Unraveling mucin type o-glycosylation signatures of colorectal cancer

Madunić. K.

COLORECTAL CANCER, BUT NOT HEALTHY COLON, EXPRESSES SPECIFIC CORE 2 SIALYLATED *O*-GLYCANS

Katarina Madunić[1], Oleg A. Mayboroda[1], Tao Zhang[1], Julia Weber[4], Geert-Jan Boons[4], Hans Morreau[3], Ronald van Vlieberghe[2], Tom van Wezel[3], Guinevere S.M. Lageveen-Kammeijer[1], Manfred Wuhrer[1]

[1] *Leiden University Medical Center, Center for Proteomics and Metabolomics, Postbus 9600, 2300 RC Leiden, The Netherlands*
[2] *Leiden University Medical Center, Department of Surgery, Postbus 9600, 2300 RC Leiden, The Netherlands*
[3] *Leiden University Medical Center, Department of Pathology, Postbus 9600, 2300 RC Leiden, The Netherlands*
[4] *Department of Chemical Biology and Drug Discovery, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands*

Manuscript submitted

Chapter 6

ABSTRACT

Current immunotherapy for colorectal cancer (CRC) shows limited patient benefit due to the relative paucity of the targeted molecules. Changes in protein glycosylation are a hallmark of CRC, and glycans are expressed widely on the cell surface, which makes them promising targets. Here we present an in-depth study of CRC mucin type *O*-glycosylation, derived from epithelial cancer regions compared to healthy colon mucosa from the same patients. We found a total of 100 tumor associated carbohydrate antigens (TACAs) which were solely observed in the cancer region. Core 2 O-glycans with Lewis, sialyl-Lewis and otherwise sialylated 6-arms were most discriminating, being found in most of the cancers whilst being absent in the neighboring healthy colon mucosa. These specific O-glycan structures present promising new targets for development of innovative cancer immunotherapeutics.

INTRODUCTION

Colorectal cancer (CRC) is one of the leading malignancies worldwide with over 900,000 deaths in 2020[1]. Conventional therapeutic strategies include chemotherapy, radiation therapy, and surgery. However, despite the recently introduced and highly successful pre-symptomatic population screening, many CRC cases are still detected at an advanced stage, leading to an unsuccessful treatment[2]. In the past decade, new CRC therapies emerged, such as specific inhibitors and antibodies targeting soluble proteins and cellular receptors[3]. These include the biologics anti-VEGF (bevacizumab), anti-EGFR (cetuximab and panitumumab), anti-PD-1 immune checkpoint inhibitors (pembrolizumab, nivolumab), BRAF V600E inhibitors (several combinations of vemurafenib, irinotecan, and cetuximab or panitumumab)[4], with many more in development. Unfortunately, these treatments are only efficient for a limited subset of patients as many patients develop therapy resistance[5]. Therefore, the identification of new specific molecular targets and the development of new targeted therapies for CRC are essential for providing better treatment for a larger group of patients.

Various studies have shown altered glycosylation patterns of proteins and lipids as a hallmark of cancer[6,7]. The changes in glycosylation are originating from a shift in the expression of glycosyltransferases, modifications of their enzymatic activity, mislocalization in the endoplasmic reticulum or Golgi apparatus, availability of substrates or nucleotide sugar donors as well as competition between the enzymes[6]. The characteristic glycan alterations in cancer include specific aberrant expression of incomplete carbohydrate structures or *de novo* expression of carbohydrate antigens also known as tumor-associated carbohydrate antigens (TACAs)[8]. Of note, aberrant cancer glycosylation has already found its way into the clinics as the CA19-9 serological marker for e.g. pancreatic cancer and CRC which is known to be the sialyl-Lewis A (sLe^A) carbohydrate antigen[9].

The expression of the heavily glycosylated intestinal mucins changes in CRC. It has been demonstrated that MUC2 and MUC4 expression is downregulated in CRC, whereas MUC1, MUC5AC and MUC17 are overexpressed[10], showing reduced expression of core 3 and core 4 structures with increased expression of *O*-linked truncated core 1 glycans, (sialyl-)T antigen, (sialyl-)Tn, and sialyl-Lewis X

(sLe^X)antigen [7], and several of these motifs are being evaluated as targets of immunotherapeutic monoclonal antibodies[11], despite their often limited specificity.

While the presence of these carbohydrate motifs in tumors is well-known, information on the underlying TACA structures is vastly lacking. The majority of the studies rely on detecting abnormal expression of glycosyltransferases in cancer and detection of terminal epitopes on all glycan types, and no differentiation is made between the *N*- and *O*-linked glycoproteins or glycolipids, nor are the specific TACA structures revealed[12–18].

In recent years, sensitive mass spectrometry (MS) based glycomics workflows have been developed enabling untargeted screening of potential glycan markers derived from complex biological samples. However, tumor tissue heterogeneity is a major obstacle for cancer glycomics studies. The tumor microenvironment is populated by non-cancer cells such as immune cells or fibroblasts which can hamper the detection of specific cancer associated glycan species. This issue can be largely overcome by sampling the areas of interest with laser capture microdissection (LCM) prior to glycan release as demonstrated in hepatocellular carcinoma by Hinneburg *et al.*[19,20].

In this study we focused on decoding the tumor specific *O*-glycan signatures of CRC. The *O*-glycome was determined from both epithelial regions of primary tumors and metastatic sites and was compared to healthy colon tissue from the same patients. Furthermore, we established a high throughput workflow that sequentially releases the *N*-and *O*-glycans from LCM formalin fixed paraffin embedded (FFPE) tissues, followed by analysis using porous graphitized carbon liquid chromatography (PGC-LC)-MS/MS in negative ion mode, enabling a powerful separation of isomeric species as well as in-depth structural characterization of TACAs. This approach allowed us to identify CRC specific TACAs as potential targets of innovative cancer immunotherapy such as anti-TACA antibodies, adoptive T-cell therapies and antibody-drug conjugates.

MATERIALS AND METHODS

MATERIALS AND REAGENTS

Hematoxylin (cat. nr. GHS232), sodium dodecyl sulfate solution 20% (cat. nr. 05030), trifluoroacetic acid (TFA; cat. nr. 1.38178.0050), tris(hydroxymethyl)amino-methane (cat. nr. 252859; lot#BCBM2559V), sodium borohydride (cat. nr. 452882; lot nr.

STBD8912V), hydrochloric acid (cat. nr. 258148; lot nr. SZBD3100V), DL-dithiothreitol (DTT; cat. nr. D0632; lot nr. SLBW0160), ammonium bicarbonate (cat. nr. 09830; lot nr. BCBQ6426V,) cation exchange resin Dowex 50W X8 (cat. nr. 217492; lot nr. MKCH2513), ammonium acetate (cat. nr. A1542) and polyvinylpirrolidone mw 40, 000 (cat. Nr PVP40, lot.nr. BCBM0949V) were purchased from Sigma Aldrich (St. Louis, MO). Ethanol (EtOH; cat. nr. 100983.1000), NaCl; cat. nr. 1.06404.1000) and methanol (MeOH; cat. nr. 1.06009.2500) were purchased from Merck (Darmstadt, Germany). Acetonitrile LC-MS grade (cat. nr. 01203502) was obtained from Biosolve (Valkenswaard, The Netherlands). Glacial acetic acid (cat. nr. A6283; lot nr. SZBG2660H) and potassium hydroxide (cat. nr. P1767) were purchased from Honeywell Fluka (Charlotte, NC). PNGase F (Flavobacterium meningosepticum recombinant in E. coli; Cat No. 11365193001) was obtained from Roche (Mannheim, Germany). SPE bulk sorbent Carbograph (cat. nr. 1769; lot nr. 5122145) was acquired from Grace Discovery sciences (Columbia, TN). MultiScreen® HTS 96-multiwell plates (pore size 0.45m) with high protein-binding membrane (hydrophobic Immobilon-P PVDF membrane) and 96-well PP Microplate (cat. nr. 651201; lot nr. E1708385) were purchased from Millipore (Amsterdam, The Netherlands), 96-well PP filter plate (cat. nr. OF1100) was obtained from Orochem technologies (Naperville, Il). MembraneSlide 1.0 PEN (cat.nr. 415190-9041-000) and Adhesive Cap 500 µL tubes (cat.nr. 415190-9211-000) were purchased from Carl Zeiss Microscopy (Göttingen, Germany). All buffers were prepared using Milli-Q water (mQ) generated from a ELGA system (Millipore, Amsterdam, Netherlands), maintained at ≥ 18 MΩ. Chemically synthesized glycopeptide standards were provided by Utrecht University (Weber et al.; manuscript in preparation).

