



Universiteit  
Leiden  
The Netherlands

## Multilayer cancer glycomics

Wang, D.

### Citation

Wang, D. (2023, May 17). *Multilayer cancer glycomics*. Retrieved from <https://hdl.handle.net/1887/3618440>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3618440>

**Note:** To cite this publication please use the final published version (if applicable).

## Chapter 3

# High diversity of glycosphingolipid glycans of colorectal cancer cell lines reflects the cellular differentiation phenotype

Di Wang <sup>1</sup>, Katarina Madunić <sup>1</sup>, Tao Zhang <sup>1</sup>, Oleg A. Mayboroda <sup>1</sup>, Guinevere  
S.M. Lageveen-Kammeijer <sup>1</sup>, Manfred Wuhrer <sup>1,\*</sup>

<sup>1</sup> Leiden University Medical Center, Center for Proteomics and Metabolomics, Postbus  
9600, 2300 RC Leiden, The Netherlands

Reprinted and adapted with permission from Mol. Cell. Proteomics, 2022, 21, 100239, DOI:  
10.1016/j.mcpro.2022.100239

Colorectal cancer (CRC)-associated changes of protein glycosylation have been widely studied. In contrast, the expression of glycosphingolipid (GSL) patterns in CRC has, hitherto, remained largely unexplored. Even though GSLs are major carriers of cell surface carbohydrates, they are understudied due to their complexity and analytical challenges. In this study, we provide an in-depth analysis of GSL glycans of 22 CRC cell lines using porous graphitized carbon nano-liquid chromatography coupled with electrospray ionization-mass spectrometry. Our data revealed that the GSL expression varies among different cell line classifications, with undifferentiated cell lines showing high expression of blood group A, B and H antigens while for colon-like cell lines the most prominent GSL glycans contained (s)Le<sup>A/X</sup> and Le<sup>B/Y</sup> antigens. Moreover, the GSL expression correlated with relevant glycosyltransferases that are involved in their biosynthesis as well as with transcription factors (TFs) implicated in colon differentiation. Additionally, correlations between certain glycosyltransferases and TFs at mRNA expression level were found, such as *FUT3*, which correlated with *CDX1*, *ETS2*, *HNF1A*, *HNF4A*, *MECOM* and *MYB*. These TFs are upregulated in colon-like cell lines pointing to their potential role in regulating fucosylation during colon differentiation. In conclusion, our study reveals novel layers of potential GSL glycans regulation relevant for future research in colon differentiation and CRC.

### 3

### 3.1 Introduction

Colorectal cancer (CRC) was the fourth most commonly diagnosed cancer (6.1%) and the second leading cause of cancer death (9.2%) worldwide in 2018 [1]. Risk factors of CRC comprise genetics, aging and lifestyle including low-fiber and high-fat diet, obesity, lack of physical exercise, as well as smoking [2]. Currently, therapeutic solutions for individual CRC patients lack efficiency due to the heterogeneity of the disease, an asymptomatic clinical course as well as cancer diagnosis at late stages [3]. Therefore, biomarker discovery is essential to allow early diagnosis and treatment of CRC and to gain a better understanding of the disease progression [4].

Various studies have shown that an altered glycosylation could be associated with the malignant transformation in colon cancer [4]. Notably, CRC is characterized by a higher abundance of  $\beta$ 1-6 branching as well as increased (poly-)N-acetylglucosamine extensions, (truncated) oligomannose types and decreased bisected N-glycans. For the O-glycans a higher expression of core 1 and core 2 O-glycans were found in CRC, an increase of (sialyl)Tn-antigen, (sialyl)T-antigen, as well as a decrease of core 3 and core 4 O-glycans. Moreover, sialyl Lewis<sup>A</sup> (sLe<sup>A</sup>) has been found to be highly expressed in the colon and rectum during neoplastic transformation as well as cancer progression. This antigen is known to be involved in the adhesion of cancer cells to E-selectin present on endothelial cells, playing an important role in the metastasis of the colon cancer cells [5]. In addition, sLe<sup>A</sup> has been correlated with a worse prognosis within each stage and is considered as a prognostic factor for patients with advanced CRC [6, 7].

