Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

"Oncologia, ematologia e patologia"

Ciclo XXXIII

Settore Concorsuale: 06/A2

Settore Scientifico Disciplinare: MED/04

IMPACT OF GLYCOSYLTRANSFERASES ON THE PHENOTYPE, SIGNALING AND TRANSCRIPTOME OF COLORECTAL CANCER CELL LINES

Focus on the role of glycosyltransferases B4GALNT2 and FUT6 and their cognate Sd^a and sLe^x antigens

Presentata da: Dott.ssa Michela Pucci

Coordinatore Dottorato — Chiar.ma Prof.ssa Manuela Ferracin

Supervisore — Char.mo Prof. Fabio Dall'Olio

Esame finale anno 2021

ABSTRACT

Many different glycoproteins and their tumor-associated carbohydrate antigens serve as engines for cancer progression and represent important biomarkers potentially useful for therapeutic intervention. Among the carbohydrate structures modulated in colorectal cancer, of particular interest are the Sd^a antigen, synthesized by the β1,4-N-actylgalactosaminyltransferase B4GALNT2, and sLe^x antigen, whose last biosynthetic step is mediated mainly by the fucosyltransferase FUT6. Sialyl Le^x antigen is often overexpressed in CRC and is associated with worse prognosis whereas B4GALNT2 and its associated Sd^a antigen are dramatically downregulated in CRC but their role in tumor progression and development is not fully clear.

To identify correlations between B4GALNT2 expression and clinical parameters, "The Cancer Genome Atlas Database" (TCGA), which contains transcriptomic and clinical data of hundreds of patients, was interrogated. Transcriptomic data showed a dramatic down-regulation of B4GALNT2 mRNA in CRC, compared with normal samples. Patients with higher B4GALNT2 mRNA in CRC samples displayed longer survival, suggesting a strong relationship between high expression of B4GALNT2 and lower malignancy. This association was not observed in other malignancies. Analysis of mechanisms regulating B4GALNT2 down-regulation in CRC revealed that methylation can play a relevant role in B4GALNT2 downregulation in colon carcinogenesis as well as miRNA expression. Methylation data of CRC patients from TCGA pointed to reduced methylation of the intronic site as a key factor in accounting for the general reduction of B4GALNT2 expression in CRC. On the other hand, among miRNAs targeting B4GALNT2 miR-204–5p appears to be the most plausible candidate inhibiting the glycosyltransferase in CRC.

In order to clarify the mechanisms linking the B4GALNT2/Sd^a expression level to CRC phenotype, three different CRC cell lines were permanently modified to express B4GALNT2 cDNA: the LS174T cell line, in which the constitutively expressed sLe^x antigen was partially replaced by Sd^a upon B4GALNT2 expression; the SW480/SW620 pair, which were derived from the primary tumor and a lymph node metastasis respectively, both lacking Sd^a and sLe^x antigens.

In LS174T cells, the expression of B4GALNT2 had little or no effect on the capacity to heal a wound or the ability to form clones in standard growth conditions while it significantly reduced the ability to grow in poor adherence conditions and the

expression of ALDH, a well-known marker of stemness. In SW620 cells, B4GALNT2 expression impacted on all the six key aspects of malignancy considered in this study, including proliferation rate, clonogenic ability on solid support, anchorage independent growth in soft agar, spheroids formation, ability to heal a wound and ALDH expression. On the other hand, in SW480 cells the expression of B4GALNT2 left unchanged the proliferation rate and the wound healing ability.

The common feature emerging from the analysis of the three cell models is the ability of B4GALNT2 to reduce characteristics associated with stemness.

To clarify the impact of sLe^x on CRC phenotype, the SW480/SW620 pair were permanently transfected to express FUT6 cDNA. In both cell lines, overexpression of FUT6 and its associated sLe^x antigen boosted the clonogenic ability in standard growth conditions. Conversely, the growth in soft agar and the capacity to close a wound were enhanced only in SW620 cells.

Transcriptome analysis of the cell lines CRC transfected either with B4GALNT2 or FUT6 showed a relevant impact of both enzymes on the modulation of gene expression. A "B4GALNT2 signature" common to all three cell lines identified the down-regulation of the *SPON2* gene, associated with CRC malignancy, as a potential pivotal change.

Overall, current data may help to personalize therapies for CRC patients according to the B4GALNT2 levels and support a causal effect of this glycosyltransferase on reducing malignancy independently of sLe^x inhibition. Additionally, these data indicate a multifactorial nature of B4GALNT2 regulation, with DNA methylation and miRNA expression playing relevant but not exclusive roles.

ABBREVIATIONS

A|

ACTB - β-actin
ADCC - Antibody dependent cellular cytotoxicity
ALDH - Aldheyde dehydrogenase 1
ANOVA - Analysis of Variance
APC - Adenomatous polyposis coli
Asn - Asparagine

B|

Bcl-2 - B-cell lymphoma 2
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
CMP - Cytidine monophosphate
COAD – Colorectal adenocarcinoma
CRC- Colorectal cancer
CSC - Cancer stem cell
C1GalT1 - Core 1 synthase

D|

DEAB - N, N-diethylaminobenzaldehyde (ALDH inhibitor)
DNA - Deoxyribonucleic acid
DOC - Sodium deoxycholate
Dol-P - Dolichol phosphate

E|

EGFR - Epidermal growth factor receptor
EMT - Epithelial to mesenchymal transition
EpCAM - Epithelial cell adhesion molecule
ER - Endoplasmic reticulum
ERβ - Estrogen receptor-β
ESCs - Embryonic stem cells

F|

FACS - Fluorescence-activated cell sorting
FTA - Phosphotungstic acid
Fuc - Fucose
FUT - Fucosyltransferase

G|

Gal - Galactose
GalNAc - N-acetylgalactosamine
GAPDH - Glyceraldehyde 3-phosphate dehydrogenase
GD - Gangliosides
GDP - Guanosine diphosphate
Glc - Glucose
GlcNAc - N-acetylglucosamine
GlcNAcT: GlcNAc-transferase
GlcNAcT-III - N-acetylglucosaminyltransferase 3
GlcNAcT-V - N-acetylglucosaminyltransferase 5
GT - Glycosyltransferase

H|

HBE - High B4GALNT2 expressers
HNPCC - Hereditary nonpolyposis colon cancer

I|

IBD - Inflammatory bowel disease
IGF2R - Insulin growth factor type 2 receptor
IgM - Immunoglobulin M
iPSC - induced pluripotent stem cells

K|

KRAS - Kirsten rat sarcoma viral oncogene homolog

L|

LBE - Low B4GALNT2 expressers
LGR5 - Leucine-rich repeat-containing G protein coupled receptor 5
LLO - Lipid-linked oligosaccharide
LOH - Loss of heterozygosity

M|

Man - Mannose
MAPK - Mitogen activated protein kinase
MGAT - Alpha-Mannoside Beta-1,6-NAcetylglucosaminyltransferase
MLH1- MutL homolog 1
MMR - DNA mismatch repair
mRNA - Messenger RNA
MS - Mass spectrometry
MSH2 - MutS homolog 2
MSI - Microsatellite instability
MSI-H - Microsatellite instability high
MSI-L - Microsatellite instability low
MSS - Microsatellite stable
MYC- Myelocytomatosis viral oncogene

N|

Neu5Ac -N-acetylneuraminic acid or sialic acid

O|

OST - Oligosaccharyltransferase

P|

PBS-T - Phosphate buffer saline with 0.1% Tween-20
PCR - Polymerase chain reaction
PI3K - Phosphatidylinositol 3-kinase
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|

SDS - Sodium dodecyl sulphate
Ser/Thr - Serine/threonine
sLea - Sialyl lewis a
sLex - Sialyl lewis x
SOX2 - (sex determining region Y)-box 2
ST - Sialyltransferase
sT - Sialyl T
STAT3 - Signal transducers and activator of transcription 3
sTn - Sialyl Tn

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

INDEX

CHAPTER I- INTRODUCTION 1

1.Colorectal cancer 1

1.1 Epidemiology 1

1.2 Molecular pathogenesis of colorectal cancer 2

1.3 CRC diagnosis, staging classification and treatment 7

2. Glycosylation 9

2.1 Site-specific structural diversity in protein glycosylation 11

2.2 Cell biology of glycosylation 12

2.2.1 N-linked glycosylation 12

2.2.2 O-linked glycosylation 15

2.3 Glycosylation in cancer 16

2.3.1 β1,6 branching 17

2.3.2 T, Tn and sialyl-Tn antigens 18

2.3.3 Core fucosylation 19

2.3.4 Sialyl Lewis antigens 20

2.3.5 B4GALNT2 enzyme and Sd^a antigen 21

3. Cancer stem cells *25*

4. Epigenetic mechanisms of gene expression regulation 26

4.1 DNA methylation 27

4.2 miRNA 28

CHAPTER II- MATERIALS AND METHODS 31

1.Analysis of TCGA Database 31

3.SMART App 32

5.Cell Lines 32

6.Slot Blot Analysis of Carbohydrate Antigens 34

7.Enzymatic activity 34

8.Doubling Time Assay 35

9.Clonogenic Assay 35

10.Soft Agar Growth Assay 35

11.Tridimensional (3D) Culture 36

12.Wound-Healing Assay 36

13.ALDEFLUOR Assay 37

14.Total RNA extraction 37

15.Transcriptomic Analysis 37
16.Statistical analysis 38
AIMS 39
CHAPTER III – RESULTS 40
3.1 Clinical implications of glycosyltransferases expression in CRC: survey of TCGA database 40
3.2 Oncogenes and tumor suppressor genes expression poorly correlates with COAD patients survival 42
3.3 Clinical implications of B4GALNT2 expression in CRC: Survey of TCGA database 44
3.4 Several glycogenes are differentially modulated in HBE and LBE 49
3.5 Methylation partially controls the expression of *B4GALNT2* 50
3.6 miR-204-5p regulates B4GALNT2 expression in CRC 52
CHAPTER IV – RESULTS 54
4.1 Phenotypic impact of B4GALNT2 expression on colon cancer cells 54
4.2 Impact of B4GALNT2 expression on the transcriptome of LS174T colon cancer cells 59
4.3 B4GALNT2 expression regulates the transcriptional response to 3D culture 65
CHAPTER V – RESULTS 72
5.1 Transfection of SW480 and SW620 with FUT6 and B4GALNT2 cDNAs 72
5.2 Phenotypic changes induced by B4GALNT2 and FUT6 expression 73
5.3 Impact of B4GALNT2 expression on the transcriptome of SW480 and SW620 colon cancer cells 78
CHAPTER VI - DISCUSSION AND CONCLUSIONS 98
TAKE HOME MESSAGES 107
REFERENCES 108

CHAPTER I- INTRODUCTION

1. Colorectal cancer

1.1 Epidemiology

Colorectal cancer (CRC) represents the third most frequently diagnosed type of malignancy and the fourth predominant cause of deaths related to cancer in the world[1,2]. New cases are expected to increase by 60% reaching more than 2.2 million new while deaths will rise up to 1.1 million by 2030[1]. Global patterns are widely divergent but strongly connected to human development, reflecting the implementation of western lifestyles that are associated with elevate risk[3]. Among the lifestyle factors identified as part of westernization process are common the alcohol consumption, a diet with low usage of fruits and vegetables but high amounts of red and processed meats, physical inactivity, obesity, and smoking[4,5,6]. However, the frequency is rising up quickly in those countries with low and middle incomes, especially because of social and economic changes[7]. Globally, only a small proportion of CRC burden is associated with family history due to the low frequency of these cases[8]. In high-income countries the incidence rate has been stabilizing as many factors have been established as protective including the regular use of aspirin, the use of estrogens after menopause and potentially vitamin D assumption as well as enhancements in early detection and prevention[3,9].

CRC survival rates drop-off with ageing and is highly dependent on the stage of the disease at the time of diagnosis: 90% 5-year survival for locally confined CRC (early stage), 70% 5-year survival when cancer is disseminated into the surrounding tissues and in the lymph nodes, and 10% 5-year survival in case of distant metastasis (advanced stage). 5-year survival is about 35% when the stage is unknown[2].

1.2 Molecular pathogenesis of colorectal cancer

Colorectal cancer arises through the cumulative effects of genetic and epigenetic modifications that favor the transition of normal colonic mucosa into invasive cancer[10,11,12]. Different molecular pathways have been demonstrated to contribute to colorectal carcinogenesis pointing to the heterogeneous nature of CRC[13,14]. These pathways are characterized by distinct models of genetic instability, firstly proposed by Fearon and Vogelstein[15]. This model comprehends three important features: first, colorectal neoplasia results from activation of oncogenes and inactivation of tumor suppressor genes due to mutational events; second, at least 5 different mutated genes are required for cancer to develop; third, the addition of different genetic alterations rather than their order is mainly the driving force of the biologic behavior of the disease[16]. Two major carcinogenesis pathway models have been identified in CRC: chromosomal instability (CIN) pathway and microsatellite instability (MSI) pathway[17].

The CIN pathway counts for 85% of colorectal cancer cases. In this model are contemplated alterations of the number and the structure of the chromosomes and mutations in many checkpoint proteins[18].

A crucial mutation affects the tumor suppressor gene *APC* (Adenomatous Polyposis Coli), located on chromosome 5q21[19], which is implicated in the Wnt pathway[20]. *APC* mutations lead to accumulation of cytoplasmic β-catenin which permits its translocation into the nucleus and subsequently the binding to the transcription factors TCF (T-cell Factor)/ LEF (Lymphoid Enhancer Factor) resulting in alterations of the expression of genes involved in proliferation, differentiation, migration, adhesion and apoptosis of colon cells[20]. APC also controls the cell cycle progression and stabilization of microtubules, potentially contributing to the primary chromosomal instability[19]. These *APC* changes are mainly present in adenomas, precursor lesions leading to benign colorectal cancer[18,15]. The early mutations of CIN pathway are followed by consecutive events that promote new mutations and facilitate the progression of the tumor from benign stages to malignant stages[15]. The transition from adenoma to carcinoma is favored primarily by mutations of *KRAS* (Kirsten rat sarcoma viral oncogene homolog), a proto-oncogene encoding the GTPase protein that regulates the transduction and propagation of extracellular signals. This leads to a permanently active state of the protein that allows the cell to

elude the apoptotic process and obtain a growth advantage[15]. More than 90% of mutations in *KRAS* gene occur at codon 12 and codon 13, conferring a different oncogenic phenotype. In fact, codon 13 mutations are more involved in the adenoma-carcinoma transition whereas codon 12 mutations predispose colorectal cancer cells to local invasion and metastasis[15]. *KRAS* mutations are followed by the inactivation of *TP53* (tumor protein p53), a tumor suppressor gene mediating the cell-cycle arrest, which can be activated by multiple cellular stresses[21]. In most cases, the two *TP53* alleles are inactivated by a combination of a missense mutation that silences the transcriptional activity of p53 and a 17p chromosomal deletion that eliminates the second *TP53* allele. The inactivation of TP53 gene often occurs with the transition from large adenomas to invasive carcinomas[18]. The *TP53* loss is often accompanied by the LOH (Loss of Heterozigosity) of chromosome *18q* (65.4%), where the tumor suppressor genes *SMAD2*, *SMAD4* and deleted in colorectal cancer (*DCC)* genes are located[22]. LOH of *18q* has been correlated with high metastatic potential of cancer cells, resulting in a strongly negative prognosis of CRC patients[23].

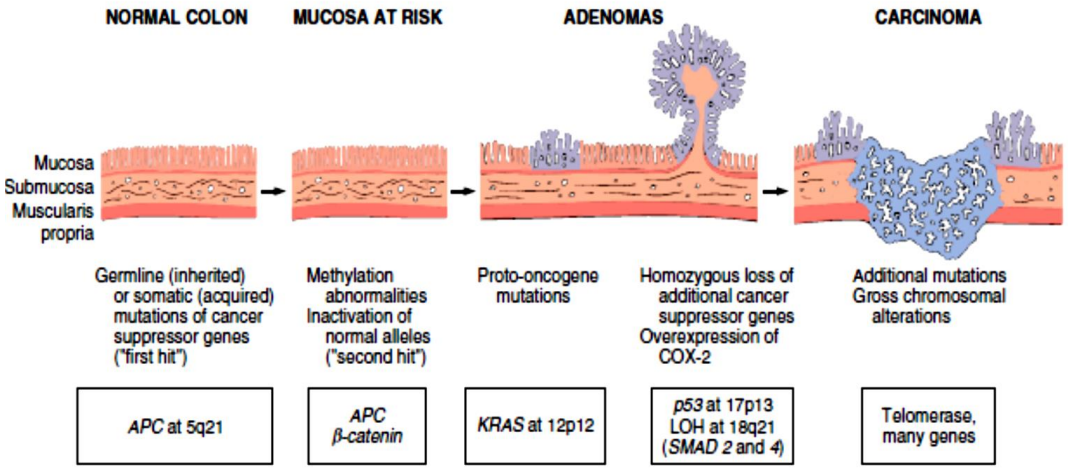


Figure 1. Molecular and morphologic alterations in the adenoma-carcinoma sequence. An early mutation in *APC* gene is followed by additional alterations including *KRAS* oncogene and other tumor suppressor genes such as *TP53*, *SMAD2* and *SMAD4*. Adapted from "Robbins Basic Pathology 9th Edition" (2012) by Kumar, V. *et al.*

The second main CRC pathway is the MSI that represents a form of genomic instability responsible of 15% of sporadic colorectal cancer cases and more than 95% of Hereditary Non Polyposis Colorectal Cancer (HNPCC) syndrome forms[16]. MSI

origins from the inactivation of the DNA Mismatch Repair (MMR) system[23]. The MMR system is comprised of many interacting proteins including the human MutS homologue (MSH) 2 and human MutL homologue (MLH) 1, functioning as a proofing machine that identifies and repairs mismatched nucleotides. This process increases the fidelity of DNA replications. However, the genes *MLH1* and *MSH2* can undergo mutations in the germline leading to instability of short repeat DNA sequences known as microsatellites[15]. Short tandem repeats are small tracts of DNA, distributed through the genome every 30-60 kilobases on the order of hundreds counts and frequently in humans composed of dinucleotide repeats such as (CA)n or (CACACACACAC . . .)[23]. Microsatellite alleles are present in two copies in most individuals. However, during cell replication, often in these areas, errors may occur on DNA strand causing a block of DNA polymerase. The MMR enzymes fix the errors that are missed by the proofreading activity of DNA polymerase preserving the genomic integrity. On the other hand, a defective MMR system will leave the genome with microsatellites that are either longer or shorter than the parent cell and this phenomenon is termed microsatellite instability (MSI). MSI serves as a marker for the loss of DNA MMR function[24]. A panel of 5 microsatellites markers (D2S123, D5S346, D17S250, BAT26 and BAT25) was suggested as a guideline to define MSI level: MSI-H (MSI-H), defined as the presence of instability in 30% of the markers; MSI-Low (MSI-L), defined as the presences of instability in 10%–29% of markers; microsatellite stable (MSS), defined as no unstable markers[23].

Microsatellite sequences can be present in intergenic regions as well as in the coding regions of genes regulating the apoptotic process, cell signaling and cell cycle. Usually, in sporadic MSI cancers mutations in *KRAS* and *TP53* genes occur less frequently than *V600E* mutations in *BRAF* (Murine sarcoma viral oncogene homolog B1) oncogene, a member of the *RAF* family mediating the cellular response to the growth signal through the RAS-RAF-MAP kinase axis[25]. Furthermore, >80% of MSI-CRCs harbor mutations of *TGFBR2* (TGF-β Receptor 2)[23]. *TGFBR2* mutations occur in adenomas either with high-grade dysplasia or progressing to adenocarcinoma, and are common in the late and metastatic passages of MSI-H CRCs. In addition, *SMAD2* and *SMAD4* genes, part of the TGF-β pathway, are frequently mutated in MSI-H CRCs[26]. Loss of function of *SMAD2* participates, independently of *SMAD4*, to deactivation of TGF-β signaling. Mutations of *SMAD4* leads to unregulated growth induced by TGF-β which may contribute to poor

prognosis in CRC[27]. Another alteration originating MSI-H CRCs is the mutation of the 2 polyadenine (A8) tracts in exon 10 of *ACVR2* (activin type 2 receptor). The *ACVR2* gene encodes for a transmembrane receptor whose activation is followed by the phosphorylation of SMAD2 and SMAD3 proteins causing differentiation and growth inhibition. These mutations have been found only in MSI-H CRCs and are frequently associated to *TGFBR2* mutations.

Another mutational target in the MSI-H CRCs is *BAX*, a pro-apoptotic tumor suppressor gene. In 50% of CRCs cases, *BAX* is inactivated by homozygous frameshift mutations that facilitate the escape of cells from intrinsic apoptosis mechanisms. These mutations, like those of *TGFBR2*, can be present in the early stage of neoplastic progressions. MSI-H CRCs have been correlated with a better prognosis than MS stable CRCs, despite the presence of mutations in *TGFBR2* and *BAX* genes.

Additionally, in MSI-H CRC other genes are target of mutations at a lower frequency including the MMR genes *MSH3* (36.5%) and *MSH6* (17.5%), *IGF2R* (Insulin Growth Factor Type 2 Receptor) (22%), *BLM* gene (16%), *PIK3CA* (15%), *PTGS2* (G protein-coupled receptor of Prostaglandin-endoperoxide synthase 2) (33%) and Cyclin D1 gene (*CCND1)* (28%).

Clinically, MSI cancers share specific traits: frequently restricted to proximal colon, undifferentiated with a mucinous phenotype, high number of infiltrating lymphocytes and a better overall survival/prognosis comparing with patients with CIN positive CRC. Morfologically, MSI cancers arise from serrated adenomas or hyperplastic polyps, unlike CIN cancers that originate from the classical adenoma-carcinoma sequence[23].

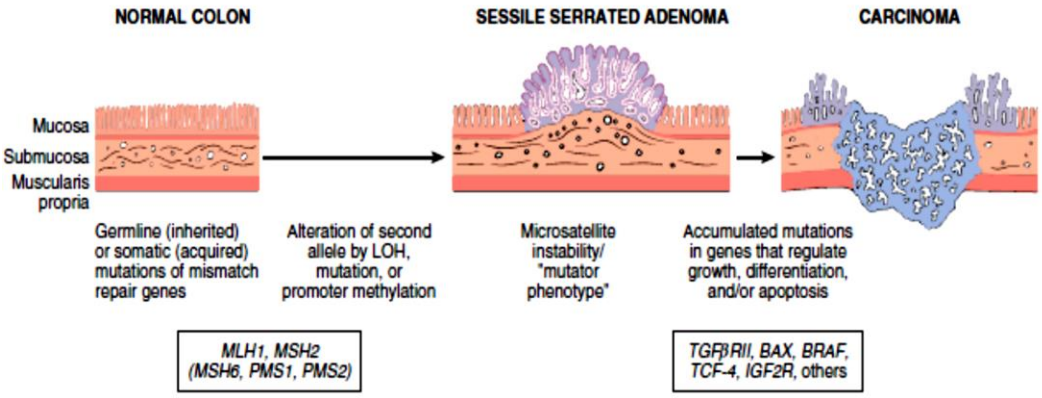


Figure 2. Molecular and morphological alterations in microsatellite instability model. It is characterized by mutations of DNA mismatch repair genes (*MLH1, MSH2*) leading to the accumulation of mutations in short repeat DNA sequences known as microsatellites. These mutations may involve genes controlling cell survival, proliferation and cell growth such as *TGFBR2*, apoptosis regulator *BAX* and oncogene *BRAF*. Adapted from "Robbins Basic Pathology 9th Edition" (2012) by Kumar, V. *et al.*

A distinct molecular subgroup of CRC displays a CpG island methylator phenotype (CIMP), consisting of the aberrant hypermethylation of CpG dinucleotide sequences present in the promoter regions of genes controlling cell cycle regulation, DNA repair, invasion and adhesion, leading to their loss of expression[28]. CIMP is identified in approximately 20%–30% of CRC and is clinically similar to MSI. An early event that is correlated with the progression of histologic grades is the silencing of the *CDKN2A* tumor suppressor gene, encoding cyclin inhibitor *p16*, whose loss of function causes uncontrolled cell proliferation, leading to neoplastic transformation[29].

The CIMP phenotype can be also divided into CIMP-high and CIMP-low groups, depending on the number of methylated markers. The *BRAF* oncogene mutation is often found in CIMP-high CRC and is associated with increase of cell growth, promotion of carcinogenesis, and high mortality. However, CIMP-high tumors, despite *BRAF* mutation, are associated with decreased colon cancer mortality. Noteworthy, *BRAF V600E* mutations occur in 90% of CRC cases with sessile serrated adenoma (SSA) lesions and are never found in the conventional adenomas[25]. In the serrated pathway *BRAF* mutations are early events that leads to a state of dormancy known as senescence. These mutations happen either in early hyperplastic polyps (the serrated precursors) or in the advanced dysplastic serrated polyps. The SSA polyps and the *BRAF* mutation frequently are associated with CIMP-high and

MSI-H features; in sporadic settings, CIMP-high microsatellite unstable CRCs derive from the serrated pathway[29,30].

1.3 CRC diagnosis, staging classification and treatment

Colorectal cancer diagnosis is performed histologically by sampling of areas of the colon suspected for tumor development, typically during colonoscopy. It is confirmed by microscopical examination of a tissue sample. However, the invasiveness of the technique limits the execution resulting in an incomplete exam for some patients. In those cases, an appropriate CRC diagnosis may be realized by Computerized axial tomography scan (CAT scan). This test reveals the presence of metastases of the chest, abdomen and pelvis. Other potential imaging tests such as Positron emission tomography (PET) and Magnetic resonance imaging (MRI) may be used in certain cases. The latter is often used for rectal lesions to determine its local stage and to facilitate preoperative planning[5].

Staging of CRC cancer is referred to TNM (tumor-node-metastasis) system which considers how much the initial tumor has spread and the presence of metastases in lymph nodes and more distant organs[31,32].

TNM classification from the International Union Against Cancer (UICC) (Figure 3A) is based on the depth of local invasion (T stage), lymph node involvement (N stage) and presence of distant metastasis (M stage). Information of these categories is combined into an overall stage definition (stage I, II, III or IV)[5].

Recent technical advances in molecular profiling of tumors have provided gene expression data that enable the molecular subtyping of CRC[33]. Four different Consensus Molecular Subgroups (CMS) have been defined: CMS1, CMS2, CMS3, CMS4 (Figure 3B). CMS1 is characterized by an upregulation of immune genes and is highly correlated with MSI-H. CMS2 represents the canonical pathway of carcinogenesis as defined by the adenoma-carcinoma sequence with a higher expression of the epidermal growth factor receptor (EGFR) and its ligands such as amphiregulin and epiregulin as well as an overexpression of the human epidermal growth factor receptor 2 (EGFR2). CMS3 is identified by metabolic dysregulation with augmented glutaminolysis and lipidogenesis. CMS4 is characterized by the activation of the tissue growth factor (TGF)-β pathway and by the epithelial–mesenchymal transition (EMT) which leads to a general resistance to chemotherapy.[34]

Stage	TNM Classification	Five-Year Survival
		%
I	T1–2, N0, M0	>90
IIA	T3, N0, M0	60–85
IIB	T4, N0, M0	
IIIA	T1–2, N1, M0	25–65
IIIB	T3–4, N1, M0	
IIIC	T (any), N2, M0	
IV	T (any), N (any), M1	5–7

Primary tumor (T)

TX: Primary tumor cannot be assessed

Tis: Carcinoma in situ

T1: Tumor invades submucosa

T2: Tumor invades muscularis propria

T3: Tumor penetrates muscularis propria and invades subserosa

T4: Tumor directly invades other organs or structures or perforates visceral peritoneum

Nodal status (N)

NX: Regional lymph nodes cannot be assessed

N0: No metastases in regional lymph nodes

N1: Metastases in one to three regional lymph nodes

N2: Metastases in four or more regional lymph nodes

Distant metastases (M)

MX: Presence or absence of distant metastases cannot be determined

M0: No distant metastases detected

M1: Distant metastases detected

B

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

Figure 3. TNM and Consensus Molecular Subtype (CMS) classification of colorectal cancer. A: TNM defines the stage T as the invasion depth, stage N in case of lymph node involvement and stage M in case of metastasis[5]. **B**: CMS defines the subtype 1 as MSI immune, subtype 2 as canonical, subtype 3 as metabolic and subtype 4 as mesenchymal.[34]

Histologically, CRC can be also categorized depending on the grade of maintenance of normal glandular architecture and cytological characteristics (well differentiated, moderately differentiated or poorly differentiated)[12]. About 20% of CRC cases are poorly differentiated, presenting a poor prognosis. CRC can present also a mucinous phenotype, characterized by the presence of a prominent intracellular accumulation

of mucins. This cancer features turn out in a very aggressive cancer with a poor prognosis[35].

As regards the treatment, surgery is the most common option, depending on several characteristics of the tumor such as its location, existence and extent of metastasis. If cancer is present only in a single polyp, it can be surgically 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 tumor. The most common drug used in CRC chemotherapy is 5-fluorouracil[4].

2. Glycosylation

Eukaryotic cells are covered on their surface with a dense group of covalently attached sugars (monosaccharides) or sugar chains (oligosaccharides), generically referred to as "glycans." Thus the sugar coat on the cell surface is called "glycocalyx"[36]. The enzymatic process that produces the glycosidic linkages of saccharides to other saccharides, proteins or lipids is known as **glycosylation** and represents not only the most abundant posttranslational modification (PTM), but also by far the most structurally diverse[37]. In nature the spectrum of all glycan structures, defined glycome, is vast. In humans, its size is greater than the number of proteins encoded by the genome, one percent of which encodes proteins that are involved in the synthesis or modification of glycans. Indeed, unlike protein sequences, which are primary gene products, glycan structures are secondary gene products[38]. The main glycans are constituted by ten monosaccharide "building blocks": glucose (Glc), galactose (Gal), N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), fucose (Fuc), xylose (Xyl), mannose (Man), 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. Being localized on the external surface of cellular and secreted macromolecules, many glycans can mediate

or modulate cell–cell, cell–matrix, and cell–molecule interactions crucial for the development and function of a complex multicellular organism, including normal embryonic development, differentiation, growth, cell signaling, intracellular trafficking and localization, disease development, rate of degradation and membrane rigidity[39].

Glycan can also play a role mediating interactions between organisms (e.g., between host and a parasite or a pathogen). Additionally, protein-bound glycans are abundant within the nucleus and cytoplasm, where they can serve as regulatory switches, and undergo to rapidly turnover. The glycans linked to proteins can modulate the intrinsic properties of the modified protein as its solubility, proper folding, functional group orientation and protection from proteases[40].

Glycans are mostly classified according to the nature of the linkage to the aglycone (protein or lipid). Glycoconjugates can be divided in five main groups: glycoproteins, proteoglycans, glycolipids (glycosphingolipids), glycosylphosphatidylinositol (GPI)- linked proteins and O-GlcNAc glycoproteins (Figure 4).

This PhD project is mainly focused on glycoproteins. Glycoproteins are produced through the specific action of glycosyltransferases (GTs), a large family of enzymes that transfer monosaccharide residues to an acceptor substrate using nucleotide-sugar donor as activated donor substrates. GTs are specific for a nucleotide-sugar donor but can recognize more than one different acceptor. They are defined accordingly to the sugar they transfer and the type of linkage they catalyze, being localized mainly in endoplasmic reticulum (ER) and Golgi apparatus (GA). Besides GTs, glycosidases are also implicated in the metabolism of glycoproteins, acting on the hydrolysis of glycosidic linkages[41].

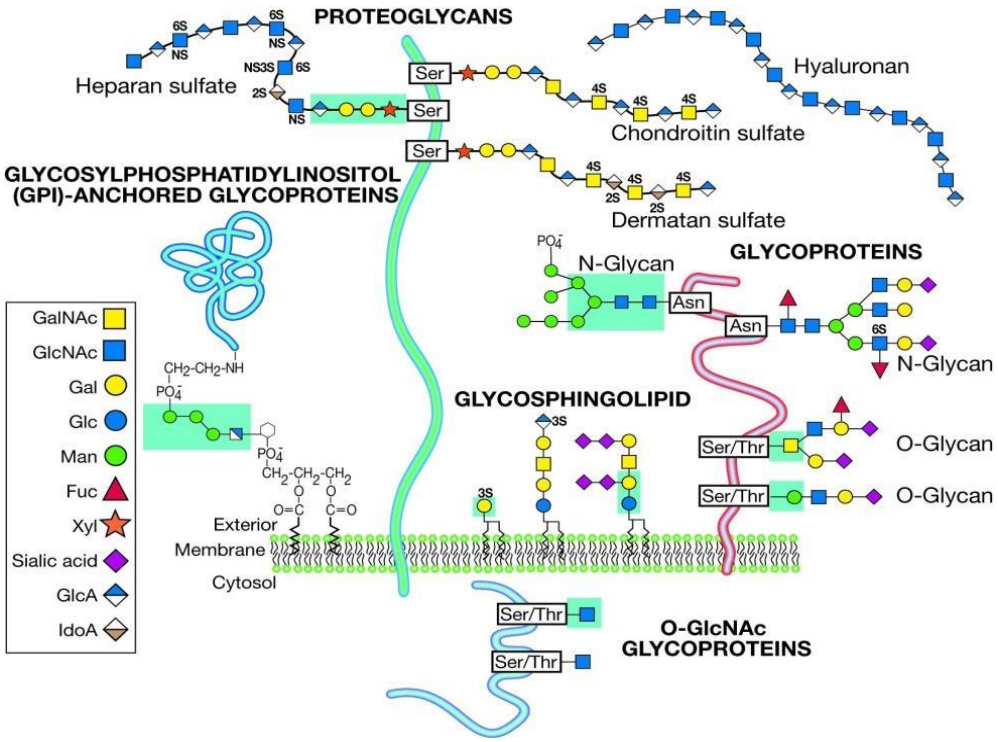


Figure 4. Common classes of animal glycoconjugates. Proteins can be N- linked to Asp (N-glycans) or O- linked to Ser/Thr (O-glycans) of a polypeptide backbone. Proteoglycans are glycoconjugates that present one or more glycosaminoglycans such as chondroitin sulfate, heparan sulfate and keratan sulfate. An exception is the hyaluronan, a glycosaminoglycan found as a free sugar chain. Glycosphingolipids are the most abundant on the cell plasma membrane, made of glycans linked to a lipid ceramide. Glycosylphosphatidylinositol (GPI)-linked proteins are attached in the outer layer of the plasma membrane by a glycan covalently linked to phosphatidylinositol. Several cytoplasmic and nuclear proteins contain O-linked N-acetylglucosamine (O-GlcNAc)[42].