FFPE TISSUE SECTIONING AND STAINING

Anonymized human CRC paraffin tissue blocks were from the Department of Pathology at Leiden University Medical Center (LUMC, Leiden, The Netherlands). All samples were handled in a coded fashion, according to the national ethical guidelines (“Code for Proper Secondary Use of Human Tissue”, Dutch Federation of Medical Scientific Societies). Paired primary tumor (T1-T12) and adjacent normal colon mucosa (C1-C12) from 12 patients were selected for the cohort. Additionally, 6 metastatic cancers obtained from liver metastases of different patients were included (M15- M21). The tissues were cut into 5 µm thick sections from paraffin blocks using

a microtome and mounted on glass slides for hematoxylin and eosin (HE) staining, or 10 µm thick sections on membrane PEN slides for LCM. All slides were dried overnight at 37 ˚C and stored at 4 ˚C.

Slides used for LCM were deparaffinized using xylene (three times for 5 min) and washed with absolute EtOH (two times for 2 min). The slides were then rehydrated in a series of EtOH by submerging the slides in 85% EtOH, followed by 70% EtOH and distilled water. Hematoxylin was applied for 20 sec. The slides were then washed with demineralized water for 2 min and dehydrated with increasing EtOH series by submerging in 70% EtOH, followed by 85% EtOH and finally absolute EtOH. The slides were briefly air dried and stored at 4 ˚C. The slides used for pathologist annotation were stained with HE following the routine protocols.

PATHOLOGIST ANNOTATION

Tumor and healthy epithelial tissue regions from HE slides were annotated by a pathologist by careful inspection under the microscope. Differentiation grade was determined by assessing glandular formation in the HE histological slides.

LASER CAPTURE MICRODISSECTION

Tissue sections mounted on PEN membrane slides for LCM were carefully inspected under the microscope. Target areas marked on HE slides by the pathologist were located, and the laser was positioned in this area (**Supplementary Fig. S2**). To ensure that comparable amounts of tissue were used for each sample, cells were counted in three different regions of the tissue in an area of approx. 2500 µm^2. An average of the count was used to extrapolate the area needed to be cut in order to obtain samples containing approximately 20,000 cells. The large difference in cell size between hepatic and colon cells was taken into account by performing the LCM on 10 µm thick tissue sections, enabling the extraction of sufficient material for the analysis. LCM was performed using PALM RoboSoftware and collected in adhesive cap 500 µL tubes. Considering the healthy colon mucosa is surrounded by infiltrated lymphocytes that could not be separated from the epithelial cells successfully, we also extracted lymphocyte follicles from different normal colon tissues and pooled them together to obtain a potentially confounding immune cell glycan profile (IM). Moreover, we extracted the stroma controls for the tumors with high stroma infiltration that might

have contamination coming from cancer associated fibroblasts and immune cells (ST4, ST6, ST7, ST11, ST12, SM19, and SM21).

GLYCAN RELEASE AND PURIFICATION

A lysis buffer (100 µL) containing 100 mM Tris HCl, 0.1 M DTT, 100 mM NaCl, and 1% SDS was added to microdissected tissues collected in adhesive caps. The pieces were carefully collected in lysis buffer from the lid of the adhesive caps and transferred to a 1.5 mL Eppendorf tube. Prior to sonication using a Branson sonicator rod (three times for 15 sec, with output power 2/10) the tubes were placed on ice. During sonication, the samples were kept on ice and left 20 sec to cool down between each cycle. Furthermore, the samples were incubated at 99 ˚C for 60 min with mild agitation (400 rpm). After incubation, the samples were slowly cooled down at room temperature (RT). In the meantime, PVDF membrane plate was preconditioned using 100 µL of 70% EtOH, until the membrane turned opaque, followed by an additional wash using 100 µL of MQ. Upon rewetting the membrane with 5 µL of 70% EtOH, 100 µL of carefully mixed sample lysates were loaded onto the membrane wells. Additionally, a tissue lysate (containing approximately 8 x 10^4 cells) was split into three separate samples (TECH1, TECH2, and TECH3) and processed independently (different wells of the PVDF membrane plate) to check for technical variability. The plate was incubated at RT on a shaker for 20 min to ensure binding. The unbound material was removed by centrifugation at 500 *x g* and washed with 100 µL of MQ. An additional, 40 µL of 0.5% PVP-40 in MQ was added to the PVDF membrane wells to block the membrane and incubated for 5 min on a shaker. Upon removal of the blocking agent by centrifugation, the membrane wells were extensively washed, two times with 100 µL of PBS, two times with 100 µL of 10 mM ammonium bicarbonate followed by two times 100 µL of MQ. Each time the washing agent was removed by centrifugation at 500 *x g* for 1 min. Since the addition of PVP-40 increases the water flux through the membrane, the hydrophilicity is also increased[21]. Therefore, to avoid enzyme mixture going through, 10 µL of MQ was added to each well of the PVDF membrane plate to soak the membrane and fill the space underneath it. The plate was incubated on a shaker with light agitation for 5 min. The enzyme mixture (15 µL) contained 2U of *N*-glycosidase F and was added to each well followed by an incubation of 15 min at 37 °C. Finally, 15 µL of water was added on top, to prevent drying out the membrane overnight, and placed in the 37 ˚C incubator. The next steps were

performed as previously described[22]. Briefly, the following day, upon recovery of *N*-glycans, 50 µL of 0.5 M sodium borohydride in 50 mM potassium hydroxide solution was added to each well and incubated for 16 h at 50 ˚C. The samples were cleaned using self-packed cation exchange columns in 96-well plates, followed by PGC-SPE. The samples were dried and stored at -20 ˚C until analysis.

PGC-LC-MS/MS MEASUREMENTS

Prior to analysis, the samples were resuspended in 15 µL of MQ and 1 µL of each sample was pooled together in a separate vial for quality check and method optimization (QC). The QC pool and a bovine submaxillary mucin standard were used to check instrument performance each day of measurement. A total of 6 µL of each tissue sample was injected for analysis (40%). A custom-made trap column (size 30 × 0.32 mm) packed with 5 µm particle size PGC stationary phase from Hypercarb PGC analytical column (size 100 × 4.6 mm, 5 µm particle size, Thermo Fisher Scientific, Waltham, MA) was used to load the samples using 100% buffer A (10 mM ammonium bicarbonate) at a loading flow of 6 µL/min. The glycans were separated on a custom-made PGC column (100 mm × 75 µm, 3 µm particle size obtained from Thermo Fisher Scientific) at a 0.6 µL/min flow rate by applying a linear gradient from 1% to 50% of buffer B (acetonitrile, 10 mM ammonium bicarbonate) over 73 min. A constant column temperature of 45 ˚C was maintained. A part of the samples were remeasured to resolve the isomers with composition H2N2F1S1 (isomer c and f). For this purpose a gradient was used ranging from 1% to 50% of buffer B over 110 minutes at 35 ˚C. The peak ratios of the resolved peaks was taken to extrapolate the peak areas from the original measurement (**Supplementary Table S11**). The LC system was coupled to an amaZon ETD speed ESI ion trap MS using the CaptiveSpray™ source (Bruker Daltonics) with an applied capillary voltage of 1000 V in negative-ionization mode. The drying gas (N_2) temperature was set at 280˚C and the flow to 3 L/min. The nebulizer gas pressure was kept at 3 psi enriched with isopropanol as described before[22]. MS spectra were acquired in enhanced mode within a mass to charge ratio (*m/z*) range of 380-1850, target mass of smart parameter setting was set to *m/z* 900, ion charge control (ICC) to 40,000 and maximum acquisition time to 200 ms. MS/MS spectra were generated by collision-induced dissociation on the three most abundant precursors, applying an isolation width of 3 Th. The fragmentation cut-off was set to 27% with

100% fragmentation amplitude using the Enhanced SmartFrag option (30–120% in 32 ms) and ICC was set to 150,000.

