Unraveling mucin type o-glycosylation signatures of colorectal cancer

Madunić. K.

INTEGRATED GLYCOMIC AND PROTEOMIC SIGNATURES OF BUTYRATE-STIMULATED COLORECTAL CANCER CELL LINE DIFFERENTIATION

Chapter 5

K. Madunić[a], Y.M.C.A. Luijkx[b,c], O.A. Mayboroda[a], G.M.C. Janssen[a], P.A. van Veelen[a], K. Strijbis[c], T. Wennekes[b], G.S.M. Lageveen-Kammeijer[a], M. Wuhrer[a]

[a] *Center for Proteomics and Metabolomics, Leiden University, The Netherlands.*

[b] *Department Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.*

[c] *Department Biomolecular Health Sciences, Utrecht University, Utrecht, The Netherlands*

Manuscript submitted

ABSTRACT

Gut microbiota of the gastrointestinal tract provide health benefits to the human host *via* bacterial metabolites. Bacterial butyrate has beneficial effects on intestinal homeostasis and is the preferred energy source of intestinal epithelial cells, capable of inducing differentiation. It was previously observed that changes in the expression of specific proteins as well as protein glycosylation occur with differentiation. In this study, specific mucin *O*-glycans were identified that mark butyrate-induced epithelial differentiation of the intestinal cell line CaCo-2, by applying porous graphitized carbon nano-liquid chromatography with electrospray ionization tandem mass spectrometry (PGC-nanoLC-MS/MS). Moreover, a quantitative proteomic approach was used to decipher changes in the cell proteome. It was found that the fully differentiated butyrate-stimulated cells are characterized by a higher expression of sialylated *O*-glycan structures, whereas fucosylation is downregulated with differentiation. By performing an integrative approach, we generated hypotheses about glycosylation signatures of specific cell adhesion proteins. These insights pave the way for future endeavors to study the dynamic *O*-glycosylation patterns in the gut, either produced *via* cellular biosynthesis or through the action of bacterial glycosidases as well as the functional role of these patterns in homeostasis and dysbiosis at the gut-microbiota interface.

INTRODUCTION

The human gut microbiota is a complex ecology of a variety of different microorganisms. Among viruses, prokaryotes, and eukaryotes, bacteria are the most abundant inhabitants of the human gastrointestinal (GI) tract. By co-evolution with the host, a symbiotic relationship has been formed between the GI tract and its bacteria[1,2]. The maintenance of homeostasis in the GI tract depends on the complex process of epithelial cell differentiation[3]. Bacterial metabolites also play a functional role in the maintenance of this homeostasis[4]. Butyrate, a short chain fatty acid, is a bacterial metabolite produced by bacterial fermentation of dietary fibers and is known to have a beneficial effect on the intestinal homeostasis. Physiological concentrations of butyrate inhibit cell proliferation, induce differentiation and increase the rate of apoptosis for a number of tumor cell types *in vivo* and *in vitro*[5–8]. Furthermore, butyrate is the main energy source of intestinal epithelial cells and capable of upregulating the gene expression of both secreted and membrane-linked mucins[9–13]. Notably, bacterial short chain fatty acids, such as butyrate, are suspected of also extending their effect outside of the gut as they have been implicated as possible mediators of phenomena observed in the gut-brain axis[14].

The mucus layer covers the epithelial cells to prevent them from being in direct contact with the microbiota. The function of the mucosal layer as a barrier is largely maintained by the gel-forming mucins, which are large, highly *O*-glycosylated proteins secreted by the epithelial cells[15]. Furthermore, epithelial cells express membrane-linked mucins that have both a barrier and a signaling function[16]. The intracellular tails of the membrane-linked mucins can be phosphorylated and mediate signaling, thereby regulating cell-cell interactions, differentiation, and apoptosis[16]. Additionally, the mucus layer and especially its *O*-glycans represent an important nutrient source for the surrounding microbiota and thereby contribute to bacterial colonization in the human gut. Typically, the majority of the mucin dry weight is made up of *O*-glycans and is the main modification that largely determines the properties and function of the mucins[17]. A wide variety of oligosaccharide structures can be attached to the mucins as well as other *O*-glycoproteins and the composition of these structures can vary within cell types due to differential expression of glycosidases and glycosyltransferases[18].

Furthermore, membrane-linked *O*-glycoproteins are seen as important oncogenic proteins as their intracellular domain and their glycosylated extracellular domain link to pathways involved in cell differentiation and apoptosis[19]. Notably, altered glycosylation can be correlated to the levels of the bacterial metabolite butyrate that regulates specific glycogenes including galectin-1 and β-galactoside-α-2-6-sialyltransferase (ST6GAL1)[20–23].

In this study the intestinal cell line, CaCo-2 (*Ca*ncer *co*li-*2*), derived from a human colorectal carcinoma (CRC) in 1977, is investigated and became a well-established model since then to study cellular differentiation. The cell line can differentiate spontaneously or by exposure to butyrate into polarized cells with morphological and biochemical features of the mature colonic epithelium[24]. Spontaneous differentiation of CaCo-2 cells has been studied previously, including transcriptomic, proteomic and glycomic analysis[25–32].The changes in *O*-glycosylation upon butyrate differentiation have yet to be characterized and insights into these changes is crucial for a better understanding of the role of glycans in gut homeostasis. The aim of the present study was to identify specific *O*-glycan features that define butyrate-induced epithelial differentiation in relation to spontaneous differentiation to gain insights into *O*-glycan signatures of colon cancer cell line differentiation derived from the gut epithelium.

MATERIALS AND METHODS

MATERIALS

Sodium borohydride ($NaBH_4$), sodium chloride (NaCl), Dowex cation-exchange resin (50W-X8), ammonium bicarbonate (ABC), trifluoroacetic acid (TFA) Dulbecco's Phosphate Buffered Saline (DPBS), Hydrochloric acid (HCl), DL-dithiothreitol (DTT), and were purchased from Sigma-Aldrich (Steinheim, Germany). Ethanol (Reag. Ph. Eur) and Mucin from bovine submaxillary glands, type I-S were purchased from Merck (BSM; Darmstadt, Germany). TMTpro Label Reagents, 8 M guanidine hydrochloride (GuHCl), Dulbecco's modified Eagle's medium (DMEM), 0.25% trypsin/EDTA, and fetal calf serum (FCS) were obtained from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Potassium hydroxide (KOH) was obtained from Honeywell Fluka. Solid phase extraction (SPE) bulk sorbent Carbograph was obtained from Grace Discovery sciences (Columbia, USA). HPLC SupraGradient acetonitrile (MeCN) was obtained from Biosolve (Valkenswaard, The Netherlands). Peptide N-

glycosidase F (PNGase F) and complete EDTA free protease inhibitor cocktail tablets were purchased from Roche Diagnostics (Mannheim, Germany). 96-well PP filter plate was purchased from Orochem Technologies (Naperville, IL, USA). MultiScreen HTS 96 multiwell plates (hydrophobic Immobilon-P PVDF membrane) and 96-well PP Microplate were obtained from Millipore (Amsterdam, the Netherlands).

CELL CULTURE

Human colorectal adenocarcinoma Caco-2 cells were grown in Dulbecco's Modified Eagle medium (Gibco, Thermo Fisher Scientific) with 10 % FCS. Cells were subcultured at 80% confluency and maintained at 37 °C in a humidified incubator with 5 % CO_2. At day 0, when cells reached full confluency, 2 mM butyrate was added to the appropriate cells. On days 5, 7, 14, 21 and 24 cells were collected in biological triplicates. For harvesting of the cells, medium was removed and adherent cells were washed twice with DPBS and trypsinized using 0.25 % trypsin- 1 mM EDTA. To stop the trypsin activity, medium (without FCS) in a ratio of 2:5 (trypsin:medium; *v/v*) was added and cells were pelleted at 300 x g for 5 min. Cells were resuspended in DPBS and counted and aliquoted to ~2.0 x 10^6 cells per replicate and washed twice with 1 mL DPBS for 3 minutes at 100 x g. The supernatant was removed and cell pellets stored at -20 °C until further use.

PHASE-CONTRAST MICROSCOPY

Phase images were acquired using phase contrast microscopy with 40 times objective. Images were acquired through an CMEX 5000 Microscope camera.

ALKALINE PHOSPHATASE ACTIVITY

Alkaline phosphatase (ALP) activity was measured in cell lysates using a colorimetric assay according to the manufacturer's instructions (Abcam). The cells were harvested with 0.25 % trypsin- 1 mM EDTA, washed with PBS and subjected to ultra-sonification in the supplied assay buffer. The obtained cell homogenates were stored at -80 °C until assayed. Cells lysed at different growth days were incubated with 1.7 mM *p*-nitrophenylphosphate (*p*-NPP) for 60 min at 25 °C. Absorbance readings were taken at 405 nm with a microplate reader (Fluorimeter Fluorstar Omega, BMG Labtech). Results are expressed as milli units/mg protein. One unit is defined as the activity that hydrolyzes 1 micromole of *p*-NPP/min at 25 °C.

CELL LYSIS AND *O*-GLYCAN RELEASE

Three biological replicates from each time point were analyzed. Frozen cell pellets containing ~2.0 x 10^6 cells were resuspended in 100 µL of lysis buffer containing Tris HCl, EDTA, NaCl and protease inhibitor cocktail. The cells were lysed using a Branson sonicator rod at 1.5, and 25 µL of the suspension (containing ~ ~5 x 10^5 cells) were loaded onto the preconditioned PVDF membrane plate wells. BSM (10 µg) was blotted in three different wells of the same PVDF membrane plate. The denaturation as well as the *N*- and *O*-glycan release were performed as described previously[31,33]. Briefly, the proteins were denatured on membrane using guanidine hydrochloride and DTT at 60 °C. Upon removal of denaturing agents, the *N*-glycans were released by PNGase F digestion overnight at 37 °C, and recovered in MQ water. Upon recovery of *N*-glycans, the *O*-glycans were released from the same wells by reductive beta elimination, using sodium borohydride in potassium hydroxide incubating at 50 °C for 16 h. Sample desalting by cation exchange resin Dowex 50 W X8 and graphitized carbon solid phase extraction were performed in self-packed 96-well filter plates. The samples were dried after cleaning and stored at –20 °C until analysis.