2.1 Site-specific structural diversity in protein glycosylation

A singular aspect of protein glycosylation is the phenomenon of microheterogeneity. Therefore, at any specific glycan attachment site on a protein synthesized by a particular cell type, several variations in the structures of the attached glycan might be found, and in some cases, the glycan may be missing[42]. In fact, a polypeptide encoded by a single gene can exist in different forms defined as "glycoforms," each constituting a distinct molecular species. For some glycoproteins the microheterogeneity at a particular site may be quite restricted, while for other sites it may be wide, even within the same glycoprotein species. Mechanistically, microheterogeneity might result from the rapidity with which multiple, sequential, partially competitive glycosylation and deglycosylation reactions occur in the endoplasmic reticulum (ER) and Golgi system, through which a newly synthesized

glycoprotein passes, along with the lack of a template for directing the synthesis and the accessibility of glycans at a site to the modifying enzymes[42].

2.2 Cell biology of glycosylation

The biosynthesis of major classes of eukaryotic glycans takes place within ER and Golgi compartments. Proteins originating in the ER are either co-translationally or post-translationally modified with glycans at various points along their way to their final destinations[42]. Oligosaccharides are linked to proteins by two main types of linkages. In the first, referred to as N- glycosylation, a GlcNAc residue is linked to the amide side chain of asparagine. In the second, referred to as O-glycosylation, a GalNAc residue is linked to the hydroxyl group of serine or threonine[36].

2.2.1 N-linked glycosylation

N-linked glycosylation is the most studied form of protein glycosylation in eukaryotic organisms (about 90% of eukaryotic proteins carry N-glycans)[43,44].

N-glycosylation biosynthesis (Figure 5) 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[44]. LLO biosynthesis is executed by a set of GTs that are encoded by asparagine linked glycosylation (ALG) genes. The formed glycan is then transferred 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-Ser/Thr tripeptide where X can be any amino acid except proline, that can accept a N-glycan. These steps are followed by a series of reactions trimming N-glycans: trimming of two Glc residues by α-glucosidases I and II originates $Glc_{Man_9GlcNAc_2}$ structure that serves as a ligand for two chaperones, calnexin and calreticulin, helping the protein folding. Misfolded proteins are translocated back to the cytosol to be degraded in the proteasomes[44].

Proteins that are properly folded pass to the Golgi complex where the first process is the demannosylation by the Golgi α-mannosidase I, forming the $Man_5GlcNAc_2$ structure, the main substrate for N-acetylglucosaminyltrasferase-I (GnT-I, product of the *MGAT1* gene). As the pathway progresses through the Golgi complex, the

GlcNAc1Man5GlcNAc2 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 added of Gal, Fuc, sialic acid, and sulfate to the antennae by the action of a number of GTs resulting in a heterogeneous group of mature glycoconjugates[44].

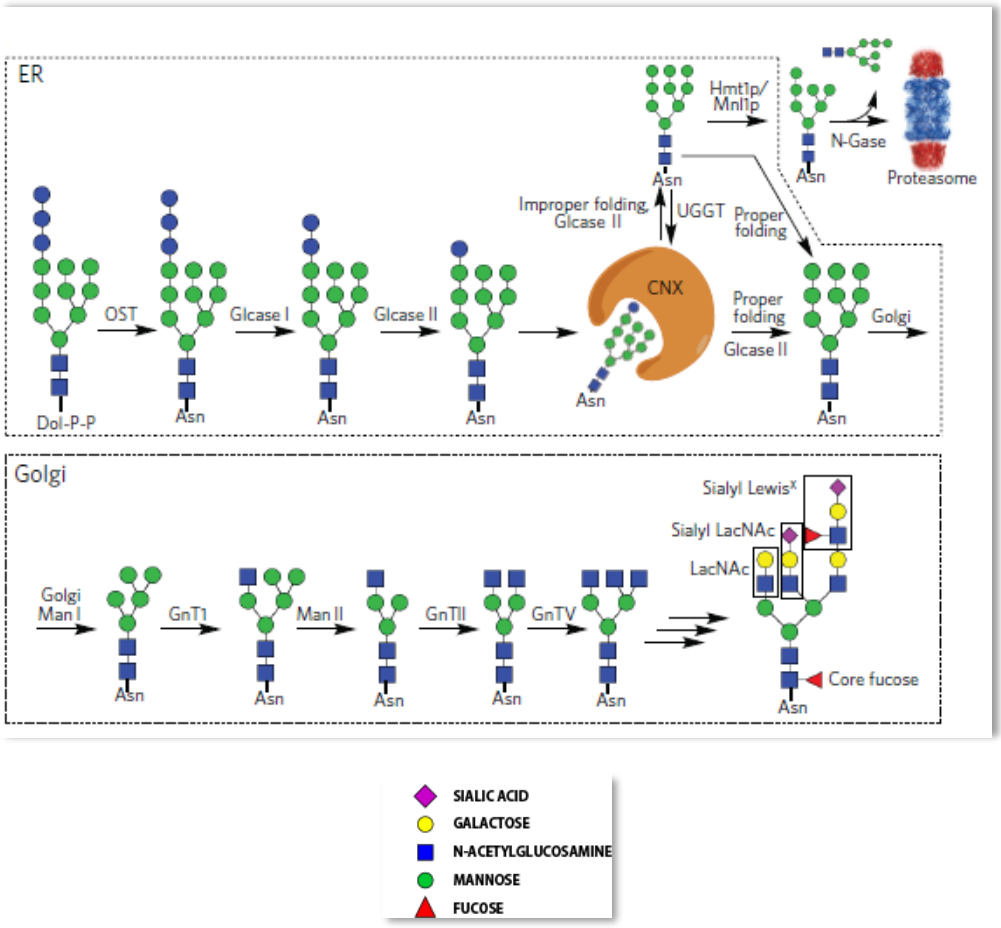


Figure 5. Schematic representation of N-linked glycoproteins biosynthesis. The entire oligosaccharide Glc3Man9GlcNAc2 is transferred to the asparagine residue of a nascent polypeptide chain by the oligosaccharyltransferase (OST) complex. After the action of glucosidase I and II (Glcase I and II), protein glycosylation contributes to the quality control of protein biosynthesis, through the chaperones calnexin (CNX) and calreticulin. Unfolded glycoproteins are moved to cytosol, where a N-glycanase (N-Gase) removes the N-glycans, and lately to the proteasome for degradation. Properly folded proteins are transported to the Golgi, where glycosyltransferases and glycosidases modify the various antennae of the glycans to give more complex structures. UGGT, UDP– glucose-glycoprotein

glucosyltransferase; Hmt1p, HnRNP methyltransferase 1; Mnl1p, mannosidase-like protein 1; Man, α-mannosidase; GnT, N-acetylglucosaminyltransferase[45].

N-glycans share a common sugar core structure (Manα1,6(Manα1,3)Manβ1,4GlcNAcβ1,4GlcNAcβ1-Asn) and are divided in three main types: high mannose, in which only mannose residues are linked 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. They also include some hybrid and complex type glycan determinants (Figure 6): 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 fucosylated structures, where a fucose residue is linked to the first GlcNAc of the chain by the fucosyltransferase VIII[46].

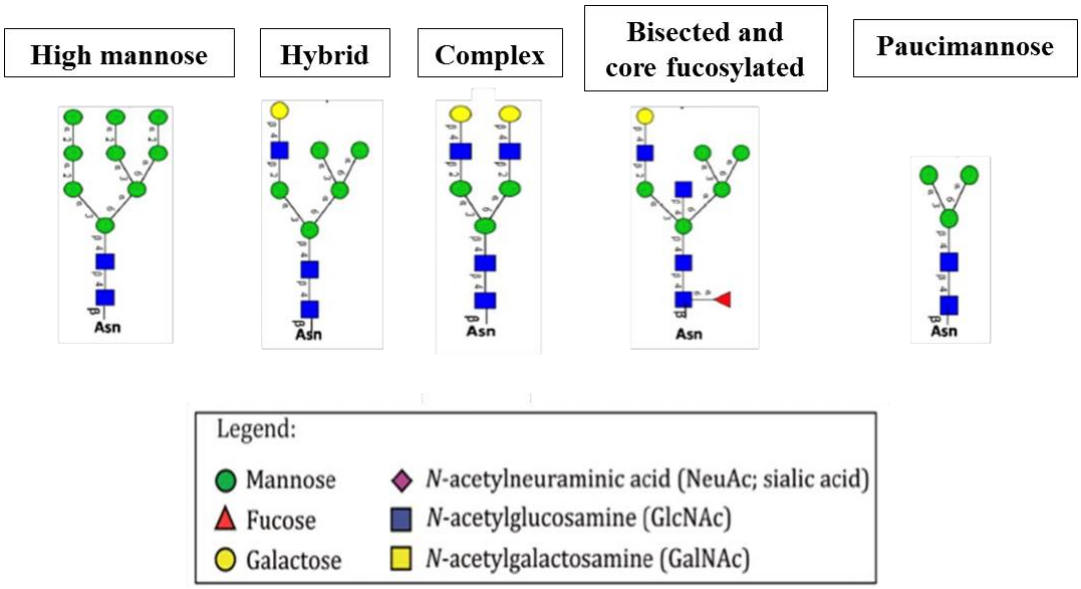


Figure 6. Main types of N-glycans in vertebrates. Types of N-glycans present in a mature glycoprotein: high mannose, hybrid and complex. All types share a common core glycan structure that can be elongated by core fucosylation, bisecting GlcNAc and other glyco determinants. Paucimannose structures are characterized by truncated glycans[46].

The glycosyltransferases involved in the N-glycan synthesis in the ER are mainly multitransmembrane proteins placed in the ER membrane whereas the glycosyltransferases in Golgi compartments are generally type II membrane proteins with a small cytoplasmic amino-terminal domain, a single transmembrane domain,

and a large lumenal domain elongated with a stem region extending from the membrane and a globular catalytic domain. The stem region is often cut off by the action of signal peptide peptidase-like proteases, particularly SPPL-3, leaving the catalytic domain into the lumen of the Golgi apparatus and allowing its secretion. Therefore, many extracellular form of glycosyltransferases are present in tissues and sera[44].

Glycans are regularly turned over by degradation and the enzymes mediating this process cleave glycans either at the outer (nonreducing) terminus (exoglycosidases) or internally (endoglycosidases)[42]. Some terminal monosaccharide units such as sialic acids could be removed and new units added during endosomal recycling, without degradation of the underlying chain. The final complete degradation of most eukaryotic glycans generally takes place in the lysosome through multiple glycosidases. Once degraded, the individual monosaccharide units are exported from the lysosome to the cytosol for reutilization. In contrast to the relatively slow turnover of glycans stemmed from the ER-Golgi pathway, the nuclear and cytoplasmic O-GlcNAc monosaccharide modifications are quite dynamic[42].

2.2.2 O-linked glycosylation

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[47]. They are present at many epithelial surfaces including respiratory, reproductive and gastro- intestinal tracts, playing an important role in the protection against pathogens[48]. Mucin-type O-glycosylation, the most common type of O-glycosylation, involves the attachment of a GalNAc residue to serine (Ser) or threonine (Thr) of a nascent protein[49]. This first step initiates in Golgi apparatus and is controlled by a large family of up to 20 genes (*GALNT1-GALNT20*) encoding polypeptide-N-acetyl-galactosaminyltransferase (ppGalNAcT). Subsequently, O-linked GalNAc residues can be further modified or extended by specific GTs, resulting in several heterogeneous structures[49]. In literature are described eight O-GalNAc glycan core structures, being those from core 1 to core 4 the most common ones (Figure 7).

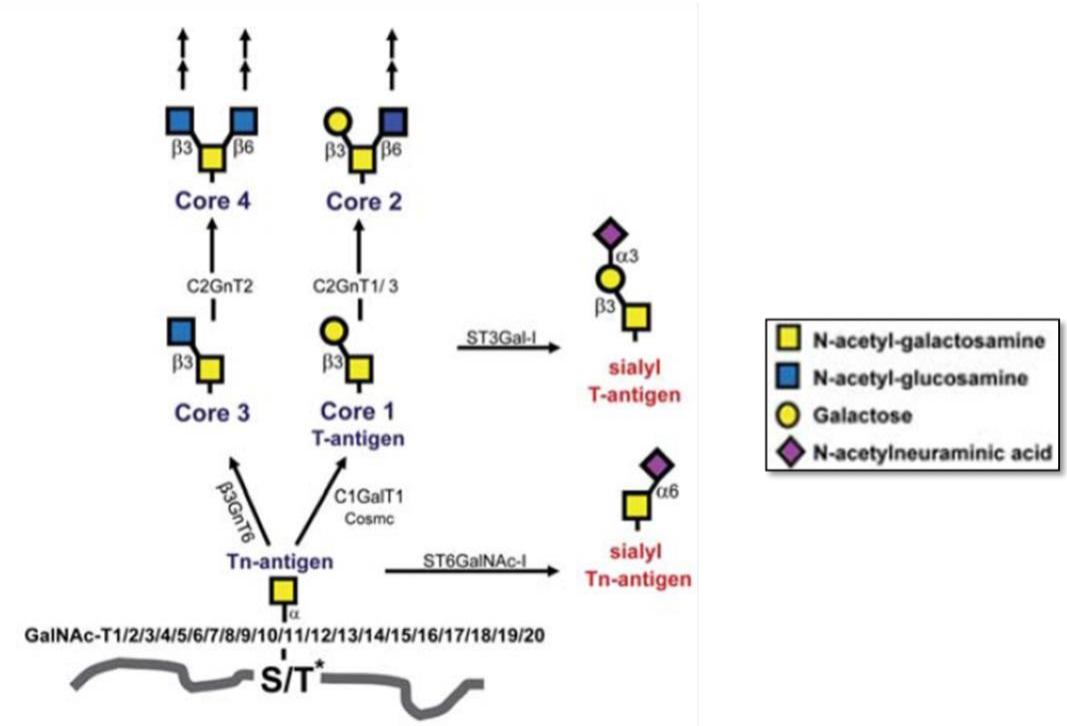


Figure 7. Common O-GalNac glycan cores structures and their biosynthetic pathway. Biosynthesis is initiated by up to 20 ppGalNAcTs forming the Tn antigen, which may be elongated by the core 1 synthase (C1GALT1) or core 3 synthase (B3GNT6), forming the T antigen and core 3 structures, respectively. Both 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 various complex O-GalNac glycans[49]. C2GNT2, core 4 synthase; C2GNT1/3, core 3 synthase; ST3GAL1, β-galactoside α-2,3-sialyltransferase 1; ST6GALNAC1, GalNAc α-2,6- sialyltransferase 1.

2.3 Glycosylation in cancer

Altered glycosylation represents a hallmark of cancer cells, associated with malignant transformation and tumor progression[40]. Some major factors that affect protein glycosylation in tumor cells are: level of expression of specific glycosyltransferases; localization of glycosyltransferases in the secretory compartments and other cellular compartments, such as the nucleus and mitochondria; expression of specific molecular chaperones that regulate protein folding and quality control of glycoproteins and glycosyltransferases; levels of expression of specific glycosidases in the processing pathway; availability of protein substrates and levels of nucleotide sugars; competition between glycosyltransferases for similar glycan acceptors[50]. Numerous glycosylation alterations have been described in cancer including the incomplete synthesis and expression of truncated

glycan structures, increased expression of complex branched N-glycans, *de novo* expression of terminal sialylated glycans, and altered fucosylation[40].

In this section will be discussed some of the most relevant cancer-associated glycosylation changes.

2.3.1 β1,6 branching

One of the most consistently alterations following neoplastic transformation is a shift toward the synthesis and expression of larger Asn-linked oligosaccharides because of the addition of β1,6-linked lactosamine antennae[51]. The increase of β1–6 branching structures is due to the enhanced expression of N-acetylglucosaminyltransferase 5 (GlcNAcT-V)[51] (Figure 8), resulting from the enhanced expression of MGAT5 gene[42]. These structures have been found to play a causative role in tumor growth and metastasis. Studies conducted in MGAT5 deficient mice revealed that growth rate of breast cancer resulted decreased and metastasis formation was almost completely inhibited[52]. Cells derived from animal models lacking Mgat5 expression exhibited increased contact inhibition and substratum adhesion than Mgat5-expressing cells. Nevertheless, the relationship between β1,6-branching and increased growth and metastasis is probably due to more than one mechanism. The sugar chains produced by MGAT5 are distributed on various cell surface molecules, including growth-promoting receptors (such as PDGFR and EGFR) and receptors with arrest and morphogenic activity (such as TGF-βR and CTLA-4), and are used as a ligand by galectin-3 which, consequently, forms a lattice which stabilizes the receptors on the cell surface. However, growth-promoting receptors express an average higher number of N-linked glycans than receptors with arrest/morphogenic activity. As a consequence, the galectin-3-mediated stabilization of membrane receptors favors highly-branched, growth promoting receptors. MGAT5 expression is regulated by the Ras pathway, thus explaining its close association with cancer. However, in many circumstances MGAT5 activity is counteracted by that of a competing enzyme, N-acetylglucosaminyltransferase 3 (GlcNAcT-III), encoded by the gene *MGAT3*, which catalyzes the formation of glycans with a bisecting GlcNAc β1,4-linked to the innermost Man residue of the core. This modification suppresses the processing and

elongation of N-glycans and decreases N-glycan branching structures. In fact, overexpression of GlcNAcT-III in several cancer types inhibits the function of growth factor receptors, reducing cancer metastasis[51].

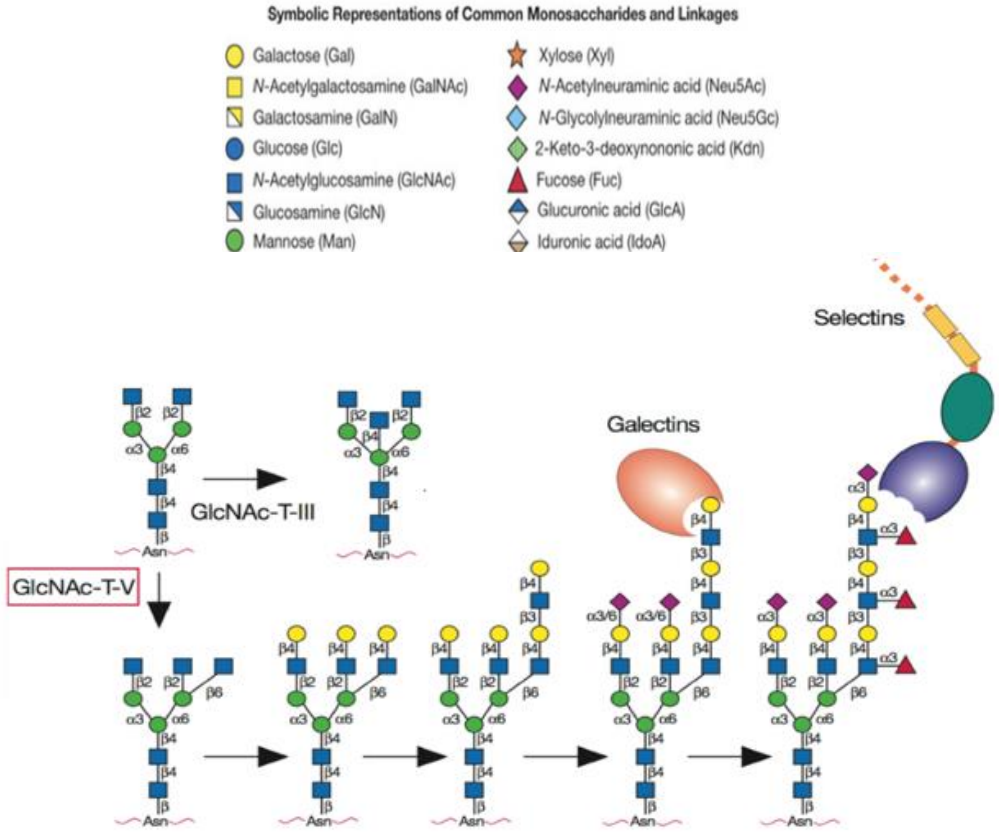


Figure 8. Bisected and branched N-glycan structures. Neoplastic progression is accompanied by an increase of GlcNAcT-V activity that leads to the formation of branched N-glycans. These structures can be further elongated and recognized by galectins and/or selectins, lectins playing an important role in cancer progression. The formation of bisected N-glycans by the expression of GlcNAcT-III is also shown[42].

2.3.2 T, Tn and sialyl-Tn antigens

In many cancers truncation of O-glycosylation pathways leads to expression of simple O-glycans[47]. These truncated glycan structures include the T , Tn and Sialyl-Tn (STn) antigens[50]. The Tn antigen is formed by a GalNAc linked to Serine or Threonine (GalNAcα1-O-Ser/Thr). This sugar can be substituted by α2,6-linked sialic acid, leading to the formation of sialyl-Tn antigen (NeuAcα2–6GalNAcα1-O-Ser/Thr), or by a β1,3-linked galactose, forming the Thomsen-Friedenreich (T) antigen (Galβ1-3GalNAcα1-O-Ser/Thr), or by a β1,3-linked GlcNAc, forming the core 3 structure[53]. The β1,3-galactosyltransferase which synthesizes the T antigen

(T-synthase) is peculiar because it requires the presence of a molecular chaperone, the product of the gene Cosmc that, in the endoplasmic reticulum, binds to T synthase preventing its ubiquitination and degradation in the proteasome[50]. In normal colonic tissues, T antigen is not expressed because it is masked by sialylation but it is highly expressed in colon carcinoma and in liver metastases[54]. It was showed that the presence of cancer cells expressing the T antigen induce the expression of galectin-3 by endothelial cells. This carbohydrate structure interacting with galectin-3 might mediate both the homotypic aggregation of cancer cells and the docking of tumor cells to endothelial cells. The homotypic aggregation protects cancer cells from anoikis induced by the absence of adhesion to extracellular substrates[50]. Inhibition of T antigen with specific antibodies was found to reduce lung metastasis formation by breast cancer cells[55]. However, in breast cancer, besides T antigen accumulation, there is an overexpression of ST3Gal1 which synthesizes sialyl T antigen. In animal models tumor progression was linked to the increased expression of the enzyme, suggesting the possibility that it acts as a tumor promoter[55]. Sialyl-Tn antigen, mainly synthesized by ST6GalNAc1, is expressed by many malignancies, including stomach, liver, pancreas. A general mechanism possibly explaining the over-expression of Tn and sTn antigens in cancer is the somatic inactivation of the gene Cosmc[53]. T and sialyl Tn antigens are carried mainly by a high molecular weight splice variant of CD44 and MUC1 in colon cancer, by MUC2 in gastric cancer and MUC1 in breast cancer[50].

2.3.3 Core fucosylation

Fucosylation is one of the most important types of glycosylation in cancer[56,57]. This reaction is regulated by several enzymes known as fucosyltransferases that may differ between glycoproteins and glycolipids but share the donor substrate, GDP-fucose[56]. The α1-6 fucosyltransferase VIII (FUT8) catalyzes the addition of a fucose to the innermost GlcNAc residue of N-glycan structures resulting in core fucosylation. Increases in core fucosylation have been reported in hepatocellular carcinomas (HCC), specifically, fucosylation of a-fetoprotein (AFP)[58]. Fucosylated AFP is a well-known tumour marker for hepotacarcinomas, but sometimes is also increased in benign liver diseases such as chronic hepatitis and liver cirrhosis. The molecular mechanism underlying the production of fucosylated AFP in HCC is

complicated. The enhancement of FUT8 is insufficient for the production of fucosylated AFP in HCC. A donor substrate, GDP-fucose, is a more important regulatory factor for the fucosylation in HCC[57].

The presence of core fucose in N-glycans also regulates the process of antibody dependent cellular cytotoxicity (ADCC). Indeed, core fucose deletion from human immunoglobulin IgG1 enhances ADCC activity[59]. Additionally, core fucosylation may be involved in the modulation of the activity of growth factor receptors, integrins and cadherins. The absence of core fucosylation inhibits TGF-β/Smad2/3 signaling and epithelial-mesenchymal transition of renal epithelial cells is reduced[60].

2.3.4 Sialyl Lewis antigens

Sialyl Lewis antigens are terminal structures that can be found on N-glycans, O-glycans and in glycosphingolipids. They are frequently overexpressed in carcinomas and the degree of their overexpression has been correlated with tumor progression and poor prognosis in (among others) colorectal, lung and renal cancer[61,62,63]. Sialyl Lewis x (sLe^x) derive from the α1,3-fucosylation of a α2,3-sialylated type 2 chain, while the sLe^a antigen derives from the α1,4-fucosylation of a α2,3-sialylated type 1 chain (Figure 9)[64]. The sLe^a tetrasaccharide is a tumor biomarker detected by the monoclonal antibody CA19.9 widely used for the clinical management of patients with gastrointestinal cancers[65].

The terminal steps for the Sialyl Lewis structures biosynthesis include the action of sialyltransferases (STs) and fucosyltransferases (FUTs). ST3GALs transfer a sialic acid residue in α2,3 linkage to a galactose. While α-2,3 sialylation of type 1 chains can be mediated only by ST3GAL3, sialylation of type 2 chains can be mediated by ST3GAL3, ST3GAL4 and ST3GAL6[64]. FUTs transfer a fucose residue to an acceptor, normally galactose or GlcNAc. There are five 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, interact with selectins expressed by the endothelial cells, regulating the metastatic cascade by favoring the arrest of tumor cells on endothelium[66]. Among all selectins, E-selectin is the major receptor involved in adhesion events during metastasis, although P- and L-selectin can also contribute to that process[67,68]. Indeed, studies in

animal models showed a decrease in tumor metastasis after the inhibition of P-selectin-mediated interactions of platelets with sLe^x/sLe^a antigens present on the surface of cancer cells[69].

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 may be due to other mechanisms[70,71] (see section "B4GALNT2 enzyme and Sd^a antigen").

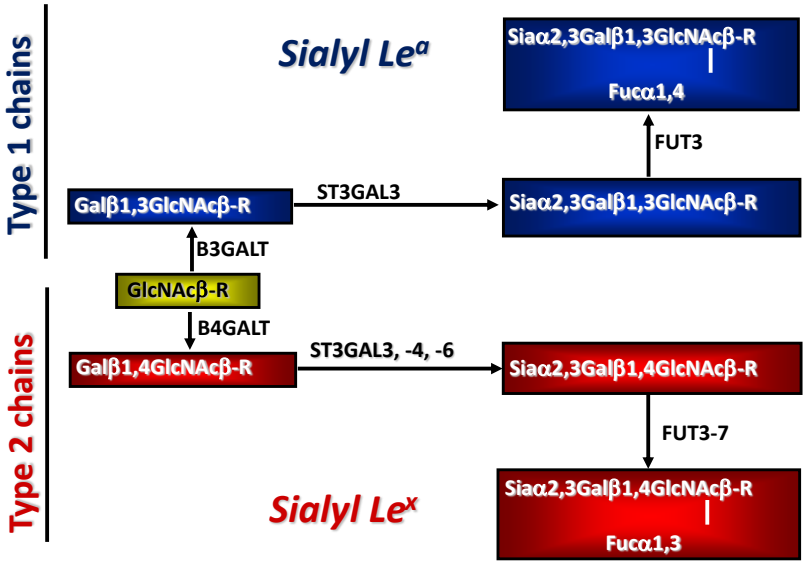


Figure 9. Structures and glycosyltransferases involved in the biosynthesis of Sialyl Lewis antigens. Sialyl Lewis x (type 2 chain structure) and Sialyl Lewis a (type 1 chain structure) are synthetized by sequential enzyme reactions, ending with the action of sialyltransferases and fucosyltransferases[64].

In this project, the focus is on the enzyme B4GALNT2 that synthesizes the Sd^a antigen and indirectly influences the expression of sLe^x antigen, affecting malignant transformation.

2.3.5 B4GALNT2 enzyme and Sd^a antigen

The Sd^a antigen belongs to the "non ABO" histo-blood group system, is expressed on erythrocytes and identified in secretions of almost 95% of individuals with Caucasian origin[72]. It is formed by an α2,3 sialylated type 2 chain to which a GalNAc residue is β1,4 linked to Gal (Figure 10). The enzyme that catalyzes the addition of the GalNAc residue is 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[71]. Several lectins have been used for the study of Sd^a antigen including *Vicia villosa* B4 lectin and *Helix pomatia* lectin[73,74].

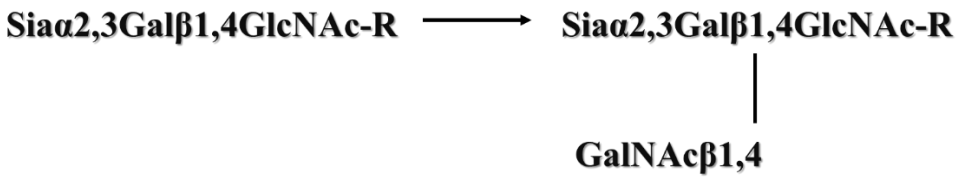


Figure 10. Last biosynthetic step of Sd^a biosynthesis. The addition of β1,4-linked GalNAc residue to Gal, mediated by B4GALNT2, leads to the formation of Sd^a antigen.

The enzyme B4GALNT2 was firstly documented in guinea pig kidney and finally identified in the colon of several species including human, rat and pig[75]. *B4GALNT2* gene maps in chromosome 17 and contains 11 exons. Two main transcripts have been identified in humans, only differing from the exon 1(Figure 11)[71], which exists in a short exon 1 (1S, 38 base pairs) and in a long form (1L, 253 base pairs). These two transcripts harbor a translational start site, 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 LS174T[76,77]. 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[78].

In the genomic sequences upstream of exons 1L and 1S there are CpG islands, thus suggesting that DNA methylation can contribute for the regulation of *B4GALNT2* gene expression[79]. In fact, *B4GALNT2* methylation was found in gastric cancer cases and in the majority of gastric and CRC cell lines[80]. Anti-DNA methylation treatment in cell lines induced a weak expression of B4GALNT2 and corresponding Sd^a antigen[80].

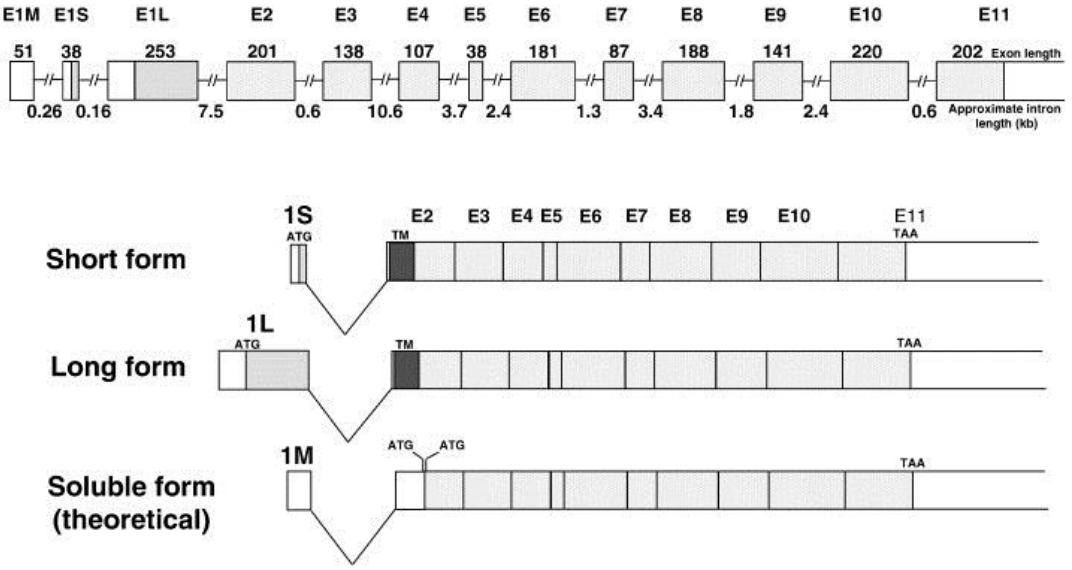


Figure 11. Organization of the human *B4GALNT2* gene and its transcripts. The gene is comprised of 11 exons with at least three alternative first exons: exons 1S, 1L and 1M. Exon length (in bp) is reported above the exons. Numbers below the introns indicate the approximate intron length expressed in kb. The coding regions are shown in gray. The length of exon 11 refers to the coding portion only. The short and long transcript forms derive by the alternative presence of exon 1S or 1L. Both contain a translational start codon and give rise to two transmembrane proteins differing in the amino-terminal portion. The predicted transmembrane domain is presented in dark gray. The "soluble form" is originated by exon 1M missing a translational start codon. Two possible ATG starting codons of this putative transcript are inside exon 2, around the end of the transmembrane encoding sequence[76].

Although only partially understood, the role of B4GALNT2 and Sd^a antigen appears to be wide and different in different tissues and organs. Examples are provided by the regulation of hemostasis in a murine model, by acting on the clearance of the Von Willebrand factor[81], a role in embryo attachment[82] and a role in preventing muscle degeneration in a mouse model of Duchenne muscular dystrophy[83,84]. However, one of the most likely role of the Sd^a antigen is to prevent the cell surface attachment of microorganisms expressing receptors for α2,3-sialylated glycans[83]. 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[85].

In colorectal cancer, B4GALNT2 activity and Sd^a antigen are dramatically reduced compared to normal colon mucosa[86,87]. As described before, the selectin ligand sLe^x antigen is 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[77,88] (Figure 12A). This hypothesis is further strengthened by the fact 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 by a α1,3 fucose linked to a GlcNAc 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[89], with a concomitant reduction of the metastatic ability. In colon specimens, mucins from normal colonic mucosa express high levels of Sd^a and low levels of sLe^x[90]. Since FUT6 is the main fucosyltransferase responsible for sLe^x biosynthesis in colonic tissues and is nearly unchanged in cancer[71], 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 12B)[70]. Accordingly, a significant linear relationship between sLe^x and the FUT6/B4GALNT2 ratio was demonstrated in normal colon but not in cancer[71].

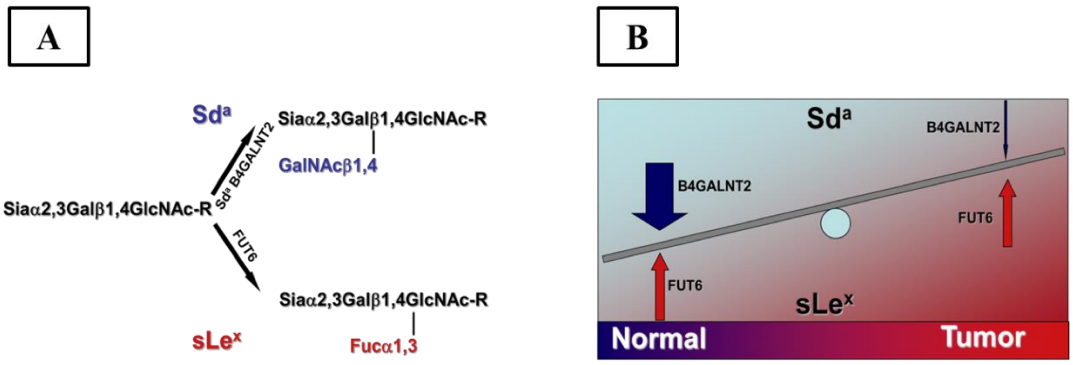


Figure 12. Biosynthetic pathway and expression of Sd^a and sLe^x in CRC. **A**: Biosynthetic pathway of sLe^x and Sd^a antigens, showing the competition between B4GALNT2 and FUT6. **B**: In CRC, the increase of sLe^x expression is not correlated with an increase in fucosyltransferase expression (arrows with same shape) meanwhile the reduced B4GALNT2 expression in cancer tissues (slight arrow) is responsible to change the equilibrium towards sLe^x expression[71].