DATA PROCESSING

Extracted ion chromatograms including all observed charge states (1^- and 2^-) of the first three isotopes were used to integrate the area under the curve (AUC) for each individual glycan isomer using Compass DataAnalysis software (v.5.0). Peaks were manually picked and integrated. Relative quantitation was performed using the total area of all glycans within one sample as reference (100%). Glycan structures were identified by manual inspection of MS/MS spectra following known *O*-glycan biosynthetic pathways and available literature[23–27]. In the structural interpretations all hexoses (H) are assumed to be galactose, all deoxyhexoses a fucose (F), all internal *N*-acetylhexosamines (N) an *N*-acetylglucosamine except for the N2S1b, confirmed to be core 5 by comparison with bovine submaxillary mucin glycans standard. All terminal *N*-acetylhexosamines attached to a galactose also substituted with a *N*-acetylneuraminic acid (S) are assumed to be *N*-acetylgalactosamine as part of the Sda/Cad antigen. All terminal *N*-acetylhexosamines attached to a galactose also substituted with a fucose are assumed to be *N*-acetylgalactosamine as part of blood group antigen A. Sulphate modification position is not determined due to partial sulphate migration, and no presence of diagnostic fragment ions for the position on the Gal or GlcNAc residue. Therefore, its position in the glycan structures can be either on the Gal or the neighboring GlcNAc residue. Structures of the selected TACAs (**Fig. 1**) were confirmed by synthesized standards (J. Weber e*t al., manuscript in preparation,* 2022).

STATISTICS

Data analysis and visualization were performed in ‘’R’’ software using the following packages: tidyverse, readxl, caret, gridExtra, EnhancedVolcano, ggpubr, pcaMethods, Rcpm, ggrepel, tidyHeatmap, stringr and ComplexUpset. An imputation of the minimum positive value (0.0001) was performed to enable use principal component analysis (PCA). Variables with near zero variance were excluded before computing the PCA model using nearZeroVar function from caret package. Differences between groups were tested using Wilcoxon-Mann-Whitney non-parametric statistical test. P-values were adjusted for multiple testing using the “Benjamini-Hochberg” method.

TRANSCRIPTOMIC DATA

TCGA expression data was downloaded via the firebrowse.org website. The laser capture microdissected data GEO: GSE21815 was downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21815.

RESULTS AND DISCUSSION

A HIGH THROUGHPUT WORKFLOW FOR GLYCAN ANALYSIS FROM LASER CAPTURE MICRODISSECTED COLON TISSUES

First a high throughput workflow was established which enables the characterization of *O*-glycans from LCM FFPE tissue slides (**Supplementary Fig. S1**). This workflow is based on previously established protocols for glycan analysis from complex biological samples[22,28] as well as a formerly developed glycomic workflows using FFPE tissue sections[19]. Briefly, specific regions were extracted of the FFPE tissue sections by LCM, such as normal colon mucosa (**Supplementary Fig. S2a** and **b**), cancer epithelia (**Supplementary Fig. S2c** and **d)** and cancer stroma. Tissue lysates from approximately 2 x 10^4 up to 2.5 x 10^4 cells were used for protein immobilization on PVDF membrane plate wells. The *N*-glycans were released prior to *O*-glycan release to avoid interference. *O*-glycans were released from the immobilized proteins, purified and analyzed by PGC-LC-MS/MS. To ensure protein solubilization from formalin fixed material and the binding of the solubilized proteins to the PVDF membrane, the amount of detergent used for the lysis was decreased to 1%, compared to the 4% used in the original FFPE tissue glycomics approach[19]. This way we avoid the tedious chloroform/methanol protein precipitation step, as this gave unreproducible protein yields for small sample amounts. Regarding *O*-glycan detection, we used isopropanol enriched nitrogen gas to increase sensitivity and the quality of the MS/MS spectra, as described previously[22]. As depicted in **Supplementary Fig. S3** the method showed very good precision.

GLYCOSYLATION SIGNATURES IN COLORECTAL CANCER

The presence of TACAs was investigated using paired primary tumors from 12 CRC patients (T1–T12) and normal colon from the same patients (C1-C12). Additionally, six metastatic CRCs from liver metastases of six different patients (M15-M21) were analyzed. Clinical data related to the samples used in our cohort is available in **Supplementary Table S1** and **S2**. From the 12 cancer samples, eight were

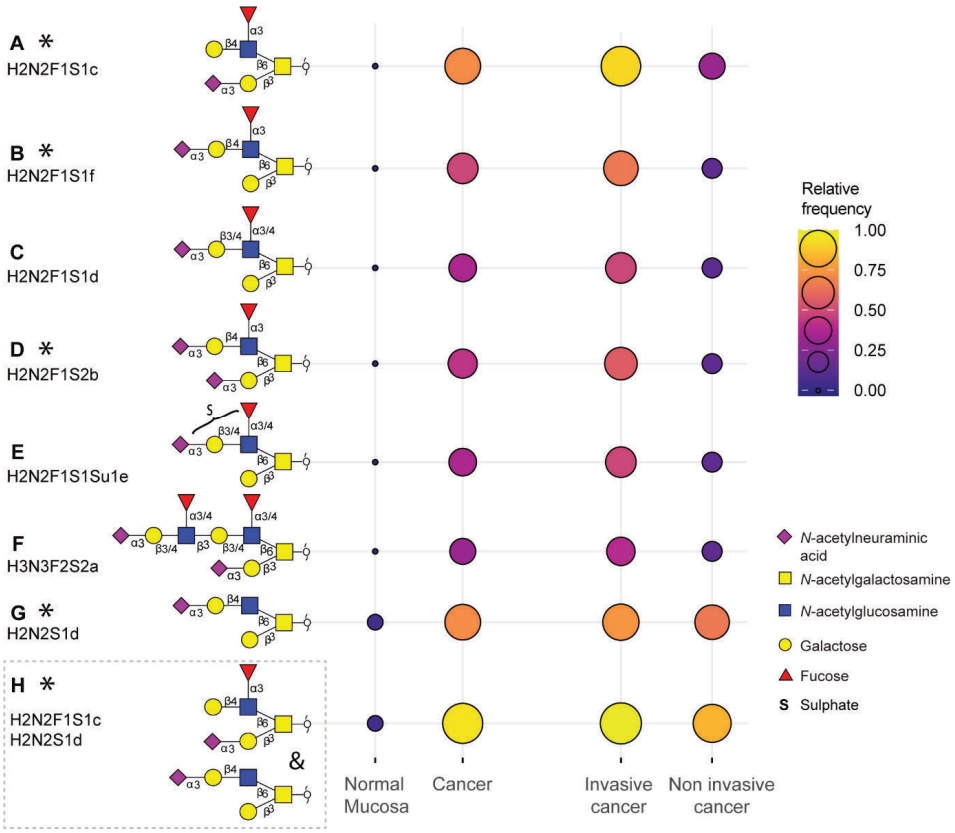

Fig. 1. Tumor associated carbohydrate antigens identified in colorectal cancer. *A subset of the O-glycans overexpressed in cancer which is detected in more than 33% of the cancers and undetectable in normal colon mucosa samples* ***(a, b, c, d, e, f)*** *or detected in maximum one normal mucosa* ***(g)****. Either one of the two TACAs in* ***(h)*** *is detected in 94% of the cancers. The selected TACAs show higher specificity for invasive cancer (Dukes stages C and D, n=12) than non-invasive cancer (Dukes stages A and B, n = 6) (right panel). Structures confirmed by synthesized standards (J. Weber et al., manuscript in preparation, 2022) are labelled with an asterisk. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

adenocarcinoma (AC), one was neuroendocrine carcinoma (NEC), and two were mucinous adenocarcinoma (MUC) of which one was partly signet-ring cell adenocarcinoma (T2). During the microdissection of the epithelial tumor regions and mucosa regions from normal colon it was ensured that the regions were well separated from the surrounding submucosal and muscle layers of the colon and tumor stroma. In addition, the stroma regions for a subset of stroma high cancers were also collected. *O*-glycosylation profiles were obtained for those regions separately.