PGC-LC-MS/MS ANALYSIS

The *O*-glycan samples were then reconstituted in 20 µL of MQ water, and 2 µL were injected for analysis. Analysis was performed using a PGC nano-LC Ultimate 3000 UHPLC system (Thermo Fisher Scientific, Sunnyvale, CA) coupled to an amaZon ETD speed ion trap (Bruker Daltonics, Bremen, Germany). The samples were loaded using 100 % buffer A (10 mM ammonium bicarbonate) at a loading flow of 6 µL/min on 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. Afterwards, the *O*-glycans were separated at a 0.6 µL/min flow rate on a custom-made PGC column (100 mm × 100 µm, 3 µm particle size obtained from Thermo Fisher Scientific) by applying a linear gradient from 1 % to 50 % buffer B (MeCN, 10 mM ammonium bicarbonate) over 73 min. During the procedures a constant column temperature of 45˚C was maintained. To continue, 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 flow rate was set to 3 L/min and the drying gas (N_2) temperature was set at 280 ˚C. The nebulizer gas pressure was kept at 3 psi. The

nanoBooster™ bottle (Bruker Daltonics) was filled with methanol, as a dopant solvent[34]. MS spectra were acquired in enhanced mode within a mass to charge ratio (*m/z*) range of 380-1850. The maximum acquisition time was set to 200 ms mass, the ion charge control (ICC) to 40,000, and the target of smart parameter setting was set to *m/z* 900. MS/MS spectra were generated by collision-induced dissociation on the three most abundant precursors, applying an isolation width of 3 Thomson. Additionally, ICC was set to 150,000 and the fragmentation cut-off was set to 27 % with a 100 % fragmentation amplitude using the Enhanced SmartFrag option (30 – 120 % in 32 ms). To integrate area under the curve (AUC) for each individual glycan isomer, extracted ion chromatograms of the first three isotopes were used in Compass DataAnalysis software (v.5.0). Peaks were manually picked and integrated. Total area normalization was employed for relative quantification of *O*-glycan species. Identification of *O*-glycan species was performed by comparison with PGC retention time, MS/MS spectra, and a BSM standard.

QUANTITATIVE PROTEOMICS USING TMT LABELLING

Cell lysis, digestion and TMT labeling was performed as described previously[35]. In short, ~1.0 x 10^6 Caco-2 cells were lysed in SDS lysis buffer (SDS [5 %], Tris-HCl [100 mM, pH 7.6]) at 95°C for 4 min. Protein concentration was determined by Pierce BCA protein assay (Thermo Fisher Scientific). Around 100 µg of protein was used for subsequent reduction with 5 mM TCEP, alkylation with 15 mM iodoacetamide and quenching with 10 mM DTT. Protein lysates were purified by methanol-chloroform precipitation. The resulting protein pellets were resuspended in 40 mM HEPES pH 8.4 and digested with trypsin (10 µg) for 16 hrs at 37 °C. Peptide concentration was measured with Pierce BCA assay.

The 10 different conditions together with common reference samples were arranged in triplicate in three TMTpro 13-plex experiments. Of each peptide preparations, 10 µg was dissolved in 25 µl of HEPES buffer (40 mM, pH 8.4) and incubated with 40 µg of one of 13 amino reactive TMTpro Label for 1h at ambient temperature. Excess TMT label was quenched by addition of 6 µL 5 % hydroxylamine and incubation for 15 min at ambient temperature. Labeled peptide samples were then mixed and freeze-dried. TMT-labeled mixtures were dissolved in 1 mL 10 mM ammonium bicarbonate and fractionated on 1cc C18 SPE cartridges (Oasis HLB, Waters) using 5 %, 10 %, 15 %, 20 %, 25 % and 35 % MeCN in 10 mM ammonium bicarbonate.

TMT-labeled peptides were dissolved in water/formic acid (100/0.1 *v/v*) and subsequently analyzed twice by on-line C18 nanoHPLC MS/MS with a system consisting of an Ultimate3000nano gradient HPLC system (Thermo, Bremen, Germany), and an Exploris480 mass spectrometer (Thermo). Fractions were injected onto a cartridge precolumn (300 µm × 5 mm, C18 PepMap, 5 um, 100 A, and eluted via a homemade analytical nano-HPLC column (50 cm × 75 µm; Reprosil-Pur C18-AQ 1.9 um, 120 A (Dr. Maisch, Ammerbuch, Germany). The gradient was run from 2 % to 40 % solvent B (20/80/0.1 water/acetonitrile/formic acid (FA); *v/v/v*) in 120 min. The nano-HPLC column was drawn to a tip of ~10 µm and acted as the electrospray needle of the MS source. The mass spectrometer was operated in data-dependent MS/MS mode for a cycle time of 3 s, with a HCD collision energy at 30 V and recording of the MS2 spectrum in the orbitrap, with a quadrupole isolation width of 1.2 Da. In the master scan (MS1) the resolution was 120,000, the scan range 400-1500, at standard AGC target at maximum fill time of 50 ms. A lock mass correction on the background ion *m/z* at 445.12 was used. Precursors were dynamically excluded after n = 1 with an exclusion duration of 45 s, and with a precursor range of 20 ppm. Charge states 2-5 were included. For MS2 the first mass was set to 120 Da, and the MS/MS scan resolution was 45,000 at an AGC target of 200 % at maximum fill time of 60 ms.

In a post-analysis process, raw data were first converted to peak lists using Proteome Discoverer version 2.2 (Thermo Electron), and submitted to the Uniprot database (Homo sapiens, 20596 entries), using Mascot v.2.2.07 (www.matrixscience.com) for protein identification. Mascot searches were with 10 ppm and 0.02 Da deviation for precursor and fragment mass, respectively, and trypsin as enzyme. Up to two missed cleavages were allowed. Methionine oxidation and acetyl on protein N-terminus were set as a variable modification; carbamidomethyl on Cys, TMTpro on N-terminus and Lys were set as a fixed modification. Protein FDR was set to 1 %. Normalization was performed on total peptide amount.

STATISTICAL ANALYSIS

Data analysis and visualization was performed using in-house developed “R” scripts. To enable use of principal component analysis, imputation of minimum positive number (0.0001) was performed. Differences between groups were tested using two-way ANOVA analysis of variance both for glycomics and proteomics datasets.

RESULTS

CELL DIFFERENTIATION ASSAY

O-glycosylation of CaCo-2 cells was analyzed at different time points during the process of spontaneous and butyrate-induced differentiation. Cells were proliferated to confluency, which is marked as day 0, after which cells start their differentiation. Differentiation of CaCo-2 cells was evaluated by a colorimetric assay measuring alkaline phosphatase (AP) activity, which is a marker for late stage differentiation, indicative for the presence of an established brush border[36]. AP activity levels of both spontaneous differentiation and induced differentiation showed a continuing increase

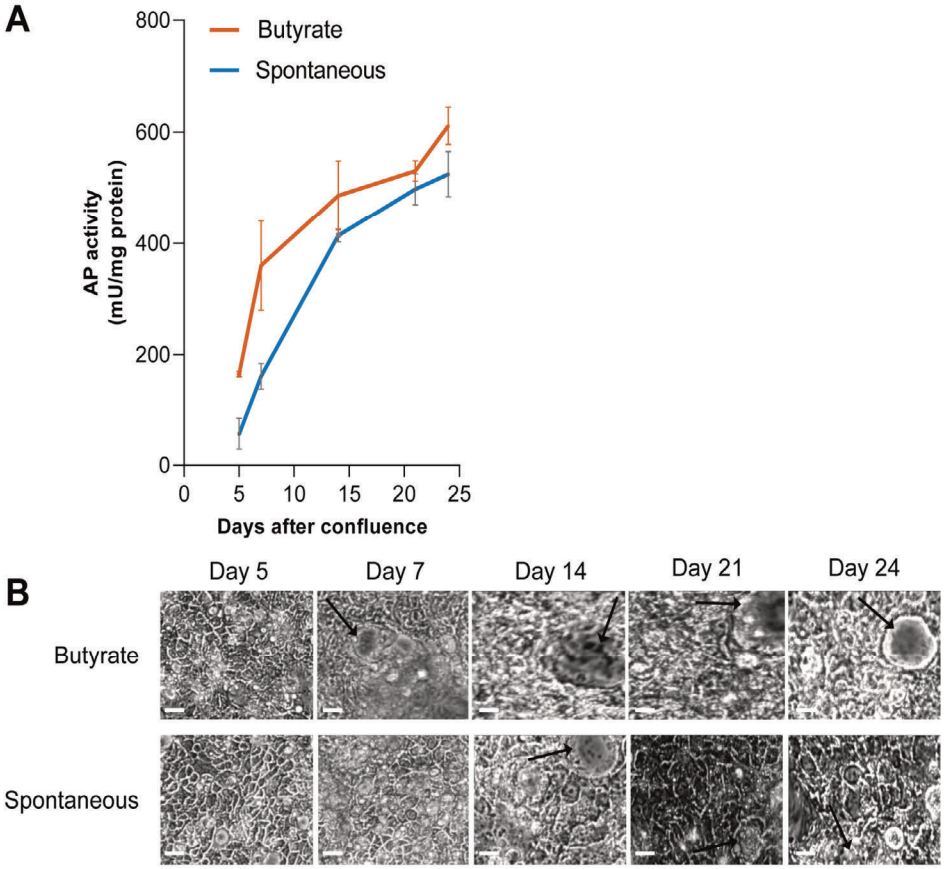

Figure 1. Biochemical and morphological validation of CaCo-2 differentiation. a) *Alkaline phosphatase (AP) activity during CaCo-2 growth. Stimulated by 2 mM butyrate (orange line) or spontaneous differentiation (blue line). Standard deviation is indicated by error bars (n = 3).* ***b)*** *Phase contrast images of CaCo-2 cells on culture dishes at different stages of growth. Arrow; dome-like structure. Original magnification; 40x. White scale bar = 50 µm.*

(**Figure 1a**). A significantly higher level of AP activity for the induced differentiation was observed for days 5 and 7, compared to the spontaneous differentiation.

In addition to AP activity levels as a validation for differentiation, phase images were taken of the apical surface to observe differentiation induced "dome" formation (**Figure 1b**). CaCo-2 cells imaged at different growth times started displaying dome formation at day 7 for the butyrate-treated cells, whereas the spontaneously differentiated cells showed dome formation much later (day 14). After having established that CaCo-2 cells were differentiated, our next aim was to assess the effect of butyrate on the *O*-glycomic profile of these CaCo-2 cells.