3. Cancer stem cells

Colorectal cancer as well as other types of tumors is frequently composed of heterogeneous cell types, and tumor initiation and growth are driven by a small subset of cells, termed cancer stem cells (CSC) or tumor-initiating cells[91].

Stem cells are described as a population of cells capable to self-renew indefinitely, form single cell derived populations and differentiate into various cell types[92]. Stem cells found in human tissues are generally multipotent and can originate a restrict number of cell types unlike embryonic stem cells (ESCs) that are pluripotent and can originate any cell type[93,94]. Recently, great emphasis has been given to the discovery that somatic cell can undergo de-differentiation through the expression of specific transcription factors such as Oct4 (octamer-binding transcription factor 4), SOX2 (sex determining region Y box 2), c-Myc and KLF4 (Kruppel-like factor 4)[93]. These reprogrammed cells have been called induced pluripotent stem cells (iPSC) and share many characteristics with ESCs but present also differences, for example different DNA methylation[95]. Stem cells have been recognized also in cancer tissues where they acquire the ability to survive from conventional treatment and escape from the immune system. Therefore, they can cause recurrence of cancer because even few surviving cancer stem cells (CSC) are sufficient to form a new tumor[92]. At present, the mechanism of CSC development is still controversial[96]. One theory suggests that they can be due to oncogenic mutations accumulating within adult stem cells, which leads to their uncontrolled proliferation, retaining stemness. The second theory considers cellular dedifferentiation from a cancer cell into a stem-like state[97]. Regardless the mechanism of development, CSCs are identified by several universal markers; however, a specific marker has not yet been found common to different cancer types. In colon cancer Oct4, SOX2, c-Myc, and KLF4 (that can dedifferentiate cells in iPSC) together with NANOG (Homeobox protein NANOG) are overexpressed, conferring the cells a stem cell like phenotype[94]. Other most used markers for CSC identification are CD133 and CD44. CD133 (Prominin-1), a five transmembrane domain glycoprotein, is a well-known cell surface marker that is expressed in HSCs and progenitor cell subpopulation[98]. N-linked glycan modification on CD133 regulates its cell surface localization. It has been reported that cells CD133+ isolated from primary CRC were able to originate tumors in mice. CD44, a transmembrane glycoprotein that mediates lymphocyte homing and HA

(hyaluronan)-dependent cell adhesion an present in various cells types, including hematopoietic system, is overexpressed since early events of colorectal cancer development because its expression is regulated by the Wnt pathway, often altered in CRC[99]. CD44 is a sLe^x carrier and a selectin ligand[100]. Single CD44+ cells are able to form tumor spheres with stem cell characteristics that give rise to tumors when injected in mice[99,101].

There are other markers under study for the identification of colorectal CSCs. SALL4 (Sal-like protein 4) controls self-renewal and pluripotency in embryonic stem cells, evidences suggest that it is regulated by the Wnt pathway[102]. Moreover, it has been found in plasma of patients with local CRC[103]. ABCG2 is a member of ATP binding cassette superfamily, therefore involved in drug resistance[104], and is associated with proliferation and maintenance of CSCs, as well as tumor formation[105]. STAT3 is a transcription factor correlated with increased proliferation and invasion of cancer cells[106], its presence has been reported in tumor-initiating CD133+ cells[106]. EpCAM and LGR5 are other potential CSC markers. EpCAM (Epithelial cell adhesion molecule) is principally expressed on tumors of epithelial origin and is enriched in colon tumors relative to normal colon. EpCAM High/CD44+ cells can originate tumors if injected in mice[92]. LGR5 (Leucine-rich repeat-containing G-protein coupled receptor 5) is a newly identified marker whose knockdown causes tumor regression, while its recovery induces tumor growth and recurrence[107]. Finally, aldehyde dehydrogenase 1 (ALDH) is considered a typical marker of normal and cancer stem cells. Its expression is increased in CRC correlating with poor prognosis. In fact, ALDH High cells present CSCs features, such as self-renewal, *in vivo* tumor growth capacity and resistance to chemotherapy[108]. Moreover, it has been reported that patients with high ALDH before chemoradiation present recurrence after surgery, thus ALDH can predict the prognosis of patients receiving post-surgery chemoradiation[109].

4. Epigenetic mechanisms of gene expression regulation

Epigenetic mechanisms act a control system within a cell regulating gene expression and silencing[110]. This control varies between tissues and plays an important role in cell differentiation. Additionally, epigenetic modifications drive the differences in

gene expression between cells, resulting in the unique function of specific cell types[111].

The major epigenetic mechanisms include DNA methylation and miRNA expression[112,113].

4.1 DNA methylation

DNA methylation is an epigenetic mechanism involving the covalent transfer of a methyl group to the cytosine bases of DNA by DNA methyltransferases (DNMTs)[114,115]. The majority of DNA methylation occurs on cytosines that precede a guanine nucleotide or CpG sites. DNA methylation is essential for numerous cellular processes including tissue-specific gene expression regulation, genomic imprinting, and X chromosome inactivation[116]. Importantly, DNA methylation in different genomic regions may exert different influences on gene activities based on the underlying genetic sequence[112]. Within intergenic regions, one of the main roles of DNA methylation is to repress the expression of potentially harmful transposable and viral genetic elements[116]. In the genome sites containing a high density of CpG are denominated CpG islands, stretches of DNA roughly 1000 base pairs. The majority of gene promoters, about 70%, are located within CpG islands. The methylation of CpG islands results in stable silencing of gene expression. Like these sites, the methylation of regions called CpG island shores, located 2 kb far from CpG islands, is highly correlated with reduced gene expression. On the other hand, DNA methylation of the gene body is associated with a higher level of gene expression in dividing cells.[117]

There are two general mechanisms by which DNA methylation inhibits gene expression: first, modification of cytosine bases can inhibit the association of some DNA-binding factors with their cognate DNA recognition sequences; second, proteins that recognize methyl-CpG can stimulate the repressive potential of methylated DNA[28].

DNA methylation in CpG-rich promoters of genes is a common feature of human cancers.[118] Indeed, aberrant hypermethylation of gene promoters is recognized as a major mechanism associated with inactivation of tumor-suppressor genes in cancers, and it is involved in almost all the critical steps of oncogenesis[28]. In colorectal cancers, epigenetic changes in selected genes are tightly related to neoplastic

transformation, and aberrant DNA methylation appears to arise very early in the colon (initially in mucosa of normal appearance), and may be part of the age-related defect in sporadic colorectal cancers[118].

Kawamura *et al.* showed that the CpG islands of the *B4GALNT2* gene encoding the enzyme responsible for the synthesis of the Sd^a structure, were heavily methylated and this methylation was closely correlated with the transcriptional silencing of the *B4GALNT2* gene[80]. In another work Wang *et al.* demonstrated the role of DNA methylation in the promotor region of *B4GALNT2* in the suppression of the Sda gene using as models gastrointestinal cancer cell lines, event that was substantially relieved by treatment with the DNA methylation inhibitor, 5-aza-2'-deoxycytidine (5-AZA-CdR, decitabine)[79].

4.2 miRNA

miRNAs have been described as another important epigenetic mechanism that influences gene expression[110]. MicroRNAs (miRNAs) are a class of endogenous small, noncoding, RNA fragments, 22-nt long, that are processed from larger (80-nt) precursor hairpins by the RNase III enzyme Dicer into miRNA:miRNA* duplexes[119]. One strand of these duplexes associates with the RNA induced silencing complex (RISC), whereas the other is generally degraded. The miRNA–RISC complex can bind to the 3-untranslated region (UTR) of the target mRNA and repress gene expression by inhibiting translation or inducing RNA degradation[120].

RNA interference (RNAi) is one of the processes by which miRNAs regulate gene expression[120]. Each small RNA forms a gene-silencing ribonucleoprotein, specific for the DNA target sequence according to the level of complementarity. Thus, the effect of miRNAs on their target genes is based on the degree of homology between the sequences of the miRNA and the target gene. The homology that controls the specificity of miRNAs is dependent on 6-7 nucleotides that bind to the 3-UTR of their target mRNAs[120]. Many combinations of miRNAs and mRNAs are possible, as miRNAs can either bind completely to a complementary sequence of mRNAs or incompletely because some nucleotides can have a mismatch complement base. The number of miRNA and mRNA pairs is increased because one miRNA can target multiple genes, and one gene can also be targeted by multiple miRNAs[121]. miRNAs regulate numerous physiological processes, including development, epithelial-

mesenchymal transition (EMT), regulation of homeostasis and metabolism[21]. However, many miRNAs appear to be deregulated in many diseases, including cancer. They are involved in cell transformation from normal to malignant status, including in colorectal cancer (CRC)[21]. miRNAs seem to influence genes involved in the initiation and progression of CRC, including some of the known frequently inactivated genes *APC, TGFBR2, TP53, SMAD4, PTEN*, constitutively activated *KRAS* or overexpressed *MYC*[122]. In addition to protein-coding genes and mRNAs, there are also miRNAs that can regulate CRC tumor-initiating cells, such as miR-34a5,6, miR-106b7, miR-1408 , miR-146a9, miR-18310, miR-20010, miR-20310, miR-21511, miR-302b12, miR-32813, miR-36314, miR-37115 and miR-45116. In CRC, miRNAs are thus involved in the regulation of many features of cellular transformation[113]. Furthermore, miRNA deregulation is also correlated to angiogenesis, proliferation and migration of cancer cells in CRC, hence contributing to cancerogenesis and invasion[123]. miR-494, miR-598 and miR-17-3p promote cell proliferation, migration and invasion. miR-106a and miR-7 affect apoptosis of CRC cells or resistance to apoptosis. miR-221 and miR-214 reduce autophagy in CRC cells. miR-192/215 and miR-19b-1 control some metabolic pathways. miR-508 induces the stem like/mesenchymal subtype in CRC by affecting the expression of cadherin CDH1 and the transcription factors ZEB1, SALL4 and BMI1[122]. miR-21-5p seems to have epigenetic effects in CRC by blocking the activation of DNA demethylation. miRNAs can also control genes of signaling pathways in CRC[21]. NF-kB regulates immune response and inflammation processes, and is associated with multiple miRNAs such as miR-150-5p, miR-195-5p and miR-203a in carcinogenesis. miRNAs have shown great clinical value in the diagnosis, treatment and prognosis of CRC[124]. Developing appropriate miRNA biomarkers is essential for early stage CRC diagnosis. The differential miRNAs and miRNA regulated genes were screened in early stage CRC tissues, precancerous lesions and colonic intraepithelial neoplasia by RNA sequencing[125]. miR-548c-5p, miR-548i and miR-548am-5p were found as the most differentially expressed miRNAs with regard to lymph node metastasis. miRNAs are correlated with molecular histological markers, such as Ki-67 and CD34, which is useful to determine cell proliferation and angiogenesis in CRC development[123]. Furthermore, miRNAs could be helpful in CRC treatment permitting also to overcome the resistance to cancer therapy[126]. For instance, miR-214 enhances CRC radiosensitivity by inhibiting autophagy in CRC

cells. The overexpression of miR-143 is related to the oxidative stress and cell death in CRC cells, which might elude resistance of CRC cells to oxaliplatin. miR-195 is able to desensitize CRC cells to 5-fluorouracil (5-FU). Several miRNAs are potential biomarkers for CRC detection[124]. However, no single miRNA alone has been identified as an ideal CRC biomarker up to now. Similarly to other gene or protein cancer markers, some miRNAs are predictive but not specifically for one kind of cancer[125]. For example, miR-18a is reported to be a tumor suppressor by inhibitng *CDC42* in CRC. However, miR-18a is also a candidate biomarker for breast cancer and lung cancer, highly expressed in benign breast samples than normal controls and correlating with poor prognosis in patients with non-small cell lung cancer[126]. Similarly, miR-155 inhibits colorectal cancer progression and metastasis, while it is significantly overexpressed in breast cancer and cervical cancer with potential as a biomarker[125].

CHAPTER II- MATERIALS AND METHODS

1. Analysis of TCGA Database

Gene expression data and clinical information for 623 colorectal adenocarcinoma samples and 51 normal colonic tissues were downloaded from the TCGA database using the Firebrowse website. RNA-Seq by Expectation Maximization (RSEM)-normalized data for the colon adenocarcinoma (COAD) cohort were matched with clinical data from the Clinical Pick Tier1 archive. B4GALNT2 mRNA expression was compared with stage, microsatellite stability (MS) status, response to treatment, histological type, and survival. Since the samples did not present a normal distribution of B4GALNT2 expression, non-parametric statistical tests were used. The Mann–Whitney test was used to analyze the difference of B4GALNT2 expression in normal and tumor tissues of mucinous vs. non-mucinous histological type. The Kruskal–Wallis test was used to evaluate B4GALNT2 mRNA expression across cancer stages and MSS/MSI groups. Identification of highly expressed genes in the high and low B4GALNT2 expressers was performed through two-way ANOVA and Bonferroni's multiple comparison test.

2. OncoLnc Database

OncoLnc is a tool for identifying survival correlations, and for downloading clinical data coupled to expression data for mRNAs, miRNAs, or long noncoding RNAs (lncRNAs)[127]. OncoLnc contains survival data for 8,647 patients from 21 cancer studies performed by The Cancer Genome Atlas (TCGA), along with RNA-SEQ expression for mRNAs and miRNAs from TCGA, and lncRNA expression from MiTranscriptome beta. OncoLnc analyses include Cox regression results as well as mean and median expression of each gene. For the Cox regression results, in addition to p-values, OncoLnc stores the rank of the correlation. The rank is calculated per cancer, per data type.

3. SMART App

SMART (Shiny Methylation Analysis Resource Tool) App is a web-based tool to explore and interpret the DNA methylation data across 33 cancer types from TCGA. The SMART App integrates multi-omics and clinical data with DNA methylation and provides key interactive and customized functions including CpG visualization, pan-cancer methylation profile, differential methylation analysis, correlation analysis and survival analysis for users to analyze the DNA methylation in diverse cancer types in a multi-dimensional manner.

4. CSmiRTar: Condition-Specific microRNA targets database

CSmiRTar (Condition-Specific miRNA Targets) provides computationally predicted targets of 2588 human miRNAs from four most widely used miRNA target prediction databases (miRDB, TargetScan, microRNA.org and DIANA-microT) and implements (i) a tissue filter to search the miRNA targets expressed in a specific tissue, (ii) a disease filter to search the miRNA targets related to a specific disease, and (iii) a database filter to search the predicted miRNA targets supported by multiple existing databases.

5. Cell Lines

In this project, three main cell lines derived from colorectal cancer were used: LS174T (ATCC® Number: CL-188™), SW480 (ATCC® CCL-228TM) and SW620 (ATCC® CCL-227TM). LS 174T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM), SW480 and SW620 cells were cultured in Leibovitz's L-15 medium, 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. LS174T cells were kept in an incubator with a humidified atmosphere of 5% CO2 at 37ºC, SW480 and SW620 cells were kept in culture in absence of CO2 in a humidifier incubator at 37ºC while

The cell line LS174T was established from a stage II colorectal adenocarcinoma in a 58 years old Caucasian female. The cell lines SW480 and SW620 were derived from the primary colorectal adenocarcinoma of a 50 year old Caucasian man, Dukes stage B, and from its lymph node metastasis, Dukes' type C, respectively.

The construction of B4GALNT2 transfectants was reported previously by our group[77]. 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).

LS174T cells were transfected using the calcium phosphate method with either an expression vector for the short form of B4GALNT2 cDNA cloned in pcDNA3 or with the empty vector and selected with 0.4 mg/mL G418. Resistant clones were isolated with cloning cylinders, expanded and screened for B4GALNT2 activity. The procedure generated two B4GALNT2-expressing clones S2 and S11 and the polyclonal negative control Neo population. B4GALNT2 enzymatic activity was measured as the difference between the incorporation of radioactive GalNAc on fetuin and the incorporation of radioactive GalNAc on asialofetuin.

In SW480 and SW620 cell lines FUT6 transfection was performed with a FUT6 pcDNA.1 expression plasmid and an empty pcDNA3.1 for G418 resistance in a 10:1 ratio. B4GALNT2 transfection was performed with the cDNA of the short form of *B4GALNT2* cloned in pcDNA3.1 described above. Mock transfections with empty pcDNA.3.1 to obtain negative control Neo transfectants were set in parallel. Cells were selected with 1mg/mL G418. G418-resistant polyclonal populations, were cloned. Single clones were isolated, expanded and screened for the expression of the Sd^a or sLe^x antigens. For both SW480 and SW620, the following cell populations were used: SW480 or SW620 Neo, which are a negative control formed by mock-transfected, G418-resistant cells; SW480 or SW620 FUT6, which are a pool of three clones, highly expressing FUT6 and sLe^x antigen; SW480 or SW620 B4GALNT2, which are a pool of three clones, highly expressing B4GALNT2 and Sd^a antigen.

6. Slot Blot Analysis of Carbohydrate Antigens

Cells were collected by trypsinization and homogenized in ice cold water. The protein concentration of the homogenates was determined using the Lowry method. Thirty micrograms of protein homogenates were spotted on a nitrocellulose membrane in a final volume of 100 µL using a slot- blot apparatus. A wash with 150 µl of phosphate buffer saline with 0,1% tween-20 (PBS-T) was performed in order to ensure the transfer of all the samples. All successive incubations were done on an orbital shaker. After the passage of all of the liquids underneath, the membrane was incubated for 1 hour at room temperature with 1% BSA in PBS, as blocking solution. Then 3 washes were performed, 5 minutes each, with PBS-T for 1 hour at room temperature with the primary antibody. Next, the membrane was washed 3 times, as described before, for 1 hour at room temperature with the secondary antibody. Membranes were washed again 3 times.

For the B4GALNT2-transfected cells were used an anti-Sd^a KM694 antibody, kindly provided by Kyowa Hakko Kogyo Co. Ltd., Tokyo, Japan, diluted 1:2000 in BSA 0,1% in PBS-T as primary antibody and an anti-IgM conjugated with horseradish peroxide, diluted 1:10000 in 0,1% BSA in PBS-T as secondary antibody.

For the FUT6 transfection, were used an anti-sLe^x diluted 1:500 in BSA 0,1% in PBS-T as primary antibody and an anti-IgG conjugated with horseradish peroxide, diluted 1:10000 in 0,1% BSA in PBS-T as secondary antibody.

The reaction was performed using Westar ηC 2.0 from 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). Densitometric analysis was performed using Kodak 1D software.

7. Enzymatic activity

B4GALNT2 enzyme activity was assessed as the difference between the incorporation of [^{3}H]-GalNAc on fetuin and asialofetuin, as previously described by our group[86]. Briefly, the assay mixture contained in a final volume of 25µL: 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 homogenates as the 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. Subsequently, the samples were boiled for 20 minutes with 1 M HCl. At the end, the samples were resuspended and read with the scintillation counter Guardian 1414 Liquid Scintillation Counter (PerkinElmer) after the addition of 3.5 ml of scintillation liquid.

8. Doubling Time Assay

In 6-well plates, aliquots of 2 × 105 cells for well were seeded in duplicate. After 24 h incubation at 37 ◦C, the cells of 2 wells were harvested and counted. This number of cells was considered T0. Pairs of wells were harvested and counted 48 h later (T48). The doubling time (DT) was calculated by using the following formula: DT = (48) × 0.3/Log (N°cellsT48/N°cellsT0). Statistical analysis was performed by using one-way analysis of variance (ANOVA) and Dunnett's multiple comparisons test.

9. Clonogenic Assay

About 50 cells diluted in 2 mL of complete DMEM or L-15 medium were seeded in triplicate in 6-well plates and incubated at 37°C. After 15 days, the plates were washed with phosphate buffered saline (PBS) and the colonies were fixed and stained for one hour at room temperature with a solution containing formaldehyde 4% and Crystal Violet 0.005% in PBS. Photographs of the wells were taken without magnification and the colonies visible at naked-eye were counted. Statistical analysis was performed using one-way ANOVA and Dunnett's multiple comparisons test.

10. Soft Agar Growth Assay

One milliliter of a 0.5% agar solution in complete DMEM or L-15 was dispensed in each well of a six-well plate and allowed to solidify. On top of this layer of agar, 1

mL of a 0.3% agar solution in complete DMEM or L-15 medium containing 1 × 104 cells per well was dispensed in triplicate. The plates were incubated for two weeks at 37° C in a humidified incubator. To evaluate the number of colonies formed, the plates were fixed and colored for one hour at room temperature with a solution containing formaldehyde (4%) and crystal violet (0.005%) in phosphate buffered saline (PBS, 20 mM phosphate buffer pH 7.5, 0.15 mM NaCl). Pictures of LS174T cells were taken at 4X magnification and colonies were counted. As regards SW480 and SW620 cell photographs of the wells were taken without magnification and the colonies visible to the naked-eye were counted. Statistical analysis was performed using the non-parametric Kolmogorov–Smirnov test for LS174T cells. One-way ANOVA and Dunnett's multiple comparisons test were for the statistical analysis of SW480 and SW620 cells.

11. Tridimensional (3D) Culture

10.000 cells were seeded in six-well plates whose bottoms were coated with 0.5% agar in complete DMEM or L-15 medium. Spheroid growth was monitored every 2–3 days. Owing to their non-adherent condition, it was impossible to quantitate spheroids by counting. Thus, cells were quantitatively collected, pelleted by centrifugation and homogenized. The protein concentration of the homogenate was assessed by Lowry assay and its volume was measured. The number of cells was calculated using the protein concentration of a homogenate obtained from a known number of cells grown in standard conditions as a reference. The statistical analysis was performed using one-way ANOVA and Dunnett's multiple comparisons test.

12. Wound-Healing Assay

The wound-healing assay was performed using Culture-Insert 2 Well (Ibidi). This consists in a 2 well silicone insert with a defined cell-free gap of approximately 500 μm, which gives the possibility to plate the cells in the two wells and to evaluate their ability to close the wound once the insert has been removed. Aliquots of 5 × 10^4 cells were seeded in each well. When the cells reached confluency, the insert was removed and the healing of the wound was measured by taking pictures every 24 h with Nikon Eclipse TS100 inverted microscopy at a 4x magnification and a

Digital C-Mount camera Sony Colour. The area free of cells was measured using the MRI Wound Healing Tool of ImageJ. The statistical analysis was performed using two-way ANOVA and Tukey's multiple comparisons test.

13. ALDEFLUOR Assay

ALDEFLUOR (Stem Cell Technologies) was activated following the manufacturer's instructions and added to 5×10^5 aliquots of cells. Half of the cell suspension was treated with DEAB, a specific ALDH inhibitor used as a negative control. After 45 min at 37° C, cells were washed and suspended in ALDEFLUOR buffer. The fluorescent signal was acquired with a FACSCalibur flow cytometer and Cell Quest Pro software. On a dot plot with FL1 (green fluorescence) on the X axis and side scatter (SSC) on the Y axis, we set the fluorescence of the DEAB sample (negative control) and defined the area for ALDH-positive cells. Cells included in this area were considered ALDEFLUOR-positive.

14. Total RNA extraction

Total RNA extraction was performed according to Chomczynski & Sacchi method[128], 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 running RNA samples on a 1% agarose gel.

15. Transcriptomic Analysis

Transcriptomic analysis of RNA from LS174T Neo and S2/S11 cells grown either in standard 2D conditions or in 3D conditions (as spheroids) and SW480 and SW620 cells transfected with FUT6 or B4GALNT2 and their respective Neo negative controls was performed in duplicate using Agilent whole human genome oligo microarray (G4851A). Statistical analysis was performed using a moderated *t*-test, and the false discovery rate was controlled with the multiple testing correction Benjamini–Hochberg with Q = 0.05. Pathway analysis of differentially expressed

genes was determined using the web-based software MetaCore (GeneGo, Thomson Reuters). Gene function was studied through an extensive literature search.

16. Statistical analysis

The GraphPad Prism 6 software was used to perform statistical analysis, using the different tests described above.

AIMS

1) **Identification of correlations between B4GALNT2 expression and clinical parameters.** This was accomplished through an *in silico* survey of the "The Cancer Genome Atlas Database" (TCGA) which contains mRNA expression values and clinical data of hundreds of cancer specimens and normal tissues.
2) **Study of the mechanisms linking B4GALNT2/Sd^a expression to CRC phenotype.** To this scope three cell models have been transfected with B4GALNT2 cDNA: LS174T cell line, constitutively expressing the sLe^x antigen, in which B4GALNT2 expression leads to both Sda expression and sLex inhibition; SW480/SW620 pair, not expressing sLex, in which B4GALNT2 expression leads to Sda expression but not sLex inhibition.
3) **Evaluation of the impact of sLex expression on the phenotype of colon cancer cells of different malignancy.** It was performed through the transfection of SW480/SW620 pair with FUT6 cDNA. The first cell line is from a primary tumor, the second from a metastasis of the same patient.
4) **Analysis of the impact of glycosyltransferase expression on the transcriptome of colon cancer cells.** This was carried out through microarray analysis of B4GALNT2 and FUT6 transfectants of the three cell lines.
5) **Search for a "B4GALNT2 signature" on the transcriptome common to the three cell lines.** This was done through a bioinformatic comparison of the three B4GALNT2 transfectants and their respective mock-counterparts.
6) **Study of the mechanisms regulating B4GALNT2 expression in CRC.** It was done through data mining and analysis of The Cancer Genome Atlas (TCGA) methylation and miRNA data.

CHAPTER III - RESULTS

Analysis of B4GALNT2 expression in colorectal cancer patients: TCGA data mining

Note: Results presented in this section were taken from the published manuscript:

- Pucci, Michela; Malagolini, Nadia; Dall'Olio, Fabio "Glycosyltransferase B4GALNT2 as a Predictor of Good Prognosis in Colon Cancer: Lessons from Databases" *Int. J. Mol. Sci.* **2021**, 22, 4331

3.1 Clinical implications of glycosyltransferases expression in CRC: survey of TCGA database

To assess the impact of glycosyltransferase expression on colon cancer progression, the survival probability of a TCGA cohort of colon adenocarcinoma (COAD) patients was analyzed as a function of the expression of glycosyltransferases relevant for the biosynthesis of cancer-associated carbohydrate structures. The TCGA survey included the following enzymes: GALNT1, GALNT8, ST6GALNAC1, C1GALT1, ST3GAL1, ST3GAL2, ST6GALNAC2, B3GNT6, GCNT1, ST6GALNAC6, MGAT3, MGAT5, FUT8, B4GALT1, B3GNT5, B3GALT5, ST3GAL3, ST3GAL4, ST3GAL6, ST6GAL1, ST6GAL2, FUT3, FUT4, FUT5, FUT6. In Figure 13 are shown the Kaplan-Meier plots relative to the survival of the patients falling in 15^{th} upper percentile- 15^{th} lower percentile of the expression of each glycosyltransferase. Surprisingly, only patients with higher expression of B4GALNT2 displayed a significant longer overall survival (Figure 13). Within the first 1000 days the two group of patients displayed very similar survival curves, while long-term survivals belonged exclusively to the high-B4GALNT2-expressers. Collectively, these data suggest a substantial effect of B4GALNT2 level in CRC patients, especially its relation with a better prognosis and a better response to treatment.

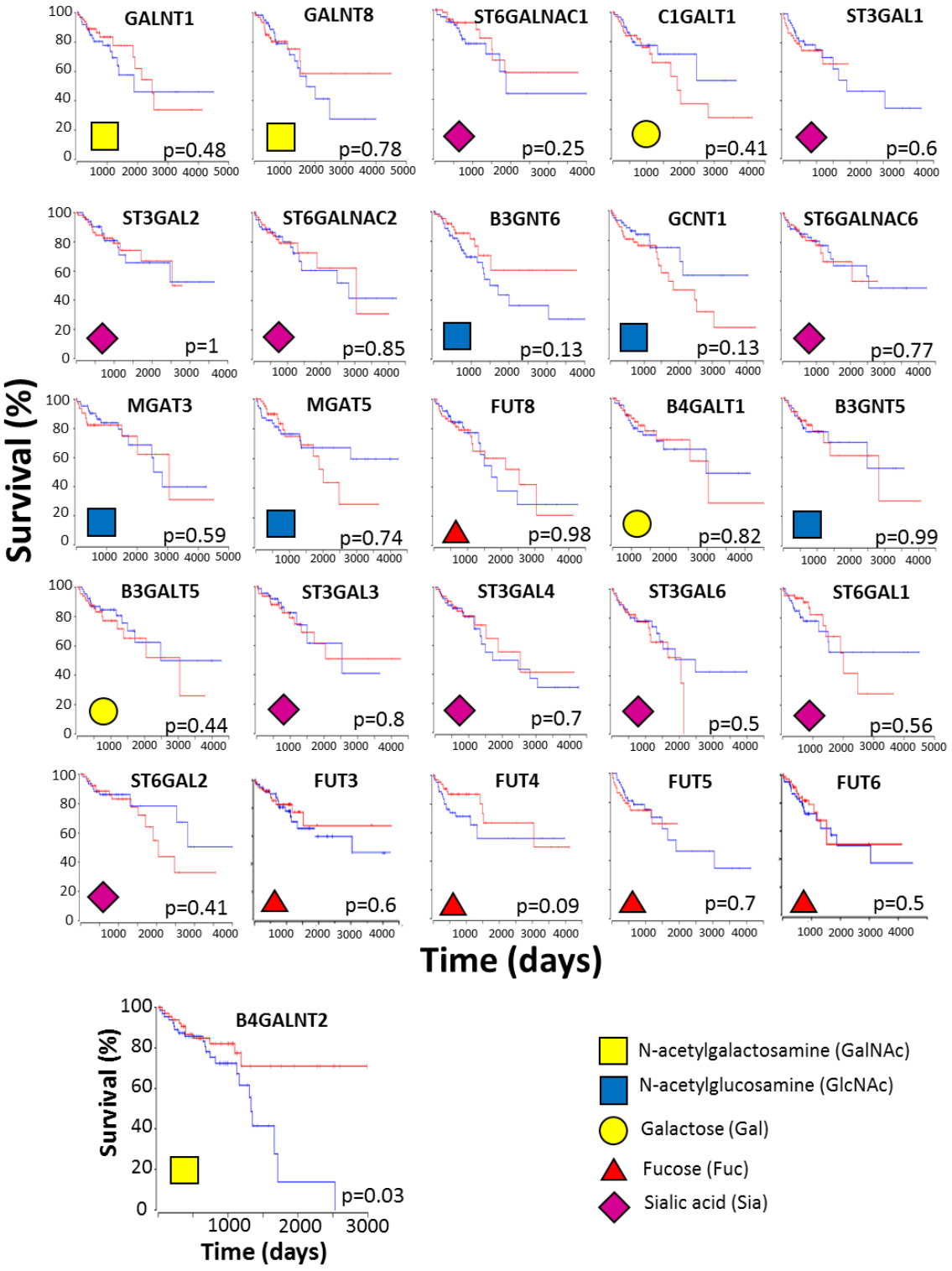


Figure 13. Kaplan-Meier survival curves of colonadenocarcinoma patients relative to glycosyltransferases expression. Kaplan-Meier analysis was performed on Oncolnc website (http://www.oncolnc.org/) dividing patients into two groups (15th lower / 15th higher percentile). Logrank p-value is shown on the right bottom of each plot.

3.2 Oncogenes and tumor suppressor genes expression poorly correlates with COAD patients survival

As shown in Figure 14, B4GALNT2 expression is a significant predictor of long survival in CRC TCGA cohort. To extend this observation to a more general context, the study investigated whether tumor suppressors and oncogenes known to play fundamental roles in cancer and, in particular in CRC, were better predictors of patients' survival. The relationship between the high/low expression of several oncogenes and tumor suppressor genes and patients' survival in the COAD TCGA cohort was analyzed. In Figure 14 are shown the Kaplan-Meier survival curves of COAD patients falling in the 15% upper or 15% lower level of expression of genes known to promote or suppress tumor growth and in particular COAD growth. Survival curves have been ordered according to the p value and boxed in red or blue according to the recognized role as tumor promoting or tumor suppressing activity of the genes. A statistically significant ($p \leq 0.05$) association with survival was shown by genes *SMAD6, TERT, EGFR, CDKN2A, CTNNB1* and *PIK3CA*. Genes whose association with survival displayed p values $0.05 \leq p \leq 0.1$ included *CCNE1*, *SMAD2*, *CDH1*, *TP53* and *BRAF*. These data reveal that the level of expression of only a few oncogenes and tumor suppressor genes is associated with patients' prognosis.

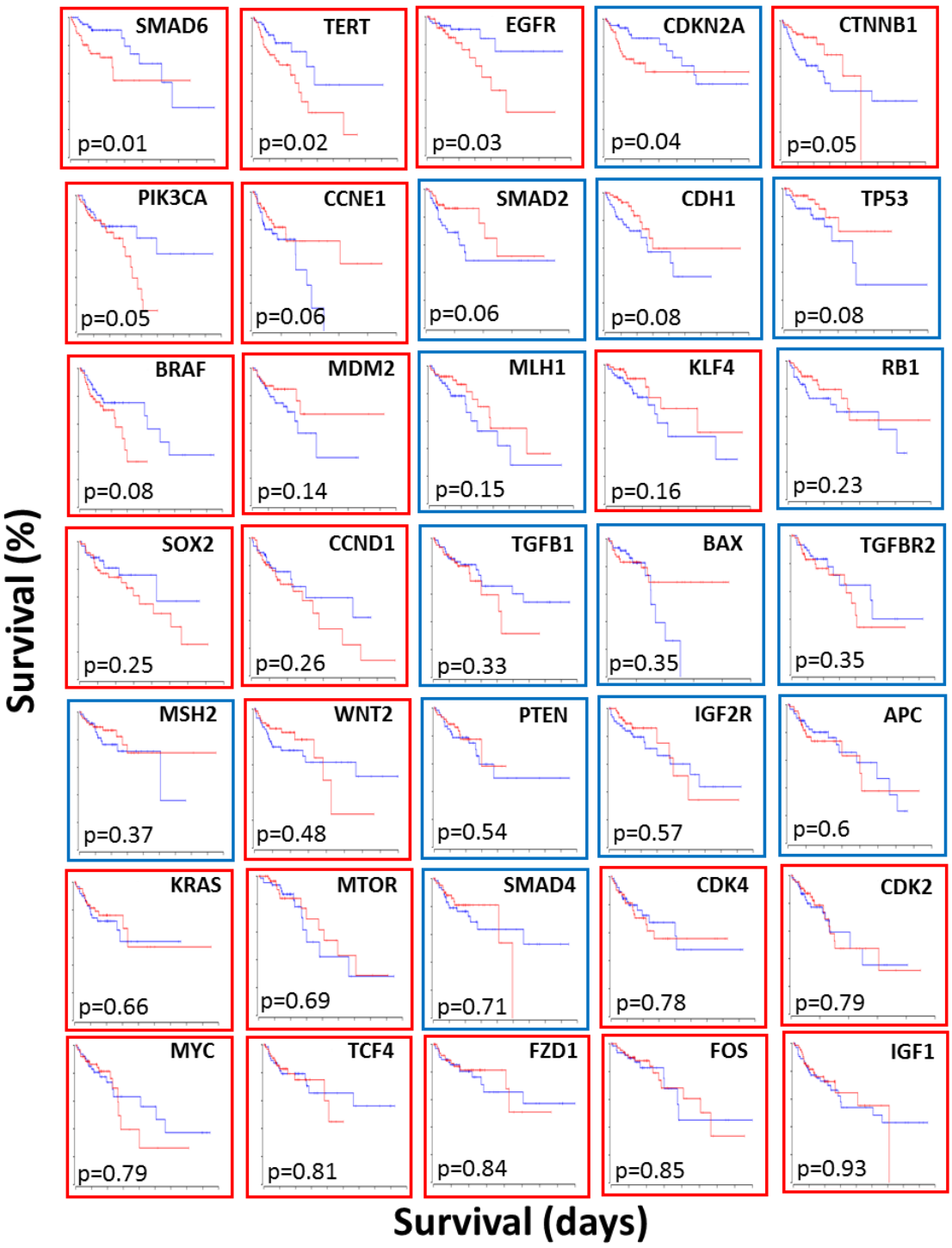


Figure 14. Kaplan Meier plots of COADREAD patients according to the expression level of oncogenes and tumor-suppressor genes. Survival curves were created on OncoLnc website using the 15% high percentile (red lines) and 15% low percentile (blu lines) of a gene. Graphs are shown in order to the increasing p value and boxed in red or blue according to the recognized role as tumor promoting or tumor suppressing activity of the gene.