Glycosphingolipids (GSLs) play essential roles in various biological processes including cellular development, differentiation and cell fate decisions [8]. Taking this into account, it is not surprising that alterations of the GSL glycosylation have been found for malignant transformation including a decrease of disialylated gangliosides and globosides (except for Gb3 [Gal $\alpha$ 1-4Gal $\beta$ 1-4Glc $\beta$ 1-ceramide]) [4, 9] Another study revealed specific changes of GSL glycans in CRC tumor tissue compared with controls, including the increase of fucosylation and the reduction of acetylation, sulfation, globosides and disialylgangliosides [10]. Misonou *et al.*, found enhanced fucosylation and sialylation of GSLs in CRC tumors when compared to control tissues [11]. Gangliosides are involved in the regulation of cell adhesion

and motility, further resulting in the initiation of metastasis in many cancers [12]. In addition, Gb3 and certain gangliosides, such as GD3 (NeuAc $\alpha$ 2-8NeuAc $\alpha$ 2-3Gal $\beta$ 1-4Glc $\beta$ 1-ceramide) and GM2 (GalNAc $\beta$ 1-4[NeuAc $\alpha$ 2-3]Gal $\beta$ 1-4Glc $\beta$ 1-ceramide), contribute to the tumor-associated angiogenesis [12-14]. Moreover, enhanced expression of Gb3 has been found in colorectal adenocarcinomas and their metastases in comparison with normal tissue [15, 16].

3 Importantly, CRC cell lines are widely used model systems as they recapitulate the molecular alterations and genomic profiles of primary tumors as demonstrated by genomic studies, drug sensitivity screening, as well as DNA-, RNA- and protein-profiling [17-19]. Previously, 34 CRC cell lines have been studied based on integrated data analysis of mutations, RNA and protein expression [19]. The cell lines were grouped into two distinct clusters, the colon-like cell lines expressed gastrointestinal differentiation markers and undifferentiated cell lines showed an upregulation of epithelial-mesenchymal transition (EMT) and transforming growth factor  $\beta$  signaling [19]. Moreover, the analysis of the *O*-glycome of 26 CRC cell lines revealed a high abundance of I-branched and (s)Le<sup>A/x</sup> antigens for colon-like cell lines while undifferentiated cell lines were dominated by truncated  $\alpha$ 2-6 core sialylated glycans in the absence of Lewis structures [20].

To date, there is only limited information on the expression and regulation of GSL glycans in CRC cell lines; thus, we performed an in-depth analysis of GSL glycans of 22 CRC cell lines to investigate the GSL glycan expression. For this analysis, the GSL glycan head groups were enzymatically released from their lipid portion, followed by analysis using porous graphitized carbon (PGC) nano-liquid chromatography (LC) electrospray ionization tandem mass spectrometry (MS) in negative ionization mode. Associations between glycosylation features and gene expression levels of glycosyltransferases (GTs) and transcription factors (TFs) were explored, shedding light onto the potential regulation of glycosylation pathways in colon differentiation.

## 3.2 Experimental Procedures

### 3.2.1 Materials

TFA and chloroform were obtained from Merck. Methanol (MeOH) of ultra LC-MS grade was purchased from Actuaall Chemical. NaBH<sub>4</sub>, HCl, ammonium bicarbonate

and Dowex cation-exchange resin (50W-X8) were obtained from Sigma-Aldrich. KOH and 100% glacial acetic acid were from Honeywell Fluka. C18 Ziptips were purchased from Millipore and TopTip (microspin column; empty column) from Glygen Corporation. Reverse phase (RP) tC18 solid-phase extraction (SPE) cartridges (50 mg) were purchased from Waters, and SPE bulk sorbent Carbograph was from BGB Analytik USA LLC. 2-Propanol was obtained from Biosolve Chimie. Acetonitrile (MeCN) of LC-MS grade was from Biosolve. Endoglycoceramidase I (EGCase I),  $\alpha$ 1-3/4 fucosidase,  $\alpha$ 1-2/4/6 fucosidase O,  $\alpha$ 2-3 neuraminidase S, purified bovine serum albumin, 10  $\times$  GlycoBuffer 1 and 1  $\times$  EGCase I reaction buffer were purchased from New England Biolabs. Dulbecco's modified Eagle's medium (DMEM) and HEPES-buffered RPMI1640 medium were purchased from Gibco (Thermo Fisher Scientific). Fetal bovine serum (FBS) was from Bodinco BV. Penicillin/streptomycin from Invitrogen, 0.5% trypsin-EDTA solution 10 $\times$  was ordered from Santa Cruz Biotechnology and T75 cell culture flasks were obtained from Greiner-Bio One B.V. Ultrapure water generated by an ELGA PURELAB system was used for all solvent preparations and washing steps.