GLYCOMIC ANALYSIS

The analysis of *O*-glycans can be challenging in comparison to *N*-glycans, due to the lack of specific enzymes that can release the intact *O*-glycans from glycosylated proteins. In this study, we used a previously established protocol that allowed the chemical release of *O*-glycans from cell lysates via reductive β-elimination in 96-well plate format[31]. *O*-glycans of the differentiating CaCo-2 cells were analyzed at days 5, 7, 14, 21 and 24 post confluency. To further validate the robustness of this approach, full technical triplicates of the cell lysate (5 x 10^5 cells) from the same biological replicate (5 days, spontaneous differentiation) were processed and analyzed independently by porous graphitized carbon nano-liquid chromatography coupled to mass spectrometry (PGC-nanoLC–ESI-MS/MS). The mean relative area of the 11 most abundant glycan species and corresponding standard deviations are shown in **Supplementary Information 1, Figure S1a**, demonstrating the low technical variability of the workflow. Parallel to this, the PGC-nanoLC–ESI-MS performance was assessed by releasing *O*-glycans from 10 µg of bovine submaxillary mucin (BSM) standard and measuring them across five days, as illustrated in **Supplementary Information, Figure S1b**. Overall, the method showed very good precision.

To explore the *O*-glycomic profiles a principal component analysis (PCA) was performed on relative abundances of individual glycans detected in the glycomic profiles (**Supplementary Information 2, Table S1).** As illustrated in **Supplementary Information 1**, **Figure S2a**, the model showed narrow clustering of most biological replicates, although substantial differences were observed within replicates for time points 24 days (butyrate) and 5 days (spontaneous). Different time points of

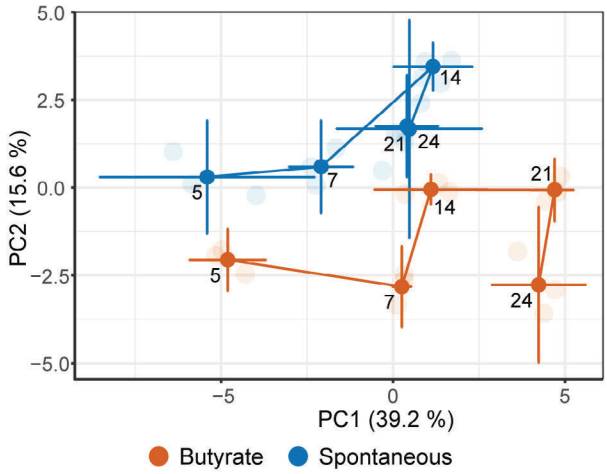

Figure 2. The geometric trajectory visualization of glycomic changes with differentiation. *A distinction between the butyrate-stimulated group (orange) and spontaneous differentiation (blue) is observed in the PCA model based on relative abundance (%) of different O-glycans. The separation between different time points in the two groups is illustrated as a trajectory. The top two principal components (PC) explain 54.8% of the variation within the data. The PCA scores from different biological replicates (faded color) were averaged to create the trajectory.*

differentiation were separated along the PC1 (39.2 %). A clear separation was apparent between the butyrate-stimulated samples and spontaneously differentiated samples along PC2 (15.6 %). A relative higher abundance of fucose containing *O*-glycans was found for the samples clustering in the upper left corner of the loadings plot, whereas sialylated *O*-glycans were enriched in the lower right part of the plot (**Supplementary Information, Figure S2b**). Further analysis of the scores plot (**Figure 2**) showed that for days 5 and 7 both groups are different from one another as well as from days 14, 21 and 24, as no overlap is observed of the representations. Additionally, when focused on the difference overtime within the butyrate and non-stimulated group (**Figure 2**), the scores plot indicates that the CaCo-2 *O*-glycome undergoes progressive changes when cells are grown from confluency to 7 days post-confluence, at which point these cells are not fully differentiated yet. Upon further differentiation, a distinct difference between the butyrate-stimulated group trajectory and the spontaneous differentiation trajectory can be observed. For the spontaneous differentiated CaCo-2 cells the representations of days 14, 21 and 24 are close to one another. This indicates that the *O*-glycome appears to stabilize in the later differentiation phases. However, the butyrate-stimulated cells are different from each

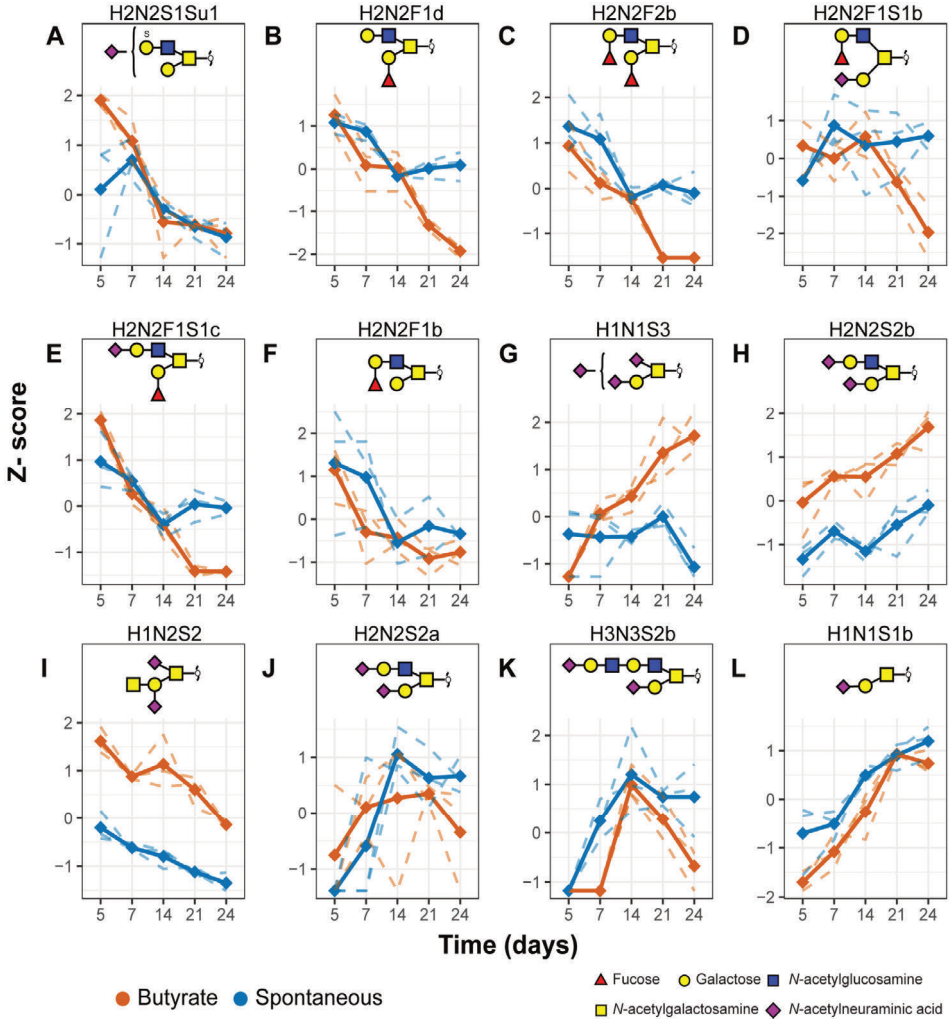

Figure 3. Differentiation induces significant downregulation (a-f; i) or upregulation (g, h, j, k. l) of CaCo-2 O-glycans as well as some butyrate-specific changes. *Spaghetti plots of the selected glycan abundances that show significant difference between time points selected from the analysis of variance (ANOVA). The dashed lines represent the scaled z-scores of the measured values of each biological replicate, whilst the continuous lines represent the z-scores of the mean values per biological replicate. Su: sulfation.*

other between days 14, 21 and 24, which indicates that the butyrate-stimulated cells still undergo progressive changes in their *O*-glycome in the late differentiation phases. To support this visual interpretation and identify the glycosylation signatures changing with differentiation, two-way ANOVA analysis of variance was performed on relative abundances of individual *O*-glycans (**Supplementary Information 2, Table S2).**

Specific *O*-glycans that show statistically significant changes with time (butyrate-stimulated *versus* spontaneous differentiation) are summarized in **Supplementary Information 1, Figure S3** and **Supplementary Information 2, Table S3.** The *O*-glycans that show a difference between the groups (butyrate-stimulated and spontaneous differentiation) are listed in **Supplementary Information 2, Table S4**. Moreover, the glycans that show a significant change with both butyrate stimulation and time are listed in **Supplementary Information 2, Table S5.**

O-glycans that show a clear consecutive pattern with differentiation were selected for visualization (**Figure 3**). **Figure 3a-f** show the *O*-glycans that are significantly downregulated with differentiation and are carriers of terminal blood group antigens (H2N2F1b, H2N2F1d, H2N2F2b, H2N2F1S1b, H2N2F1S1c), as well as sulfation (H2N2S1Su1). All *O*-glycans that show an increase overtime carry terminal *N*-acetylneuraminic acids and no fucosylation (H2N2S2b, H2N2S2a, H3N3S2b, H1N1S1a, H1N2S2, H1N1S3). A glycan with a disialyl motif, with composition H1N1S3 (**Figure 3g**) is upregulated in the butyrate stimulated cells. Notably, the Cad antigen (GalNAcβ1-4(Neu5Acα2-3)Galβ1-3[Neu5Acα2-6]GalNAc) present on *O*-glycans with composition H1N2S2 (**Figure 3i**) shows downregulation with differentiation. However, it is significantly higher in the butyrate-stimulated group compared to the spontaneous differentiation. Overall, we observed a downregulation of blood group antigen H fucosylation, and upregulation of terminal sialylation with differentiation.