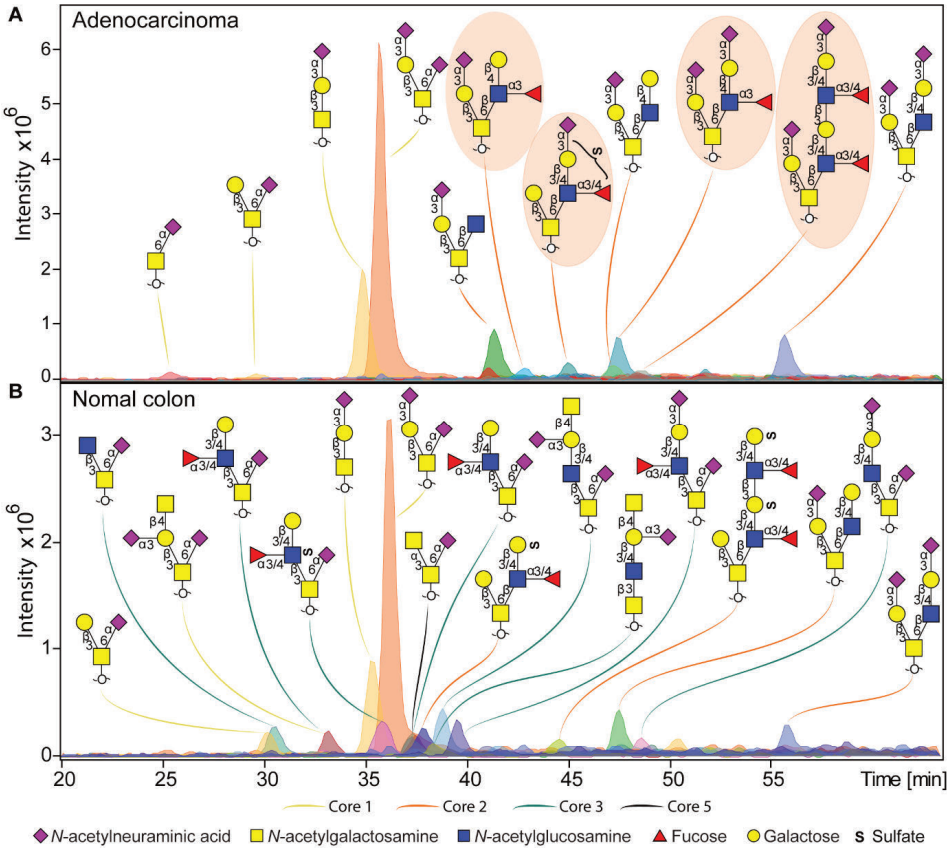

Fig. 2. Example of O-glycan chromatographic profile from (a) adenocarcinoma and (b) normal colon mucosa from the same patient (T8 vs C8). a) *The adenocarcinoma from patient 8 is characterized by specific expression of core 2 glycans, carrying terminal Le^{X} and sialyl-Lewis$^{X/A}$ antigens, or just terminal α2-3 sialylation linked to the galactose. Presence of TACAs (Fig. 1) are circled with orange background.* ***b)*** *Normal colon mucosa from the same patient shows expression of a diversity of core 3 structures, in most cases carrying an α2-6 linked sialic acid linked to the core GalNAc. Core 3 glycans are carrying terminal Sda antigens, as well as Le$^{X/A}$ and sulpho-Le$^{X/A}$ epitopes. The TACAs observed in adenocarinoma are not detected in the normal colon mucosa from the same patient. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

Overall, a total of 172 *O*-glycans were detected in the analyzed tissues (**Supplementary Table S3** and **S4**). From those, 108 were observed in the cancer and stroma, with no expression in normal colon mucosa (**Supplementary Table S5**). From those, 100 *O*-glycans were solely detected in cancer (TACAs), whereas only 20 *O*-glycans were shared between cancer and normal colon mucosa (**Supplementary**

Fig. S5). The majority of the TACAs have a core 2 structure carrying sialyl-Lewis$^{X/A}$ (sLe$^{X/A}$) antigens or an α2-3-linked sialic acid attached to the galactose (Gal) (**Supplementary Fig. S4, S5, S6b** and **S7**). Seven core 2 *O*-glycans that were not detected in normal mucosa, but exclusively found in more than six cancers (relative frequency > 33%) are depicted in **Fig. 1a-g** and listed in **Supplementary Table S6** and **S7**. MS/MS spectra of the selected TACAs are shown in **Supplementary Fig. S8.** Additional confirmation was found by performing standard addition experiments with available synthesized standards as exemplified in **Supplementary Fig. S9.** Sialylated core 2 *O*-glycan with terminal Le$^{X/A}$ antigen (**Fig. 1a**) was absent from normal mucosa and found in 72% of the cancers. One core 2 glycan with terminal α2-3 sialylation of Gal on the 6 arm is found in one normal mucosa sample, and in 72% of the cancers (**Fig. 1g**). In particular, at least one of the two TACAs depicted in **Fig. 1h** were detected in 94% of the cancers in our study. Interestingly, the only sample in which those glycans were not detected is a poorly differentiated adenocarcinoma, specifically classified as solid type, without glandular formation, usually associated with better patient prognosis[29]. An example of TACA expression is illustrated in **Fig. 2a**, which shows the expression of four specific TACAs (labeled with an orange background) in the adenocarcinoma tissue from patient 8, while those *O*-glycans were not detected in the normal colon mucosa (**Fig. 1b**). This pattern of TACA expression in cancer regions but not in presumably unaffected epithelium of the same patient was consistently observed across the entire dataset. Therefore, our data provides strong evidence for existence of highly specific *O*-glycans which are not present in the normal colon mucosa tissue. Moreover, significantly different *O*-glycomic traits (summarized in **Supplementary Fig. S4**) were found between CRC and normal mucosa (**Fig. 3**; **Supplementary Table S8**, **S9** and **S10**). While core 2 glycans are overexpressed, core 3 *O*-glycans are downregulated in cancer (**Fig. 3a** and **f**). A different expression is also observed in regard to sialylation, where terminal α2-3 sialylation is overexpressed, and α2-6 sialylation is downregulated in cancer (**Fig. 3b** and **g**). A core 2 *O*-glycan with terminal α2-3 sialylation on Gal is detected in the normal colon mucosa, (**Fig. 3e**) but it is strongly overexpressed in cancer. Differences were found also in regard to antigen expression, namely terminal sLe$^{X/A}$ and terminal α2-3 sialylation were upregulated in cancer in the context of core 2 *O*-glycans (**Fig. 3c** and **d; Supplementary Fig. S7** and **S10a**). Although the presence of sLe$^{X/A}$ was associated with several epithelial cancers[30–33], no specific targets were previously

