3S1/H4N5F1	1.11	22.24
23	6.85	H6N2/H5N2/H4N2	5.52	4.90
24	6.97	H5N4	0.52	40.44
25	7.10	H4N3F1S1	1.79	5.84
27	7.34	H5N4F1	3.28	24.27
28	7.45	H4N5S1/H5N3S1/H5N5F1	0.55	3.98
29	7.50	H5N5F1/H4N4F1S1	1.64	7.10
30	7.59	H7N2	0.13	5.95
32	7.77	H7N2/H5N4F1S1/H4N5F1S1	4.08	5.79
33	7.88	H5N4S1/H5N4F1S1	6.18	5.08
36	8.15	H5N5S1/H6N5	0.87	15.79
37	8.25	H5N4F1S1/H6N3S1	4.47	2.04
38	8.35	H5N4F1S2/H3N6F3	0.90	4.99
39	8.43	H5N4S2/H5N5F1S1	0.55	4.66
40	8.49	H5N5F1S1/H5N4S2/H8N2	2.08	1.99
41	8.64	H8N2	4.89	3.70
42	8.74	H5N4F3S1	0.33	12.57
43	8.84	H5N4S2	11.91	9.58
44	9.02	H5N4F2S1/H6N5S1/H5N5S2/H6N5F1S1	1.20	6.35
45	9.09	H5N4F1S2	0.60	8.93
46	9.19	H5N4F1S2/H6N5F1S1	1.88	38.16
47	9.33	H9N2/H5N5F1S2/H6N5F1S1	5.55	18.62
49	9.78	H5N5F1S2	0.39	57.99
50	9.88	H6N5S2	0.68	19.84
51	9.99	H6N5F1S2	0.53	2.07
52	10.04	H6N5F1S2	1.11	42.96
54	10.39	H6N5S3	0.89	5.21
55	10.51	H6N5F1S2/H6N5S3	0.83	12.38
56	10.68	H6N5F1S3	0.41	51.16
57	10.82	H6N5S3	1.26	18.64
58	10.93	H6N5F1S3	2.42	3.32

Sample 11 - Polyp				
UHPLC-FLR-ESI-MS/MS				
Peak id	**GU**	**Composition**	**Average % Area**	**CV**
10	**3.70**	H2N2F1	0.63	16.91
11	**4.20**	H3N2	1.05	14.96
13	**4.64**	H3N2F1	1.85	16.70
14	**5.02**	H4N2	0.54	13.28
15	**5.09**	H4N2/H3N3F1	0.22	13.99
16	**5.19**	H3N3F1	0.54	16.00
17	**5.31**	H3N4	0.22	26.37
18	**5.44**	H4N2F1	0.14	21.77
19	**5.69**	H3N4F1/H3N5	2.95	16.20
21	**5.91**	H5N2	0.27	8.73
22	**6.02**	H5N2/H4N3F1/H3N5F1	7.96	14.05
25	**6.35**	H4N5	0.65	11.25
26	**6.41**	H5N2F1/H4N3F1/H4N5	0.52	1.04
27	**6.47**	H4N4F1/H5N3	0.94	1.84
28	**6.54**	H3N4F2/H4N4F1	1.30	4.92
29	**6.72**	H4N5F1	2.48	1.62
30	**6.87**	H6N2	6.81	7.59
31	**7.01**	H5N4	0.20	6.91
32	**7.15**	H4N3F1S1	3.41	6.84
33	**7.21**	H5N3F1/H4N5F3	1.29	5.37
35	**7.36**	H5N4F1	4.45	11.11
36	**7.47**	H4N5S1/H5N5F1	1.19	12.68
37	**7.52**	H5N5F1/H4N5F1S1/H4N5F1/H4N4F1S1	4.08	18.00
38	**7.62**	H5N3F2	1.55	6.36
39	**7.79**	H7N2/H4N5F2/H5N4F1S1	6.10	0.89
40	**7.88**	H5N4F1S1/H5N4S1/H5N3F1S1	3.64	6.07
41	**7.93**	H5N4S1/H5N4F1S1	1.91	5.34
42	**8.03**	H5N5F1S1/H5N4F1	1.84	6.31
43	**8.10**	H5N4F2/H6N3F1	1.37	14.86
44	**8.18**	H5N5S1/H5N4F2	1.65	7.65
45	**8.27**	H5N4F1S1/H5N5F1S1/H6N3S1	3.22	2.34
46	**8.38**	H3N6F1S1/H5N4F1S2	3.59	8.23
48	**8.56**	H4N5F3/H5N4S2	3.71	36.20
49	**8.67**	H8N2	6.61	1.24
50	**8.87**	H5N4S2/H5N4F1S1	7.42	1.83
51	**9.05**	H5N4F2S1	3.07	17.57
52	**9.23**	H5N4F1S2/H5N5F2S1	2.01	24.28
53	**9.37**	H9N2/H5N5F1S2	7.70	25.46
58	**10.43**	H6N5S3	0.47	98.99
61	**10.86**	H6N5S3	0.30	61.34
62	**10.97**	H6N5F1S3	0.72	34.30