STRUCTURAL IDENTIFICATION OF GLYCAN SPECIES

Due to reports from the literature that provide contradictory *O*-glycan assignments, we made additional efforts to support the assignment of the specific *O*-glycan structures. As the *O*-glycomic profile of BSM has been very well characterized by both NMR and tandem MS, we used this as a reference standard[37–39]. At *m/z* 895.34, two isomers with composition H2N2F1 were observed in our Caco-2 samples (**Supplementary Information 1, Figure S4a;** isomer *1* and *2a*). The major isomer expressed in Caco-2 cell line (2a), matched with the major isomer in BSM standard (**Supplementary Information 1, Figure S4b;** *2b*) by retention time and MS/MS spectrum which allowed us to confirm that the major *O*-glycan isomer (*2a*) in the Caco-2 samples carried the terminal blood group H antigen[38]. The MS/MS spectra of this isomer (**Supplementary Information 1, Figure S4d** and **e**) shows the presence of a highly abundant Z-ion (*m/z* 569.20 (H1N2) indicating the occupancy of the 6'-antenna of the core 2 *O*-glycan.

On the other hand, isomer *1* was found solely in Caco-2 cells, and the MS/MS spectra (**Supplementary Information 1, Figure S4c)** revealed the presence of terminal fucose linked to a type 2 *N*-acetyllactosamine on the 6' antenna, indicated by the characteristic cross-ring fragment of the β1-4-galactosylated GlcNAc (*m/z* 409.02). Previous studies have reported an α1-2-linked terminal L-fucose, as part of blood group antigen H[40]. In contrast, a recent study has reported the presence of an unconventional terminal α1-6-linked fucose[30]. Since this glycan was not present in the BSM standard, it was not possible to determine the linkage of the terminal fucose residue and further experiments will be needed. Additionally, two isomers with composition H2N2F2 at *m/z* 1041.40 were observed in our Caco-2 samples (**Supplementary Information 1, Figure S5a,** isomer *1* and *2a*). Isomer *2a* matched by retention time and MS/MS spectra with the major isomer of the BSM standard (**Supplementary Information 1, Figure S5b** and **d;** isomer *2b*), indicating a blood group antigen H related terminal α1-2-fucose[30]. Isomer *1* could not be identified by MS/MS and did not provide insightful differences in the fragmentation patterns due to low abundance. Interestingly, a recent study reported the presence of two α1-6- linked terminal fucosides carried by a major isomer with composition H2N2F2 in Caco-2[30]. The structural identification of an *O*-glycan with composition H1N1F1S1 at *m/z* 821.30 is illustrated in **Supplementary Information 1, Figure S6.** Based upon fragments *m/z* 530.23 and 512.14 this glycan carries a blood group antigen type 3. The signal at *m/z* 495.25 indicated an α2-6-linked sialic acid on the innermost GalNAc, which is in contrast to a previously reported annotation where the fucose was stated to be linked to the innermost GalNAc, whereas the sialic acid was linked to terminal galactose in α2-6 linkage[30]. Structural elucidation of the *O*-glycan with composition H1N1S3 at *m/z* 1257.44 is shown in **Supplementary Information 1, Figure S7**. Two sialic acids appeared to be linked to each other, indicated by the presence of a characteristic fragment ion at *m/z* 581.18. While our data did not provide sufficient evidence to deduce the location of this disialyl motif, it has been previously described that this occurs on the α2-6-linked sialic acid linked to the innermost GalNAc[41]. Moreover, structural elucidation of a glycan with Cad antigen, upregulated in the butyrate stimulated cells, was depicted in **Supplementary Information 1**, **Figure S8**. Overall, we observed a downregulation of terminal blood group H antigen expression and an upregulation of both galactose α2-3-sialylation and core GalNAc α2-6-sialylation with differentiation.

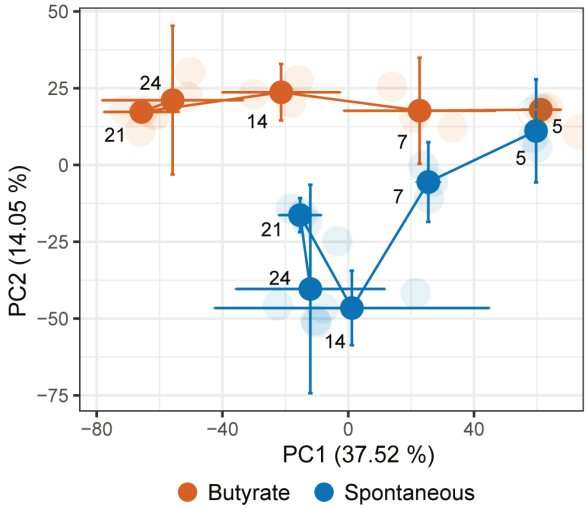

Figure 4. The geometric trajectory visualization of proteomic changes with differentiation. *PCA model based on abundance of different proteins shows a separation between the butyrate-stimulated group (orange) and spontaneous differentiation (blue). The separation between different time points in the two groups is illustrated as a trajectory. The top two principal components (PC) explain 51.57 % of the variation within the data. The PCA scores from different biological replicates (faded color) were averaged to create the trajectory.*

Proteomic analysis

Butyrate has been extensively used for stimulation of differentiation. However, to our knowledge a quantitative proteomics approach has not been employed to study the effect as well as differences between spontaneous differentiation and butyrate-stimulated differentiation in the Caco-2 cell line. Therefore, we performed quantitative bottom-up proteomics on the very same set of samples analyzed for the profiling of the *O*-glycome. The samples were digested with trypsin, isotopically labeled with a tandem mass tag (TMT), and analyzed by LC-MS/MS. With this approach a total of 5050 proteins could be confidently quantified (mascot score >21, unique peptides >= 1; **Supplementary Information 3, Table S2**).

To explore specific proteome variations that correlate to the differentiation of butyrate-stimulated samples, principal component analysis (PCA) was used (**Supplementary Information 1, Figure S9**). The geometric trajectory in **Figure 4** (based on the PCA model) illustrates changes in the cell proteome with differentiation as well as the differences in butyrate stimulated cells. At day 5, both groups cluster closely together, whereas the changes start to be more prominent at day 7 culminating at day 14. In the

non-stimulated cells the proteome appears to stabilize after day 14. In contrast, butyrate stimulation induced significant changes in the proteome until day 21. Proteins that show statistically significant changes (Two-way ANOVA, Bonferroni corrected p-value < 0.05) with time, with butyrate stimulation, as well as with both time and with butyrate stimulation are listed in **Supplementary Information 3, Tables S4, S5** and **S6**, respectively. Interestingly, significant upregulation of alkaline phosphatases (ALPP, and ALPG) and cytokeratin 20 (KRT20) are present only in the butyrate stimulated cells (**Supplementary Information 1, Figure S10**), Specific transcription factors known for their role in regulation of colon differentiation also showed a change (**Supplementary Information 1, Figure S11)**. Hepatocyte nuclear factor 1α (HNF1A) decreased in abundance over time in both groups, however, in the butyrate stimulated cells it continued to decline after 14 days. Hepatocyte nuclear factor 4α (HNF4A) showed an upregulation after 5 days, with a decline after 7 days in spontaneous differentiation, whereas in the butyrate stimulated cells in continued to rise until 14 days, followed by a decline. Transcription factors GATA6 and FOXA1 showed a continuous decrease in abundance both in spontaneous and butyrate stimulated cells. Although not statistically significant, transcription factor CDX2 decreased over time in both groups. All transcription factors which show a change with differentiation are listed in **Supplementary Information 3, Table S9.** Moreover, proteins that show the most prominent upregulation with butyrate stimulation are protocadherin Fat 1 (FAT1), galectin-1 (LGALS1), and phosphoenolpyruvate carboxykinase, (PCK1) (**Supplementary Information 3, Table S5**). The majority of the proteins changing specifically with butyrate stimulation are related to cell metabolism (hsa01100-Kegg pathways).

In order to relate the *O*-glycomic changes to changes in the abundance of specific *O*-glycoproteins, we continued our analysis of the proteomic data by selecting the proteins involved in the GO ontologies-Glycosylation, GO-Monosaccharide transport, as well as *O*-glycoproteins from different cell lines discovered by the Simple cell approach[42]. From this selection, 120 proteins showed a statistically significant change with time, whereas only 36 proteins showed a significant change with both time and butyrate stimulation (**Supplementary Information 3, Table S7** and **S8**, respectively). Interestingly, the only mucin that showed an upregulation over time was MUC13 (**Supplementary Information 3, Table S7)**. Additionally, changes were detected in

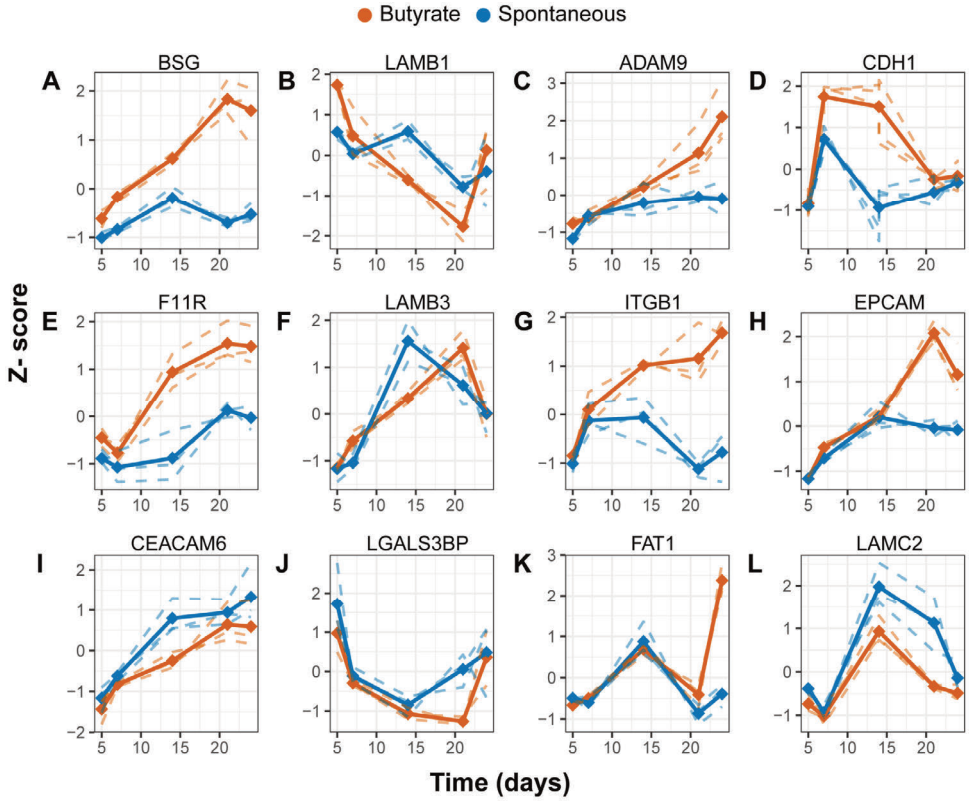

Figure 5. Differentiation induces significant changes in the abundance of cell adhesion O-glycoproteins. *Spaghetti plots of the protein abundances that show significant difference between groups selected from the analysis of variance (ANOVA). The dashed lines represent the scaled Z-scores of the measured values of each biological replicate, whilst the continuous lines represent the Z-scores of the mean values per biological replicate.*

monosaccharide transport related proteins including solute carrier family facilitated monosaccharide transporters (SLC2A1, SLC2A3, SLC2A5) as shown in **Supplementary Information 1, Figure S12**. Also, the abundances of polypeptide GalNAc transferases 2 and 3 (GALNT2, GALNT3 respectively) were increased over time (**Supplementary Information 1, Figure S12**). However, other glycosyltransferases involved in the biosynthesis of *O*-glycans did not show a significant change or were not detected with our proteomics workflow. Interestingly, the abundance of GDP-mannose 4,6-dehydratase (GMDS) was downregulated with differentiation, however, the change was not statistically significant.