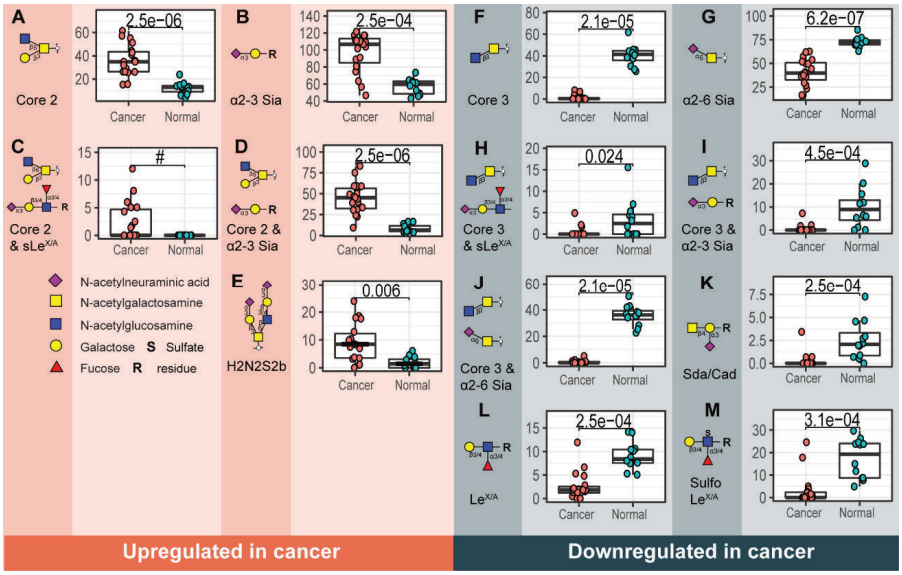

Fig. 3. Structural O-glycan features that differentiate between cancer and normal colon mucosa. *Statistically significant upregulation in cancer is found for various O-glycan features;* ***a)*** *Core 2 O-glycans,* ***b)*** *α2-3 sialylation,* ***c)*** *core 2 O-glycans with sLe$^{X/A}$ and* ***d)*** *core 2 O-glycans with terminal α2-3 sialylation linked to a galactose together with* ***e)*** *one individual O-glycan with composition H2N2S2. Whereas,* ***f)*** *core 3 O-glycans,* ***g)*** *α2-6 sialylation of the core GalNAc,* ***h)*** *core 3 O-glycans with sLe$^{X/A}$ antigen,* ***i)*** *core 3 O-glycans with terminal α2-3 sialylation,* ***j)*** *core 3 O-glycans with terminal α2-6 sialylation,* ***k)*** *Sda antigen,,* ***l)*** *Le$^{X/A}$ and* ***m)*** *sulfo-Le$^{X/A}$ antigen show statistically significant downregulation in cancer. Differences between groups were tested using Wilcoxon-Mann-Whitney non-parametric statistical test. Correction for multiple testing was made using the Benjamini-Hochberg method.* [#] *no p-value is reported as there is no variance in the control group. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

identified. Notably, the sLe$^{X/A}$ antigens and terminal α2-3 sialylation were also carried by core 3 structures in normal colon mucosa and mucinous adenocarcinomas but showed a downregulation in cancer (**Fig. 3h** and **i; Supplementary Fig. S10b** and **Table S10**). *O*-glycans carrying this antigen but with a core 4 structure did not show a difference between cancer and normal mucosa (**Supplementary Fig. S10c** and **Table S10**). Based on our results we suggest that the specificity of the immunotherapeutic antibodies should move towards TACAs with specific (s)Le$^{X/A}$ epitopes carried by core 2 *O*-glycans, and not those that are present on core 3 or core 4. Interestingly, a previous study evaluated the influence of glycan cores for the specificity of the sLe^{X} antibodies, and revealed that commonly used KM93 antibody binds to sLe^{X} on core 2

O-glycans, and not on core 3 *O*-glycans[34]. This reveals the potential for development of specific antibodies targeting a single core-type *O*-glycan in combination with cancer associated $sLe^{X/A}$ epitopes.

GLYCOSYLATION SIGNATURES IN NORMAL COLON MUCOSA

In the normal colon mucosa nearly 80 *O*-glycans were detected and 26 of them were specific for the mucosa, belonging mainly to *O*-glycans with a core 3, carrying an α2-6 sialylation on the core GalNAc, terminal Sda antigen ([GalNAcβ1-4(Neu5Acα2-3)Gal-R) and (sulfo)-$Le^{X/A}$ antigens, and showing statistically significant downregulation in cancer (**Fig. f-m)**. A total of 12 *O*-glycans were shared between cancer, mucosa and microenvironment controls, which were mainly sialylated core 1 *O*-glycans, such as ubiquitous sialyl-3T, and disialyl-T (**Supplementary Fig. S7** and **S12**) and core 2 types with terminal α2-3 sialylation on Gal (**Fig. 3e; Supplementary Fig. S5** and **S7**). Previous studies on the glycosylation of secreted mucus from normal colon mucosa reported expression of very similar glycosylation features, namely a high expression of core 3 structures carrying (sulfo)-$Le^{X/A}$ and Sda antigens. Yet the abundance of sialylated core 1 *O*-glycans was lower than observed here[35,36]. Potentially the glycosylation of the secreted mucins could differ from the cell glycosylation, however, those signatures could also partly originate from the immune cells infiltrated into the healthy colon mucosa, which could not be fully excised using LCM.

GLYCOSYLATION OF SPECIFIC CANCER TYPES

The observed glycomic signatures were further explored by taking into account various cancer types (**Supplementary Fig. S11** and **S12**). No clear clustering was found of cancers with the same differentiation grade, tumor stroma ratio and location origin in the colon (**Supplementary Fig. S13** and **S14**). The tumors with high immune infiltration did not show a specific glycan signature (**Supplementary Fig. S14b**). Moreover, MSI tumors, known for high immune infiltration, did not cluster together on the PCA plot (**Supplementary Fig. S13a**). Sialyl-Tn antigen was detected in both cancer and healthy mucosa (**Supplementary Fig. S12**), however, upregulated in cancer (no significance). It showed the highest expression in T11, a poorly differentiated stage 4 MSI tumor, the two mucinous MSI adenocarcinomas T2 and T3, and neuroendocrine carcinoma T5. Previous reports found that the expression is higher in poorly differentiated and mucinous colon adenocarcinomas, but no

associations with MSI were investigated[37]. It is yet to be determined whether MSI status is associated with expression of specific glycosyltransferases leading to higher expression of α2-6 sialylated Tn and T antigens.

Mucinous adenocarcinomas (T2 and T3) expressed 43 specific *O*-glycans (**Supplementary Fig. S11**), mostly carrying $sLe^{X/A}$ antigen, on core 2 but also core 4 *O*-glycans (**Supplementary Fig. S11** and **S12**). Both tumors were microsatellite instable (MSI) which are known to behave as low grade cancers[38]. Core 4 *O*-glycan expression can be explained by a strong downregulation of *ST6GALNAC1* and *3* in cancer, leading to the branching of the core 3 precursors in cancers with expression of core 4 synthase *GCNT3*[39]. Neuroendocrine carcinoma (T5) showed specifically high expression of Sd^a antigen, compared to other types with low or no expression (**Supplementary Fig. S12**).

Cancers with lymph node invasion (LNI) or invasion to distant organs, including Dukes stage C and D, showed a higher expression of the selected TACAs on core 2 *O*-glycans compared to cancers without invasion (Dukes stage B) (**Fig. 1; Supplementary Table S6**). Due to their absence or limited expression in the normal colon mucosa, and high specificity for invasive cancer, those TACAs can serve as promising targets for treatment of invasive CRC (**Fig. 2**).