3.3 Clinical implications of B4GALNT2 expression in CRC: Survey of TCGA database

In order to investigate the clinical implication of B4GALNT2 in CRC patients, TCGA survey was carried out through collection of the main clinical information of 623 colorectal cancer patients, including 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). Some of these clinical data were not available for some CRC specimen. In 51 cases gene expression data of normal colonic mucosa (matching samples with those of CRC) were also accessible.

The survey of TCGA database revealed a relationship between B4GALNT2 gene expression and clinical parameters of CRC. As shown in Figure 15 A, the mean level of B4GALNT2 mRNA in CRC tissues is very low compared to normal tissues, albeit extremely variable. No significant correlation was found between B4GALNT2 expression and stage or microsatellite stability status (Figure 15 B, C). However, B4GALNT2 expression was significantly high in the therapy responder (Figure 15 D) and non-mucinous subtype groups (Figure 15 E). Yet, it was compared the level of B4GALNT2 mRNA in patients either affected or not affected by mutations in genes relevant for CRC carcinogenesis such as the tumor suppressor genes *TP53*. Interestingly, it was found a significant correlation between high B4GALNT2 expression and wild-type *TP53* (Figure 15 F).

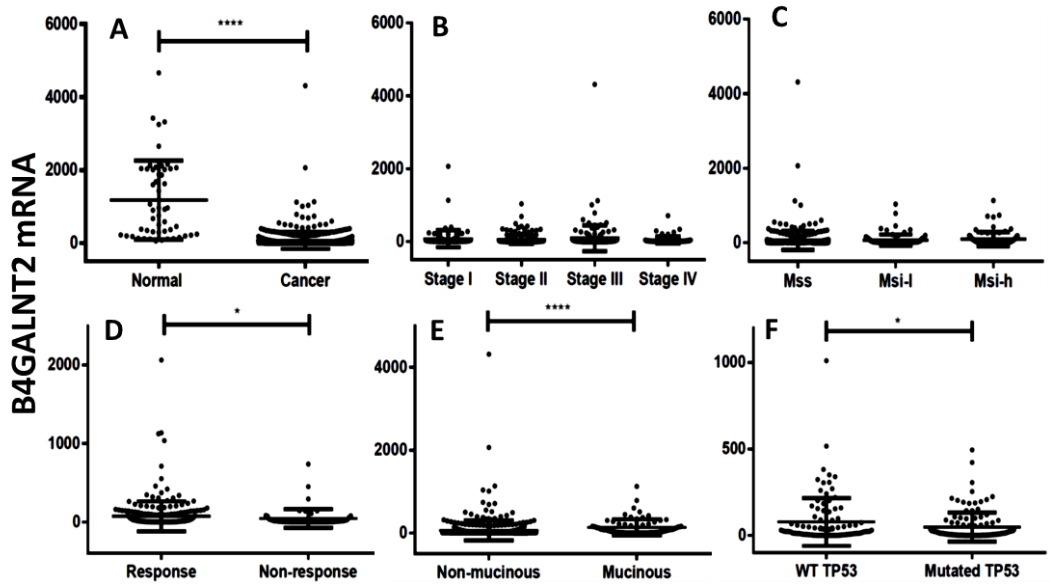


Figure 15. The Cancer Genome Atlas (TCGA) data. (**A**) Expression level of B4GALNT2 mRNA in normal mucosa and colorectal cancer (CRC) specimens. (**B**–**F**) Expression of B4GALNT2 mRNA in CRC specimens grouped according to stage (**B**), microsatellite stability status (**C**), response to therapy (**D**) subtypes (**E**), and TP53 mutation (**F**). MSS: microsatellite stable; MSI-l: microsatellite instable-low; MSI-H: microsatellite instable-high. * $p \leq 0.05$; **** $p \leq 0.0001$.

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 1 represents the genes statistically modulated between high and low B4GALNT2 expressers in CRC.

In the cohorts of HBE and LBE the mean±SD levels of *B4GALNT2* expression was 0±0 and 367±501, respectively. 614 genes displayed a significantly different expression level: 451 were highly expressed in HBE; 163 genes had an opposite behavior. The gene expression ratio between high/low expresser ranged from 200 to -11. Genes showing the most remarkable changes, selected for a ratio higher than 10.0 or lower than -4.0, have been characterized by an extensive literature search, in particular for their role in cancer (Table 1). A color tag was assigned to the putative tumor promoting- or tumor-restraining role of the change as follows: green for higher expression of tumor-restraining genes or lower expression of tumor-promoting genes in HBE; *vice versa* for red. Only genes with a recognizable role in cancer were reported. High *B4GALNT2* expression was associated with 27 tumor

restraining and 10 tumor promoting changes, suggesting its association with a low-malignancy molecular signature.

Table 1. Genes differentially modulated in high B4GALNT2 expressers (HBE) and low B4GALNT2 expressers (LBE)

Gene	Ratio	Gene role	PubMed	
CLCA1	203	Involved in mucus secretion and as a tumor suppressor. Suppresses CRC malignancy.	28974231	
ZG16	151	Involved in protein trafficking. Sequentally reduced from adenoma to CRC.	29661177	
ITLN1	62	Lectin recognizing microbial carbohydrates. Protective in CRC	31893510	
CLCA4	51	Involved in mediating chloride conductance. Down-regulated genes in CRC.	32027181	
SPINK4	48	Serine Peptidase Inhibitor. Its down-regulation is associated with poor survival in CRC.	31888570	
CA1	45	Carbonic anhydrase. Predictive biomarker in CRC.	32031891	
MAGEA1	37	Involved in transcriptional regulation, acts as an oncogene in some cancers.	30509089	
PYY	33	Inhibitis intestinal mobility. Decreased expression is associated with CRC.	11825654	
GUCA2B	32	Regulator of intestinal fluid transport. Tumor suppressor in CRC.	29788743	
CA4	27	Stimulates the ion transporter activity of SLC4A4. Predictive biomarker in CRC.	32031891	
MS4A12	25	Involved in signal transduction. Promotes malignant progression in CRC.	18451174	
BEST2	23	Anion channel. Methylation marker for early detection and prognosis of CRC.	22496748	
HEPACAM2	23	Required for centrosome maturation. Associated with good prognosis.	29659199	
TMIGD1	22	Controls cell-cell adhesion and proliferation. Tumor suppressor in CRC.	33129760	
CLDN8	16	Claudin 8. Component of tight junctions. Down-regulated in CRC	21479352	
B3GNT6	14	Synthesizes core 3 O-linked chains. Down-regulation associated with malignancy in CRC	28745318	
KIF19	13	Microtubule-dependent motor protein. Higher expression associated with longer survival	28901309	
CSAG2	13	Chondrosarcoma-Associated Gene 2/3 Protein. Necessary for tumorigenesis.	32761762	
FCGBP	12	Maintens of the mucosal structure. High expression is associated with better prognosis	31268166	
CDKN2BAS	12	CDKN2B Antisense RNA 1. Promotes progression of ovarian cancer	32572907	
REG1B	11	Regenerating Islet-Derived Protein 1-β. Its silencing inhibits CRC growth.	25768000	
IGJ	11	Joining Chain Of Multimeric IgA And IgM. Down-regulated in CRC	31749922	
LEFTY2	10	Member of the TGF-β superfamily. Negative regulator of endometrial cell proliferation.	27497669	
FUT5	10	Fucosyltransferase 5. Promotes the development of CRC	28771224	

MUC2	10	Secreted mucus forming mucin. Suppresses CRC migration and metastasis.	28725043	
PLIN1	-4	Modulator of adipocyte lipid metabolism. Inhibits breast cancer cell proliferation.	27359054	
PCP4	-4	Functions as a modulator of calcium-binding by calmodulin. Anti-apoptotic peptide.	25153723	
IGF2	-4	Possess growth-promoting activity. Overexpression is associated with poor prognosis.	24080445	
SLC14A1	-4	Urea channel. Cancer stem cell marker.	29329541	
FREM1	-5	Extracellular matrix protein. Associated with better prognosis in bladder cancer.	33058542	
CASQ2	-5	Calsequestrin. High expression associated with poor survival in bladder cancer.	31991631	
CPLX2	-6	Involved in exocytosis. Associated with poor prognosis in lung tumors.	3912489	
ADIPOQ	-6	Adiponectin. Anti-inflammatory adipokine. Lower expression in CRC.	27061803	
WIF1	-7	Inhibits WNT actvities. Hypermethylation is associated with a favorable clinical outcome.	31830937	
CHRNB2	-9	Cholinergic Receptor Nicotinic Beta 2 Subunit. Down-regulated in gastric cancer.	30175534	
AP3B2	-10	Involved in protein sorting. Low expression is associated with long survival in rectal cancer.	29050227	
TSIX	-11	XIST Antisense RNA. Dysregulates cancer pathways in multiple tumor contexts.	29617668	

Genes differentially modulated in HBE and LBE cohorts were analyzed by the false discovery rate two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli. Only genes showing up-regulation ≥10 or down-regulation ≤-4 and with a recognized role in cancer are reported. "Ratio" refers to the HBE/LBE ratio. When the expression was higher in LBE, the HBE/LBE ratio was expressed preceded by a "minus" sign. The role of the gene was deduced from Genecards website. The red or green labels indicate putative tumor-promoting or tumor-restraining changes, respectively.

To establish the prognostic potential of genes modulated with respect to *B4GALNT2*, the survival curves of the 15 top highly expressed genes and the 10 less expressed genes in HBE were obtained (Figure 16). The predictive potential of the highly regulated genes was very good, while genes poorly expressed in HBE lacked any association with prognosis. In particular, the Kaplan-Meier curves of the 15 highly regulated genes (A) were relatively similar, with strong expresser patients displaying a more or less pronounced tendency to better prognosis (the red curve is always above the blue curve). Four genes (*ZG16, ITLN1, BEST2* and *GUCA2B*) displayed a statistically significant relationship. The significance of *ZG16*, a gene previously shown to be associated with good prognosis in CRC, was particularly high. The *p* value of these genes was always lower than 0.5. On the other hand, genes poorly expressed in HBE displayed *p* values always above 0.5.

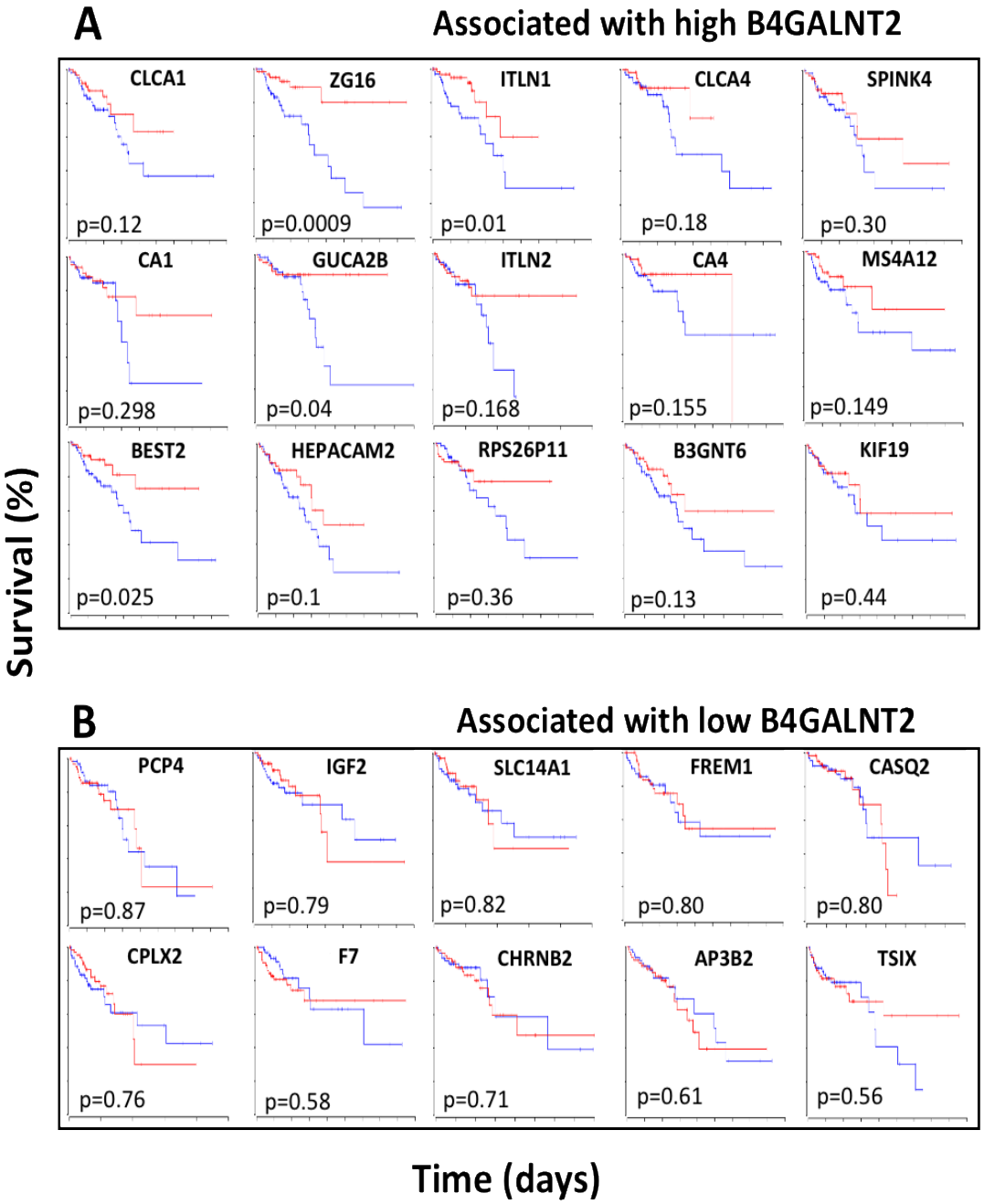


Figure 16. Kaplan-Meier survival curves of patients expressing different levels of genes in LBE and HBE cohorts. Curves were generated using the 15% higher (red lines) and 15% lower expressers (blue lines) of the indicated genes. **A:** 15 top highly up-regulated genes in HBE. **B:** 10 top down-regulated genes in HBE.

3.4 Several glycogenes are differentially modulated in HBE and LBE.

In LBE and HBE groups the expression of genes involved in the biosynthesis and recognition of glycans, as well as heavily glycosylated glycoproteins, such as mucins, and sugar binding proteins, such as galectins, appears significantly different. Table 2 shows the expression level of glycogenes differentially modulated in LBE and HBE. These genes encode proteins involved in: first steps of *O*-glycans biosynthesis (*GALNT8, B3GNT6, ST6GALNAC1, ST6GALNAC2);* ganglioside production (*ST6GALNAC6*); proteoglycan synthesis (*B3GNT7*); synthesis of sialyl Lewis antigens (*B3GALT5, ST3GAL4, FUT5)*; terminal galactose recognition (*LGALS4, LGALS9B*). Four genes encode *O*-glycoproteins (*MUC1, MUC2, MUC4, MUC5B*). Only two genes (*ST6GAL1, ST6GAL2*) that codify enzymes responsible for Sia6 LacNAc structures biosynthesis display higher expression in LBE.

Table 2. Expression level of glycogenes in HBE and LBE.

			Low B4GALNT2	High B4GALNT2			
	Transferase reaction	Gene symbol	Mean±SD	Mean±SD	Corrected p	Fold difference	Role
First steps of O-linked biosynthesis	O-GalNAC to peptide	GALNT8	58±176	545±833	0.000000	9.5	Attaches the first GalNAc residue of the O-linked chains
	β1,3 GlcNAc to GalNAc	B3GNT6	77±611	1114±1514	0.000000	14.4	Synthesizes Core 3 of the O-linked chains by attaching GlcNAc to GalNAc
	α2,6 Sia to GalNAc	ST6GALNAC1	1133±1375	6133±4623	0.000000	5.4	Synthesizes sialyl Tn by attaching Sia to GalNAc
		ST6GALNAC2	75±82	158±163	0.000011	2.1	Synthesizes sialyl T by attaching Sia to GalNAc of T antigen
Ganglioside biosynthesis		ST6GALNAC6	536±669	1350±1401	0.000001	2.5	Synthesizes higher gangliosides
Proteoglycan biosynthesis	β1,3 GlcNAc to Gal	B3GNT7	256±311	1606±2907	0.000011	6.3	Keratan sulfate biosynthesis
Biosynthesis of sialyl Lewis antigens	β1,3 Gal to GlcNAC	B3GALT5	63±125	252±330	0.000000	4.0	Synthesizes type 1 chains
	α2,3 Sia to Gal	ST3GAL4	458±657	1852±1983	0.000000	4.0	Sialylates type 2 chains
	α1,3 Fuc to GlcNAc	FUT5	30±125	326±382	0.000000	10.8	Fucosylates type 2 chains
Biosynthesis of Sia6 LacNAc structures	α2,6 Sia to Gal	ST6GAL1	3973±3054	2010±1457	0.000000	-2.0	α2,6 sialylation of glycoproteins
		ST6GAL2	120±183	44±76	0.000150	-2.7	α2,6 sialylation of soluble substrates
Terminal galactose recognition	Galectins	LGALS4	15371±10268	31046±16948	0.000000	2.0	Galectin 4, expressed in the gut, underexpressed in CRC
		LGALS9B	68±136	154±162	0.000067	2.2	Highly similar to Galectin 9
O-glycoproteins	Mucins	MUC1	2484±3238	5033±4103	0.000002	2.0	Membrane bound mucin with multiple functions
		MUC2	5933±18245	61677±81704	0.000000	10.4	Secreted mucus forming mucin
		MUC4	1024±2379	5351±4933	0.000000	5.2	Membrane and secreted mucin
		MUC5B	6333±13428	16484±23100	0.000161	2.6	Gel-forming mucin

"Ratio" indicates the ratio between gene expression in HBE/LBE. When the expression was higher in LBE, the HBE/LBE ratio was expressed with "minus" sign.

3.5 Methylation partially controls the expression of *B4GALNT2*

To elucidate mechanisms regulating B4GALNT2 expression, gene methylation was investigated in colorectal cancer patients through the SMART (Shiny Methylation Analysis Resource Tool) App, a web-based tool that allows a comprehensive analysis of DNA methylation data of TCGA project.

The segment plot in Figure 17A shows the CpGs associated to *B4GALNT2* gene and their genomic locations along with transcripts. Genomic sites covered by the methylation probes include the CpG island as well as a Northern shore (N-shore, upstream the island), a Southern shore (S-shore, downstream the island) and an

intronic (open-sea) site located between exons 6 and 7. Differential analysis of tumor and normal samples revealed that methylation in both the N-shore and S-shore was never statistically different between normal and tumor tissues, although in the latter the methylation level was more heterogeneous among patients (Figure 17B). On the other hand, in both normal and the vast majority of cancer tissues the extent of methylation was very low in seven locations within the island (cg01147550-cg18208707 and cg02445664). In the same region positions cg20233029 and cg03167683 displayed a small but significantly reduced methylation in tumor tissues. The "open sea" site cg043380107, located at the intron, displayed a highly significant and very heterogeneous methylation decrease in cancer. Correlation analysis of B4GALNT2 expression with methylation status of the 16 sites in tumor tissues (Figure 17C) indicated that in some cases methylation results in enhancement, rather than inhibition, of gene expression. Indeed, methylation of the intronic site cg043380107 is associated with increased, rather than decreased, B4GALNT2 expression. Exception for the first two sites in the N-shore, in all the remaining sites low methylation is required for high *B4GALNT2* expression, although many samples displaying very low methylation failed to express *B4GALNT2* (Figure 17B).

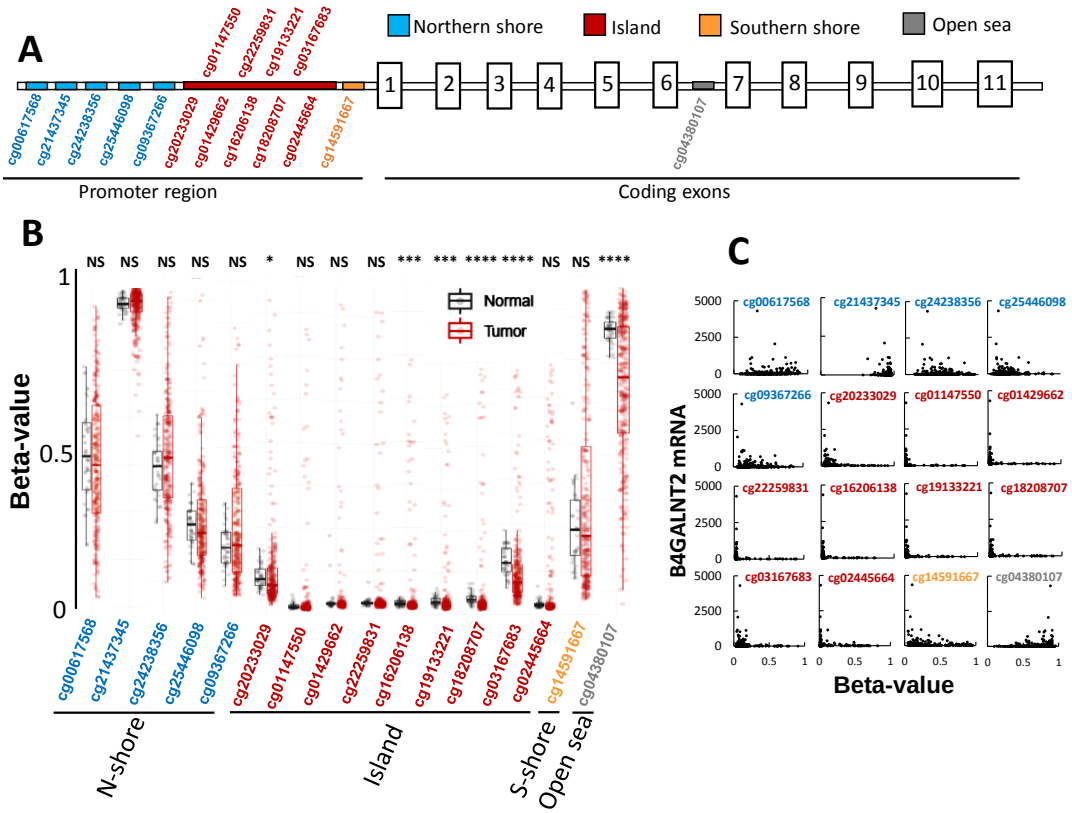


Figure 17. DNA methylation of B4GALNT2 promoter region. A: segment plot highlighting the promoter region and coding exons of the *B4GALNT2* gene. The approximate position of the probes is indicated. **B:** Methylation level of the different probes. *p<0.05; ***p<=0.001; ****p<0.0001. **C:** Correlation between B4GALNT2 expression level and methylation of specific positions in tumor tissues.

3.6 miR-204-5p regulates B4GALNT2 expression in CRC

Gene expression can be post-transcriptionally regulated by microRNAs (miRNAs), small non-coding RNAs of ~22nt, through suppressing mRNA translation or inducing mRNA degradation. Thus, the potential role of miRNAs in the regulation of *B4GALNT2* expression was investigated. Interrogation of CSmiRTar database that integrates miRNA-target interactions and their functional roles in various biological processes provided a list of miRNA potentially targeting *B4GALNT2* in colorectal cancer. The study considered only miRNA supported by at least two of the four miRNA target prediction databases and with a "normalized miRNA score, NMR" >0.2 (Figure 18A). To understand the role of these miRNA on *B4GALNT2* expression, their mean expression level was determined in the LBE and HBE groups. In consideration of the lower number of miRNA data available for TCGA patients, LBE and HBE subjects expressing a level of *B4GALNT2* mRNA lower or higher than 20 were considered, respectively. Among miRNAs targeting *B4GALNT2*, five displayed little and non-significant differences between the two groups while miR-204-5p was 2.7 fold less expressed in HBE than in LBE (p=0.01) (Figure 18A). Correlation analysis of *B4GALNT2* with single miRNAs expression (Figure 18B) revealed that miR-204-5p was not expressed in all the HBE patients, although several patients not expressing miR-204-5p failed to express *B4GALNT2*.

A

Expression of miRNA potentially targeting B4GALNT2 in LBE and HBE cohorts

miRNA	Mean NMS*	LBE	HBE	LBE/HBE Ratio	*p**
hsa-miR-105-5p	0.664	23.2	24.7	0.9	0.43
hsa-miR-145-5p	0.652	2146	2022	1.1	0.37
hsa-miR-204-5p	**0.594**	**9.7**	**3.6**	**2.7**	**0.01**
hsa-miR-150-3p	0.449	2.4	2.5	0.97	0.31
hsa-miR-146b-3p	0.289	62.3	65.1	0.96	0.32
hsa-miR-134-5p	0.272	312	289	1.1	0.1

*NMS: Normalized miRNA score. Numbers represent the mean value reported in at least 2. Only miRNA with NMS> 0.2 were reported. LBE and HBE represent the expression of the miRNA in the LBE (B4GALNT2 expression <20) or HBE (B4GALNT2 expression >20) cohorts. The only significantly modulated miRNA is indicated in bold. * Student's t test for independent samples.

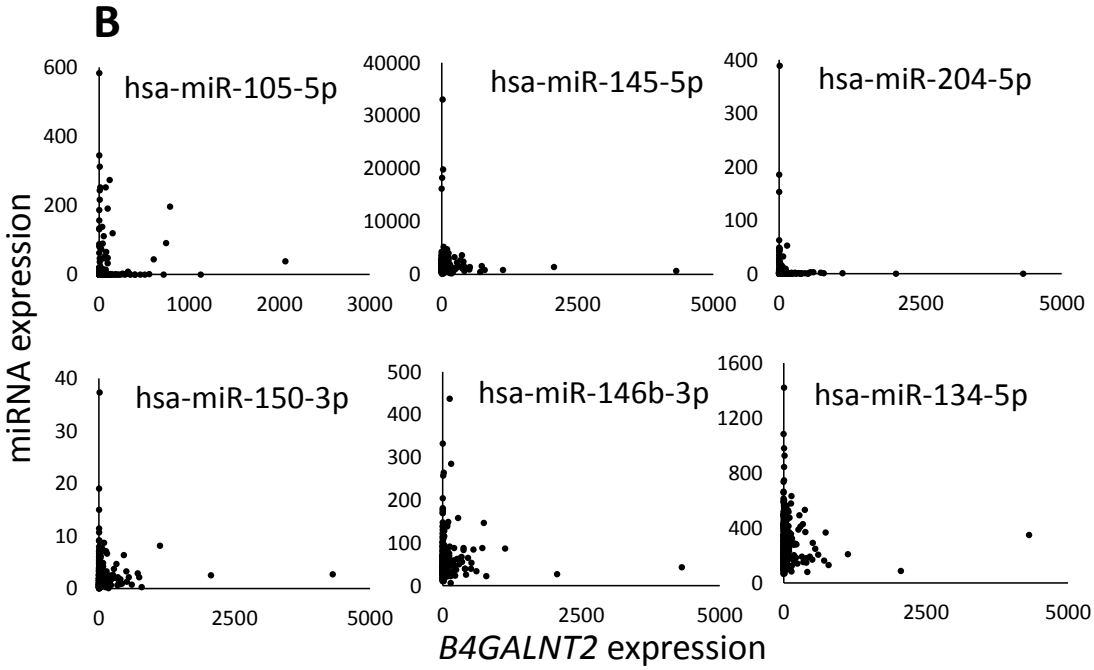


Figure 18. Correlation between B4GALNT2 and miRNA expression. A: miRNA potentially targeting B4GALNT2 obtained from CSmiRTar database. Only miRNA supported by at least two of the four miRNA prediction target databases and with a "normalized miRNA score, NMR" >0.2 were considered. **B:** correlation dot plots of *B4GALNT2* with miRNA expression

CHAPTER IV - RESULTS

Transcriptomic and phenotypic impact of B4GALNT2 expression in LS174T CRC cells

Note: Results presented in this section were taken from the published manuscript:

- Pucci M.; Gomes Ferreira I.; Orlandani M.; Malagolini N.; Ferracin M.; Dall'Olio F., "High Expression of the Sda Synthase B4GALNT2 Associates with Good Prognosis and Attenuates Stemness in Colon Cancer", CELLS, **2020**, 9, pp. 1 - 18

4.1 Phenotypic impact of B4GALNT2 expression on colon cancer cells

Data in clinical setting revealed a clear association between high B4GALNT2 and better prognosis. Thus, aiming at investigating the impact of B4GALNT2 expression on the malignant phenotype *in vitro*, in particular as a function of sLe^x expression, the colorectal cancer cell line LS174T was employed as a model due to the negligible levels of B4GALNT2 and good levels of the sLe^x antigen. LS174T cells have been either transfected with the short form of B4GALNT2 or mock-transfected. The three cell lines analyzed were Neo - a polyclonal population of mock-transfectants - and S2 and S11, two B4GALNT2-transfected clones. As shown in Figure 19A, the level of B4GALNT2 mRNA and enzyme activity in mock transfectants was nearly undetectable, while it was high in S2 and S11 clones. In S2 and S11 clones, but not in Neo cells, the Sd^a antigen was strongly expressed on high-molecular-weight proteins (Figure 19B). On the other hand, the sLe^x antigen, which is also carried by high-molecular-weight proteins, was more strongly expressed by Neo cells than by S2 and S11 clones (Figure 19B). This is due to the previously documented competition between the fucosyltransferases synthesizing sLe^x and B4GALNT2.

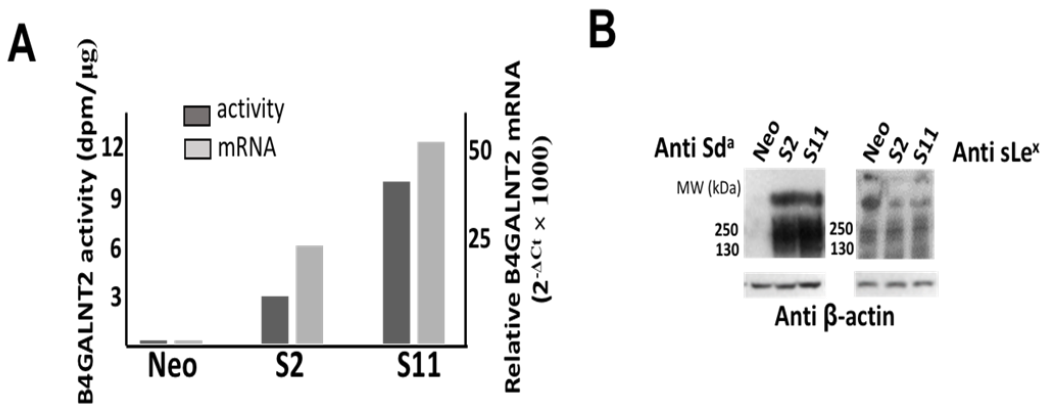


Figure 19. Biochemical characterization of B4GALNT2-transfected cell lines. A: The enzymatic activity (dark gray) of Neo cells and B4GALNT2-transfected clones S2 and S11 was measured as the difference between the incorporation of radioactive GalNAc on fetuin and asialofetuin. The mRNA (light gray) was measured by real-time RT-PCR and normalized with β-actin/GAPDH. B: Western blot analysis of Neo cells and B4GALNT2-transfected clones with anti Sd^a (left) and anti sLe^x (right) antibodies, revealing a partial replacement of the sLe^x antigen with the Sd^a.

The LS174T Neo and B4GALNT2 transfected cells were analyzed for the following typical aspects of malignant growth:

- **ANCORAGE INDEPENDENT GROWTH IN SOFT AGAR**

LS174T cells were evaluated for their ability to form colonies by a soft agar colony formation assay. This anchorage-independent growth assay is a well-established method for characterizing the ability of transformed cells to grow independently of a solid surface, and is a hallmark of carcinogenesis[129].

The rationale behind this technique is that normal and often cancer cells depend on contact with the extracellular matrix to grow and divide. Conversely, a proportion of a cancer cell population is not dependent on adhesion to extracellular matrix to grow and divide. Therefore, cancer cells able to form colonies in a semi-solid medium are considered particularly malignant[129]. In this assay, the cells were plated as a single cell within a layer of agar. Compared with mock-transfected Neo cells, S2 and S11 clones displayed a strongly reduced ability to grow in a semi-solid medium (Figure 20 A), forming 20–40% of the clones formed by Neo cells.

- **3D TUMOR SPHEROIDS**

LS174T cells were evaluated for their capacity to form spheroids. 3D tumor spheroids are self-assembled cultures of tumor cells formed in conditions where cell-cell interactions predominate over cell-substrate interactions. Multi- cellular tumor spheroids resemble avascular tumor nodules, micro-metastases, or the intervascular regions of large solid tumors with respect to their morphological features, microenvironment, volume growth kinetics and gradients of nutrient distribution, oxygen concentration, cell proliferation and drug access[130]. LS174T B4GALNT2-expressing clones were compared to Neo cells for their capacity to survive and grow in these harsh conditions. B4GALNT2 clones displayed a 60% reduction in ability to grow as spheroids in a completely liquid medium (Figure 20 B).