In regard to time, significant changes were observed for the abundances of different *O*-glycoproteins involved in cell adhesion such as laminin (LAMB3, LAMC2, LAMB1), galectin 3 binding protein (LGALS3BP), cadherin 1 (CDH1), epithelial cell adhesion molecule (EPCAM), intracellular adhesion molecule1 (ICAM1), carcinoembryonic antigen related cell adhesion molecule 6 (CEACAM6), integrin beta 1 (ITGB1), disintegrin and metalloproteinase domain containing protein 9 (ADAM9), protocadherin (FAT1) and junctional adhesion molecule A (F11R) (**Figure 5**). Interestingly, butyrate-stimulated cells induce higher abundances of cell adhesion proteins such as protocadherin (FAT 1), cadherin 1 (CDH1) and EPCAM than spontaneously differentiated cells (**Figure 5** and **Supplementary Information 3, Table S8**).

DATA INTEGRATION

To identify associations between the proteomic and glycomic changes with differentiation we used a data integration approach (MixOmics)[43]. For this purpose, only the above-mentioned proteins of interest were included in the analysis. Due to the complexity of the model, the proteome and *O*-glycome differentiation signatures over time were integrated separately for the spontaneous and butyrate-stimulated cells. The clustered image map (CIM) indicates that changes on the glycome level highly associates with the changes on the proteome level in spontaneous differentiation, revealing potential glycosylation changes of specific proteins or relationship with specific enzymes and monosaccharide transporters (**Figure 6**). Among others, cathepsin D (CTSD) was significantly downregulated with differentiation, which associated with fucosylated glycomic signatures H2N2F1d, H2N2F2b and H2N2F1S1c. Interestingly cell adhesion glycoprotein cadherin 1 (CDH1) showed strong upregulation at day 7 together with *O*-glycan signatures such as H3N3, and sialylated glycans H3N3S2a, H1N2S1, H2N2F1S1b, H2N2S2b. At day 14 prominent upregulation was seen in the abundance of other cell adhesion proteins such as EPCAM, ICAM1, CEACAM6, which showed a similar expression pattern as sialylated glycans H1N1S1b and H2N2S2a. Moreover, an upregulation of SLC2A1, a monosaccharide transporter of a wide range of aldoses is seen at this stage of differentiation, as well as glutamine-fructose-6-phosphate aminotransferase 1 (GFPT1) and phosphoacetylglucosamine mutase (PGM3) which catalyzes the conversion of GlcNAc-6-P into GlcNAc-1-P during the synthesis of uridine UDP-

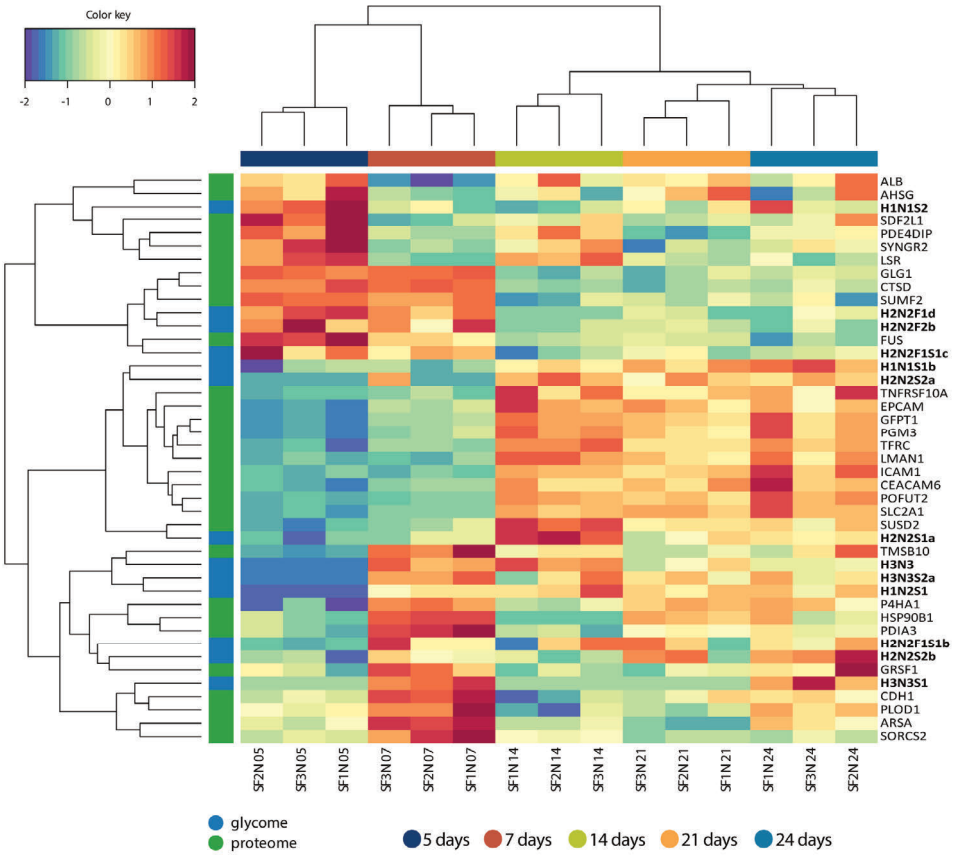

Figure 6. Multi-omics signature of spontaneous differentiation. *Clustered Image Map illustrating changes in both proteome and glycome with differentiation based on a combination of the PLS (Partial Least Square Regression) generalized for the multiple matched datasets and LASSO (least absolute shrinkage and selection operator) based variable selection. The model was tuned for the maximal correlation.*

GlcNAc. The changes in the later time points are very similar as for day 14 due to the stabilization that was observed in the proteome and glycome of spontaneously differentiated cells. On the other hand, butyrate stimulated differentiation induces significant changes on the proteome level even until day 21, and less prominent changes until day 24 (**Figure 7**). Downregulation of, among others, apolipoproteins E and A1, alpha-2-HS-glycoprotein (AHSG), albumin (ALB), lipolysis-stimulated lipoprotein receptor (LSR), sodium/ascorbate transporter (SLC23A1), chitobiosyldiphosphodolichol beta-mannosyltransferase (ALG1), beta-hexosaminidase is observed in spontaneously differentiated cells. On the other hand,

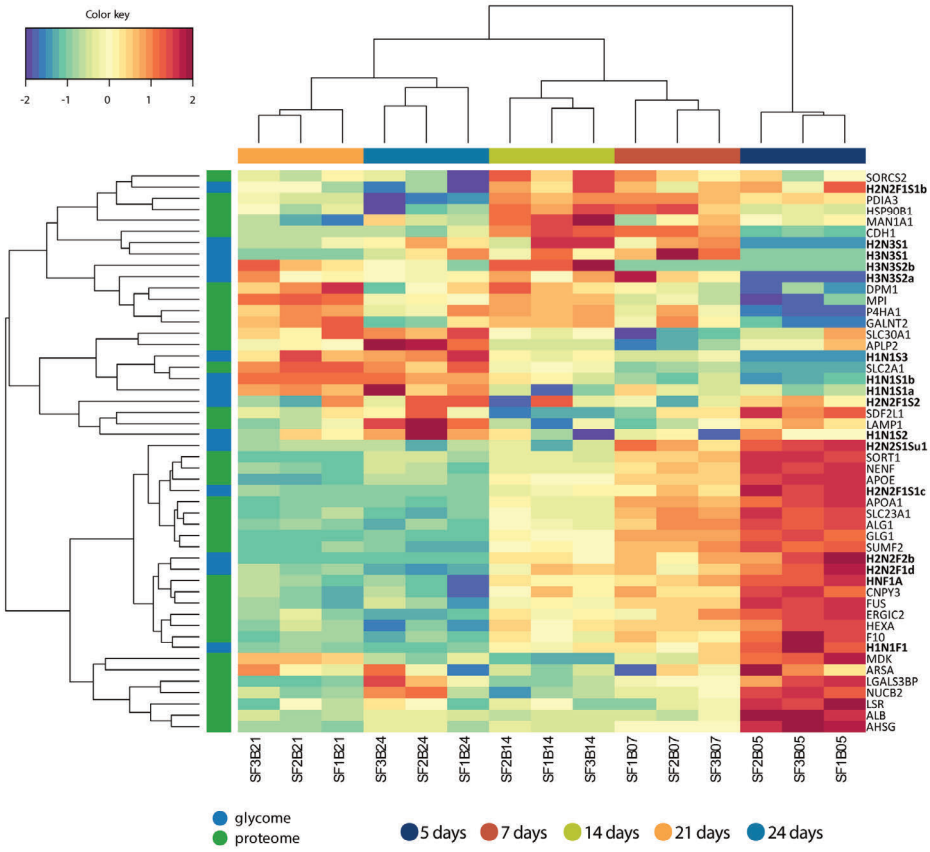

Figure 7. Multi-omics signature of butyrate stimulated differentiation. *Clustered Image Map illustrating changes in both proteome and glycome with differentiation based on a combination of the PLS (Partial Least Square Regression) generalized for the multiple matched datasets and LASSO (least absolute shrinkage and selection operator) based variable selection. The model was tuned for the maximal correlation.*