PATHWAYS OF *O*-GLYCAN BIOSYNTHESIS IN COLORECTAL CANCER

We created a biosynthetic model (**Fig. 4**) explaining the glycosylation changes that occur in cancer compared to normal colon mucosa by integrating the glycomics results with the online available transcriptomics[40] (**Supplementary Fig. S15**). Downregulation of *B3GNT6*, core 3 synthase that adds β1-3-linked GlcNAc to the core GalNAc, leads to downregulation of core 3 structures, whereas upregulation of *C1GALT1* leads to upregulation of core 1 structures in cancer. Previously, expression of core 3 synthase was associated with better patient prognosis in pancreatic[41], prostate cancer[42] and colon cancer[43]. This competition between core 1 and core 3 synthase as well as core α2-6 sialylation in the context of sialyl-Tn has been described previously in colon cancer cells[44]. We did not observe statistically significant upregulation in the expression of sialyl-Tn antigen in cancer although the trend was present. The main enzyme responsible for the biosynthesis of sialyl-Tn antigen *ST6GALNAC1* is downregulated in cancer, however, a higher expression of sialyl-Tn antigens can be

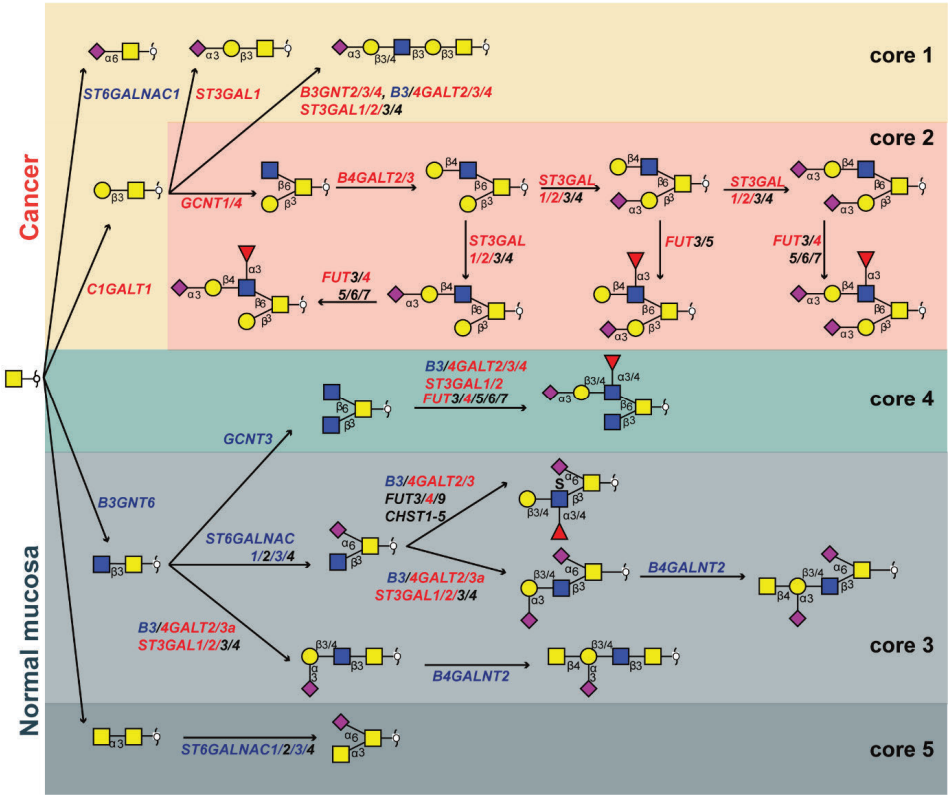

Fig. 4. Proposed biosynthetic model explaining the differences in glycosylation of colorectal cancer and normal colon mucosa. *The most abundant structures in cancer (upper red and yellow panel) and normal colon mucosa (lower green and gray panels) are depicted together with genes encoding for the GTs involved in their biosynthesis. The biosynthetic pathways of different core structures are labelled with different colors. Pathways upregulated in cancer are marked in orange and yellow, whereas pathways downregulated in cancer are marked in light and dark gray. Core 4 O-glycans show no statistically significant difference between normal mucosa and cancer, marked in green. Glycosyltransferase genes upregulated in cancer are labelled with red and those downregulated in cancer are displayed in blue. Blue square: N-acetylglucosamine, green circle: mannose, yellow circle: galactose, red triangle: deoxyhexose, pink diamond: N-acetylneuraminic acid.*

due to a mutation in *C1GALT1C1* gene encoding for a chaperone (Cosmc) required for the activity of core 1 synthase which could lead to a blockage of the alternative pathway[45]. On the other hand, upregulation of *GCNT1* and *GCNT4* (both core 2 synthases that add the GlcNAc to the core GalNAc in the β1-6 position, creating the 6-arm), leads to overexpression of core 2 structures. The upregulation of *GCNT1* and *4* is seen in the microdissected CRC tumors, but not in the TCGA dataset

(**Supplementary Fig. S15**). Addition of a galactose residue to the 6-arm by activity of β1-4-galactosyltransferases is upregulated in cancer [46,47], leading to the biosynthesis of the pentasaccharide with α2-3-linked sialic acid (**Fig. 1g; H2N2S1d**), a TACA with high specificity for cancer and a precursor for other highly specific TACAs (**Fig. 1b-e**). Additionally, the addition of another LacNAc creates a sialyl-dimeric $Le^{X/A}$ antigen glycan which is also specifically found in cancer (**Fig. 1f**). The fucosyltransferase *FUT4* involved in the biosynthesis of sLe^X antigen is overexpressed in colon cancer (**Supplementary Fig. S15**) and it was previously associated with lung adenocarcinoma metastasis and poor patient prognosis[48,49]. It has also been previously described that the upregulation of $sLe^{X/A}$ antigens in CRC is due to downregulation of *B4GALNT2* which adds β1-4-linked GalNAc creating the Sda antigen in the normal colon mucosa[50,51]. This enzyme shows a downregulation in The Cancer Genome Atlas (TCGA) dataset, and only a trend could be observed towards the same direction in LCM dataset (**Supplementary Fig. S15**). Additionally, the expression of 6-sulfo-Le^X and sialyl-6-sulfo-Le^X were previously associated with normal colon epithelia which is in contrast to the expression of sLe^X in CRC[13]. Additionally, a lower expression of $Lewis^A$ type antigens as well as sLe^A was found in CRC compared to normal colon mucosa due to a downregulation of β1-3-galactosyltransferase B3GALT5[52]. The elongation of core 1 structures is in competition with the core 2 biosynthetic pathway and the biosynthesis of sialyl-3T antigen (H1N1S1b). While this antigen is also detected in normal mucosa it shows an overexpression in cancer *(***Supplementary Fig. S12**) due to an upregulation of *ST3GAL1 (***Supplementary Fig. S15)** which is previously associated with lymph node invasion in CRC[53]. Finally, we detected core 5 structure H1N2S1b in the normal colon mucosa, and this has been described before[36]. However, with our PGC-LC-MS/MS approach we could confidently distinguish only one core 5 *O*-glycan (N2S1b) from isomeric core 3 glycans as this glycan was previously characterized by NMR from bovine submaxillary mucin[27]. It is unknown whether any of the remaining core 3 isomers originate from core 5.

In summary, our findings show that the downregulation of core 3 synthase (B3GNT6) in cancer leads to the overexpression of core 1 truncated *O*-glycans such as sialyl3-T by action of *ST3GAL1*, or branched *O*-glycans by the action of core 2 synthases *GCNT1* and *GCNT4.* The addition of a β1-4 galactose and α2-3 sialic acid on the 6

arm of the *O*-glycans forms a specific pathway in cancer, starting from a sialylated precursor (**Fig. 1g**) leading to biosynthesis of fucosylated TACAs carrying terminal sLe^X antigen by the addition of a α1-3-linked fucose by *FUT4* overexpressed in cancer.

FUTURE PERSPECTIVES

Analysis of specific molecular signatures from patient derived material is fundamental for understanding molecular mechanisms of disease onset and progression. FFPE tissue sectioning is part of standard care in pathology and many tissues are widely available and well preserved at room temperature for a long period of time. Yet, tissue heterogeneity poses a problem for glycomics and transcriptomics analysis masking cell specific signatures. Dissecting specific regions of the tissue by LCM, largely overcomes this issue and enables analysis of glycosylation signatures from specific cells of interest[54]. However, it is essential that further improvements are made in regard to sensitivity of our current MS methods, making single cell glycomics analysis possible to avoid any contamination from different cell populations.

It is of great importance to determine the specificity of different glycosyltransferases for core structures, different arms of branched glycans as well as expression of TACAs on different glycan classes. Targeting glycans may have several benefits compared to proteins. Namely, the TACAs are expressed on the cell surface, are directly accessible to therapeutics and can be carried by multiple proteins, reflecting the overall glycosylation phenotype of the cell, providing a broader tumor targeting strategy[55]. Ideally, the expression of TACAs is absent or limited in normal colon mucosa, however, some of the TACAs have been reported as important for the leukocyte extravasation through the endothelium, albeit only at very low relative abundance[56,57] serving as selectin ligands on P-selectin glycoprotein ligand-1 (PSGL-1). Cancer cells carrying the same ligands may employ this physiological mechanism for successful metastasis to distant organs which is why it is important to target these antigens, particularly in invasive carcinoma[30].