- **CLONOGENIC ASSAY OF CELLS *IN VITRO***

LS174T cells were analyzed for their ability to grow into a colony from a single cell in standard growth conditions (with adhesion to a solid substrate). This assay essentially tests every cell in the population for its ability to undergo ‘‘unlimited’’ division[131]. It was performed by seeding 50 Neo, S2, or S11 cells in standard conditions. After 15 days the number of growing colonies was similar in Neo cells and B4GALNT2 clones (Figure 20 C).

- **WOUND HEALING ASSAY**

To evaluate whether the B4GALNT2 expression could modify the ability of cells to proliferate and migrate (an important feature associated with malignant transformation[132]), wound healing assay was performed with LS174T Neo population and S2/S11 clones. It was observed that the capacity to heal a scratch wound was not significantly affected by B4GALNT2 expression clones (Figure 20 D).

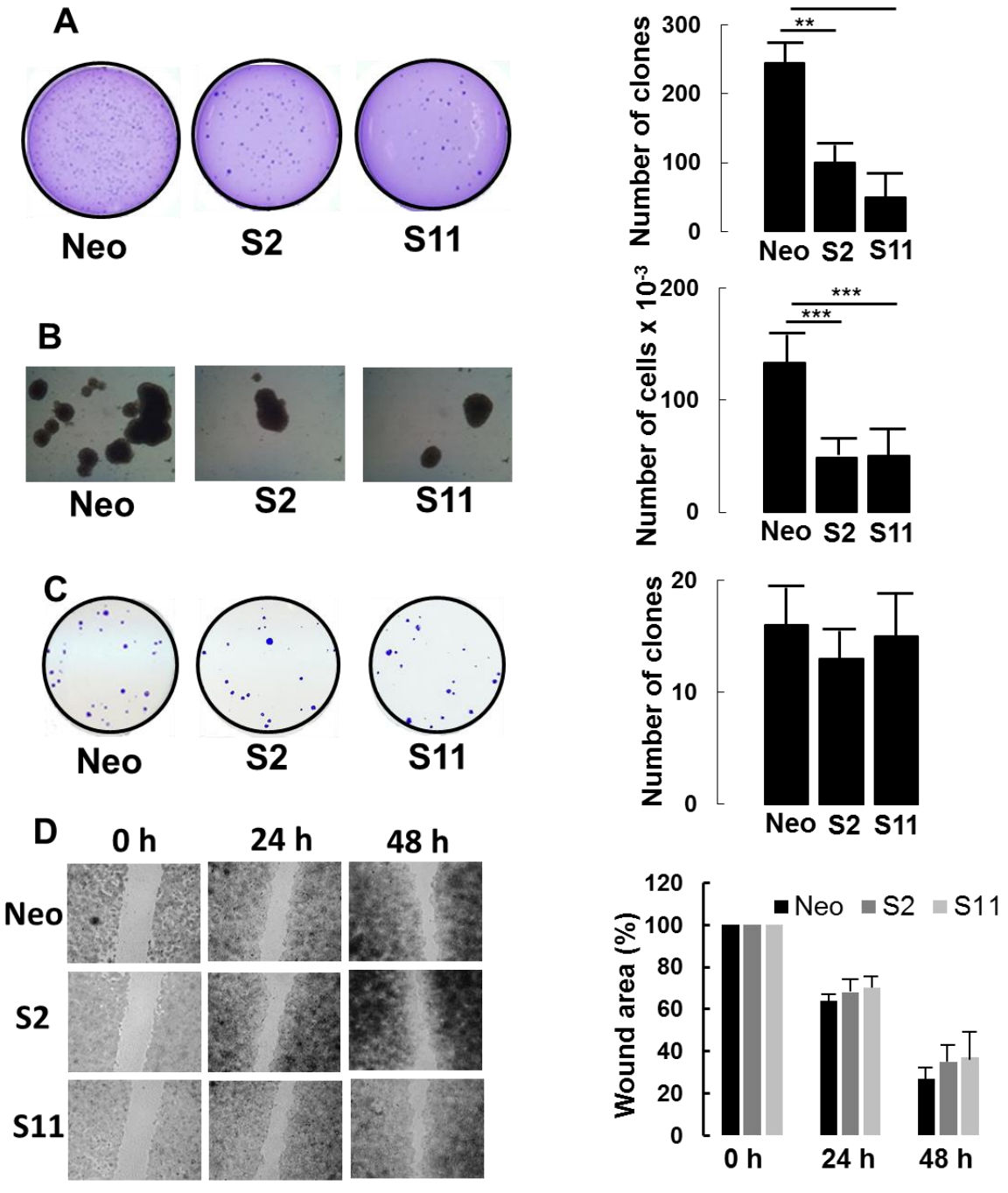
- **ALDEFLUOR ASSAY**

The study of the phenotype *in vitro* revealed that B4GALNT2 expression leads to a dramatic inhibition of the ability to grow in poor or no adherence, pointing to a

specific effect of B4GALNT2 in regulating this property. The ability to survive and proliferate without the intracellular signals generated by the mechanosensors is documented to be intimately associated with stemness.

To investigate the relationship between B4GALNT2 expression and stemness, the three LS174T cell lines were analyzed for the expression of aldehyde dehydrogenase (ALDH), reported to be a stem-cell and cancer-initiating cell marker in many tissues, including colon tissue[108].

In a typical experiment (Figure 20 E), cells were incubated with the ALDH substrate ALDEFLUOR, either in the presence or in the absence of DEAB (a specific ALDH inhibitor) to provide a negative control. While the percentage of ALDH-positive cells in LS174T Neo was about 40%, it was around 30% in the two B4GALNT2 clones, consistent with a marked reduction in the number of cancer stem cells (CSC).


A
Neo
S2
S11
B
Neo
S2
S11
C
Neo
S2
S11
D
0 h
24 h
48 h
Neo
S2
S11
Number of clones
300
200
100
0
**
**
Neo S2 S11
Number of cells x 10-3
200
100
0
***
***
Neo S2 S11
Number of clones
20
10
0
Neo S2 S11
Wound area (%)
120
100
80
60
40
20
0
Neo S2 S11
0 h
24 h
48 h

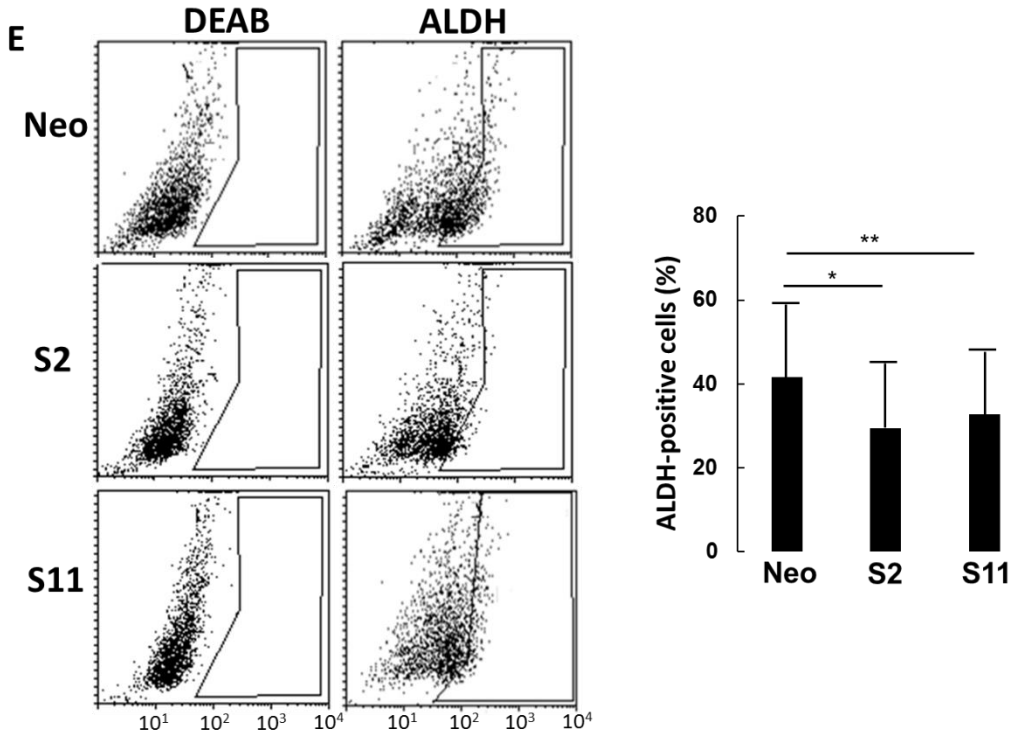


Figure 20. Phenotypic characterization of B4GALNT2-expressing cells and mock-transfectants. **A**: Growth in 0.33% soft agar. Photographs were taken without magnification and the colonies visible to the naked-eye were counted. **B**: Spheroids formation assay. The aspect of the spheroid is shown. The total amount of protein was calculated and taken as a measure of the cells grown in 3D conditions. **C**: Colony formation assay in standard conditions of growth. **D**: Wound healing assay. The free area of the wound was quantitated by ImageJ and normalized to the free area of the same cell line at 0 h, which was taken as 100%. Graphs report the quantification of the healing process at each time point. The microphotographs were taken at a 4x magnification. **E**: ALDEFLUOR assay. Cells were incubated with ALDEFLUOR either in the presence or in the absence of the inhibitor N,N-diethylaminobenzaldehyde (DEAB). Gates excluding all of the cells labelled in the presence of DEAB were set. Cells included in the gate in the absence of DEAB, were considered to be ALDH positive. Histograms report the percentage of ALDH positive cells ±SD. All experiments were repeated at least three times. *$p \leq 0.05$, **$p \leq 0.01$, *** $p \leq 0.001$.

4.2 Impact of B4GALNT2 expression on the transcriptome of LS174T colon cancer cells

To understand the origin of the dramatic effect of B4GALNT2 on the phenotype of LS174T cells and, in particular, on the ability to grow in non-adherent conditions, the impact of B4GALNT2 and of 3D growth in liquid medium on the transcriptome of LS174T cells was investigated by microarray analysis.

RNA preparation and analysis of the Neo population and S2/S11 clones grown in standard conditions or as 3D spheroids was performed in duplicate. Using

microarray technology, the mean level of B4GALNT2 expression was found to be 3 in Neo and 230 in S2/S11 cells. Figure 21A illustrates a heat-map graph that reports the modulation of 142 genes showing a fold change ≥ 2 in LS174T S2 and S11, compared with Neo cells, grown in standard conditions. Panel B in Figure 21 shows genes modulated by 3D culture in Neo and in B4GALNT2-expressing cells.

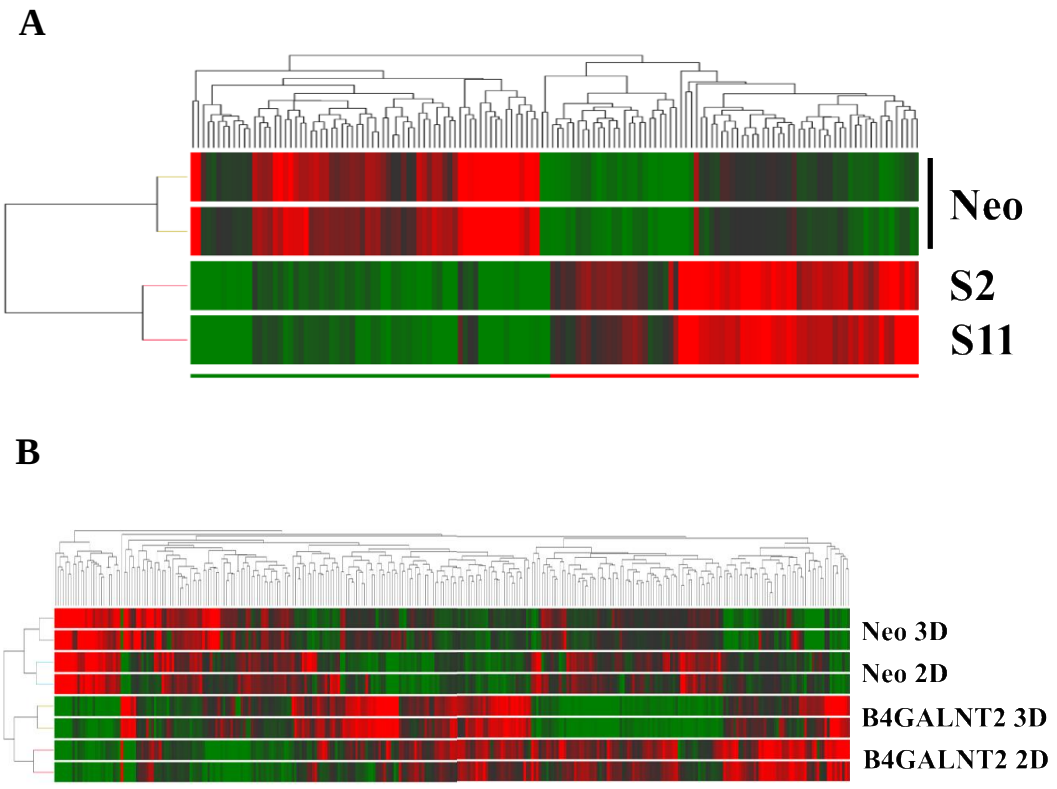


Figure 21. Heatmaps of gene expression analysis. A: B4GALNT2-expressing and control Neo LS174T cells grown in standard 2 D conditions. **B**: Cells grown in 3D conditions or in standard 2D conditions. The genes that are differentially expressed are reported. Genes (columns) and samples (rows) were grouped by hierarchical clustering (Manhattan correlation). High- and low- expression was normalized to the average expression across all samples. Differences were analyzed by the moderated t-test. Corrected p-value cut-off: 0.15; multiple test correction used: Benjamini-Hochberg.

The most relevant pathways modulated by B4GALNT2 in LS174T cells identified by GeneGo Metacore analysis are shown in Table 3. The expression of the glycosyltransferase affects mainly the stem cell pathways, blood coagulation, main growth factor signaling cascades, cell adhesion, cytoskeleton remodeling and G protein-coupled receptors signaling.

Table 3. Networks and Networks objects modulated by B4GALNT2 expression in LS174T transfectants.

Networks	Networks Objects
Stem cell pathways	SOX2, FGFR3, HEY2, IGF1, c-Kit, MEF2C, MLRC, MyHc
Blood coagulation	MyHC, Coagulation factor V, PAR1
Main growth factor signaling cascades	FGFR3, IGF-1
Chemoresistance pathways	c-Kit, IGF-1
Cell adhesion	Nidogen, IGF-1, MyHC, MRLC
Cytoskeleton remodeling	MyHC, MRLC
G protein-coupled receptors signaling	G_(i)-specific peptide GPCRs, G_(q)-specific peptide GPCRs

Pathway map visualization was performed using MetaCore pathway analysis by GeneGo.

To restrict the study to the most biologically relevant genes, a more in-depth analysis revealed 25 genes to be modulated by B4GALNT2 by a fold change ≥ 4 (Table 4). Among these genes, four displayed up-regulation (*CD200, NGFRAP1, SKAP1,SLC14A1*) and 21 displayed down-regulation (*FAM26F, FAM110B, ALX1, F5, NMT, MYH3, MBOAT2, ROR1, RAI1, FMO3, PEG10, NINL, ARMC4, MID2, SOX2 28, LGALS2, NPTX, GALC, STARD3NL, ZNF22, NID1*). In the Table 4 each gene was associated to its function in cancer by functional annotation. It was also attributed a cancer-promoting activity or a cancer-restraining activity to many of the modulated genes through an intense search in literature. In addition, a violet or yellow label was assigned on the basis of the putative tumor-promoting or tumor-restraining change (violet for up-regulation of tumor-promoting or down-regulation of tumor-restraining genes and *vice versa* for the yellow label).

As highlighted in Table 4 by the label color, only three changes were putatively tumor-promoting and 12 were tumor-restraining.

Table 4. Genes highly modulated by B4GALNT2 in LS174T cells.

Gene Symbol	Expression		Fold Change S2/S11 Vs Neo	p value S2/S11 Vs Neo	GeneName	Function in cancer	PMID	
	Neo	S2/S11						
CD200	2	27,0	16,8	0,0411	CD200 molecule	Possible colon cancer stem cell marker	27574016	
NGFRAP1	39	456,3	11,6	0,0383	nerve growth factor receptor (TNFRSF16) associated protein 1	Overexpression inhibits growth of breast tumor xenografts.	26408910	
SKAP1	138	912,3	6,6	0,0231	src kinase associated phosphoprotein 1	Modulates TCR signaling.	18320039	
SLC14A1	2	10,8	5,2	0,0360	solute carrier family 14 (urea transporter), member 1 (Kidd blood group)	Potential tumor suppressor in lung cancer	22223368	
FAM26F	8	2,0	-4,1	0,0195	family with sequence similarity 26, member F	Little or no information		
FAM110B	9	2,1	-4,5	0,0142	family with sequence similarity 110, member B	Promotes growth of prostate cancer cells	21919029	
ALX1	12	2,6	-4,6	0,0167	ALX homeobox 1	Promotes EMT and invasion in ovarian and lung cancer.	26722397 23288509	
F5	12	2,6	-4,7	0,0331	coagulation factor V (proaccelerin, labile factor)	Little or no information		
INMT	9	1,8	-4,7	0,0142	indolethylamine N-methyltransferase	Negatively associated with prostate cancer progression	22075945	
MYH3	1198	238,9	-5,0	0,0167	myosin, heavy chain 3, skeletal muscle, embryonic	Little or no information		
MBOAT2	14	2,5	-5,4	0,0383	membrane bound O-acyltransferase domain containing 2	Little or no information		
ROR1	12	1,8	-6,4	0,0163	receptor tyrosine kinase-like orphan receptor 1	Associated with ovarian cancer stem cells	25411317	
RAI14	51	7,7	-6,6	0,0190	retinoic acid induced 14	Overexpressed in gastric cancer, associated with worse prognosis.	29654694	
FMO3	14	1,8	-7,7	0,0253	flavin containing monooxygenase 3	Involved in de-toxification of drugs.	16800822	
PEG10	44	5,3	-8,4	0,0233	paternally expressed 10	Enhances cell invasion by upregulating β-catenin, MMP-2 and MMP-9	25199998	
NINL	244	28,4	-8,6	0,0339	ninein-like	High expression associates with poor prognosis in prostate cancer	30637711	
ARMC4	15	1,7	-8,7	0,0196	armadillo repeat containing 4	Can be mutated in gastric cancer.	26330360	
MID2	32	2,1	-15,0	0,0152	midline 2	In breast cancer associates with BRCA1 and promotes growth.	26791755	
SOX2	28	1,7	-16,5	0,0163	SRY (sex determining region Y)-box 2	Associated with motility and a cancer stem cell phenotype in CRC	29228716 30518951	
LGALS2	362	21,5	-16,8	0,0142	lectin, galactoside-binding, soluble, 2	Elevated in plasma of CRC patients. Promotes adhesion to endothelia.	21933892	
NPTX1	42	2,4	-17,3	0,0123	neuronal pentraxin I	Anti proliferative in colon cancer	29345391	
GALC	49	2,0	-24,9	0,0077	galactosylceramidase	Unclear		
STARD3NL	98	3,6	-27,4	0,0346	STARD3 N-terminal like	Little or no information		
ZNF22	83	1,9	-44,6	0,0077	zinc finger protein 22	Little or no information		
NID1	459	5,1	-89,4	0,0306	nidogen 1	Promotes EMT and metastasis in ovarian, breast and lung cancer.	28416770 28827399	

Corrected p value was calculated using the multiple test correction Benjamini-Hochberg. $p<0.05$, fold change mean S2/S11 B4GALNT2 vs Neo ≥ 4. The red line separates up-

regulated genes from down-regulated genes. The violet or yellow labels indicate putative tumor-promoting or tumor-restraining changes, respectively.

Yet, the investigation intended to examine whether those genes that were found to be up-regulated in LS174T S2 and S11 cells were also up-regulated in patients showing high B4GALNT2 levels in cancer tissues and *vice versa* for genes displaying down-regulation in S2/S11 cells. To this aim, it was considered the same cohorts of patients shown in Figure 14, comprising 15% of non-expressers and 15% of high expressers. For the 25 genes showing modulation by B4GALNT2 reported in Table 4, it was determined the mean level of expression in the non-expressers and in the high-expressers cohorts, respectively (Table 5), from TCGA. Out of the 25 genes, one was not expressed (*SLC4A1*); 13 showed a difference between non-expressers and high-expressers, consistent with the hypothesized role of B4GALNT2 in regulating gene expression (*CD200, NGFRAP1, FAM110B, F5, INMT, MYH3, RAI14, FMO3, NINL,SOX2, NPTX1, STARD3NL, NID1*); for six genes (*CD200, MYH3, NINL, SOX2, NPTX1, STARD3NL*) the change was statistically significant.

Table 5. Gene expression comparison between TCGA cohort (Non- and High-B4GALNT2 expressers and microarray analysis of LS174T cells (S2/S11 comparison with Neo).

		Non-B4GALNT2 expressers	High-B4GALNT2 expressers		
	Gene name	**Mean±SD**	**Mean±SD**	**Consistency**	**p**
Genes up-regulated in LS174T S2/S11	CD200	171±144	250±236	Yes	≤0.01
	NGFRAP1	850±641	843±5655	Yes	N.S.
	SKAP1	220±176	205±192	No	
	SLC4A1	Not expressed	Not expressed		
Genes down-regulated in LS174T S2/S11	FAM26F	66±87	96±84	No	
	FAM110B	37±33	35±42	Yes	N.S.
	ALX1	3±13	4±8	No	
	F5	325±968	180±627	Yes	N.S.
	INMT	132±153	118±120	Yes	N.S.
	MYH3	41±125	16±16	Yes	≤0.05
	MBOAT2	498±282	660±405	No	
	ROR1	27±39	25±31	No	
	RAI14	765±420	661±586	Yes	N.S.
	FMO3	86±620	26±25	Yes	N.S.
	PEG10	136±401	245±685	No	
	NINL	168±203	119±123	Yes	≤0.05
	ARMC4	4±13	6±8	No	
	MID2	146±122	150±104	No	
	SOX2	107±281	32±117	Yes	≤0.01
	LGALS2	77±141	128±210	No	
	NPTX1	39±95	15±29	Yes	≤0.01
	GALC	644±525	663±503	No	
	STARD3NL	697±255	646±253	Yes	≤0.1
	ZNF22	625±239	638±267	No	
	NID1	1558±993	1544±1496	Yes	N.S.

The cohorts of non-expressers (Mean ± SD =0±0) and high-expressers (Mean ± SD 367±69) represent the 15% lower and higher percentiles of the TCGA cohort. The column "Consistency" indicates whether the difference in gene expression of no- or high

B4GALNT2 expressers was consistent with that observed by microarray analysis of LS174T model. Genes showing statistically significant consistent difference are indicated in bold ($p \leq 0.05$ Student's t test for independent samples). N.S.= non significant.

4.3 B4GALNT2 expression regulates the transcriptional response to 3D culture

Owing to the markedly reduced ability to adapt to non-adherent growth displayed by B4GALNT2-expressing cells, the research focused on which genes were modulated by 3D culture in LS174T cells and which genes displayed a differential response to 3D culture conditions in B4GALNT2-expressing cells S2/S11. Many genes were modulated by 3D culture conditions, regardless of B4GALNT2 expression. Among these, 106 displayed a fold change ≥ 4 as shown in Table 6.

Table 6. Genes highly modulated by 3D culture in LS174T cells.

Gene Symbol	2D	3D	Fold Change 3D Vs 2D	Corrected p value 3D Vs 2D	Description	Role	Broad cellular function
ATP4A	2	31	13,5	0,0020	ATPase, H+/K+ exchanging, alpha polypeptide	Catalyzes the hydrolysis of ATP coupled with the exchange of H(+) and K(+) through the plasma membrane	Energy metabolism
NDUFA4L2	357	4056	11,4	0,0016	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4-like 2	Respiratory electron transport	
ANGPTL4	42	428	10,1	0,0010	angiopoietin-like 4, transcript variant 1	Regulates glucose homeostasis, lipid metabolism, and insulin sensitivity	
OLAH	4	39	10,1	0,0055	oleoyl-ACP hydrolase, transcript variant 2	Contributes to the release of free fatty acids from fatty acid synthase	
CA9	803	5805	7,2	0,0004	carbonic anhydrase IX	Hypoxia response	
CHGA	5	28	5,3	0,0023	chromogranin A (parathyroid secretory protein 1)	Precursor of vasostatin, pancreastatin, and parastatin	
PPP1R3G	47	238	5,1	0,0008	protein phosphatase 1, regulatory subunit 3G	Involved in the regulation of hepatic glycogenesis	
LCN15	15459	77132	5,0	0,0028	lipocalin 15	Transporter of glucose and other small molecules	
PFKFB4	734	3649	5,0	0,0003	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	Induced by hypoxia. Involved in glycolysis	
ALDOC	1145	5455	4,8	0,0004	aldolase C, fructose-bisphosphate	Glycolytic enzyme	
LIPF	3	15	4,6	0,0214	lipase, gastric, transcript variant 2	Involved in the digestion of dietary triglycerides	
FABP1	2585	11605	4,5	0,0107	fatty acid binding protein 1, liver	Binds fatty acids and other hydrophobic ligands	
PGM1	1472	6277	4,3	0,0215	phosphoglucomutase 1, transcript variant 1	Glycolytic enzyme	
EGLN3	135	563	4,2	0,0002	egl-9 family hypoxia-inducible factor 3	Induced by hypoxia. Adds hydroxyl groups on prolyl residues of HIF-1alpha	
PGK1	17615	70132	4,0	0,0013	phosphoglycerate kinase 1	Glycolytic enzyme	
INSM1	3	37	11,4	0,0048	insulinoma-associated 1	Transcriptional repressor	Transcription regulation
MEX3B	4	16	4,5	0,0071	mex-3 RNA binding family member B	May be involved in post-transcriptional regulatory mechanisms	
WT1	109	443	4,0	0,0011	Wilms tumor 1, transcript variant D	Transcription factor. Tumor suppressor	
GLI1	95	23	-4,1	0,0033	GLI family zinc finger 1, transcript variant 1	Transcriptional activator	
TCEANC	23	6	-4,1	0,0282	transcription elongation factor A (SII) N-terminal and central domain containing, transcript variant 2	Transcription regulation	
RUNX2	32	7	-4,4	0,0055	runt-related transcription factor 2, transcript variant 1	Transcription factor	
BARX1	9	2	-4,5	0,0002	BARX homeobox 1	Homeobox transcription factor	
EGR2	57	10	-5,8	0,0005	early growth response 2, transcript variant 1	Transcription factor	
HOXC9	23	3	-8,2	0,0082	homeobox C9	Homeobox transcription factor	
FOS	13400	1163	-11,5	0,0013	FBJ murine osteosarcoma viral oncogene homolog	Transcription factor	
TCF4	173	12	-14,0	0,0024	transcription factor 4, transcript variant 2	Transcription factor	
EGR1	19510	1167	-16,7	0,0002	early growth response 1	Transcriptional regulator. Mediates response to hypoxia.	
EGR3	57	3	-19,6	0,0027	early growth response 3, transcript variant 1	Transcriptional regulator.	
FOSB	2295	114	-20,2	0,0002	FBJ murine osteosarcoma viral oncogene homolog B	Transcription factor	
NPSR1	25	252	10,2	0,0071	neuropeptide S receptor 1, transcript variant 2	G-protein coupled receptor	Cell signaling
GPR133	3	17	6,5	0,0022	G protein-coupled receptor 133	Orphan membrane receptor. Transduces intracellular signals	
P2RY6	3	19	6,3	0,0051	pyrimidinergic receptor P2Y, G-protein coupled	G-protein-coupled receptor that responds to extracellular purine and pyrimidine nucleotides	
NPSR1	96	570	5,9	0,0179	neuropeptide S receptor 1	G-proteins coupled receptor	
KIT	1318	7499	5,7	0,0021	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog, transcript variant 1	Receptor tyrosine kinase. Activates multiple signaling pathways	
DKK1	95	424	4,5	0,0099	dickkopf WNT signaling pathway inhibitor 1	Inhibits beta-catenin-dependent Wnt signaling	
HTR1D	6	25	4,2	0,0289	5-hydroxytryptamine (serotonin) receptor 1D, G protein-coupled	G-protein coupled receptor for serotonin	
ADORA2A	3	12	4,1	0,0184	adenosine A2a receptor	Adenosine receptor mediated by G proteins activating adenylyl cyclase	
GAL	107	423	4,0	0,0081	galanin/GMAP prepropeptide	Ligand of G-protein coupled receptors, involved in smooth muscle contraction	
TGFB2	12	3	-4,0	0,0143	transforming growth factor, beta 2, transcript variant 2	Involved in many cellular processes	
TAC4	24	6	-4,0	0,0061	tachykinin 4 (hemokinin)	Neurotransmitter	
TAS2R46	14	3	-4,3	0,0124	taste receptor, type 2, member 46	May play a role in sensing the composition of the gastrointestinal content. G protein coupled receptor	
OR51E2	19	4	-4,8	0,0113	olfactory receptor, family 51, subfamily E, member 2	G-protein-coupled olfactory receptor	
RTP3	93	13	-7,3	0,0064	receptor (chemosensory) transporter protein 3	Intracellular transporter	
FGD5	5	41	7,7	0,0022	FYVE, RhoGEF and PH domain containing 5	May play a role in regulating the actin cytoskeleton and cell shape.	Cytoskeleton
KIF19	67	302	4,5	0,0030	kinesin family member 19	Microtubule-dependent motor protein	
CFAP74	7	2	-4,1	0,0012	cilia and flagella associated protein 74	Plays a role in cilium movement	
CFAP58	8	2	-4,4	0,0007	cilia and flagella associated protein 58	Plays a role in cilium movement	
GSN	11	2	-4,7	0,0087	gelsolin	Plays a role in actin polimerization. Plays a role in ciliogenesis	
SLIT1	18	4	-5,0	0,0047	slit homolog 1 (Drosophila)	Molecular guidance cue in cellular migration	
PHACTR3	9	2	-5,1	0,0040	phosphatase and actin regulator 3, transcript variant 1	Actin binding	
CLMN	13	2	-5,5	0,0051	calmin (calponin-like, transmembrane)	Actin binding	
AKR1B15	213	2803	13,1	0,0001	aldo-keto reductase family 1, member B15	NADPH-dependent reductase acting on various aromatic and non aromatic compounds	Detoxification
AKR1B10	351	4503	12,8	0,0001	aldo-keto reductase family 1, member B10	NADPH-dependent reductase acting on various aromatic and non aromatic compounds	
AKR1C1	923	7236	7,8	0,0001	aldo-keto reductase family 1, member C1	NADPH-dependent reductase acting on various aromatic and non aromatic compounds	
CYP1A1	252	2141	8,5	0,0003	cytochrome P450, family 1, subfamily A, polypeptide 1	Detoxification and synthesis of cholesterol, steroids and other lipids	
CYP1A2	193	911	4,7	0,0010	cytochrome P450, family 1, subfamily A, polypeptide 2	Detoxification and synthesis of cholesterol, steroids and other lipids	
MT1H	994	139	-7,1	0,0020	metallothionein 1H	Binds heavy metals. Protects from metals and free radicals toxicity	
MT1F	4249	955	-4,4	0,0015	metallothionein 1F	Binds heavy metals. Protects from metals and free radicals toxicity	
IL37	54	793	14,7	0,0026	interleukin 37, transcript variant 1	Suppressor of inflammatory and immune responses	Immunity and inflammation
DMBT1	3	28	8,4	0,0020	deleted in malignant brain tumors 1, transcript variant 2	Putative tumor suppressor. May play roles in mucosal defense system and cellular immune defense	
IL36G	3	20	5,8	0,0039	interleukin 36, gamma, transcript variant 1	Member of the interleukin 1 cytokine family	
CCL20	37	198	5,3	0,0011	chemokine (C-C motif) ligand 20, transcript variant 1	Chemotactic for lymphocytes	
FAM19A2	41	178	4,3	0,0131	family with sequence similarity 19 (chemokine (C-C motif)-like), member A2	Neurotrophic factor involved in neuronal survival	
ULBP1	43	8	-5,7	0,0054	UL16 binding protein 1	Ligand of a Natural Killer cells receptor	
IL11	13	2	-6,0	0,0052	interleukin 11, transcript variant 1	Stimulates antibody production by B cells	

SLC7A14	6	47	7,5	0,0237	solute carrier family 7, member 14	Aminoacid transporter	**Membrane transport**
STC1	2	11	4,6	0,0122	stanniocalcin 1	May play a role in the regulation of intestinal calcium and phosphate transport	
TCAF2	15	62	4,2	0,0014	family with sequence similarity 115, member C	Negative regulator of the plasma membrane cation channel	
KCNIP4	12	53	4,4	0,0275	Kv channel interacting protein 4, transcript variant 5	Potassium channel	
KCNG1	282	57	-4,9	0,0006	potassium channel, voltage gated modifier subfamily G, member 1	Potassium channel subunit	
KCNG1	36	5	-7,6	0,0065	potassium voltage-gated channel, subfamily G, transcript variant X2	Potassium channel	
ITGAX	4	57	15,1	0,0035	integrin, alpha X (complement component 3 receptor 4 subunit), transcript variant 2	Alpha chain of integrins. Binds fibrinogen	**Cell adhesion**
PPFIA4	9	99	11,5	0,0041	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 4	May regulate the disassembly of focal adhesions	
WISP2	4	38	9,5	0,0088	WNT1 inducible signaling pathway protein 2	inhibits the binding of fibrinogen to integrin receptors	
CDH15	6	36	6,3	0,0123	cadherin 15, type 1, M-cadherin (myotubule)	Calcium-dependent cell adhesion proteins	
PIGZ	357	1594	4,5	0,0009	phosphatidylinositol glycan anchor biosynthesis, class Z	Involved in GPI anchor biosynthesis	
PCDHB5	18	4	-4,1	0,0092	protocadherin beta 5	Potential calcium-dependent cell-adhesion protein.	
BMP2	6	95	15,2	0,0059	bone morphogenetic protein 2	Plays a role in bone and cartilage development	**Extracellular matrix**
MMP13	2	12	5,5	0,0014	matrix metallopeptidase 13 (collagenase 3)	Degradation of extracellular matrix	
SPP1	4	27	6,1	0,0025	secreted phosphoprotein 1, transcript variant 1	Binds to hydroxyapatite. Forms an integral part of the mineralized matrix	
SPRR1A	13	52	4,1	0,0014	small proline-rich protein 1A	Protein of the keratinocyte membrane	
SPRR2A	3	27	10,6	0,0005	small proline-rich protein 2A	Protein of the keratinocyte membrane	
SPRR2D	6	27	4,3	0,0004	small proline-rich protein 2D	Protein of the keratinocyte membrane	
CABP7	6	148	26,6	0,0012	calcium binding protein 7	Negatively regulates Golgi-to-plasma membrane trafficking	**Intracellular transport**
FAM71A	12	3	-4,2	0,0064	family with sequence similarity 71, member A	Important for integrity of the Golgi	
UNC13D	299	68	-4,4	0,0040	unc-13 homolog D (C. elegans)	Regulates assembly of recycling and late endosomal structures	
PPP2R2C	2	26	11,1	0,0002	protein phosphatase 2, regulatory subunit B, gamma, transcript variant 1	Negatively controls cell growth and division	**Cell growth**
LMO2	16	4	-4,0	0,0377	LIM domain only 2 (rhombotin-like 1), transcript variant 1	Crucial role in hematopoietic development	
CYR61	2166	184	-11,8	0,0021	cysteine-rich, angiogenic inducer, 61	Promotes cell proliferation, chemotaxis, angiogenesis and cell adhesion	
ASB2	2	20	8,4	0,0038	ankyrin repeat and SOCS box containing 2, transcript variant 2	Subunit of a E3 ubiquitin ligase complex that mediates the degradation of actin-binding proteins.	**Ubiquitination**
RNF183	110	580	5,3	0,0010	ring finger protein 183	E3 ubiquitin ligase	
CDRT1	13	3	-4,3	0,0068	CMT1A duplicated region transcript 1, transcript variant 1	Ubiquitin ligase	
NDRG1	3283	31299	9,5	0,0002	N-myc downstream regulated 1, transcript variant 2	Involved in stress response. Necessary for p53-dependent caspase activation and apoptosis	**Stress response and apoptosis**
UNC5B	1139	244	-4,7	0,0023	unc-5 homolog B (C. elegans), transcript variant 1	Netrin receptor activating apoptosis in the absence of its ligand	
WFDC2	47	210	4,5	0,0267	WAP four-disulfide core domain 2	Broad range protease inhibitor	**Protease regulation**
PAH	214	870	4,1	0,0030	phenylalanine hydroxylase	phenylalanine metabolism	**Aminoacid metabolism**
GDAP1L1	101	22	-4,7	0,0074	ganglioside induced differentiation associated protein 1-like 1, transcript variant 2	Expressed after neuron differentiation induced by ganglioside GD3	**Unclear**
SCGB2A1	6	27	4,6	0,0172	secretoglobin, family 2A, member 1	May bind androgens and other steroids	
EDN1	216	26	-8,4	0,0029	endothelin 1, transcript variant 1	Vasoconstricting peptide	
FAM167A	18	4	-4,7	0,0443	family with sequence similarity 167, member A	**No information**	
ANKFN1	10	2	-4,6	0,0010	ankyrin-repeat and fibronectin type III domain containing 1		
KIAA0825	12	3	-4,3	0,0054	KIAA0825		
TMEM45A	4	17	4,2	0,0131	transmembrane protein 45A		
CLIP4	2	14	6,1	0,0054	CAP-GLY domain containing linker protein family, member 4, transcript variant 1		
C4orf47	31	203	6,5	0,0022	chromosome 4 open reading frame 47		
C4orf51	2	15	8,5	0,0002	chromosome 4 open reading frame 51		

Corrected p value was calculated using the multiple test correction Benjamini-Hochberg ($p < 0.05$, fold change 3D vs 2D ≥ 4). Information on the gene role was obtained from: https://www.genecards.org

A search for genes differentially modulated by 3D culture depending on the expression or non-expression of B4GALNT2 yielded a list of 31 genes, 13 of which showed up-regulation in response to 3D culture only in S2/S11 cells, while the remaining 18 showed down-regulation in response to 3D culture only in S2/S11 cells (Table 7).