butyrate stimulated differentiation induces subunit alpha (EXA) and hepatocyte nuclear factor 1-alpha (HNF1A) is observed, which are correlated with fucosylated *O*-glycomic signatures such as H2N2F1S1c, H2N2F2b, H2N2F1d and H1N1F1. Furthermore, it also induces downregulation of galectin-3-binding protein (LGALS3BP) which promotes integrin-mediated cell adhesion. Remarkably, at the same time, these cells showed upregulation of another cell adhesion protein cadherin 1 (CDH1), which is associated with the upregulation of glycans H2N2F1S1b, H2N3S1, H3N3S1, H3N3S2b, H3N3S2a. Moreover, butyrate stimulation induced upregulation of proteins involved in the monosaccharide metabolism, such as mannose-6-phosphate isomerase (MPI) involved in the synthesis of the GDP-mannose as well as solute

carrier family 2, facilitated glucose transporter member 1 (SLC2A1). Upregulation of proteins involved in *N*-glycan biosynthesis was also induced by butyrate stimulation such as mannosyl-oligosaccharide 1,2-alpha-mannosidase IA (MAN1A1) and dolichol-phosphate mannosyltransferase subunit 1 (DPM1). Upregulation of polypeptide GalNAc transferase 2 (GALNT2) shows a similar pattern to the upregulation of the following sialylated glycans H1N1S3, H1N1S1b, H1N1S1a, H2N2F1S2 and H1N1S2 all showing increasing abundances with cell differentiation.

The associations found with HNF1A triggered us to explore if other transcription factors associate with the *O*-glycomic changes. Therefore, we performed the integrative analysis selecting only the *O*-glycans and the transcription factors that show a significant change with differentiation (**Supplementary Information 3, Table S9**). Downregulation of fucosylation in spontaneous differentiation showed associations with FOXA1, FOXP4, GATA6, STAT1 and HNF1A transcription factors. Specifically, HNF4A expression was associated more with the expression of H2N2S1Su1 and H3N3S1 glycomic signatures (**Supplementary Information 1, Figure S13)**. Similar to HNF1A, that showed associations with glycomic signatures carrying terminal blood group H (H2N2F2b, H2N2F1S1c) (**Figure 7**), transcription factor HNF4A associated with H2N2F1S1b in the butyrate stimulated cells (**Supplementary Information 1, Figure S14)**. On the other hand, upregulation of sialylated species associated with NF-Kappa-B Transcription Factor P65 (RELA) and MYC promoter-binding protein 1 (ENO1) in butyrate stimulated cells.

DISCUSSION

In this study we observed changes in the Caco-2 *O*-glycome correlated to differentiation over time. Both butyrate-stimulated and spontaneous differentiation were accompanied by the elevation of galactose α2-3-sialylation, and a downregulation of the terminal blood group H antigen over time. However, it should be noted that there are contradictory reports in literature regarding the correlation of the terminal blood group H-antigen expression with differentiation. Namely, Murakami *et al.* reported no detection of the H type blood group antigen, while Vincente *et al.* reported diffused expression of H type antigen, Amano *et al.* reported higher H type 1 antigen expression with differentiation and Xu *et al.* reported no changes in the *O*-linked fucosylation[30,44–46]. It is important to note that there were considerable

differences in the experimental design regarding induction of differentiation, days in culture as well as detection methods. Moreover, it was previously postulated that Caco-2 cells can form different subpopulations during growth and differentiation, therefore considerable differences between laboratories have previously been reported for identical experimental conditions[47,48]. Therefore, we first validated that the cells showed changes in the morphology related to cell differentiation (**Figure 1b**), as well as changes in the activity of alkaline phosphatase (**Figure 1a**), which is a well-established marker of epithelial differentiation. Thereafter we aimed at deciphering differences between spontaneous and butyrate stimulated changes with cell differentiation focusing on the *O*-glycome and proteome by MS.

Our study on the relative abundances of individual *O*-glycans showed that the Caco-2 *O*-glycome for spontaneous differentiation stabilizes after day 14, this is in agreement with a proteomics analysis on the spontaneous differentiation of Caco-2 cells previously published by Stierum *et al.*[25]. The *O*-glycome for the butyrate-stimulated group behaved significantly differently and did not stabilize after day 14. Noticeable, in the spontaneously differentiated cells the expression of terminal blood group antigens carrying an α-1,2-fucoside stabilized at day 14, whereas most of these α-1,2-fucoside-carrying antigens in the butyrate-stimulated cells continued decreasing even after day 14. Fucoside moieties on glycans are increasingly recognized as critical attributes for cell-cell interactions and signaling processes[49]. The decrease of type 2 blood group H fucosylated glycans with a subsequent increase in type 1 blood group H and Lewis B (Le^B) glycans was previously correlated to the spontaneous differentiation of Caco-2 cells[46,50]. This effect was linked to the changes in the activity of specific enzymes responsible for type 1 blood group H biosynthesis such as β1-3-galactosyltransferase and α1-2-fucosyltransferase[51]. We did not find the blood group H type 1 antigens on *O*-glycans, which might indicate that other types of glycans are carriers of the antigen such as glycosphingolipids or *N*-glycans. For further studies it would be interesting to investigate the implications of this butyrate-induced fucose downregulation on cell properties. At the same time, upregulation of sialylated glycan species is induced upon differentiation, which was previously correlated to differentiation assessing the expression of corresponding glycosyltransferases such as ST3GAL6[52–54]. Previously, we compared the *O*-glycome of 26 CRC cell lines and revealed associations with differentiation where the well-differentiated cells expressed

Lewis-type fucosylation, whereas the undifferentiated cells, such as the Caco-2 cell line, showed expression of *O*-linked blood group H antigens[31]. However, it should be taken into account that, although developed from colon carcinoma, Caco-2 cells showed quite specific glycosylation, very rich in terminal blood group antigen expression compared to other CRC cell lines[31]. Moreover, they differentiate into small intestinal enterocyte-like cells, with a consequent loss of colonocyte properties[55]. This could explain why there was not an increase in Lewis type fucosylation observed in this study upon differentiation.

The *O*-glycan carrying a Cad antigen showed a downregulation over time for both groups, correlating to differentiation, but a significantly higher expression was observed for the butyrate-stimulated cells compared to the spontaneously differentiated cells. Structures with terminal Cad and Sda epitopes have previously been described as characteristic for normal colon mucin tissue[56]. However, in line with our findings, most studies report a downregulation of the Cad antigen with differentiation in CRC cell lines[32,56]. β1-4-GalNAc-transferase 2 is the *N*-acetylgalactosaminyltransferase (B4GALNT2) responsible for the addition of the terminal GalNAc residue in a β1-4 linkage to the sialylated galactose. It has previously been observed that the expression of the Cad antigen and the B4GALNT2 gene are regulated by promoter DNA methylation[57]. Recent literature reported the capability of butyrate to trigger DNA demethylation in CRC cell lines[58]. Promoter demethylation triggered by butyrate, followed by differential expression of the B4GALNT2 gene and corresponding Cad antigen, could explain the significantly higher expression of the Cad antigen in the butyrate stimulated group compared to the spontaneous differentiation group.

Another interesting *O*-glycan that shows an upregulation in the butyrate-stimulated samples compared to the spontaneously differentiated samples (after day 14), is the H1N1S3 *O*-glycan. Tandem MS revealed that this *O*-glycan carries a di-sialic acid (DiSia) unit (**Supplementary Information, Figure S7**), potentially with an α2-8 glycosidic linkage[59]. Previously an upregulation of α2-8-sialyltransferase *ST8SIA6* mRNA expression was reported in Caco-2 cells that underwent spontaneous differentiation[60]. Details regarding the biological function of this DiSia motif remains unknown, making it an interesting structure for future studies. Moreover, an increase in core 2 sialylated *O*-glycans was observed in differentiated cells, which is in line with

previously reported changes in the expression of α2-3-sialyltransferases (ST3GAL4 and ST3GAL6) as well as core 2 synthases (GCNT4 and GCNT3)[60]. Additionally, the previously reported upregulation of B3GNT3[60] could be responsible for the upregulation of the H3N3S2b glycan, by adding another LacNAc to the 6 arm of the *O*-glycan.

BSM and its *O*-glycan repertoire has been very well characterized by both tandem MS and NMR[38,39] and served in our study as a valuable reference standard allowing us to assign specific glycan linkages. We were able to confirm that the dominant isomers of *O*-glycans with composition H2N2F2 and H2N2F1 are carrying α1-2 linked terminal fucoses (**Supplementary Information, Figure S4** and **S5**, respectively) whereas a recent study reported the presence of an unconventional terminal α1-6-linked fucose[30]. However, it was not possible to fully determine the linkages of all *O*-glycan isomers, as several of them were not present in the BSM standard.