While there is no doubt that the differences in the TACAs patterns between the cancer and healthy mucosa are a consequence of changes in the glycosylation machinery of the cells, it remains unclear whether the TACAs are protein specific, and if the differences in their expression is related to mucin expression. With the aim to increase immunogenicity and specificity of glycan targets, they are often coupled to protein

carriers, either adjuvants or pathological carriers such as mucins. Aberrant expression of Tn antigen was found on MUC1 membrane glycoprotein, and anti-Tn and T antigen carrying MUC1 expressed on showed efficiency in phase I clinical trial on ovarian, breast and cervical cancer[58] but no effect on patient outcome in phase II clinical trial[59]. Another monoclonal antibody targeting Tn-antigen on MUC1 showed promising preclinical results for targeting breast cancer, however, the results need further validation by clinical trials[60]. Additionally, genetically modified T-cells expressing chimeric antigen receptors (CAR-s) targeting Tn glycoforms of MUC1 showed promising cytotoxicity in xenograft models of T-cell leukemia and pancreatic cancer[61]. Our study showed that core 2 sialylated or core 2 $sLe^{X/A}$ carrying *O*-glycans have the highest specificity for CRC, however, more research is needed to determine if these signatures are mucin specific, and whether bispecific antibodies or CAR-T cells targeting both the protein carrier and the cancer specific glycans will increase the specificity and immunogenicity of the developed therapeutics.

CONCLUSION

In this study we present a novel panel of highly specific TACAs, based upon changes in the *O*-glycomic profile between CRC and healthy colon mucosa. These TACAs are promising new targets for development of innovative cancer immunotherapeutics and lay the foundation for the treatment of invasive CRC.

SUPPORTING INFORMATION

Supporting information is available upon request.

DATA AVAILABILITY

The data in support of the findings of this study may be found within the manuscript and in the associated supplementary files. Mass spectrometry-based glycomics raw data were deposited in the Glycopost repository accessible via the following link: https://glycopost.glycosmos.org/preview/460809305620fc81b8a4e5 (pin code: 6390). Detailed MS/MS spectra peaklists with annotations are available in the Glycoworkbench files.

ACKNOWLEDGMENTS

The authors thank Meaghan Polack for the tumor stroma ratio assessment and Stijn A.S.L.P. Crobach for a discussion about cancer slides included in this study. Additionally, we thank Carolien A.M. Koeleman, Lisa A. de Neef and Agnes L. Hipgrave Ederveen for technical assistance. This work was supported by the European Commission's Horizon 2020 programme ''GlyCoCan'' project, grant number 676421. The results shown in Supplementary Fig. S15 are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

AUTHOR INFORMATION

CONTRIBUTIONS

K.M. performed the experiments. K.M., G.L., M.W. conceptually designed the work. K.M., G.L., M.W. wrote the manuscript. R.V. assisted with laser capture microdissections. T.Z assisted with glycan MS/MS spectra annotation. T.W. assisted with study design and access to patient material. O.M. assisted with statistical analysis. H.M. annotated HE slides. J.W. and G.B. prepared glycan standards. All authors read and commented on the manuscript.

CORRESPONDING AUTHOR

Manfred Wuhrer – Center for Proteomics and Metabolomics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands; orcid.org/0000-0002-0814-4995

ETHICS DECLARATION

The authors declare no competing interests.

REFERENCES

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**, E359–E386 (2015).
2. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
3. Rosenblum, D., Joshi, N., Tao, W., Karp, J. M. & Peer, D. Progress and challenges towards targeted delivery of cancer therapeutics. *Nature Communications* vol. 9 1–12 (2018).
4. Piawah, S. & Venook, A. P. Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. *Cancer* **125**, 4139–4147 (2019).

5. Ayyar, B. V., Arora, S. & O'Kennedy, R. Coming-of-Age of Antibodies in Cancer Therapeutics. *Trends Pharmacol. Sci.* **37**, 1009–1028 (2016).
6. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
7. Holst, S., Wuhrer, M. & Rombouts, Y. *Glycosylation characteristics of colorectal cancer*. vol. 126 203–256 (Elsevier Inc., 2015).
8. Kannagi, R., Yin, J., Miyazaki, K. & Izawa, M. Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants-Hakomori's concepts revisited. *Biochimica et Biophysica Acta - General Subjects* **1780**, 525–531 (2008).
9. Chen, W.-S., Chang, H.-Y., Li, C.-P., Liu, J. M. & Huang, T.-S. Tumor beta-1,4-galactosyltransferase IV overexpression is closely associated with colorectal cancer metastasis and poor prognosis. *Clin. Cancer Res.* **11**, 8615–8622 (2005).
10. Krishn, S. R. *et al.* Mucins and associated glycan signatures in colon adenoma-carcinoma sequence: Prospective pathological implication(s) for early diagnosis of colon cancer. **374**, 304–314 (2017).
11. Costa, A. F., Campos, D., Reis, C. A. & Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends in Cancer* vol. 6 757–766 (2020).
12. Jass, J. R., Allison, L. J. & Edgar, S. G. Distribution of sialosyl Tn and Tn antigens within normal and malignant colorectal epithelium. *J. Pathol.* **176**, 143–149 (1995).
13. Izawa, M. *et al.* Expression of sialyl 6-sulfo Lewis X is inversely correlated with conventional sialyl Lewis X expression in human colorectal cancer. *Cancer Res.* **60**, 1410–1416 (2000).
14. Itzkowitz, S. H. *et al.* Expression of Tn, Sialosyl-Tn, and T Antigens in Human Colon Cancer. *Cancer Res.* **49**, (1989).
15. Fukasawa, T. *et al.* Associated expression of α2,3sialylated type 2 chain structures with lymph node metastasis in distal colorectal cancer. *Surg. Today* **43**, 155–162 (2013).
16. Ito, H. *et al.* Altered mrna expression of specific molecular species of fucosyl-and sialyl-transferases in human colorectal cancer tissues. vol. 71 556–564 1097-0215(19970516)71:4 (1997).
17. Kudo, T. *et al.* Up-regulation of a set of glycosyltransferase genes in human colorectal cancer. *Lab. Invest.* **78**, 797–811 (1998).
18. Shimodaira, K. *et al.* Carcinoma-associated Expression of Core 2 β-1,6-N-Acetylglucosaminyltransferase Gene in Human Colorectal Cancer: Role of O-Glycans in Tumor Progression. *Cancer Res.* **57**, (1997).
19. Hinneburg, H. *et al.* Unlocking Cancer Glycomes from Histopathological Formalin-fixed and Paraffin-embedded (FFPE) Tissue Microdissections. *Mol. Cell. Proteomics* **16**, 524–536 (2017).
20. Hinneburg, H., Schirmeister, F., Korać, P. & Kolarich, D. N- and O-Glycomics from Minor Amounts of Formalin-Fixed, Paraffin-Embedded Tissue Samples. in 131–145 (Humana Press, New York, NY, 2017). doi:10.1007/978-1-4939-6493-2_11.