Table 7. Genes highly modulated by 3D only in B4GALNT2-expressing LS174T cells (clones S2/S11).

	Expression Neo		Expression B4GALNT2		Fold change 3D/2D Neo	Fold change 3D/2D B4GALNT2				
GeneSymbol	2D	3D	2D	3D			Corrected p value	GeneName	Role	Broad functional category
KIZ	16,2	13,4	19,0	3,8	-1,2	-4,9	0,0086	kizuna centrosomal protein	Centrosomal protein necessary to endure the forces converging on the centrosomes during spindle formation.	Cytoskeleton and mitosis
CEP120	7,3	7,5	15,8	3,1	1,0	-5,2	0,0086	centrosomal protein 120kDa	Functions in the microtubule-dependent coupling of the nucleus and the centrosome.	
DNAH6	6,3	5,3	13,1	2,3	-1,2	-5,7	0,0163	dynein, axonemal, heavy chain 6	Member of the dynein family, which are constituents of the microtubule-associated motor protein complex.	
SGOL2	8,6	7,7	18,4	2,2	-1,1	-8,3	0,0086	shugoshin-like 2 (S. pombe)	Targets PPP2CA to centromeres, leading to cohesin dephosphorylation.	
STARD13	13,5	16,3	19,8	4,2	1,2	-4,7	0,0156	StAR-related lipid transfer (START) domain containing 13	Involved in regulation of cytoskeletal reorganization, cell proliferation and motility.	
UPK1A	4,0	3,7	1,9	12,0	-1,1	6,3	0,0086	uroplakin 1A	Member of the tetraspanin family, mediates signaling. Decreased expression is associated with CRC progression and poor prognosis. (PMID: 25197375)	Cell Signaling
OR52R1	11,2	10,3	3,3	15,5	-1,1	4,8	0,0131	olfactory receptor, family 52, subfamily R, member 1 (gene/pseudogene)	Olfactory receptors are G-protein-coupled receptors involved in perception of smell and other functions.	
TAS2R45	42,6	35,9	68,8	17,2	-1,2	-4,0	0,0247	taste receptor, type 2, member 45	Taste receptors play a role in the perception of bitterness and in sensing the chemical composition of the gastrointestinal content. Some taste receptors inhibit cancer growth and stemness. (PMID: 28467517)	
TAS2R19	41,9	31,8	70,3	17,1	-1,3	-4,1	0,0116	taste receptor, type 2, member 19		
TAS2R30	254,5	215,3	402,3	81,8	-1,2	-4,9	0,0168	taste receptor, type 2, member 30		
TNFAIP8L2	4,7	6,0	2,0	14,5	1,3	7,3	0,0319	tumor necrosis factor, alpha-induced protein 8-like 2	Promotes Fas-induced apoptosis. (PMID: 28186089)	Apoptosis
MYOD1	3,4	5,7	2,3	10,4	1,7	4,4	0,0239	myogenic differentiation 1	Mediates apoptosis through caspase 3. (PMID: 28131747)	
PPM1K	16,3	11,7	25,7	4,0	-1,4	-6,5	0,0086	protein phosphatase, Mg2+/Mn2+ dependent, 1K	Regulates the mitochondrial permeability transition pore and is essential for cellular survival.	
SDPR (CAVIN2)	4,3	1,8	18,3	2,3	-2,4	-8,1	0,0106	serum deprivation response	Role in caveolar biogenesis and morphology. Metastasis suppressor and activator of apoptosis. (PMID: 26739564).	
PHF20L1	7,4	7,1	24,2	2,7	1,0	-8,9	0,0089	PHD finger protein 20-like 1	Predicted to be involved in regulation of transcription. Stabilizes SOX2 postranslationally. (PMID: 30089852)	Transcription regulation
KLF12	5,2	3,9	12,3	2,2	-1,3	-5,7	0,0235	Kruppel-like factor 12	Inhibitor of the AP-2 alpha transcription factor. Inhibits growth and anoikis resistance of ovarian cancer cells. (PMID: 28095864)	
PCF11	3,6	5,0	6,7	1,7	1,4	-4,0	0,0086	PCF11 cleavage and polyadenylation factor subunit	It is necessary for efficient Pol II transcription termination	
CTLA4	2,1	3,7	2,4	14,2	1,7	5,8	0,0136	cytotoxic T-lymphocyte-associated protein 4	Inhibitor of T cell activation.	Immunity and inflammatio
IL1A	1,7	2,8	2,1	11,7	1,7	5,6	0,0259	interleukin 1, alpha	Involved in immune responses and inflammatory processes.	
TDO2	3,1	6,5	7,3	29,0	2,1	4,0	0,0086	tryptophan 2,3-dioxygenase	In tryptophan metabolism catalyzes the first step of the kynurenine pathway. Increased kynurenine may suppress antitumor immune responses.	
FSIP2	8,3	5,8	17,0	2,9	-1,4	-5,9	0,0365	fibrous sheath interacting protein 2	Protein associated with the sperm fibrous sheath.	Fertilization
SPACA1	5,5	6,4	4,1	18,2	1,2	4,5	0,0293	sperm acrosome associated 1	Localizes to the acrosomal membrane of spermatozoa, playing a role in acrosomal morphogenesis and in sperm-egg fusion.	
USP11	6,1	7,0	3,3	13,3	1,1	4,0	0,0352	ubiquitin specific peptidase 11	Encodes a cysteine protease that cleaves ubiquitin from ubiquitin-conjugated protein substrates.	Ubiquitinatio n
ST13	6,9	12,8	2,7	11,0	1,9	4,1	0,0090	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)	Mediates the association of the heat shock proteins HSP70 and HSP90.	Protein folding
HIST4H4	81,2	193,8	101,6	406,8	2,4	4,0	0,0365	histone cluster 4, H4	Component of the nucleosome.	Chromatin structure
TRAPPC2	10,3	7,1	8,4	1,9	-1,5	-4,4	0,0086	trafficking protein particle complex 2	May play a role in vesicular transport from endoplasmic reticulum to Golgi	Intracellular transport
C8orf74	6,3	5,3	10,1	2,3	-1,2	-4,4	0,0135	chromosome 8 open reading frame 74	Little or no information	
NAALADL2	15,5	10,4	17,9	4,1	-1,5	-4,3	0,0196	N-acetylated alpha-linked acidic dipeptidase-like 2		
SAMD12	21,1	17,1	25,8	6,3	-1,2	-4,1	0,0323	sterile alpha motif domain containing 12		
FRG2	40,8	49,1	40,2	173,8	1,2	4,3	0,0138	FSHD region gene 2		
FRG2C	3,5	8,4	7,2	35,0	2,4	4,9	0,0124	FSHD region gene 2 family, member C		

Corrected p value was calculated using the multiple test correction Benjamini-Hochberg ($p < 0.05$, fold change B4GALNT2 3D vs B4GALNT2 2D). Information on the gene role was obtained from: https://www.genecards.org and from PubMed.

Yet, data collected from microarray analysis of LS174T cells grown in anchorage-independent conditions as spheroids were compared to gene expression data of TCGA cohorts (Non- and High- B4GALNT2 expressers). The rationale of the study is explained by the notion that 3D spheroids resemble the tumor *in vivo* with respect to its morphological features, microenvironment, volume growth kinetics, oxygen concentration and cell proliferation (Table 8).

Table 8. Gene expression comparison between TCGA cohort (Non- and High B4GALNT2 expressers and microarray analysis of LS174T cells (modulated by 3D only in S2/S11).

		Non-B4GALNT2 expressers	High-B4GALNT2 expressers		
Broad functional category	**Gene name**	**Mean±SD**	**Mean±SD**	**Consistency**	**p-value**
Cytoskeleton and mitosis	KIZ (PLK1S1)	346±302	250±154	Yes	≤0.01
	CEP120	404±116	340±135	Yes	≤0.01
	DNAH6	33±28	27±15	Yes	≤0.05
	SGOL2	316±124	273±138	Yes	≤0.01
	STARD13	529±271	355±211	Yes	≤0.01
Cell signaling	UPK1A	16±112	5±21	Yes	N.S.
Apoptosis	TNFAIP8L2	46±38	58±42	Yes	≤0.05
	PPM1K	97±133	10±45	No	
	SDPR	186±271	153±162	Yes	N.S
Transcription regulation	PHF20L1	941±296	794±349	Yes	≤0.01
	KLF12	239±199	179±128	Yes	≤0.01
	PCF11	1025±302	856±227	Yes	≤0.01
Immunity and inflammation	CTLA4	47±85	42±34	No	
	IL1A	34±86	69±215	Yes	≤0.05
	TDO2	232±885	115±156	No	

The mean level of expression in TCGA database of genes selectively modulated by 3D growth only in S2/S11 cells (Table 7) was compared in the cohorts of non-B4GALNT2 expressers and high B4GALNT2 expressers. The column "Consistency" indicates whether the difference observed in the cohorts was consistent with that reported in Table 7. Genes showing statistically significant consistent difference are indicated in bold ($p \leq 0.05$ Student's t test for independent samples). N.S. = non-significant. A few genes present in Table 7 are not present in this Table because they were not present in TCGA or not expressed.

CHAPTER V - RESULTS

Transcriptomic and phenotypic impact of B4GALNT2 expression in SW480 and SW620 CRC cells

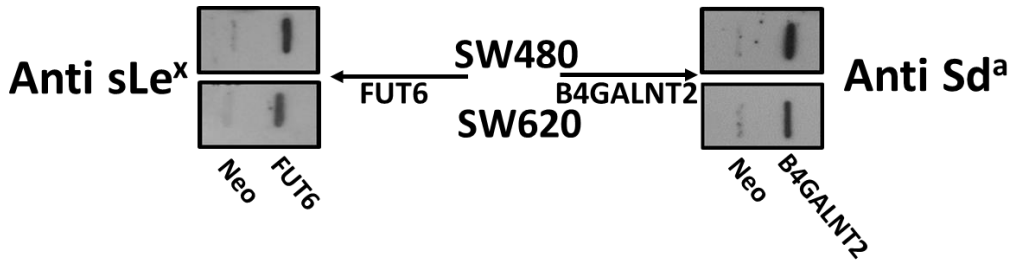
Note: Results presented in this section were taken from the published manuscript:

- Pucci, Michela; Gomes Ferreira, Inês; Malagolini, Nadia; Ferracin, Manuela and Dall'Olio, Fabio "The Sda Synthase B4GALNT2 Reduces Malignancy and Stemness in Colon Cancer Cell Lines Independently of Sialyl Lewis X Inhibition" *Int. J. Mol. Sci.* **2020**, 21, 6558

5.1 Transfection of SW480 and SW620 with FUT6 and B4GALNT2 cDNAs

Stemming from TCGA transcriptomic data that indicate CRC patients with higher B4GALNT2 expression level display longer overall survival and the significant impact of B4GALNT2 expression on the phenotype of LS174T cells, the second part of the research aimed at understanding whether the effects were due to the expression of the Sd^a antigen or inhibition of sLe^x antigen or both. To this scope, the expression of FUT6 or B4GALNT2 was forced in the CRC cell lines SW480 and SW620. The two colon cancer cell lines were chosen essentially for two reasons. First, they both lacked FUT6 and B4GALNT2 enzyme activities and sLe^x and Sd^a antigens. Second, they were derived from the primary tumor (SW480) and a lymph node metastasis (SW620) of the same patient, allowing the investigation of the effect of the two enzymes at different tumor stages. In both SW480 and SW620 cell lines, transfection of *B4GALNT2* cDNA induced the expression of Sd^a antigen, whereas introduction of FUT6 plasmid led to sLe^x expression. The mRNA level of the two glycosyltransferases, as detected by microarray analysis, was negligible in mock transfectants but well expressed in the respective glycosyltransferase transfectants (Figure 22A). In FUT6 transfectants the level of FUT6 mRNA was very high, in B4GALNT2 transfectants the level of B4GALNT2 mRNA was lower, although sufficient to ensure a good level of Sd^a antigen expression (Figure 22 B).

A



B

FUT6 mRNA *expression*			B4GALNT2 mRNA *expression*	
SW480	SW620	Transfection	SW480	SW620
4	3	Mock	7	14
11430	1390	FUT6	2	3
7	3	B4GALNT2	117	62

Figure 22. Biochemical characterization of FUT6 or B4GALNT2-transfected cell lines. **A**: Slot blot analysis of Neo cells and of FUT6- or B4GALNT2-transfected clones with anti sLe^x or anti Sd^a antibodies. **B**: mRNA expression determined by microarray analysis and expressed in arbitrary units of the two enzymes in Neo (mock transfectants) or populations transfected with FUT6 or B4GALNT2.

5.2 Phenotypic changes induced by B4GALNT2 and FUT6 expression

To unravel the relative contribution of the Sd^a and sLe^x antigens to the phenotype of the two cell lines, the following aspects of cancer cells growth and malignancy were analyzed *in vitro*.

- **DOUBLING TIME ASSAY**

In order to evaluate the incidence of the two antigens on the cellular proliferation rate of the two lines, the doubling time (DT) was measured. The aim was to evaluate the time it takes for a population to double its size. In this assay, the cells were counted 24 hours apart three times. At the end of the measurements intervals it was found that expression of B4GALNT2 reduced the growth rate of SW620, increasing the doubling time from 23 ± 3 to 28 ± 4 h (Figure 23A). Although a tendency to increased speed of growth in FUT6-expressing SW620 cells was observed, it did not reach statistical significance. On the other hand, no effect of either glycosyltransferase was observed on the doubling time of SW480.

- **CLONOGENIC ASSAY**

To investigate whether the presence of either glycosyltransferase and their cognate antigens was able to affect the number of cells able to generate a colony in standard conditions of growth it was performed a colony formation assay. A small number of mock-transfected and FUT6- and B4GALNT2- SW480 and SW620 transfected cells were seeded in standard conditions in a way that they could grow independently. After the incubation time, the number of cells able to form colonies resulted to be affected by both glycosyltransferases in both cell lines (Figure 23 B). The cell line SW620 was provided with a higher capability to form colonies, compared to SW480. However, in both cell lines FUT6 expression induced a small but significant increase of the clonogenic ability which, by contrast, was strongly impaired by B4GALNT2.

- **ANCORAGE INDEPENDENT GROWTH IN SOFT AGAR**

SW480 and SW620 cells were analyzed for their ability to grow without anchorage to a solid surface by a soft agar assay. In this study, the cells were plated as a single cell within a layer of agar to mimic the growth in absence of adhesion. After the incubation time, SW480- and SW620- FUT6 and B4GALNT2 transfected cells were compared with mock-transfected Neo cells for their ability to form colonies. In both cell lines, B4GALNT2 expression significantly reduced the formation of clones, while FUT6 induced a slight increase of clone formation only in SW620 cells (Figure 23 C).

- **3D TUMOR SPHEROIDS**

SW480 and SW620 cells transfected either with FUT6 or B4GALNT2 and mock-transfected were evaluated for their capacity to survive and grow in a liquid medium through a 3D tumor spheroids assay. This assay measures the ability of the cells to survive and proliferate without any mechanical anchorage, a condition associated with stemness, even more drastic than soft agar growth. SW480 cells formed mainly rounded spheroids with regular edges, whereas SW620 cells formed spheroids with irregular shape and many cells were found as single. B4GALNT2 expression reduced the formation of spheroids in both cell lines while FUT6 had no effect

(Figure 23 D), as resulted from protein quantification described in “Materials and methods” section.

- **WOUND HEALING ASSAY**

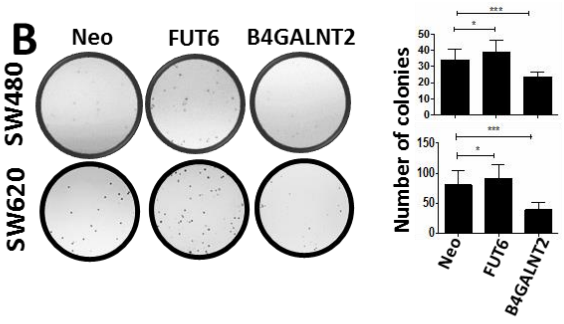
To evaluate whether the B4GALNT2 and FUT6 expression could modify the ability of cells to proliferate and migrate (an important feature associated with malignant transformation), wound healing assay was performed with SW480- , SW620- FUT6 and -B4GALNT2 cells and mock-transfected Neo cells. The ability to heal a wound in a layer of confluent cells provided an example of the differential response of the two cell lines to glycosyltransferase expression. In fact, in the cell line SW480 the expression of either FUT6 or B4GALNT2 left unaltered the ability to heal a wound. On the contrary, in the cell line SW620 the healing capability was greatly enhanced by FUT6 but reduced by B4GALNT2 (Figure 23 E).

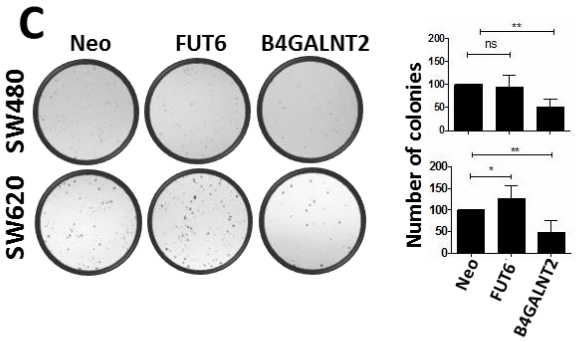
- **ALDEFLUOR ASSAY**

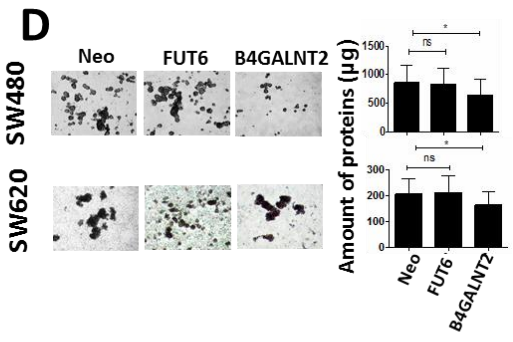
To investigate whether the expression of the two glycosyltransferases and their cognate carbohydrate antigens could affect the stemness property of SW480 and SW620 lines, mock-transfected and FUT6- and B4GALNT2- expressing cells were analyzed for the expression of aldehyde dehydrogenase (ALDH).This assay for stemness revealed that B4GALNT2 induced a marked down-regulation of the number of stem cells in both cell lines (Figure 23 F). Unexpectedly, FUT6 induced a slight reduction of ALDH positive cells, which reached statistical significance only in SW480.

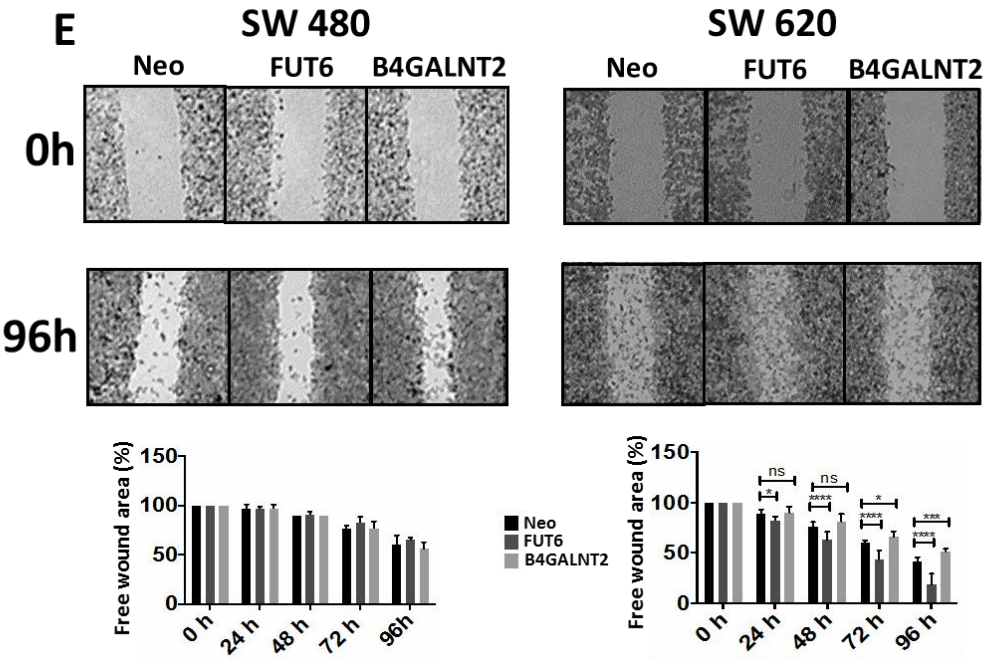
A

Doubling time (DT)						
	SW480			SW620		
	DT	SD	*p*	DT	SD	*p*
Neo	25	1		23	3	
B4GALNT2	25	2	N.S.	28	4	0.003
FUT6	26	3	N.S.	22	2	0.16









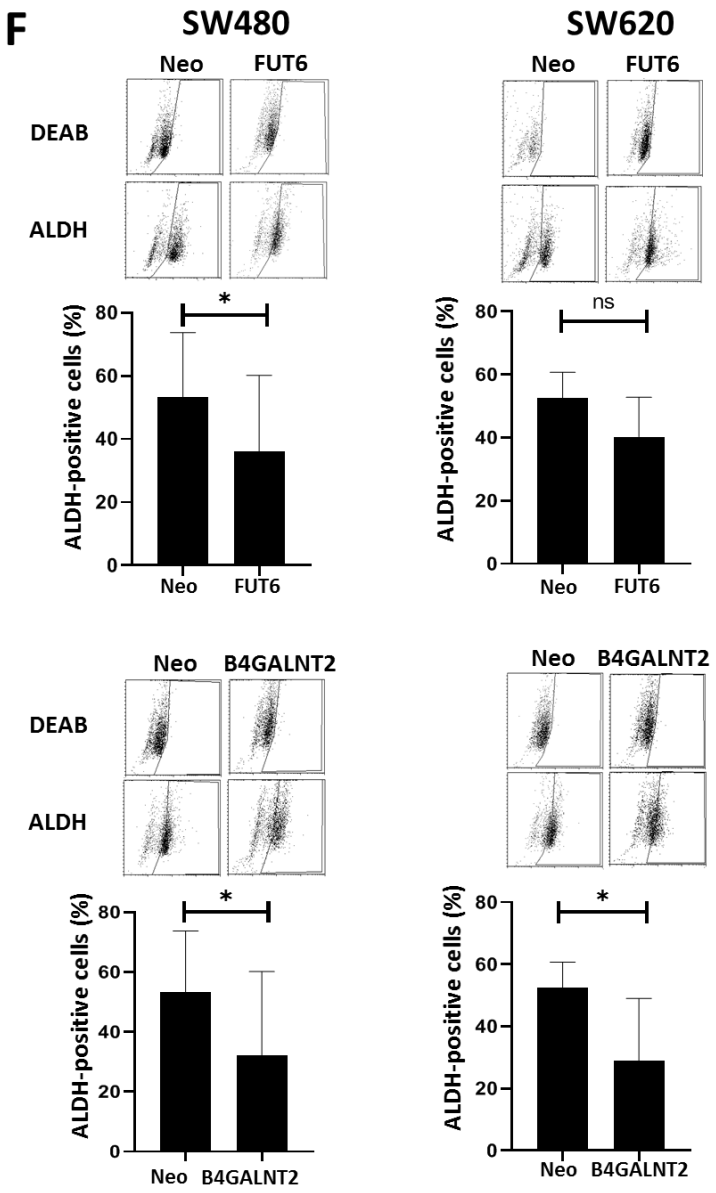


Figure 23. Phenotypic effects induced by FUT6 or B4GALNT2 expression. (A) Doubling time (DT),expressed in hours. (B) Colony formation assay in standard conditions of growth. Histograms indicate the total number of colonies. (C) Growth in 0.33% soft agar. Photographs were taken without magnification and the colonies visible to the naked-eye were counted. (D) Spheroids formation assay. The aspect of the spheroids is shown. The total amount of protein was calculated and taken as a measure of the cells grown in 3D conditions. (E) Wound healing assay. The free area of the wound was quantitated by ImageJ and normalized to the free area of the same cell line at 0 h, which was taken as 100%. The photographs show only the start (0 h) and the end point (96 h) of the healing process. Graphs in the bottom report the quantification of the healing process at each time point. The microphotographs were taken at a 4x magnification. (F) ALDEFLUOR assay. Cells were incubated with ALDEFLUOR either in the presence or in the absence of the inhibitor N,N-diethylaminobenzaldehyde (DEAB). Gates excluding all of the cells labelled in the presence of DEAB were set. Cells included in the gate in the absence of DEAB, were considered to be ALDH positive. Histograms report the percentage of ALDH positive cells ±SD in five independent experiments. Statistical analysis was performed using one-way analysis of variance (ANOVA) and Dunnett's multiple comparisons test. ns,not significant; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$.

5.3 Impact of B4GALNT2 expression on the transcriptome of SW480 and SW620 colon cancer cells

To establish whether and how the overexpression of FUT6 and B4GALNT2 could modify the gene expression profile of colorectal cancer cells, two independent RNA preparations of the six cell lines (SW480 and SW620 transfected with FUT6 or B4GALNT2 or mock-transfected) were microarray-analyzed.

Genes significantly and consistently modulated either by FUT6 or B4GALNT2 in both SW480 and SW620 cells were identified. The impact on the transcriptome of the two glycosyltransferases was assessed by Pathway Enrichment analysis and found to be very different. In fact, FUT6 modulated 1779 genes while B4GALNT2 modulated only 128 genes. Figure 24 is a heat-map of the genes modulated by FUT6 or B4GALNT2, compared with Neo in both SW480 and SW620 cell lines.

Analysis showed a very different predicted impact on cell behavior (Table 8 A and B). Both glycosyltransferases appeared to impact cell cycle, cytoskeleton, and cell adhesion.

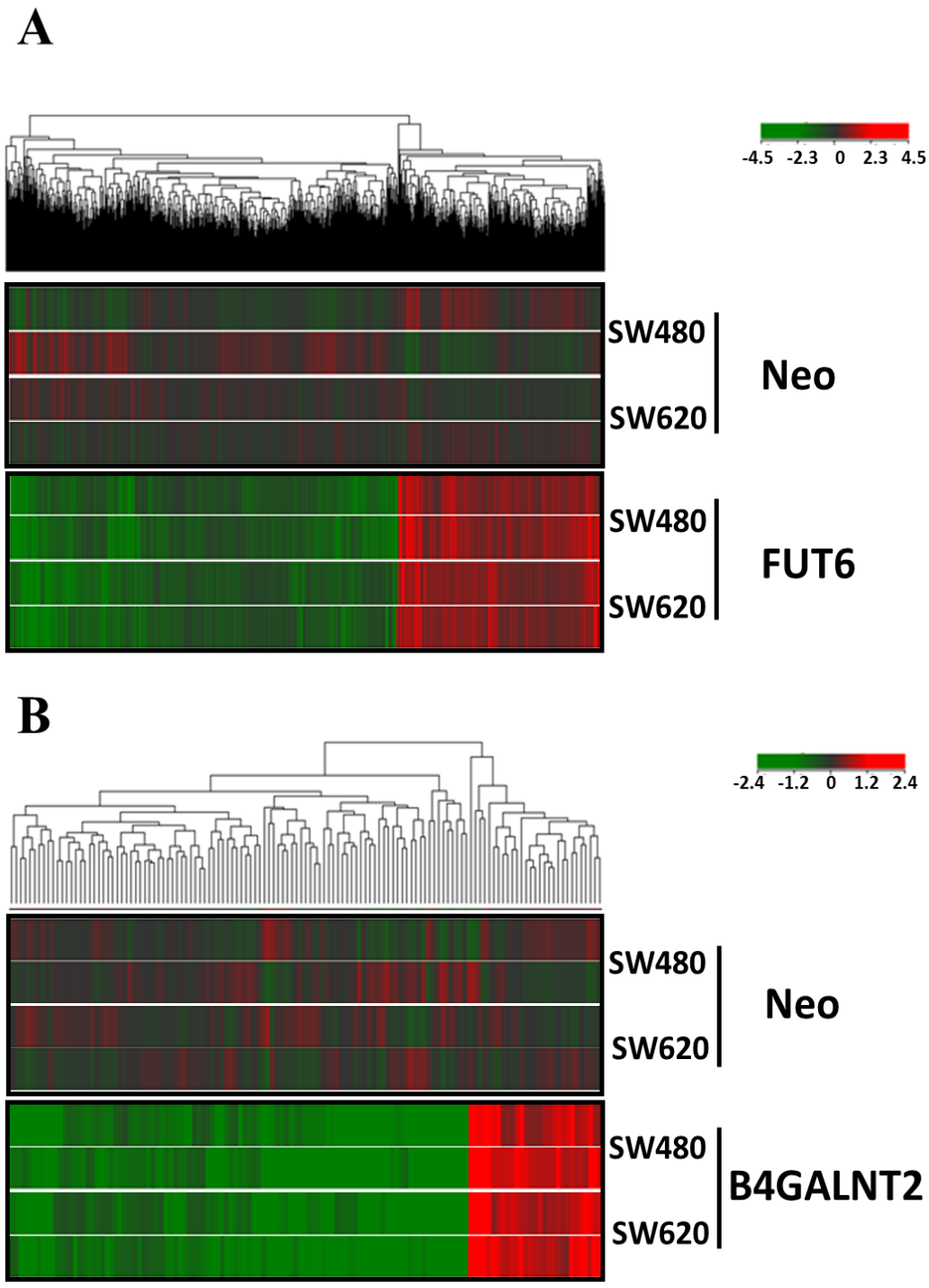


Figure 24. Heatmaps of genes modulated upon FUT6 and B4GALNT2 expression. A: Cluster analysis of SW480 and SW620 cells transfected with FUT6 and compared with control Neo. **B**: Cluster analysis of SW480 and SW620 cells transfected with B4GALNT2 and compared with control Neo. Differentially expressed genes are reported. Genes (columns) and samples (rows) were grouped by hierarchical clustering (Manhattan correlation). High- and low-expression was normalized to the average expression across all samples. Differences were analyzed by the moderated t-test. Corrected p-value cut-off: 0.05; multiple test correction used: Benjamini–Hochberg. Color codes refer to the level of up- or down-regulation.

In Table 9A and 9B are shown the ten most relevant networks modulated by FUT6 and B4GALNT2 expression.

Table 9. Networks modulated upon FUT6 and B4GALNT2 expression.