The stabilization of the *O*-glycome for spontaneously differentiated cells after day 14 is in line with the stabilization of the proteome as reported previously by Stierum *et al*., 14 days post-confluence[25]. In this study, we were able to study both the effect of butyrate stimulation and spontaneous differentiation on the same sample set used for profiling the *O*-glycome. In agreement with the dynamics of their *O*-glycomics data, the proteome of the butyrate-stimulated cells also showed further changes 14 days post-confluence, indicating an important role for butyrate-stimulated cellular regulation in the late phase of differentiation. In order to relate the previously mentioned *O*-glycomic changes to changes in the abundance of specific *O*-glycoproteins, we continued our analysis of the proteomic data by selecting the proteins involved in the GO ontologies – Glycosylation, GO Monosaccharide transport, as well as *O*-glycoproteins[42]. An upregulation of specific monosaccharide transporters SLC2A1 (GLUT1), SLC2A3 (GLUT3) and SLC2A5 (GLUT5) is seen in butyrate stimulated differentiated cells, important for uptake of glucose and other monosaccharides into the cell, which was previously associated with differentiated Caco-2 cells[61,62]. Interestingly, the only monosaccharide present in the cell culture media is glucose, therefore the availability of other monosaccharide precursors for glycan biosynthesis is dependent on their intracellular de novo biosynthesis from glucose. Changes in the expression of enzymes involved in metabolic conversion of monosaccharide precursors were observed with differentiation. Namely, hexokinase that converts

glucose into glucose-6-P (HK2), glutamine-fructose-6-phosphate aminotransferase 1 (GFPT1) and phosphoacetylglucosamine mutase (PGM3), both involved in the biosynthesis of UDP-GlcNAc, are upregulated with differentiation. These enzymes are important rate limiting enzymes of the hexosamine pathway, regulating the availability of precursors for *N*- and *O*-linked glycosylation of proteins. Studies on a molecular level for glycosyltransferases are known to be challenging, as many of them are low abundance proteins[63]. With our current method we were able to detect an increase in abundance of polypeptide GalNAc transferases 3 and 2 with differentiation (GALNT3, GALNT2 respectively), followed by a decline in the late stages of differentiation. Previously, knock out of GALNT2 and GALNT3 in human organotypic skin models caused impaired cell adhesion and decreased differentiation, respectively, implying the importance of these enzymes initiating *O*-GalNAc glycosylation for cell–extracellular matrix interactions and epithelial differentiation[64]. Additionally, proteins such as F11R, BCAM, ADAMTSL4, LGALS3BP, and laminin are found to be GALNT2 specific targets, whereas protocadherin (FAT2) and cathepsin (CTSD) are found to be a GALNT3 specific targets in skin cells[64]. Another study pinpointed apolipoprotein E (APOE), nucleobindin-2, lipolysis-stimulated lipoprotein receptor, laminin (LAMC2), coagulation factor X, and protein disulfide-isomerase A3 as GALNT3 and GALNT6 targets[65]. These observations could provide interesting perspectives to study the GALNT specific targets in the colon, and how the glycosylation of the specific proteins influences their function in cell adhesion during differentiation.

To identify associations between proteomics and glycomics changes with differentiation we continued using a data integration approach (MixOmics)[43]. The specific butyrate pattern shows higher abundances of cell adhesion proteins in later stages of differentiation. Those proteins show a similar change in abundance as sialylated glycan species indicating that specific glycosylation changes may be a consequence of changes in the abundance of specific proteins (**Figure 6**). Moreover, specific *O*-glycosylation of cell adhesion *O*-glycoproteins may play an important role in cell differentiation. The structure-function relationship of glycosylation in these proteins involved in adhesion is still to be deciphered.

Remarkably, we observed a general downregulation of abundances of transcription factors involved in epithelial differentiation such as HNF1A, GATA6, FOXA[19]. Whereas, abundance of HNF4A, the key regulator of intestinal genes in Caco-2

model[19], increased by day 14 in butyrate stimulated cells, in line with previous research where an upregulation on mRNA and protein levels was observed[66]. Moreover, it was previously shown that HNF4A also regulates the HNF1A and CDX2 promoter activity[66] and the functional interaction between HNF1A and CDX2 was demonstrated before in Caco-2 cells[67]. Although no unique peptides for CDX1 were identified in our study, the previously described target protein cytokeratin 20 showed upregulation with differentiation[68]. The activity of these transcription factors depends on a complex network of interactions, and not solely on the abundance of the protein product. The integrative analysis with transcription factors and *O*-glycans (**Supplementary Information 1, Figure S13** and **Figure S14)** reveals new potential regulators of glycosylation such as FOXA1, FOXP4 and STAT1, however, further studies are required to validate these hypotheses.

Interestingly, in the butyrate stimulated differentiation, the downregulation of HNF1A transcription factor correlated with fucosylated glycomics signatures such as H2N2F1S1c, H2N2F2b, H2N2F1d and H1N1F1 **(Figure 7).** HNF1A is a transcription factor expressed in several organs including the intestine and stomach. Genome-wide association studies by Lauc *et al.,* identified HNF1A as a key regulator of fucosylation[69]. It was demonstrated that HNF1A activates antennary FUTs (FUT3-11 in liver cell lines) and downregulated core FUTs (FUT8). Additionally, it showed that HNF1A upregulated GDP-mannose 4,6 dehydratase (GMDS), an important enzyme in de-novo d-fucose biosynthesis pathway which increases the availability of GDP-fucose precursor for antennary fucosylation of glycans[69]. While a trend in downregulation of GMDS was seen on the proteome level, it did not meet the significance threshold. Nevertheless, our data indicates that there might be an important role of HFN1A as a master transcriptional regulator of fucosylation in CacoCaco-2 cells, where downregulation of HNF1A correlates with the downregulation of fucosylation upon differentiation. HNF1A downregulation can lead to the limitation of d-fucose availability and consequently upregulation of terminal sialylation which is in direct competition with terminal blood group H antigen biosynthesis. Unfortunately, we did not detect the CDX1 transcription factor in our proteomics study, a previously reported transcriptional regulator of Lewis type fucosylation on *N*-glycans[32,70] and the detected CDX2 showed no statistically significant change. Moreover, the Caco-2 cell line, among other undifferentiated cell

lines, previously showed a predominant expression of *O*-linked blood group H fucosylation and weak associations could be seen with HNF4A transcription factor[31]. This further supports the hypothesis that HNF1A might be a key regulator of *O*-linked glycan fucosylation in Caco-2 cells during cell differentiation. Several other associations between specific glycan structures and protein expression are reported in this study and could potentially be the start of new endeavors which may have important implications for understanding normal biological functions such as colon differentiation.[69]

In this study an integrative approach was used to generate hypotheses about relationships between cell glycome and proteome that occurs with cell differentiation. However, to validate those hypotheses, *O*-glycoproteomic analysis is needed, which still remains a challenging task in complex biological samples. Previously, a genetic engineering approach that limits the glycan diversity to a single GalNAc or SiaGalNAc enabled enrichment of cell *O*-glycoproteome by lectin affinity chromatography and revealed the complexity of the cell *O*-glycoproteome[71]. Although this approach is invaluable for discovery of *O*-glycoprotein sites, it does not give information about the glycan microheterogeneity. In this study important hypotheses were made using an integrative approach which can be exploited for future research with targeted *O*-glycoproteomic approach of specific proteins. Moreover, functional studies can be employed to validate the suggested transcriptional regulators of glycosylation, and their role in cell differentiation. Which, eventually, will provide insights about the role of glycosylation cell adhesion and the role of fucosylation and sialylation in cell differentiation.

CONCLUSIONS

Differentiated CacoCaco-2 cells are an important and widely used model cell line. However, the glycome expression might change dependent on the growth- and bacterial metabolites available during the culturing process. In this study an untargeted in-depth screening of the CacoCaco-2 cell line was performed, identifying specific *O*-glycans that mark butyrate-induced epithelial differentiation and link them to potential protein carriers. While interesting structures were identified, further studies are required to identify their role in maintaining homeostasis in the epithelium. In the future this might allow us to gain a better understanding of the constantly changing *O*-

glycome at the gut-microbiota interface as correlated to bacterial hydrolase activities and its relevance to maintaining homeostasis or role in dysbiosis.

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.

ACKNOWLEDGMENTS

This work was supported by the European Commission's Horizon 2020 program via the "GlyCoCan" project (Grant 676421), the ERC-2019-STG 852452 grant for K. Strijbis, and by the Netherlands Organization for Scientific Research (NWO) via a VIDI grant to T. Wennekes (723.014.005).

AUTHOR CONTRIBUTIONS

K.M. performed the glycomic analysis and proteomic data analysis. Y.M.C.A.L cultured the cells and performed alkaline phosphatase assay and microscopy. K.M. and O.A.M. performed the integrative data analysis and statistics. P.A.V. and G.M.C.J. performed the proteomics experiments. K.M., Y.M.C.A.L, G.S.M.L., K.S., T.W., M.W. conceptually designed the work. K.M., Y.M.C.A.L, K.S., T.W., G.S.M.L. and M.W. wrote the manuscript. All authors gave approval to the final version of the manuscript.

REFERENCES

1. Neish, A. S. Microbes in Gastrointestinal Health and Disease. (2009).
2. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. (2005).
3. Yang, B., Cao, L., Liu, B., McCaig, C. D. & Pu, J. The Transition from Proliferation to Differentiation in Colorectal Cancer Is Regulated by the Calcium Activated Chloride Channel A1. *PLoS ONE* **8**, e60861 (2013).
4. Martinez, K. B., Leone, V. & Chang, E. B. Microbial metabolites in health and disease: Navigating the unknown in search of function. (2017).

5. McIntyre, A., Gibson, P. R. & Young, G. P. Butyrate production from dietary fibre and protection against large bowel cancer in a rat model. *Gut* **34**, 386–391 (1993).
6. Lazarova, D. L., Chiaro, C. & Bordonaro, M. Butyrate induced changes in Wnt-signaling specific gene expression in colorectal cancer cells. *BMC Research Notes* **7**, 226 (2014).
7. Han, R., Sun, Q., Wu, J., Zheng, P. & Zhao, G. Sodium Butyrate Upregulates miR-203 Expression to Exert Anti-Proliferation Effect on Colorectal Cancer Cells. *Cellular Physiology and Biochemistry* **39**, 1919–1929 (2016).
8. Zuo, L., Lu, M., Zhou, Q., Wei, W. & Wang, Y. Butyrate suppresses proliferation and migration of RKO colon cancer cells though regulating endocan expression by MAPK signaling pathway. *Food and Chemical Toxicology* **62**, 892–900 (2013).
9. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. (2016).
10. Willemsen, L. E. M., Koetsier, M. A., Van Deventer, S. J. H. & Van Tol, E. A. F. Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E1 and E2 production by intestinal myofibroblasts. *Gut* **52**, 1442–1447 (2003).
11. Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. The role of the gut microbiota in nutrition and health. (2012).
12. Klepinina, L. *et al.* Colon cancer cell differentiation by sodium butyrate modulates metabolic plasticity of Caco-2 cells via alteration of phosphotransfer network. *PLoS ONE* **16**, e0245348 (2021).
13. Gaudier, E., Rival, M., Buisine, M. P., Robineau, I. & Hoebler, C. Butyrate enemas Upregulate Muc genes expression but decrease adherent mucus thickness in mice colon. *Physiological Research* **58**, 111–119 (2009).
14. Silva, Y. P., Bernardi, A. & Frozza, R. L. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. (2020).
15. Birchenough, G. M. H., Johansson, M. E. V., Gustafsson, J. K., Bergström, J. H. & Hansson, G. C. New developments in goblet cell mucus secretion and function. (2015).
16. Van Putten, J. P. M. & Strijbis, K. Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer. (2017).
17. Paone, P. & Cani, P. D. Mucus barrier, mucins and gut microbiota: The expected slimy partners? (2020).
18. Bergstrom, K. S. B. & Xia, L. Mucin-type O-glycans and their roles in intestinal homeostasis. (2013).
19. Yin, S. *et al.* Functional genomics analysis of human colon organoids identifies key transcription factors. *Physiol. Genomics* **52**, 234–244 (2020).
20. Gaudier, E. *et al.* Butyrate regulation of glycosylation-related gene expression: Evidence for galectin-1 upregulation in human intestinal epithelial goblet cells. *Biochemical and Biophysical Research Communications* **325**, 1044–1051 (2004).
21. Shah, S. *et al.* n-Butyrate reduces the expression of β-galactoside α2,6- sialyltransferase in Hep G2 cells. *Journal of Biological Chemistry* **267**, 10652–10658 (1992).