21. Guo, Y. W. *et al.* Effect of PVP Hydrophilic Additive on the Morphology and Properties of PVDF Porous Membranes. *Adv. Mat. Res.* **981**, 891–894 (2014).
22. Madunić, K., Wagt, S., Zhang, T., Wuhrer, M. & Lageveen-Kammeijer, G. S. M. Dopant-Enriched Nitrogen Gas for Enhanced Electrospray Ionization of Released Glycans in Negative Ion Mode. *Anal. Chem.* acs.analchem.1c00023 (2021) doi:10.1021/acs.analchem.1c00023.
23. Robbe, C. *et al.* Evidence of regio-specific glycosylation in human intestinal mucins: Presence of an acidic gradient along the intestinal tract. *J. Biol. Chem.* **278**, 46337–46348 (2003).
24. Holmén Larsson, J. M., Karlsson, H., Sjövall, H. & Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756–766 (2009).
25. Savage, A. V., Donohue, J. J., Koeleman, C. A. M. & van den Eijnden, D. H. Structural characterization of sialylated tetrasaccharides and pentasaccharides with blood group H and Lex activity isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **193**, 837–843 (1990).
26. Savage, A. V., D'Arcy, S. M. T. & Donoghue, C. M. Structural characterization of neutral oligosaccharides with blood group A and H activity isolated from bovine submaxillary mucin. *Biochem. J* **279**, 95–103 (1991).
27. Savage, A. V., Donoghue, C. M., D'Arcy, S. M., Koeleman, C. A. M. & van den Eijnden, D. H. Structure determination of five sialylated trisaccharides with core types 1, 3 or 5 isolated from bovine submaxillary mucin. *Eur. J. Biochem.* **192**, 427–432 (1990).
28. Zhang, T. *et al.* Development of a 96-well plate sample preparation method for integrated: N - And O -glycomics using porous graphitized carbon liquid chromatography-mass spectrometry. *Molecular Omics* **16**, 355–363 (2020).
29. Kazama, Y. *et al.* Microsatellite instability in poorly differentiated adenocarcinomas of the colon and rectum: relationship to clinicopathological features. *J. Clin. Pathol.* **60**, 701–704 (2007).
30. Trinchera, M. *et al.* Selectin Ligands Sialyl-Lewis a and Sialyl-Lewis x in Gastrointestinal Cancers. *Biology* **6**, 16 (2017).
31. Burdick, M. D., Harris, A., Reid, C. J., Iwamura, T. & Hollingsworth, M. A. Oligosaccharides expressed on MUC1 produced by pancreatic and colon tumor cell lines. *J. Biol. Chem.* **272**, 24198–24202 (1997).
32. Gomes, C. *et al.* Carcinoembryonic antigen carrying SLe X as a new biomarker of more aggressive gastric carcinomas. *Theranostics* **9**, 24 (2019).
33. 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).
34. Löfling, J. & Holgersson, J. Core saccharide dependence of sialyl Lewis X biosynthesis. *Glycoconj. J.* **26**, 33–40 (2009).

35. Mihalache, A. *et al.* Structural Characterization of Mucin O-Glycosylation May Provide Important Information to Help Prevent Colorectal Tumor Recurrence. *Front. Oncol.* **5**, 217 (2015).
36. Robbe C. Capon C. Coddeville B. & Michalski, J. C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem. J* **384**, 307–316 (2004).
37. Grosso, M., Vitarelli, E., Giuffrè, G., Tuccari, G. & Barresi, G. Expression of Tn, sialosyl-Tn and T antigens in human foetal large intestine. *Eur. J. Histochem.* **44**, 359–363 (2000).
38. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *Journal of Gastrointestinal Oncology* vol. 3 153–173 (2012).
39. Yeh, J. C., Ong, E. & Fukuda, M. Molecular cloning and expression of a novel β-1,6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I branches. *J. Biol. Chem.* **274**, 3215–3221 (1999).
40. Kogo, R. *et al.* Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res.* **71**, 6320–6326 (2011).
41. Doi, N. *et al.* Clinicopathological significance of core 3 O-glycan synthetic enzyme, β1,3-N-acetylglucosaminyltransferase 6 in pancreatic ductal adenocarcinoma. *PLoS One* **15**, e0242851 (2020).
42. Lee, S. H. *et al.* Core3 O-glycan synthase suppresses tumor formation and metastasis of prostate carcinoma PC3 and LNCaP cells through down-regulation of α2β1 integrin complex. *J. Biol. Chem.* **284**, 17157–17169 (2009).
43. Iwai, T. *et al.* Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4572–4577 (2005).
44. Barrow, H., Tam, B., Duckworth, C. A., Rhodes, J. M. & Yu, L. G. Suppression of Core 1 Gal-Transferase Is Associated with Reduction of TF and Reciprocal Increase of Tn, sialyl-Tn and Core 3 Glycans in Human Colon Cancer Cells. *PLoS One* **8**, e59792 (2013).
45. Ju, T. *et al.* Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res.* **68**, 1636–1646 (2008).
46. Isshiki, S. *et al.* Cloning, Expression, and Characterization of a Novel UDP-galactose:β-N-Acetylglucosamine β1,3-Galactosyltransferase (β3Gal-T5) Responsible for Synthesis of Type 1 Chain in Colorectal and Pancreatic Epithelia and Tumor Cells Derived Therefrom *. *J. Biol. Chem.* **274**, 12499–12507 (1999).
47. Salvini, R., Bardoni, A., Valli, M. & Trinchera, M. β1,3-Galactosyltransferase β3Gal-T5 Acts on the GlcNAcβ1→3Galβ1→4GlcNAcβ1→R Sugar Chains of Carcinoembryonic Antigen and Other N-Linked Glycoproteins and Is Down-regulated in Colon Adenocarcinomas *. *J. Biol. Chem.* **276**, 3564–3573 (2001).
48. Lu, H. H. *et al.* Fucosyltransferase 4 shapes oncogenic glycoproteome to drive metastasis of lung adenocarcinoma. *EBioMedicine* **57**, 102846 (2020).

49. Park, S. *et al.* Altered expression of fucosylation pathway genes is associated with poor prognosis and tumor metastasis in non-small cell lung cancer. *Int. J. Oncol.* **56**, 559–567 (2020).
50. 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).
51. Pucci, M., Ferreira, I. G., Orlandani, M. & Malagolini, N. High Expression of the Sd a Synthase B4GALNT2 Associates with Good Prognosis and Attenuates Stemness in Colon Cancer. **3**, (2020).
52. Miyazaki, K. *et al.* Loss of disialyl Lewisa, the ligand for lymphocyte inhibitory receptor sialic acid-binding immunoglobulin-like lectin-7 (Siglec-7) associated with increased sialyl Lewisa expression on human colon cancers. *Cancer Res.* **64**, 4498–4505 (2004).
53. Schneider, F. *et al.* Overexpression of sialyltransferase CMP-sialic acid:Galbeta1,3GalNAc-R alpha6-Sialyltransferase is related to poor patient survival in human colorectal carcinomas. *Cancer Res.* **61**, 4605–4611 (2001).
54. Lageveen-Kammeijer, G. S. M., Kuster, B., Reusch, D. & Wuhrer, M. High sensitivity glycomics in biomedicine. *Mass Spectrom. Rev.* e21730 (2021) doi:10.1002/mas.21730.
55. Houvast, R. D. *et al.* Targeting glycans and heavily glycosylated proteins for tumor imaging. *Cancers* vol. 12 1–26 (2020).
56. Lo, C. Y. *et al.* Competition between core-2 GlcNAc-transferase and ST6GalNAc-transferase regulates the synthesis of the leukocyte selectin ligand on human P-selectin glycoprotein ligand-1. *J. Biol. Chem.* **288**, 13974–13987 (2013).
57. Wilkins, P. P., McEver, R. P. & Cummings, R. D. Structures of the O-glycans on P-selectin glycoprotein ligand-1 from HL- 60 cells. *J. Biol. Chem.* **271**, 18732–18742 (1996).
58. Fiedler, W. *et al.* A phase i study of PankoMab-GEX, a humanised glyco-optimised monoclonal antibody to a novel tumour-specific MUC1 glycopeptide epitope in patients with advanced carcinomas. *Eur. J. Cancer* **63**, 55–63 (2016).
59. Ledermann, J. *et al.* A double-blind, placebo-controlled, randomized, phase 2 study to evaluate the efficacy and safety of switch maintenance therapy with the anti-TA-MUC1 antibody PankoMab-GEX after chemotherapy in patients with recurrent epithelial ovarian carcinoma. *Ann. Oncol.* **28**, v626 (2017).
60. Lavrsen, K. *et al.* Aberrantly glycosylated MUC1 is expressed on the surface of breast cancer cells and a target for antibody-dependent cell-mediated cytotoxicity. *Glycoconj. J.* **30**, 227–236 (2013).
61. Posey, A. D. *et al.* Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity* 44, 1444–1454 (2016).