A: FUT6 modulated

Networks	Network Objects
Cell cycle_Mitosis	Cyclin B1, Cyclin B, Cyclin B2, Histone H3, PBK, Cyclin A, PLK1, Securin, CENP-H, SIL, Separase, HZwint-1, CENP-F, CAP-G/G2, Aurora-A, CDC25, CDC25C, Tubulin beta, KNSL1, CAP-E, AF15q14, HEC, CENP-E, TPX2, SPBC25, ASPM, MAD2a, Survivin, BUB1, CAP-C, Actin, Histone H1, CENP-A
Cell cycle_Core	CDC45L, Cyclin B1, Cyclin B, Cyclin B2, Cyclin A, PLK1, Securin, RPA3, CENP-H, Separase, CAP-G, Aurora-A, CDC25C, CAP-E, p18, HEC, p21, CENP-E, ORC6L, CKS2, MAD2a, Survivin, BUB1, CAP-C, CENP-A
Cytoskeleton_Spindle microtubules	Cyclin B1, Cyclin B, Cyclin B2, KIF4A, DEEPEST, PLK1, Securin, CENP-H, GTSE1, Separase, HZwint-1, CENP-F, Aurora-A, Tubulin beta, KNSL1, Tau (MAPT), HEC, MKLP2, CENP-E, CKS2, MAD2a, BUB1, CENP-A
Development_Regulation of angiogenesis	MMP-9, FOXM1, IL-8, PKC, CD13, Oct-3/4, TRIP6, TrkB, GLI-1, WT1, DBH, Cathepsin B, Ephrin-B, Ephrin-A, PLC-beta, Galpha(i)-specific peptide GPCRs, c-Myc, IL8RB, Galpha(q)-specific peptide GPCRs, PI3K reg class IA, STAT5, IL-15, Ephrin-A receptors, p21, Plasminogen, Angiostatin, Plasmin, Ephrin-B receptor 4, Ephrin-B receptors, IP3 receptor, IL-1RI, Ihh, Hedgehog, EGFR, PLAUR (uPAR), EDNRB
Cell cycle_G2-M	FOXM1, Cyclin B1, Cyclin B, Cyclin B2, Histone H3, MYRL2, MRLC, Cyclin A, Cyclin A2, PLK1, Securin, GTSE1, Claspin, CAP-G, CAP-G/G2, Aurora-A, CDC25, CDC25C, RGC32, KNSL1, CAP-E, c-Myc, p21, Rad51, BLM, CKS2, MAD2a, BUB1, EGFR, CAP-C, Histone H1.5, Histone H1, FANCD2
Cell cycle_S phase	CDC45L, Cyclin B1, Cyclin B, Cyclin B2, Histone H3, Cyclin A, Cyclin A2, Histone H4, PLK1, Securin, RPA3, Separase, DRF1, PDS5, RGC32, PRIM2A, p21, ORC6L, Rad51, AHR, DDX11, BUB1, Histone H1.5, Histone H1, Sgo1
Cell cycle_Meiosis	Cyclin B1, HSP70, Cyclin A, GCNF, PARD3, PLK1, Securin, SMC1L2, FANCG, RAD54L, Separase, CDC25C, Tubulin beta, c-Myc, PP2A regulatory, PI3K reg class IA, Rad51, BLM, RAD54B, EGFR
Development_Neurogenesis_ Synaptogenesis	FGF7, APOE, Syntaxin 1A, TrkB, ErbB3, nAChR alpha, WNT, Ephrin-B3, Ephrin-B, Neurexin beta, Ionotropic glutamate receptor, Kainate receptor, NT-4/5, Neuregulin 2, NMDA receptor, Frizzled, MAGI-1(BAIAP1), FGFR2, Synaptotagmin VII, Synaptotagmin, Ephrin-B receptors, X11, FGFR4, Endophilin A3, Actin, NR1
Cell adhesion_Attractive and repulsive receptors	5T4, Semaphorin 3A, MENA, SLIT1, c-Fes, UNC5B, AF-6, Ephrin-B3, Ephrin-B, Ephrin-A, Ephrin-A3, Ephexin, Tau (MAPT), L1CAM, PI3K reg class IA (p55-gamma), PI3K reg class IA, Ephrin-A receptors, Ephrin-A receptor 3, Collagen XIII, Ephrin-B receptor 4, Ephrin-B receptors, RHO6, Actin, Integrin, Intersectin
Development_Neurogenesis_Axonal guidance	AHNAK, Syntenin 2, APOE, Semaphorin 3A, PKA-reg (cAMP-dependent), PARD3, CRMP4, TrkB, MENA, SLIT1, c-Fes, Ryanodine receptor 1, UNC5B, Ephrin-B3, Ephrin-B, Ephrin-A, Ephrin-A3, PLC-beta, NT-4/5, L1CAM, PI3K reg class IA, Ephrin-A receptors, Ephrin-A receptor 3, Guanine deaminase, Ephrin-B receptor 4, Ephrin-B receptors, RHO6, IP3 receptor, Actin, Integrin

Ten most relevant networks altered by FUT6 expression. Pathway map visualization was performed using MetaCore pathway analysis by GeneGo.

B: B4GALNT2 modulated

Networks	Network Objects
Cell adhesion_Cell-matrix interactions	Mindin, Galectin-7, CD44 (EXT), CD44 (ICD), CD44 soluble, CD44
Cell cycle_S phase	Rad51, MCM10, ORC6L, RGC32
Development_Neurogenesis_Axonal guidance	DISC1, Mindin, Semaphorin 3B, PLC-beta, Netrin-1
Cytoskeleton_Intermediate filaments	Tubulin beta 2, Tubulin beta, Kinesin heavy chain
Reproduction_GnRH signaling pathway	mGluR8, Galpha(i)-specific metabotropic glutamate GPCRs, PLC-beta, PLC-beta1
Cytoskeleton_Regulation of cytoskeleton rearrangement	SPTBN(spectrin1-4), Tubulin beta 2, Tubulin beta, CD44
Reproduction_Gonadotropin regulation	mGluR8, Galpha(i)-specific metabotropic glutamate GPCRs, PLC-beta, PLC-beta1
Cytoskeleton_Cytoplasmic microtubules	Tubulin beta, Kinesin heavy chain, KIF5A
Reproduction_Feeding and Neurohormone signaling	CD44, PLC-beta, PLC-beta1, AKR1C1
Transport_Bile acids transport and its regulation	MRP3, AKR1C1

Ten most relevant networks altered by B4GALNT2 expression. Pathway map visualization was performed using MetaCore pathway analysis by GeneGo.

To obtain more detailed information on the possible phenotypic effects of the expression of the two glycosyltransferases, the analysis was restricted to the genes showing a level of expression either in Neo or in glycosyltransferase-transfected cells ≥ 50 and a fold change ≥ 3 for FUT6 (owing to the much greater number of modulated genes) (Table 10) or ≥ 2 for B4GALNT2 (Table 11). The main role of each gene was deduced from the GeneCards web site (https://www.genecards.org/). Out of the 63 FUT6-modulated genes reported in Table 10, 10 displayed up-regulation, whereas 53 were down-regulated. Among the 45 genes included in Table 11, 11 showed up-regulation while 34 displayed down-regulation in B4GALNT2 expressing SW480 and SW620 cells.

Table 10. Genes up- or down-regulated in both SW480 and SW620 cells in response to FUT6 expression.

Gene symbol	Mean Neo	Mean FUT6	Gene Name	Function
KCNJ2	14.1	54.3	potassium channel, inwardly rectifying subfamily J, member 2	Allows potassium to flow into the cell
GPAT2	42.8	158.6	glycerol-3-phosphate acyltransferase 2, mitochondrial	Involved in processing step during piRNA biosynthesis.
CPLX2	19.3	68.9	complexin 2	Positively regulates a late step in exocytosis of various cytoplasmic vesicles
APOH	57.6	185.4	apolipoprotein H (beta-2-glycoprotein I)	Binds to various kinds of negatively charged substances such as heparin, phospholipids, and dextran sulfate
SNORA30	90.3	284.3	small nucleolar RNA, H/ACA box 30	Small nucelolar RNA
OXR1	143.2	446.1	oxidation resistance 1	May be involved in protection from oxidative damage
MYH7B	3383.7	10495.5	myosin, heavy chain 7B, cardiac muscle, beta	Involved in muscle contraction
CENPI	27.3	84.3	centromere protein I	Involved in accurate chromosome alignment and segregation
CAPN15	254.5	767.0	calpain 15	May function as a transcription factor
SNORA62	272.5	815.1	small nucleolar RNA, H/ACA box 62	Small nucelolar RNA
CD55	338.7	114.6	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	Inhibits complement activation
FXYD4	54.0	18.1	FXYD domain containing ion transport regulator 4	Modulates the properties of the Na,K-ATPase
BRSK2	93.3	31.0	BR serine/threonine kinase 2	Plays a role in the regulation of the mitotic cell cycle progress and the onset of mitosis. Regulates reorganization of the actin cytoskeleton
TMEM255B	274.5	89.8	transmembrane protein 255B	Little or no information
CFAP70	161.0	52.2	cilia and flagella associated protein 70	Little or no information
XDH	156.0	50.5	xanthine dehydrogenase	Key enzyme in purine degradation
MCAM	1107.7	355.3	melanoma cell adhesion molecule	Plays a role in cell adhesion
CXCL8	231.1	73.8	chemokine (C-X-C motif) ligand 8	Chemotactic factor
C11orf96	5360.3	1701.6	chromosome 11 open reading frame 96	Little or no information

NCF2	130.0	41.2	neutrophil cytosolic factor 2	Subunit of the NADPH oxidase complex found in neutrophils, which produces superoxide to kill bacteria
RASGEF1A	128.6	40.7	RasGEF domain family, member 1A	Guanine nucleotide exchange factor specific for RAS
FAM228B	347.8	109.9	family with sequence similarity 228, member B	Little or no information
CAPN5	1312.9	413.2	calpain 5	Calcium-dependent cysteine protease involved in signal transduction
SLIT1	60.3	18.8	slit homolog 1 (Drosophila)	Acts as molecular guidance cue in cellular migration
BEX2	9397.4	2922.5	brain expressed X-linked 2	Regulator of mitochondrial apoptosis and G1 cell cycle. Regulates transcription. Tumor suppressor.
WDR78	53.2	16.5	WD repeat domain 78	Little or no information
CAPN8	72.3	22.2	calpain 8	Involved in membrane trafficking in mucus cells
SCEL	1132.9	343.4	sciellin	May function in the assembly or regulation of proteins in the cornified envelope
PTPN13	109.7	33.2	protein tyrosine phosphatase, non-receptor type 13 (APO-1/CD95 (Fas)-associated phosphatase)	Tyrosine phosphatase which regulates negatively FAS-induced apoptosis
TPSAB1	63.6	18.8	tryptase alpha/beta 1	Tryptases are trypsin-like serine proteases
BHLHE41	489.9	143.2	basic helix-loop-helix family, member e41	Transcriptional repressor involved in the regulation of the circadian rhythm
FILIP1	58.2	16.9	filamin A interacting protein 1	By acting through a filamin-A/F-actin axis, it controls the start of neocortical cell migration
REPS2	59.3	17.0	RALBP1 associated Eps domain containing 2	Involved in growth factor signaling
GRB10	1072.8	304.4	growth factor receptor-bound protein 10	Binds to insulin and insulin like growth-factor receptors, inhibiting signaling
GBP3	84.3	23.8	guanylate binding protein 3	Encodes a member of the guanylate-binding protein (GBP) family
TMEM159	512.0	137.6	transmembrane protein 159	Little or no information
HRK	708.6	188.6	harakiri, BCL2 interacting protein	Promotes apoptosis by interacting with the apoptotic inhibitors BCL-2 and BCL-X(L)
DIP2C	149.4	39.4	DIP2 disco-interacting protein 2 homolog C (Drosophila)	May be a transcription factor binding
SERPINE2	942.6	244.0	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 2	Inhibits serine proteases

CPLX1	234.5	59.5	complexin 1	Positively regulates a late step in exocytosis of various cytoplasmic vesicles
TFPI	58.0	14.4	tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)	Serine protease inhibitor that regulates the tissue factor (TF)-dependent pathway of blood coagulation
ADRBK2	208.1	51.5	adrenergic, beta, receptor kinase 2	Phosphorylates the agonist-occupied form of the β-adrenergic receptor
ZIC5	54.9	13.5	Zic family member 5	May act as a transcriptional repressor
MIA	75.7	18.5	melanoma inhibitory activity	Growth inhibitor
EPAS1	99.7	24.1	endothelial PAS domain protein 1	Transcription factor involved in the induction of oxygen regulated genes
PEAR1	57.1	13.4	platelet endothelial aggregation receptor 1	Platelet receptor that signals upon the formation of platelet-platelet contacts
CREB5	102.7	23.9	cAMP responsive element binding protein 5	Binds to the cAMP response element and activates transcription
AHNAK2	98.3	22.9	AHNAK nucleoprotein 2	May play a role in calcium signaling by associating with calcium channel proteins
ZNF462	125.8	28.7	zinc finger protein 462	May be involved in transcriptional regulation
C16orf45	198.0	43.5	chromosome 16 open reading frame 45	Little or no information
AKAP12	1432.1	303.7	A kinase (PRKA) anchor protein 12	Associates with protein kinases and phosphatase, serving as a scaffold protein in signal transduction
PRDM13	177.8	36.3	PR domain containing 13	Little or no information
BEST1	177.2	34.8	bestrophin 1	Forms calcium-sensitive chloride channels
NTN4	58.7	11.5	netrin 4	Netrins are laminin-related proteins
GPR126	65.2	12.4	G protein-coupled receptor 126	G-protein coupled receptor which is activated by type IV collagen
ANTXR2	175.4	32.6	anthrax toxin receptor 2	Necessary for cellular interactions with laminin and the extracellular matrix
TIMP3	138.5	25.4	TIMP metallopeptidase inhibitor 3	Inactivates metalloproteases
TUBB2B	1559.3	247.2	tubulin, beta 2B class IIb	Major constituent of microtubules
TRPV6	212.5	31.5	transient receptor potential cation channel, subfamily V, member 6	Mediates Ca(2+) uptake in various tissues, including the intestine
HES7	84.9	10.4	hes family bHLH transcription factor 7	Transcriptional repressor
HS3ST1	286.8	32.5	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	Involved in heparan sulfate biosynthesis
CALB2	130.7	14.4	calbindin 2	Calcium binding
CALCA	52.6	2.2	calcitonin-related polypeptide alpha	Calcitonin and related receptors are a family of G-protein-coupled receptors

"Mean Neo" and "Mean FUT6" represent the mean expression value of SW480 and SW620 Neo- and FUT6-expressing cells respectively. Are reported only protein coding genes showing a fold change "Mean FUT6 / Mean Neo" ≥3, a corrected p value ≤ 0.05 and a level of expression either in Neo or in FUT6 ≥ 50.

Table 11. Genes up- or down-regulated in both SW480 and SW620 cells in response to B4GALNT2 expression.

Gene symbol	Mean Neo	Mean B4GALNT2	Gene name	Function
G0S	785	1603	G0/G1 switch 2	Promotes apoptosis by preventing the formation of protective BCL2-BAX heterodimers
ORC6	468	898	Origin recognition complex, subunit 6	Coordinates chromosome replication and segregation with cytokinesis
CDCA5	7025	12027	Cell division cycle associated 5	Regulator of sister chromatid cohesion in mitosis
MCM10	268	457	Minichromosome maintenance complex component 10	Acts as a DNA replication initiation factor. Prevents DNA from damage during replication
MAF	55	91	V-maf avian musculoaponeurotic fibrosarcoma oncogene homolog	Can be a transcriptional activator or repressor. Behaves as an oncogene or a tumor suppressor
PPIH	3920	6365	Peptidylprolyl isomerase H (cyclophilin H)	Assists protein folding
RAD51	2049	3381	RAD51 recombinase	Involved in DNA repair through homologous recombination
ANKRD32 (SLF1)	354	575	Ankyrin repeat domain 32	Involved in the DNA damage response and genomic stability maintenance
SLC43A3	1721	2788	Solute carrier family 43, member 3	Putative transporter

IL18	532	829	Interleukin 18	Proinflammatory cytokine
MTA2	1835	2889	Metastasis associated 1 family, member 2	Involved in transcription regulation as repressor and activator, interacting with histones
ZNF276	138	90	Zinc finger protein 276	May be involved in transcriptional regulation.
DDR1	1881	1228	Discoidin domain receptor tyrosine kinase 1	Receptor tyrosine kinase acting as a cell surface adhesion molecule, regulating migration and proliferation
PLLP	736	471	Plasmolipin	Could participate in ion transport events
IDS	373	236	Iduronate 2-sulfatase	Lysosomal enzyme involved in the degradation of dermatan- and heparan sulfate
SPIRE2	1942	1243	Spire-type actin nucleation factor 2	Actin nucleation factor involved in intracellular vesicle transport and for asymmetric cell division during meiosis
NXN	358	232	Nucleoredoxin	Functions as a redox-dependent negative regulator of Wnt signaling and as a transcriptional regulator.
MAGED2	4889	2894	Melanoma antigen family D, 2	Regulates NaCl co-transporters
LGALS7	1245	740	Lectin, galactoside-binding, soluble, 7	Pro-apoptotic galectin

SPON2	9495	5535	Spondin 2, extracellular matrix protein	Functions as an opsonin for macrophage phagocytosis of bacteria.
ZNF83	80	47	Zinc finger protein 83	May be involved in transcriptional regulation
RNF157	66	38	Ring finger protein 157	Ubiquitin ligase preventing apoptosis. Acts as a downstream effector of the PI3K and MAPK signaling
FILIP1L	81	45	Filamin A interacting protein 1-like	When overexpressed in endothelial cells, leads to inhibition of cell proliferation and migration and an increase in apoptosis.
SRPK3	138	78	SRSF protein kinase 3	Phosphorylates the SR splicing factor SRSF1
ABCC3	2539	1440	ATP-binding cassette, sub-family C (CFTR/MRP), member 3	May act as an inducible transporter in the biliary and intestinal excretion of organic anions
RHBDF1	1118	626	Rhomboid 5 homolog 1 (Drosophila)	Regulates ADAM17 protease, releasing epidermal growth factor (EGF) receptor ligands and TNF
PTPRN2	4125	2379	Protein tyrosine phosphatase, receptor type, N polypeptide 2	Regulates PI(4,5)P2 level in the plasma membrane and actin dynamics related to cell migration and metastasis
BAIAP3	191	103	BAI1-associated protein 3	Functions in endosome to Golgi retrograde transport.
SLC4A11	1517	740	Solute carrier family 4, sodium borate transporter, member 11	Sodium-coupled borate cotransporter that is essential for borate homeostasis,

SEMA3B	8740	4612	Sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B	Inhibits axonal extension and acts as a tumor suppressor by inducing apoptosis
TUBB2A	6276	3199	Tubulin, beta 2A class IIa	Component of microtubules, key participants in processes such as mitosis and intracellular transport
AKR1C1	721	360	Aldo-keto reductase family 1, member C1	In the liver and intestine, it may have a role in the transport of bile
COL7A1	299	142	Collagen, type VII, alpha 1	May contribute to epithelial basement membrane organization and adherence
SYT13	1037	497	Synaptotagmin XIII	May be involved in transport vesicle docking to the plasma membrane
KRTAP3-2	220	89	Keratin associated protein 3-2	Member of the keratin-associated protein (KAP) family
KRT15	1421	610	Keratin 15, type I	Component of the intermediate filaments
SPTBN5	250	99	Spectrin, beta, non-erythrocytic 5	Binds actin and kinesin
EMP1	16043	6672	Epithelial membrane protein 1	Little or no information
RGCC	842	328	Regulator of cell cycle	Overexpression activates or suppresses cell cycle progression
CDHR2	322	132	Cadherin-related family member 2	Involved in cell-cell adhesion and contact inhibition in epithelial cells. Candidate tumor suppressor
PLCB1	536	227	Phospholipase C, beta 1 (phosphoinositide-specific)	Produces the second messenger molecules diacylglycerol (DAG) and

				inositol 1,4,5-trisphosphate (IP3)
IQCH	68	24	IQ motif containing H	May play a regulatory role in spermatogenesis
DISC1	119	35	Disrupted in schizophrenia 1	Positively regulates Wnt-mediated proliferation. Plays a role in the microtubule network formation
CDIP1	52	15	Cell death-inducing p53 target 1	Acts as an important p53-apoptotic effector
CD44	475	108	CD44 molecule (Indian blood group)	Receptor for hyaluronic acid and other ligands

"Mean Neo" and "Mean B4GALNT2" represent the mean expression value of SW480 and SW620 Neo- and B4GALNT2-expressing cells respectively. Are reported only protein coding genes showing a fold change "Mean B4GALNT2 / Mean Neo" ≥ 2, a p value ≤ 0.05 and a level of expression either in Neo or in B4GALNT2 ≥ 50.

A deeper analysis of gene expression data from SW480 and SW620 cells revealed that several genes appeared modulated by FUT6 expression in either SW480 or SW620 cells or both. To simplify the analysis, genes were divided by functional classes as indicated in Table 12.

Table 12. Genes modulated by FUT6 expression in either SW480 or SW620 cells or both.

Functional class	Both in SW480 and SW620	Only in SW480	Only in SW620
Apoptosis	*BEX2; PTPN13; HRK*		*RSL1D1; BIRC3; PPM1K*
Ca binding	*CALB2*		*CAB39L*
Cell adhesion	*MCAM; GPR126; ANTXR2; SLIT1; PEAR1; NTN4*	*ITGB7; NEBL*	*CDH16; DOCK4; AGR2*
Cell cycle	*BRSK2; BEX2*	*CCNI*	*TERT; ORC6; CDKN2C; TYMS; POLE4*
Chromatin remodelling		*HIST1H2AI; HIST1H2BE; HIST1H2AG; HIST1H1B; HIST1H4L*	
Cytoskeleton-cytokinesis	*MYH7B; CENPI; BRSK2; TUBB2; FILIP1*	*LLGL2; FGFR1OP; WDR1; RGCC; CEP95; TUBB2A*	*TRIM58; FRMD4A; ANLI TNNC1; KIF18A; DNM3; MICAL3; APC2; MARCKS; KIF19; TUBB2B*
DNA damage response			*RAD51AP1*
Drug metabolism		***CYB5R2***	***CYB5R2***; *CYP2J2; ADH1C*
Energy production		***DNAJC15***	***DNAJC15***
Extracellular matrix	*SCEL; HS3ST1*	*COL9A3; COL6A1*	*FMOD; SDC4*
Glycosylation			*GALNT18*
Growth factors	*MIA*	*IGFBP2; MDK*	*IHH; KITLG; WLS*
Growth factors receptors	*GRB10; GPR126; FZD6; CALCA*	*NTRK2*	*GFRA3; EFNB3; GPR160 NOTCH2; RAMP1; SMO*
Hypoxia response	*HEPAS1*	*HIF1A*	
Inflammation and immunity	*CD55; CXCL8; NCF2*	*ANXA1*	*RARRES2*
Intracellular transport	*CPLX2; CPLX1; CAPN8*	***MVB12B***; *SEZ6L2; GOLGA8A*	*RAB36; HIP1;* ***MVB12B;*** *BLOC1S4; HTT; CPE; AGR2; SPIRE2*
Ion transport	*KCNJ2; AHNAK2; TRPV6; BEST1; FXYD4*		*TMC4; AKAP7; PIEZO1; MFSD10; BSPRY; MTL5; CACNA2D4; ATP6V0A4*
Lipid metabolism		***LIPC; CYB5R2***	*PLIN4;* ***LIPC***; ***CYB5R2;*** *CYP2J2*
Mucosa protection		*TFF1; TFF3; SLPI*	*AGR2*
Nuclear structure and function		*NPIPB5*	*NOP14*
Phosphatases		*SGPP2*	*DUSP23; PPM1K*
Proteolysis	*SERPINE2; TIMP3; TPSAB1; TFPI*		*MME; WFDC2*
RNA maturation	*GPAT2; SNORA30; SNORA62*	*SNORA75; SNORA2B; SNORA13; SFPQ; CLK1*	*TRA2A*

Signal transduction	*RASGEF1A; CAPN5; AKAP12; GBP3; ADRBK2; REPS2*	*AFAP1L2; SGPP2; NGEF; PLCB1; CRABP2; PIM1; PTPRM; PTPRS*	*CHN2; SHCBP1; GPER1; CNPY1; APC2; DGKQ; MYZAP; NOTCH2NL; SQSTM1; LMTK3; ARHGEF4; PROM1; PKIB; TSPAN5; RRAGD; PTPN13; GRB10; AKT3*
Stress response	*OXR1*		
Transcription	*CAPN15; BEX2; BHLHE41; DIP2C; HEPAS1; CREB5; ZNF462; HES7; ZIC5*	*LEF1; PAX6; HIF1A; RNF187; TCEA3; TCEA2*	*PRRX1; KLF17; NELFA; ZNF581; MXD4; RCOR2; MXI1; ESSRA; MYCL; RELB; FOXD1; BCL3; ZNF22*
Translation		*EEF1A2*	
Transporters		*SLC43A3; ABCC3*	*SLC39A11; SLC29A2*
Ubiquitin proteasome pathway		*UBASH3B; OTUD1*	*NEURL3*

Only genes showing a fold change "Mean FUT6 / Mean Neo" ≥3, a *p* value ≤ 0.05 and a level of expression either in Neo or in B4GALNT2 ≥ 50 are reported. Genes up-regulated or down-regulated are marked in red or blue, respectively. Genes showing opposite regulation in the two cell lines are in bold.

As FUT6 expression, B4GALNT2 expression modulated the gene expression in either SW480 or SW620 cells or both. To simplify the analysis, genes were divided by functional classes as indicated in Table 13.

Table 13. B4GALNT2-modulated genes in either SW480 or SW620 or in both, grouped for functional classes.

Functional class	Both in SW480 and SW620	Only in SW480	Only in SW620
Apoptosis	*G0S2; CDIP1; LGALS7; RNF157; FILIP1L; SEMA3B*	*RNF130*	*EVA1A; PTPN13; CDH13*
Cell adhesion	*CD44; CDHR2; DDR1*	*ELSPBP1; PCDH9*	*NEBL; SERPINB8*
Cell cycle	*ORC6; DISC1; RGCC*		*CDK14; BRSK2*
Chromatin remodelling	*MTA2*	*ING4; BAZ2B*	*ATRX; ZNF462*
Cytoskeleton-cytokinesis	*ORC6; CDCA5; MCM10; DISC1; SPIRE2; PTPRN2; SEMA3B; TUBB2A; KRT15; SPTBN5*	*DLC1; SHROOM2*	*CENPI; ASPM; ERCC6L; NUF2; ANLN; CEP152; KIFC3; MAP1LC3B; BRSK2; SPEG2; KRT14; ARC*
DNA damage response	*MCM10; RAD51; ANKRD32 (SLF1)*		*RAD51AP1; ERCC6L; DNA2; ARHGAP11A*
Drug metabolism		*CYB5R2; SLC47A1; FMO3*	
Extracellular matrix	*COL7A1; KRTAP3-2*	*ECM2; COL9A3*	
Glycosylation		*GALNT18; GALC; AMY1C*	*FUT3*
Growth factors		*IGFBP2; ANGPTL2; ISM1; EDA; IGFALS*	*AREG*
Growth factors receptors	*RHBDF1*	*NOTCH2; ROR1; EFNA1*	
Hypoxia response			*EGLN3*
Inflammation and immunity	*IL18; SPON2*		*CD8B; NCF2; SLAMF7*
Intracellular transport	*SPIRE2; BAIAP3; TUBB2A; SYT13*	*BET1L; C16orf62; LMF1; MYRIP*	*LPHN2; KIFC3; RAB37; GOLGA7B*
Ion transport	*PLLP; MAGED2; SLC4A11*	*KCNS3; CNKSR3; STAC3; ATP6AP1L;*	*STOM*
Lipid metabolism		*CYB5R2; CROT; LMF1; FABP6*	*ELOVL2; SLCO1B3*
Lysosomal enzymes	*IDS*		
Mucosa protection			*MUC6*
Phosphatases	*PTPRN2*	*SGPP2*	*PTPN13*
Proteolysis		*TPP2*	*ADAM30; SERPINA3*
RNA maturation	*SRPK3*	*SNAR-G1; SNAR-F; SNAR-G2; SNAR-H; SNAR-D; SNAR-A3*	
Signal transduction	*PLCB1; DDR1; NXN; RNF157*	*DLC1; SGPP2; STK32C; KSR2*	*ATRNL1; DKK4; ARHGAP11A; ADRB1; RAB37; ADRB2; TBC1D*
Stress response		*HSPA1A*	
Transcription	*MAF; MTA2; ZNF276; NXN; ZNF83*	*ZNF316; LEF1; TFAP2C; KLF12; HBP1; ZFP62*	*ZSCAN20; ZNF462*
Transporters	*SLC43A3; ABCC3; AKR1C1*	*SLC47A1; XK; SLC2A8*	*SLC2A6; SLC7A2*
Ubiquitin proteasome pathway	*RNF157*	*TPP2; RNF130; OTUD1*	

Only genes showing a fold change "Mean B4GALNT2 / Mean Neo" ≥2, a *p* value ≤ 0.05 and a level of expression either in Neo or in B4GALNT2 ≥ 50 are reported. Genes up-regulated or down-regulated are marked in red or blue, respectively.

Surprisingly, it was found that transfection with either glycosyltransferase cDNA induced consistent gene modulation in one of the two cell lines. For example, the gene *ANLN*, which encodes for an actin-binding protein required for cytokinesis, was up-regulated in SW620 by expression of either B4GALNT2 or FUT6. In Table 14 are shown the ten genes consistently modulated by expression of either FUT6 or B4GALNT2 classified in functional classes.

The genes *ANLN* and *RAD51AP1* are up-regulated in SW620 by expression of either glycosyltransferase. The genes *CYB5R2*, *IGFBP2*, *SGPP2*, *LEF1* are up-regulated in SW480 by expression of either FUT6 or B4GALNT2 while *COL9A3*, *PLCB1*, *ABCC3*, *OTUD1* are down-regulated by either glycosyltransferase.

Table 14. Genes consistently modulated by expression of either FUT6 or B4GALNT2, grouped for functional classes.

Gene	Function	Functional class	Cell line	Regulation
ANLN	Actin-binding protein required for cytokinesis	Cytoskeleton-cytokinesis	SW620	Up
RAD51AP1	Participates to homologous recombination repair	DNA damage response	SW620	Up
CYB5R2	Involved in desaturation and elongation of fatty acids, cholesterol biosynthesis, drug metabolism	Drug metabolism	SW480	Up
COL9A3	Structural component of hyaline cartilage	Extracellular matrix	SW480	Down
IGFBP2	Binds to IGF, prolonging its activity	Growth factors	SW480	Up
SGPP2	Degrades the bioactive signaling molecule sphingosine 1-phosphate	Phosphatases	SW480	Up
PLCB1	Production of the second messenger molecules diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP3)	Signal transduction	SW480	Down
LEF1	Transcription factor of the Wnt signaling, activates *MYC* and *CCND1* expression and enhances proliferation of pancreatic tumor cells	Transcription	SW480	Up
ABCC3	May act as an inducible transporter in the biliary and intestinal excretion of organic anions	Transporters	SW480	Down
OTUD1	Removes ubiquitin	Ubiquitin proteasome pathway	SW480	Down

The study of B4GALNT2 expression in SW480 and SW620 cells revealed a remarkable cell specificity of the glycosyltransferase. Thus, the further investigation evaluated the effects of B4GALNT2 expression in SW480 and SW620 with those previously observed in the B4GALNT2-transfected cell line LS174T in search of a "B4GALNT2 signature", common to the three cell lines.

In Table 15, it is presented a signature of seven protein coding genes as a results of the statistical analysis performed to search a B4GALNT2 signature common to three cell lines employed for the study. It was attributed a function in cancer by an intensive search in the literature as indicated through the PMID.

Table 15. Genes consistently modulated by B4GALNT2 expression in SW480, SW620 and LS174T.

	Expression level								
	LS174T		SW480		SW620				
Gene symbol	Neo	B4GALNT2	Neo	B4GALNT2	Neo	B4GALNT2	Gene name	Function in cancer	PMID
MCOLN2	4	9	10	22	19	59	Mucolipin 2	Promotes glioma progression	27248469
PLLP	104	52	674	441	803	509	Plasmolipin	Little or no information	
FILIP1L	20	13	59	36	107	57	Filamin A interacting protein 1-like	Suppresses tumor progression by inhibiting cell proliferation and angiogenesis in colorectal cancer	7750216
FAM231A	24	15	8	5	16	8	Family with sequence similarity 231, member A	Little or no information	
SPON2	338	181	10476	5779	8579	5447	Spondin 2, extracellular matrix protein	Promotes growth and invasion of CRC	26686083
COL20A1	14	8	40	21	54	28	Collagen type XX α1	Overexpressed in glioma	31556357
BCL2L10	8	2	26	12	33	12	BCL2-like 10 (apoptosis facilitator)	Functions as tumor suppressor gene in ovarian and hepatocellular cancer	31894274 27770580

"Neo" and "B4GALNT2" refer to the mean expression value of the gene in the six cell populations. The seven genes are the only showing a statistically significant ($p \leq 0.05$) different expression level in B4GALNT2, compared with Neo in the three cell lines.

CHAPTER VI - DISCUSSION AND CONCLUSIONS

Unlike many other diseases, CRC is preventable and potentially curable when diagnosed in its early stages. However, it may remain often undetected due to the unspecific symptoms[11]. Current screening techniques are invasive or lack either sensitivity or specificity. A key point in the management of CRC is the identification of patients at higher or lower risk of recurrence and progression, to spare the most aggressive (and sometimes expensive) therapies to lower risk patients. To this aim, it is really essential to identify biological markers useful for patients' stratification. Glyco-markers are particularly suitable for this purpose. During colorectal cancer oncogenesis, many glyco-antigens undergo remarkable changes and can contribute significantly to cancer development and progression. Amongst the carbohydrate structures, the Sd^a antigen, synthesized by the glycosyltransferase B4GALNT2[77], is down-regulated in colorectal cancer.

The clinical implications of its dramatic down-regulation in CRC are investigated for the first time in this work. This was pursued through the analysis of TCGA data to search a correlation between gene expression and clinical parameters. First of all, TCGA data confirm in a large cohort of samples that mRNA of *B4GALNT2* is dramatically decreased in CRC, compared to normal tissue, consistent with a reduced B4GALNT2 enzymatic activity in CRC tissues previously reported by the group. The finding that patients expressing the highest levels of B4GALNT2 mRNA in their cancer tissues display longer survival and better response to therapy provides the first clinical demonstration of the relationship between high B4GALNT2 expression and lower malignancy. This was consistent with the observed association of high B4GALNT2 with non-mucinous phenotype and wild type *TP53* status. On the other hand, no relationship was observed with clinical stage and microsatellite instability *status*. Interestingly, *B4GALNT2* mRNA expression exhibits a prognostic predictive potential in CRC exceeding that of all the glycosyltransferases tested in this work and is even better than that of many oncogenes and tumor-suppressor genes. Among those tested, only *SMAD6*, encoding an inhibitor of TGF-β signaling, and *TERT*, that encodes the reverse transcriptase subunit of telomerase, displayed a better prognostic value than *B4GALNT2*, which equals that of *EGFR*. A possible

explanation for the lack of relationship with patients' overall survival is provided by the fact that oncogene activation or tumor-suppressor inactivation can be due to mutations, altered phosphorylation, mislocalization, rather than increased or decreased mRNA expression. Patient stratification according to *B4GALNT2* mRNA expression revealed that high B4GALNT2 expressers displayed a concomitant high level of other genes associated with positive prognosis, such as *ZG16*, *ITLN1*, *BEST2* and *GUCA2B*. These data support the notion that *B4GALNT2* is a key member of a gene signature associated with good prognosis.