22. Jung, T. H., Park, J. H., Jeon, W. M. & Han, K. S. Butyrate modulates bacterial adherence on LS174T human colorectal cells by stimulating mucin secretion and MAPK signaling pathway. *Nutrition Research and Practice* **9**, 343–349 (2015).
23. Cornick, S., Tawiah, A. & Chadee, K. Roles and regulation of the mucus barrier in the gut. (2015).
24. Buhrke, T., Lengler, I. & Lampen, A. Analysis of proteomic changes induced upon cellular differentiation of the human intestinal cell line Caco-2. *Dev. Growth Differ.* **53**, 411–426 (2011).
25. Stierum, R. *et al.* Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1650**, 73–91 (2003).
26. Pshezhetsky, A. V. *et al.* Subcellular proteomics of cell differentiation: Quantitative analysis of the plasma membrane proteome of Caco-2 cells. *Proteomics* **7**, 2201–2215 (2007).
27. Tadjali, M., Seidelin, J. B., Olsen, J. & Troelsen, J. T. Transcriptome changes during intestinal cell differentiation. *Biochimica et Biophysica Acta - Molecular Cell Research* **1589**, 160–167 (2002).
28. Turck, N. *et al.* Proteomic analysis of nuclear proteins from proliferative and differentiated human colonic intestinal epithelial cells. *Proteomics* **4**, 93–105 (2004).
29. Tremblay, E. *et al.* Gene expression profiles of normal proliferating and differentiating human intestinal epithelial cells: A comparison with the Caco-2 cell model. *J. Cell. Biochem.* **99**, 1175–1186 (2006).
30. Xu, G., Goonatilleke, E., Wongkham, S. & Lebrilla, C. B. Deep structural analysis and quantitation of O-linked glycans on cell membrane reveal high abundances and distinct glycomic profiles associated with cell type and stages of differentiation. **14**, 2020.
31. Madunić, K. *et al.* Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype. *Cell. Mol. Life Sci.* **1**, 3.
32. Holst, S. *et al.* N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homebox 1 (CDX1)/Villin mRNA Expression. *Mol. Cell. Proteomics* **15**, 124–140 (2016).
33. 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).
34. 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.
35. Paulo, J. A. & Gygi, S. P. Nicotine-induced protein expression profiling reveals mutually altered proteins across four human cell lines. *Proteomics* **17**, (2017).
36. McComb, R. B., Bowers, G. N., Jr & Posen, S. *Alkaline Phosphatase - Robert B. McComb, George N. Bowers, Jr., Solomon Posen - Google Books*. (1979).

37. 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).
38. 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).
39. 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).
40. Prieto, P. A. *et al.* Expression of human H-type α1,2-fucosyltransferase encoding for blood group H(O) antigen in chinese hamster ovary cells. Evidence for preferential fucosylation and truncation of polylactosamine sequences. *Journal of Biological Chemistry* **272**, 2089–2097 (1997).
41. Möginger, U. *et al.* Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. *Front. Oncol.* **8**, 70 (2018).
42. Steentoft, C. *et al.* Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013).
43. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
44. Carmona-Vicente, N., Allen, D. J., Rodríguez-Díaz, J., Iturriza-Gómara, M. & Buesa, J. Antibodies against Lewis antigens inhibit the binding of human norovirus GII.4 virus-like particles to saliva but not to intestinal Caco-2 cells. *Virology Journal* **13**, (2016).
45. Murakami, K. *et al.* Norovirus Binding to Intestinal Epithelial Cells Is Independent of Histo-Blood Group Antigens. *PLoS ONE* **8**, e66534 (2013).
46. Amano, J. & Oshima, M. Expression of the H type 1 blood group antigen during enterocytic differentiation of Caco-2 cells. *J. Biol. Chem.* **274**, 21209–21216 (1999).
47. Lea, T. Caco-2 Cell Line. in *The Impact of Food Bioactives on Health: in vitro and ex vivo models* (eds. Verhoeckx, K. et al.) 103–111 (Springer International Publishing, 2015). doi:10.1007/978-3-319-16104-4_10.
48. Walter, E. & Kissel, T. Heterogeneity in the human intestinal cell line Caco-2 leads to differences in transepithelial transport. *European Journal of Pharmaceutical Sciences* **3**, 215–230 (1995).
49. Li, J., Hsu, H. C., Mountz, J. D. & Allen, J. G. Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. (2018).
50. Youakim, A. & Herscovics, A. Differentiation-associated decrease in the proportion of fucosylated polylactosaminoglycans of CaCo-2 human colonic adenocarcinoma cells. *Biochemical Journal* **247**, 299–306 (1987).
51. Amano, J., Kobayashi, K. & Oshima, M. Comparative study of glycosyltransferase activities in Caco-2 cells before and after enterocytic differentiation using lectin-affinity high-performance liquid chromatography. *Arch. Biochem. Biophys.* **395**, 191–198 (2001).

52. Link-Lenczowski, P. *et al.* A switch of N-glycosylation of proteome and secretome during differentiation of intestinal epithelial cells. *Biochimica et Biophysica Acta - Molecular Cell Research* **1866**, 118555 (2019).
53. Qi, F. *et al.* ST3GAL3, ST3GAL4, and ST3GAL6 differ in their regulation of biological functions via the specificities for the α2,3-sialylation of target proteins. *FASEB Journal* **34**, 881–897 (2020).
54. Pinto, R. *et al.* CDX2 homeoprotein is involved in the regulation of ST6GalNAc-I gene in intestinal metaplasia. *Lab. Invest.* **95**, 718–727 (2015).
55. Engle, M. J., Goetz, G. S. & Alpers, D. H. Caco-2 cells express a combination of colonocyte and enterocyte phenotypes. *J. Cell. Physiol.* **174**, 362–369 (1998).
56. Malagolini, N., Dall'Olio, F. & Serafini-Cessi, F. UDP-Ga1NAc:NeuAcα2,3Galβ-R (GalNAc to Gal) β1,4-N-acetyl-galactosaminyltransferase responsible for the Sda specificity in human colon carcinoma CaCo-2 cell line. *Biochem. Biophys. Res. Commun.* **180**, 681–686 (1991).
57. Wang, H. R., Hsieh, C. Y., Twu, Y. C. & Yu, L. C. Expression of the human Sda β-1,4- N-acetylgalactosaminyltransferase II gene is dependent on the promoter methylation status. *Glycobiology* **18**, 104–113 (2008).
58. Sun, X. & Zhu, M. J. Butyrate Inhibits Indices of Colorectal Carcinogenesis via Enhancing α-Ketoglutarate-Dependent DNA Demethylation of Mismatch Repair Genes. *Molecular Nutrition and Food Research* **62**, e1700932 (2018).
59. Sato, C. & Kitajima, K. Disialic, oligosialic and polysialic acids: Distribution, functions and related disease. (2013).
60. Park, D. *et al.* Characteristic Changes in Cell Surface Glycosylation Accompany Intestinal Epithelial Cell (IEC) Differentiation: High Mannose Structures Dominate the Cell Surface Glycome of Undifferentiated Enterocytes. *Mol. Cell. Proteomics* **14**, 2910–2921 (2015).
61. Mahraoui, L. *et al.* Presence and differential expression of SGLT1, GLUT1, GLUT2, GLUT3 and GLUT5 hexose-transporter mRNAs in Caco-2 cell clones in relation to cell growth and glucose consumption. *Biochem. J* **298 Pt 3**, 629–633 (1994).
62. Mahraoui, L. *et al.* Expression and localization of GLUT-5 in Caco-2 cells, human small intestine, and colon. *Am. J. Physiol.* **263**, G312-8 (1992).
63. R, S. *et al.* Targeted Proteomics Reveals Quantitative Differences in Low Abundance Glycosyltransferases of Patients with Congenital Disorders of Glycosylation. (2020) doi:10.1101/2020.09.15.291732.
64. Bagdonaite, I., Pallesen, E. M. H., Ye, Z. & Vakhrushev, S. Y. O-glycan initiation directs distinct biological pathways and controls epithelial differentiation. *EMBO* (2020).
65. Lavrsen, K. *et al.* De novo expression of human polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) in colon adenocarcinoma inhibits the differentiation of colonic epithelium. (2017) doi:10.1074/jbc.M117.812826.
66. Boyd, M., Bressendorff, S., Møller, J., Olsen, J. & Troelsen, J. T. Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC Gastroenterol.* **9**, 68 (2009).

67. Mitchelmore, C., Troelsen, J. T., Spodsberg, N., Sjöström, H. & Norén, O. Interaction between the homeodomain proteins Cdx2 and HNF1α mediates expression of the lactase-phlorizin hydrolase gene. *Biochem. J* **346**, 529–535 (2000).
68. Chan, C. W. M. *et al.* Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1936–1941 (2009).
69. Lauc, G. *et al.* Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1α as a master regulator of plasma protein fucosylation. *PLoS Genet.* **6**, e1001256 (2010).
70. Holst, S. *et al.* N-Glycomic and Transcriptomic Changes Associated with CDX1 mRNA Expression in Colorectal Cancer Cell Lines. *Cells* **8**, 273 (2019).
71. Schjoldager, K. T.-B. G. & Clausen, H. Site-specific protein O-glycosylation modulates proprotein processing — Deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1820**, 2079–2094 (2012).