Patients expressing the highest and lowest levels of B4GALNT2 display also glycosylation genes differentially modulated. Differential expression of several glycosyltransferases predicts that cancer cells of high B4GALNT2 expressers display higher levels of mucin-type O-glycosylation (*GALNT8*) with sugar chains terminating with sialyl-Tn, (*ST6GALNAC1*), sialyl-6-T structures (*ST6GALNAC2*) and Core 3 structures (*B3GNT6*); increased biosynthesis of type 1 chains (*B3GALT5*) and increased α2,3 sialylation of type 2 chains (*ST3GAL4*), forming acceptor substrates for B4GALNT2. *ST6GAL1* and *ST6GAL2* showed reduced expression, suggesting that α2,6-sialylation could be reduced in HBE. Among non-glycosyltransferase molecules, it is worth mentioning the higher expression of galectin 4, which is associated to normal gut, and of the gel-forming mucins MUC2, MUC4 and MUC5B in HBE. These data suggest that in LBE and HBE cohorts different glycophenotypes exist.

Data collected *in silico* indicate complex mechanisms controlling *B4GALNT2* expression in colonic tissues. The promoter region of the *B4GALNT2* gene contains CpG islands suggesting that methylation can play a relevant role in B4GALNT2 down-regulation in colon carcinogenesis. However, the general down-regulation of the gene observed in the vast majority of cancer cases cannot be explained by a differential methylation of the CpG sites located in the island and shores. Rather, the intronic open-sea site located between exon 6 and 7 presents a generally reduced methylation in cancer samples. Interestingly, methylation of this site is associated with increased, rather than decreased, *B4GALNT2* expression. Thus, in some cases high methylation of specific sequences can have stimulatory effects due to the fact that it promotes interaction of distant enhancers with the regulatory regions of the gene[133]. However, in many samples *B4GALNT2* was not expressed despite a permissive methylation status, indicating that other regulatory mechanisms are implicated. The glycosyltransferase expression might also be regulated by small non

coding RNAs (miRNAs). Several miRNAs are theoretically predicted to inhibit B4GALNT2 but miR-204-5p appears to be the most plausible candidate. In fact, it is the only one down-regulated in HBE and no one of the samples expressing *B4GALNT2* above a background threshold expressed this miRNA. However, down-regulation of this miRNA does not ensure B4GALNT2 expression, since many cases lacking miR-204-5p failed to express *B4GALNT2*. Thus, *B4GALNT2* regulation appears to be multifactorial, with DNA methylation and miRNA expression playing a relevant but not exclusive role. The lack of appropriate transcription factors could be a plausible reason for the lack of *B4GALNT2* expression in the majority of CRC samples.

To understand whether high B4GALNT2 and lower malignancy were causally related, the impact of the forced expression of B4GALNT2 on the phenotype and the transcriptome of CRC *in vitro* was studied on three CRC cell lines transfected with B4GALNT2 cDNA. In the cell line LS174T, which constitutively expresses sLe^x, B4GALNT2 induced both sLe^x inhibition and Sd^a expression. Thus, the phenotypic and transcriptomic changes observed could not be separately attributed to the two antigens. To investigate separately the effects of Sd^a and sLe^x, the CRC cell lines SW480 and SW620, lacking the two antigens, were transfected with B4GALNT2 or FUT6.

On LS174T cells, B4GALNT2 expression reduced some features associated with malignancy, such as soft agar growth, spheroid formation and ALDH expression, which are all associated with stemness. On the other hand, the proliferation rate, the clonogenic ability on solid substrates and the wound healing ability were not affected by B4GALNT2 expression in this cell line.

It is commonly stated that sLe^x is the key player of malignancy while to B4GALNT2 and Sd^a only an ancillary role was attributed through the inhibition of sLe^x biosynthesis[134]. Thus, the research was conducted on the CRC cell lines SW480 and SW620, originally lacking both Sd^a and sLe^x antigens, transfected with B4GALNT2 or FUT6. This model allowed to analyze the effects on the CRC phenotype and transcriptome resulting from the *de novo* expression of the Sd^a or sLe^x antigens. The fact that the two cell lines were derived from the primary tumor (SW480) and a metastasis of the same patient (SW620), allowed to analyzed the effects of

glycosyltransferase expression on cells with a common genetic background but different malignancy.

In SW620 cells, all the six properties associated with malignancy resulted down-regulated by B4GALNT2 expression. On the contrary, in SW480 cells the proliferation rate and the wound healing capability were not affected. It appears that B4GALNT2 was able to inhibit specifically those properties, such as increased proliferation rate and ability to heal a wound, that were more pronounced in SW620 due to their metastatic origin. It could be hypothesized that in SW620 cells proliferation is supplied by additional pathways, specifically inhibited by B4GALNT2. FUT6 up-regulated the clonogenic ability in both SW480 and SW620, while the soft agar growth and the wound healing capability were up-regulated only in SW620. These data clearly show that the phenotypic changes induced by a glycosyltransferase are strongly dependent on the malignancy of the recipient cells. It could be hypothesized that this effect could be seen only if this ability is already present, as in SW620, but it could not be induced *de novo* if not present as in SW480 cells. Apparently, the overexpression of FUT6, can exacerbate the phenotype only if the cell has previously reached a given "malignant threshold". The only features commonly inhibited by B4GALNT2 in all three cell lines were the ability to grow in poor adherence conditions and ALDH expression, all related with stemness. Importantly, these effects induced by B4GALNT2 were independent of sLe^x inhibition, as they were also found in cells lacking sLe^x antigen.

Stemming from these results *in vitro* that highlight the strong impact of the B4GALNT2 expression on the phenotype of CRC cells, the project investigated also the effect of the glycosyltransferase expression on the transcriptome of transfected cell lines. The ability to modulate the transcriptome of cancer cells by a glycosyltransferase was previously reported. Nevertheless, the transcriptomic changes induced by B4GALNT2 in standard conditions of growth were surprisingly relevant.

Among the most down-regulated genes in B4GALNT2-expressing clones with known functions in cancer there were nidogen-1 (*NID1*), galectin-2 (*LGALS2*), paternally expressed gene 10 (*PEG10*), retinoic acid induced
(*RAI14*) and receptor-tyrosine-kinase-like orphan receptor 1 (*ROR1*); 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 related gene,

transcription factor SOX2, which belongs to the family of Sry related HMG Box proteins and is essential for the maintenance of self-renewal and pluripotency in embryonic stem cells as well as adult tissue progenitor cells[135]. *SOX2* overexpression is correlated with tumorigenicity and CSC phenotype[136]. This gene is involved in pathways related with stemness, as displayed in GeneGO enrichment analysis. Its down-regulation may help to explain the reduced ability to grow in non-adherent conditions and the reduced number of stem cells observed in S2/S11 cells. On the other hand, another gene related with cancer stemness (*CD200*) was found to be the most up-regulated gene in B4GALNT2 transfectant clones. *NGFRAP1* (nerve growth factor receptor-associated protein 1, also known as BEX3) which is involved in mitogenic signaling and apoptosis, was also up-regulated by B4GALNT2.
The strong predominance of down-regulated genes over up-regulated genes in B4GALNT2-expressing cells may conceivably be related to a reduced ability to perform several cellular functions. The fact that the vast majority of these genes displayed cancer-promoting activity in different systems is consistent with an attenuation of the cancer phenotype in LS174T cells. Additionally, a causative role of B4GALNT2 in the control of these genes was enforced by the observation that in high and low B4GALNT2 expressers of the TCGA cohorts, the mean expression level of these genes, including *SOX2*, was in some cases significantly consistent with that observed in LS174T cells.

Owing to the notably reduced ability to adapt to non-adherent growth displayed by B4GALNT2-expressing cells, the transcriptome of cells grown as spheroids in 3D conditions was also analyzed. The study focused on the genes modulated by 3D culture in LS174T cells as well as on those genes that displayed a differential response to 3D culture conditions in B4GALNT2-expressing cells S2/S11. The analysis showed that many genes were modulated by 3D culture conditions, regardless of B4GALNT2 expression. Among these, many genes were implicated in energy metabolism, including the glycolytic process, such as *PFKFB4, PGM1, ALDOC, PGK1* and some were part of the hypoxia response (*CA9, EGLN3, EGR1*). Some of these genes were expressed at extremely high levels and were all up-regulated including *PGK1, LCN15, FABP1, ALDOC, PGM1, CA9*. This modulation could be explained by the notion that 3D cell models partially simulate the structure of the tumor microenvironment and cancer cells adapt their metabolism activating

the hypoxia response, resulting in increased glucose uptake and fermentation to foster proliferation, growth and survival[137,138].

The genes belonging to the transcriptional regulation group, such as those encoding the transcription factors FOS and FOSB and the transcriptional regulator EGR1, were highly expressed in 2D culture and dramatically down-regulated in 3D culture as alterations of transcriptional activity evolve during progression of 3D cultures similarly to *in vivo* microenvironment. On the other hand, among the genes involved in cell signaling, the gene *KIT* resulted up-regulated in 3D culture. This gene encodes a tyrosine-protein kinase acting as cell-surface receptor and playing an important role in the regulation of survival, proliferation and migration of cells. Genes involved in cytoskeleton organization displayed a general down-regulation as disorganization of cellular architecture is a common feature of tumors, while those involved in detoxification displayed up-regulation allowing cancer cells to grow in an otherwise toxic environment[139]. The genes *PIGZ,* involved in the biosynthesis of the glycosylphosphatidylinositol (GPI)-anchor, and *NDRG1*, involved in stress response, were found to be strongly upregulated.

All the above-mentioned genes displayed 3D culture modulation regardless of B4GALNT2 expression. However, several genes displaying 3D culture modulation were identified only in S2/S11 cells; 13 genes were up-regulated while 18 were down-regulated. The most relevant change was the strong down-regulation of five genes regulating cytoskeletal organization in mitosis and motility including *KIZ, CEP120, DNAH6, SGOL2, STARD13*. This finding could explain the reduced ability to grow in non-adherent conditions observed in B4GALNT2-expressing cells. Among the genes controlling cell signaling, three genes encoding taste receptors (*TAS2R45, TAS2R19, TAS2R30*) appeared to be down-regulated. These genes are implicated in tasting bitterness but can also play a tumor-suppressive role. 3D culture augmented the propensity to apoptosis in B4GALNT2-expressing clones that can be explained through the modulation of at least three genes: *TNFAIP8L2, MYOD1* and *PPM1K*. Among the genes involved in transcriptional regulation, *PHF20L1* gene was remarkably downregulated, thus giving a contribution to explain the reduced stemness as it stabilizes the transcription factor SOX2 post-translationally. Three genes modulated by B4GALNT2 in 3D conditions were related to immunity and inflammation: *CTLA4*, a well-known inhibitory receptor of T lymphocytes; *IL1A*, an

inflammatory cytokine; *TDO2*, a gene involved in a pathway potentially suppressing anti-tumor immune responses.

If it is assumed that 3D culture conditions *in vitro* are closer to those of *in vivo* growth, then it should be expected that high B4GALNT2-expressing tumors displayed modulation of the same genes we observed modulated by 3D only in B4GALNT2-expressing cells. This was true for all the genes falling in the groups "Cytoskeleton and mitosis" and "Transcription regulation". Additionally, the gene *TNFAIPL2*, which promotes apoptosis, and the gene of the pro-inflammatory cytokine, *IL1A*, also were found to be consistently modulated in the TCGA cohort. Although IL1A plays different and even opposite roles in the tumor microenvironment, when it is expressed by CRC cells, it functions as an immunostimulatory molecule and enhances an anti-tumor immune response.

Transcriptomic analysis of SW480 and SW620 cells expressing FUT6 or B4GALNT2 showed that FUT6 was able to modulate a higher number of genes than B4GALNT2 in both cell lines. Besides many genes regulated in parallel in the two cell lines, some were modulated only in one of the two and others were consistently modulated by either glycosyltransferase in the same cell line.

Expression of FUT6 impacted on genes belonging to different functional classes in both cell lines: transcription, cell adhesion and cell signaling. However, this effect of FUT6 expression was cell-type specific. Indeed, in SW620 FUT6 expression heavily impacted on groups of genes which were poorly or unaffected in SW480, specifically genes directly involved in DNA duplication, such as the telomerase reverse transcriptase (*TERT*), the component of the DNA polymerase E-complex (*POLE4*), thymidylate synthase (*TYMS*), and the origin recognition complex member *ORC6*. Also genes related to cytoskeleton-cytokinesis, growth factor receptors, ion transport, cell signaling and transcription were more strongly affected by FUT6 in SW620 than in SW480. Unexpectedly, a few genes such as *CYB5R2*, *DNAJC15*, *MVB12B* and *LIPC* were subjected to opposite regulation by FUT6 in the two cell lines.

Expression of B4GALNT2 strongly affected several functional classes in both cell lines including apoptosis, cytoskeleton and cytokinesis and transcription. However, even for B4GALNT2 the response was strongly cell line-specific. Indeed, a large group of genes involved in cytoskeleton organization and cytokinesis was modulated only in SW620, whereas a greater number of genes belonging to the transcription and growth factors classes were modulated by B4GALNT2 only in SW480 cells. To

be noted that a group of six SNAR [Small NF90 (ILF3) Associated RNAs], non-protein coding RNAs with an unclear role in RNA function, were up-regulated only in SW480 cells.

Amongst the genes more consistently down-regulated in both cell lines by B4GALNT2 is *CD44*. It has been documented that CD44 functions as a carrier of sLe^{x}[100], and its fucosylated form allows CRC progression[140]. Thus, current data may contribute a new possible mechanism through which B4GALNT2 hinders cancer progression mediated by FUT6: besides the well reported competition for the carbohydrate substrate acceptor, through the down-regulation of one of the preferred sLe^{x} carriers. Down-regulation of CD44, considered a stemness marker[92], is consistent with the reduced stemness induced by B4GALNT2 expression.

Interestingly, a small number of genes were consistently modulated by either glycosyltransferase in the same cell line. To explain this unexpected observation it could be proposed that the presence of either the sLe^{x} or the Sd^{a} antigens on the same glycan receptor can, in some cases, produce a similar biological effect (for example the inhibition of the binding of a soluble factor).

Owing to the marked cell specific effects on the transcriptome of glycosyltransferase overexpression observed in SW480 and SW620 cells, the study intended to compare the effects of B4GALNT2 expression on the transcriptome of SW480 and SW620 with those previously observed in B4GALNT2-transfected LS174T cells. Statistical analysis was thus performed in search of a "B4GALNT2 signature", common to the three cell lines. Seven protein coding genes were identified: *SPON2, PLL, FAM321A, MCOLN2, FILIP1L, COL20A1, BCL2L10*. Among these genes, only *MCOLN2* displayed up-regulation, while the remaining were down-regulated. Literature search reports a tumor promoting activity for *MCOLN2, SPON2* and *COL20A1* and a tumor restraining activity for *FILIP1L* and *BCL2L10*, while for *PLLP* and *FAM231A* no information were available. Thus, changes of *MCOLN2, FILIP1L* and *BCL2L10* would be tumor-promoting while changes of *SPON2* and *COL20A1* would be tumor-restraining. However, except for *SPON2*, which encodes an extracellular matrix protein promoting cell motility, growth and invasion of CRC, the level of expression of these genes was so low that they can hardly be responsible for relevant biological effects induced by B4GALNT2 in all the three cell lines.

Overall, these data indicate that glycosyltransferases may induce a given phenotypic effect through the activation of multiple interconnected and converging pathways that are activated by different genes in a strongly cell-specific manner.

In the clinic, the level of B4GALNT2 expressed by CRC tissues could be used to stratify patients according to their risk of progression. This aim is very important to avoid invasive and expensive therapies to low risk patients. In this light, it is crucial to understand the ways through which B4GALNT2 and its associated Sd^a antigen have a tumor restraining effect in CRC. This study provides evidence that the tumor restraining activity was not exerted only through sLe^x inhibition, otherwise it was not working in sLe^x-negative cases. On the contrary, it is largely independent of sLe^x inhibition and is due to deep modifications of colon cancer cell biology.

TAKE HOME MESSAGES

- B4GALNT2 glycosyltransferases attenuates malignancy of CRC by reducing stemness of CRC cells.
- This effect is independent on sLe^x inhibition.
- Both FUT6 and B4GALNT2 are able to modulate the CRC transcriptome.
- The impact of glycosyltransferase overexpression is strongly cell-type specific.
- The level of B4GALNT2 could be useful to identify CRC patients at lower risk of progression.

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.* **68**, 394–424 (2018).
2. Araghi, M. *et al.* Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *Lancet Gastroenterol. Hepatol.* **4**, 511–518 (2019).
3. Siegel, R. L. *et al.* Global patterns and trends in colorectal cancer incidence in young adults. *Gut* **68**, 2179–2185 (2019).
4. Yoshihara, M. *et al.* Epidemiology of colorectal cancer. *Nihon Naika Gakkai Zasshi.* **96**, 200–206 (2007).
5. Vipperla, K. & O' Keefe, S. J. Colorectal cancer. *Metab. Hum. Dis. Organ Physiol. Pathophysiol.* 149–154 (2014). doi:10.1007/978-3-7091-0715-7_24
6. Constance M. Johnson *et al.* Meta-analyses of Colorectal Cancer Risk Factors. *Cancer Causes Control.* **24**, 1207–1222 (2013).
7. Torre, L. A. *et al.* Global Cancer Statistics, 2012. *CA a cancer J. Clin.* **65**, 87–108 (2015).
8. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
9. Kuipers, E. J. *et al.* Colorectal cancer. *Nat. Rev. Dis. Prim.* **1**, 1–25 (2015).
10. Colussi, D. *et al.* Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int. J. Mol. Sci.* **14**, 16365–16385 (2013).
11. De Rosa, M. *et al.* Genetics, diagnosis and management of colorectal cancer (Review). *Oncol. Rep.* **34**, 1087–1096 (2015).
12. Hassan, C. *et al.* Histologic risk factors and clinical outcome in colorectal malignant polyp: A pooled-data analysis. *Dis. Colon Rectum* **48**, 1588–1596 (2005).
13. Lynch, J. P. & Hoops, T. C. The genetic pathogenesis of colorectal cancer. *Hematol. Oncol. Clin. North Am.* **16**, 775–810 (2002).
14. Al-Sohaily, S. *et al.* Molecular pathways in colorectal cancer. *J. Gastroenterol. Hepatol.* **27**, 1423–1431 (2012).
15. Fearon, E. R. R. Molecular Genetics of Colorectal Cancer. *Annu Rev Pathol* **6**, 479–507 (2011).
16. Móran, A. *et. al.* Differential colorectal carcinogenesis: Molecular basis and clinical relevance. *World J. Gastrointest. Oncol.* **2**, 151–8 (2010).
17. Purnak, T. *et al.* Molecular basis of colorectal cancer. *N. Engl. J. Med.* **362**, 1246 (2010).
18. Bogaert, J. & Prenen, H. Molecular genetics of colorectal cancer. *Ann. Gastroenterol.* **27**, 9–14 (2014).
19. Parker, T. W. & Neufeld, K. L. APC controls Wnt-induced β-catenin destruction complex recruitment in human colonocytes. *Sci. Rep.* **10**, 1–14 (2020).
20. Behrens, J. The role of the Wnt signalling pathway in colorectal tumorigenesis. *Biochem. Soc. Trans.* **33**, 672–675 (2005).
21. Zoratto, F. *et al.* Focus on genetic and epigenetic events of colorectal cancer pathogenesis: Implications for molecular diagnosis. *Tumor Biol.* **35**, 6195–6206 (2014).
22. Kaprio, T. *et al.* KRAS Alleles. *Oncotarget* **18**, 139–148 (2016).
23. Kloor, M. *et al.* Clinical significance of microsatellite instability in colorectal cancer. *Langenbeck's Arch. Surg.* **399**, 23–31 (2014).

24. De'angelis, G. L. *et al.* Microsatellite instability in colorectal cancer. *Acta Biomed.* **89**, 97–101 (2018).
25. 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).
26. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
27. Xourafas, D. *et al.*The impact of somatic SMAD4 mutations in colorectal liver metastases. *Chinese Clin. Oncol.* **8**, 6–11 (2019).
28. Esteller, M. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene* **21**, 5427–5440 (2002).
29. Mojarad, E. N. *et al.*The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol. Hepatol. from Bed to Bench* **6**, 120–128 (2013).
30. Jass, J. R. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* **50**, 113–130 (2007).
31. Brenner, H. *et al.* Colorectal cancer. *Lancet* **383**, 1490–1502 (2014).
32. Sobin L.H., Gospodarowicz M.K., W. C. *TNM Classification of malignant tumours*. (2009).
33. Soneson, C. *et al.* the consensus molecular subtypes of CRC. *Nat. Med.* **21**, 1350–1356 (2015).
34. Van Cutsem, E. *et al.* The ESMO Guidelines Working Group. Metastatic colorectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **25**, iii1–iii9 (2014).
35. Kang, H. *et al.* A 10-year outcomes evaluation of mucinous and signet-ring cell carcinoma of the colon and rectum. *Dis. Colon Rectum* **48**, 1161–1168 (2005).
36. Ohtsubo, K. & Marth, J. D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **126**, 855–867 (2006).
37. Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143**, 672–676 (2010).
38. Freeze, H. H. Genetic defects in the human glycome. *Nat. Rev. Genet.* **7**, 537–551 (2006).
39. Ghazarian, H. *et al.* A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **113**, 236–247 (2011).
40. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
41. Riezman, H. & Munro, S. Essentials of Glycobiology. **10**, 552–553 (2000).
42. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).
43. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
44. Aebi, M. N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2430–2437 (2013).
45. Wolfert, M. A. & Boons, G. J. Adaptive immune activation: Glycosylation does matter. *Nat. Chem. Biol.* **9**, 776–784 (2013).
46. Sethi, M. K. & Fanayan, S. Mass spectrometry-based N-glycomics of colorectal cancer. *Int. J. Mol. Sci.* **16**, 29278–29304 (2015).
47. Vasconcelos-dos-Santos, A. *et al.* Biosynthetic machinery involved in aberrant glycosylation: Promising targets for developing of drugs against cancer. *Front. Oncol.* **5**, 1–23 (2015).
48. Mhatre V. Ho & Kelsey C. Martin, J.-A. L. Recent insights into the biological roles of mucin-type O- glycosylation. *Bone* **23**, 1–7 (2012).
49. 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).
50. Dall'Olio, F. *et al.* Mechanisms of cancer-associated glycosylation changes. *Front. Biosci.* **17**, 670–699 (2012).
51. Dennis, J. W. *et al.* β1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science (80-.).* **236**, 582–585 (1987).
52. Yoshimura, M. & Taniguchi, N. Suppression of lung metastasis of B16 mouse melanoma cells by introduction of N-acetylglucosaminyltransferase III gene. *Nippon rinsho. Japanese J. Clin. Med.* **53**, 1786–1790 (1995).
53. Ju, T. *et al.* Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res.* **68**, 1636–1646 (2008).
54. JA, F. *et al.* Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol. Oncol.* **7**, 719–731 (2013).
55. Pinho, S. *et al.* Biological significance of cancer-associated sialyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. *Cancer Lett.* **249**, 157–170 (2007).
56. Miyoshi, E. *et al.* Fucosylation is a promising target for cancer diagnosis and therapy. *Biomolecules* **2**, 34–45 (2012).
57. Miyoshi, E. *et al.* Biological function of fucosylation in cancer biology. *J. Biochem.* **143**, 725–729 (2008).
58. Tu, C. F. *et al.* FUT8 promotes breast cancer cell invasiveness by remodeling TGF-β receptor core fucosylation. *Breast Cancer Res.* **19**, 1–15 (2017).
59. 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).
60. 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. - Ren. Physiol.* **300**, 1017–1025 (2011).
61. Tozawa, K. *et al.* Positive correlation between sialyl Lewis X expression and pathologic findings in renal cell carcinoma. *Kidney Int.* **67**, 1391–1396 (2005).
62. Irimura, T. *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).
63. 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).
64. Dall'Olio, F. *et al.* Sialosignaling: Sialyltransferases as engines of self-fueling loops in cancer progression. *Biochim. Biophys. Acta - Gen. Subj.* **1840**, 2752–2764 (2014).
65. 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).
66. Läubli, H. & Borsig, L. Selectins promote tumor metastasis. *Semin. Cancer Biol.* **20**, 169–177 (2010).
67. Kannagi, R. *et al.* Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis. *Cancer Sci.* **95**, 377–384 (2004).
68. Gout, S. *et al.* Selectins and selectin ligands in extravasation of cancer cells and organ selectivity of metastasis. *Clin. Exp. Metastasis* **25**, 335–344 (2008).
69. Borsig, L. *et al.* Heparin and cancer revisited: Mechanistic connections involving platelets, P-selectin, carcinoma mucins, and tumor metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3352–3357 (2001).
70. 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).
71. Dall'Olio, F. *et al.* The expanding roles of the Sda/Cad carbohydrate antigen and its cognate glycosyltransferase B4GALNT2. *Biochim. Biophys. Acta - Gen. Subj.* **1840**, 443–453 (2014).
72. Renton, P. H. *et al.* Anti- Sda, a New Blood Group Antibody. *Vox Sang.* **13**, 493–501 (1967).
73. Kornfeld, R. The B4 Lectin from Vicia villosa Seeds Interacts with N-Acetylgalactosaniine Residues a-Linked to Serine or Threonine Residues in Cell Surface Glycoproteins. **258**, 5172–5176 (1983).
74. Blanchard, D. *et al.* Identification of a novel ganglioside on erythrocytes with blood group Cad specificty. *J. Biol. Chem.* **260**, 7813–7816 (1985).
75. Serafini Cessi, F. & Dall'Olio, F. Guinea pig kidney β-N-acetylgalactosaminyltransferase towards Tamm-Horsfall glycoprotein. Requirement of sialic acid in the acceptor for transferase activity. *Biochem. J.* **215**, 483–489 (1983).
76. Lo Presti, L. *et al.* Molecular cloning of the human b1,4 N-Acetylgalactosaminyltransferase responsible for the biosynthesis of the Sda histo-blood group antigen: the sequence predicts a very long cytoplasmic domain. **134**, 675–682 (2003).
77. Malagolini, N. *et al.* Biosynthesis and expression of the Sda and sialyl Lewis x antigens in normal and cancer colon. *Glycobiology* **17**, 688–697 (2007).
78. 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).
79. Wang, H. R. *et al.* Expression of the human Sda β-1,4- N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. *Glycobiology* **18**, 104–113 (2008).
80. Kawamura, Y. I. *et al.* DNA Hypermethylation Contributes to Incomplete Synthesis of Carbohydrate Determinants in Gastrointestinal Cancer. *Gastroenterology* **135**, 142–151 (2008).
81. 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).
82. Li, P. T. *et al.* Localization of B4GALNT2 and its role in mouse embryo attachment. *Fertil. Steril.* **97**, 1206-1212.e3 (2012).
83. Thomas, P. J. *et al.* 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).
84. Xu, R. *et al.* Overexpression of Galgt2 reduces dystrophic pathology in the skeletal muscles of alpha sarcoglycan-deficient mice. *Am. J. Pathol.* **175**, 235–247 (2009).
85. Heaton, B. E. *et al.* A CRISPR activation screen identifies a pan-avian influenza virus inhibitory host factor. **20**, 1503–1512 (2017).
86. Malagolini, N. *et al.* Expression of UDP-GalNAc: NeuAcα2,3Galβ-R β1,4(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).
87. 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).
88. Kawamura, Y. I. *et al.* Introduction of Sda carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. *Cancer Res.* **65**, 6220–6227 (2005).

89. Capon, C. *et al.* Sda-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).
90. Trinchera, M. *et al.* The biosynthesis of the selectin-ligand sialyl Lewis x in colorectal cancer tissues is regulated by fucosyltransferase VI and can be inhibited by an RNA interference-based approach. *Int. J. Biochem. Cell Biol.* **43**, 130–139 (2011).
91. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–37 (2013).
92. Dalerba, P. *et al.* Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10158–10163 (2007).
93. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).
94. Munro, M. J. *et al.* Cancer stem cells in colorectal cancer: A review. *J. Clin. Pathol.* **71**, 110–116 (2018).
95. Yamanaka, S. Induced pluripotent stem cells: Past, present, and future. *Cell Stem Cell* **10**, 678–684 (2012).
96. Baumann, M. *et al.* Exploring the role of cancer stem cells in radioresistance. *Nat. Rev. Cancer* **8**, 545–554 (2008).
97. Ricci-Vitiani, L. *et al.* Colon cancer stem cells. *J. Mol. Med.* **87**, 1097–1104 (2009).
98. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
99. Du, L. *et al.* CD44 is of functional importance for colorectal cancer stem cells. *Clin. Cancer Res.* **14**, 6751–6760 (2008).
100. Hanley, W. D. *et al.* CD44 on LS174T colon carcinoma cells possesses E-selectin ligand activity. *Cancer Res.* **65**, 5812–5817 (2005).
101. Barkeer, S. *et al.* Glycosylation of Cancer Stem Cells: Function in Stemness, Tumorigenesis, and Metastasis. *Neoplasia (United States)* **20**, 813–825 (2018).
102. Hao, L. *et al.* Expression and clinical significance of SALL4 and β-catenin in colorectal cancer. *J. Mol. Histol.* **47**, 117–128 (2016).
103. Ardalan Khales, S. *et al.* SALL4 as a new biomarker for early colorectal cancers. *J. Cancer Res. Clin. Oncol.* **141**, 229–235 (2014).
104. Tuy, H. D. *et al.* ABCG2 expression in colorectal adenocarcinomas may predict resistance to irinotecan. *Oncol. Lett.* **12**, 2752–2760 (2016).
105. Ding, X. wei *et al.* ABCG2: A potential marker of stem cells and novel target in stem cell and cancer therapy. *Life Sci.* **86**, 631–637 (2010).
106. Corvinus, F. M. *et al.* Persistent STAT3 activation in colon cancer is associated with enhanced cell proliferation and tumor growth. *Neoplasia* **7**, 545–55 (2005).
107. Shimokawa, M. *et al.* Visualization and targeting of LGR5 + human colon cancer stem cells. *Nature* **545**, 187–192 (2017).
108. Deng, S. *et al.* Distinct expression levels and patterns of stem cell marker, aldehyde dehydrogenase isoform 1 (ALDH1), in human epithelial cancers. *PLoS One* **5**, (2010).
109. Deng, Y. *et al.* ALDH1 is an independent prognostic factor for patients with stages II-III rectal cancer after receiving radiochemotherapy. *Br. J. Cancer* **110**, 430–434 (2014).
110. Gibney, E. R. & Nolan, C. M. Epigenetics and gene expression. *Heredity (Edinb).* **105**, 4–13 (2010).
111. Kagohara, L. T. *et al.* Epigenetic regulation of gene expression in cancer: Techniques, resources and analysis. *Brief. Funct. Genomics* **17**, 49–63 (2018).
112. Jin, B., Li, Y. & Robertson, K. D. DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes and Cancer* **2**, 607–617 (2011).

113. Bandres, E. *et al.* Epigenetic regulation of microRNA expression in colorectal cancer. *Int. J. Cancer* **125**, 2737–2743 (2009).
114. Keith D. Robertson. DNA methylation and human disease. *Nat. Rev. Genet.* **6**, 597–610 (2005).
115. Moore, L. D. *et al.* DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
116. Esteller, M. Aberrant DNA methylation as a cancer-inducing mechanism. *Annu. Rev. Pharmacol. Toxicol.* **45**, 629–656 (2005).
117. Deng, J. *et al.* Methylation Associated With Nuclear Reprogramming. *Nat. Biotechnol.* **27**, 353–360 (2009).
118. Bardhan, K. & Liu, K. Epigenetics and colorectal cancer pathogenesis. *Cancers (Basel).* **5**, 676–713 (2013).
119. Nieto, M. A. The Ins and Outs of the Epithelial to Mesenchymal Transition in Health and Disease. *Annu. Rev. Cell Dev. Biol.* **27**, 347–376 (2011).
120. Cummins, J. M. *et al.* The colorectal microRNAome. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3687–3692 (2006).
121. Yang, L. *et al.* MicroRNA and colorectal cancer. *World J. Surg.* **33**, 638–646 (2009).
122. Jevšinek Skok, D. *et al.* The integrative knowledge base for miRNA-mRNA expression in colorectal cancer. *Sci. Rep.* **9**, 1–9 (2019).
123. Slaby, O. *et al.* MicroRNAs in colorectal cancer: Translation of molecular biology into clinical application. *Mol. Cancer* **8**, 1–13 (2009).
124. Chen, B. *et al.* Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis. *Open Biol.* **9**, (2019).
125. Wu, W. K. K. *et al.* MicroRNA in colorectal cancer: From benchtop to bedside. *Carcinogenesis* **32**, 247–253 (2011).
126. Ren, A. *et al.* Detection of miRNA as non-invasive biomarkers of colorectal cancer. *Int. J. Mol. Sci.* **16**, 2810–2823 (2015).
127. Anaya, J. OncoLnc: Linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. *PeerJ Comput. Sci.* **2016**, (2016).
128. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159 (1987).
129. Borowicz, S. *et al.* The soft agar colony formation assay. *J. Vis. Exp.* 1–6 (2014). doi:10.3791/51998
130. Sant, Shilpa, P. A. J. The Production of 3D Tumor Spheroids for Cancer Drug Discovery - HHS Public Access. *Physiol. Behav.* **176**, 139–148 (2017).
131. Franken, N. A. P. *et al.* Clonogenic assay of cells in vitro. *Nat. Protoc.* **1**, 2315–2319 (2006).
132. Jonkman, J. E. N. *et al.* An introduction to the wound healing assay using live cell microscopy. *Cell Adhes. Migr.* **8**, 440–451 (2014).
133. Grimmer, Matthew R., Costello, J. F. Oncogene brought into the loop. *Nature* **529**, 34–35 (2016).
134. Kawamura, Y. I. *et al.* Introduction of Sda carbohydrate antigen in gastrointestinal cancer cells eliminates selectin ligands and inhibits metastasis. *Cancer Res.* **65**, 6220–6227 (2005).
135. 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**, 1–11 (2016).
136. Maurizi, G. *et al.* Sox2 is required for tumor development and cancer cells proliferation in osteosarcoma. *Oncogene* **37**, 4626–4632 (2018).
137. Liverani, C. *et al.* A biomimetic 3D model of hypoxia-driven cancer progression. *Sci.*

Rep. **9**, 1–13 (2019).
138. Jong, B. K. *et al.* Three-dimensional in vitro tissue culture models of breast cancer - A review. *Breast Cancer Res. Treat.* **85**, 281–291 (2004).
139. Collino, A. *et al.* Sustained activation of detoxification pathways promotes liver carcinogenesis in response to chronic bile acid-mediated damage. *PLoS Genet.* **14**, 1–19 (2018).
140. Pan, S. *et al.* HOTAIR/miR-326/FUT6 axis facilitates colorectal cancer progression through regulating fucosylation of CD44 via PI3K/AKT/mTOR pathway. *Biochim. Biophys. Acta - Mol. Cell Res.* **1866**, 750–760 (2019).