### 3.2.2 Cell lines and cell culture

The human CRC cell lines were provided by the Department of Surgery and the Department of Cell and Chemical Biology of Leiden University Medical Center as well as the Department of Molecular Cell Biology and Immunology of Amsterdam UMC. Cells were cultured in RPMI1640 medium, except for cell lines LS174T, SW1398 and SW948, which were cultured with DMEM. All media were supplemented with 10% FBS and 1% antibiotics [streptomycin (5 mg/ml) and penicillin (5000 IU/ml)]. Cultures were performed in a cell incubator with 5% CO<sub>2</sub> in humidified air at 37 °C. A trypsin/EDTA solution was used to detach the cells when approximately 80% confluence was reached. The medium was used to terminate the enzymatic activity. The harvested cells were counted with a TC20 automated cell counter. The cells pellets were obtained by centrifuging at 1500g for 5 min after being washed twice with 1  $\times$  PBS. The prepared cell pellets were stored at -20 °C. All CRC cell lines have been tested by means of the Promega PowerPlex Fusion System 5C autosomal STR kit. More details about cell culturing, cell origin and the characteristics of the cell lines are provided in **Supplemental Table S1**.

### 3.2.3 Extraction and purification of GSLs from cells

3 A detailed procedure for the extraction and purification of GSLs from cells was described in a previous study [21]. In brief, two million cells were lysed in 200  $\mu\text{L}$  of  $\text{H}_2\text{O}$ , followed by the addition of 500  $\mu\text{L}$  of chloroform and 350  $\mu\text{L}$  of MeOH, after each addition the plate was shaken on a horizontal shaker for 10 min, followed by sonication for 10 min. Afterward, the sample was shaken for 4 h at room temperature (RT), followed by centrifugation for 15 min at  $3000 \times g$  at  $20^\circ\text{C}$ . A 400  $\mu\text{L}$  mixture of chloroform/MeOH (2:1; v/v) and 400  $\mu\text{L}$  of MeOH/ $\text{H}_2\text{O}$  (1:1; v/v) were added sequentially to the sample after collection of the upper phase. The sample mixture was shaken overnight at 1000 rpm at RT followed by the collection of the upper phase. Another two extractions were performed after adding 400  $\mu\text{L}$  of MeOH/ $\text{H}_2\text{O}$  (1:1; v/v) to the sample, followed by a centrifugation step (15 min at  $3000 \times g$  at  $20^\circ\text{C}$ ). The collected upper phase was combined into a single glass vial and dried under a vacuum. The dried sample was dissolved in 100  $\mu\text{L}$  of MeOH and incubated for 30 min on a shaker, followed by the addition of 100  $\mu\text{L}$  of  $\text{H}_2\text{O}$  to the suspended sample. The sample was loaded onto the tC18 RP SPE cartridge which was preconditioned with the sequential addition of 1 mL chloroform/MeOH (2:1; v/v), 1 mL MeOH and 2 mL MeOH/ $\text{H}_2\text{O}$  (1:1; v/v). The loaded cartridge was washed using 2 mL of MeOH/ $\text{H}_2\text{O}$  (1:1; v/v) and elution was performed by sequentially adding 2 mL of MeOH and 2 mL of chloroform/MeOH (2:1; v/v). The combined eluate was dried under vacuum.

### 3.2.4 Enzymatic release and purification of GSL glycans

The GSL glycan head groups were enzymatically released from extracted GSLs on the basis of the manufacturer's instruction. In brief, the purified GSLs were suspended with 36  $\mu\text{L}$  of  $\text{H}_2\text{O}$ , 4  $\mu\text{L}$  EGCCase I reaction buffer and 2  $\mu\text{L}$  EGCCase I enzyme. The reaction system (42  $\mu\text{L}$ ) was incubated at  $37^\circ\text{C}$  for 36 h. The purification of released glycans was performed on tC18 RP SPE cartridge preconditioned with 2 mL of MeOH and 2 mL of  $\text{H}_2\text{O}$ . The loaded cartridge was washed twice with 200  $\mu\text{L}$  of  $\text{H}_2\text{O}$ . A sample mixture that contained the flowthrough and the solution after washing was collected and was lyophilized under a vacuum.

### 3.2.5 Reduction and purification of reduced GSL glycan alditols

The purified GSL glycans were dissolved in 40  $\mu\text{L}$  of freshly prepared 1 M  $\text{NaBH}_4$  in

50 mM KOH, followed by an incubation at 50 °C for 3 h. The reaction was quenched using 4  $\mu\text{L}$  of glacial acetic acid. The mixture was loaded onto the RP C18 Ziptips containing 50W-X8 resin, which was preconditioned with 3  $\times$  60  $\mu\text{L}$  of 1 M HCl, 3  $\times$  60  $\mu\text{L}$  of MeOH and 3  $\times$  60  $\mu\text{L}$  of  $\text{H}_2\text{O}$ . The loaded cartridge was washed twice with 40  $\mu\text{L}$  of  $\text{H}_2\text{O}$ . The flowthrough and washing were collected and dried under a vacuum. To remove residual borate, 150  $\mu\text{L}$  of MeOH was added twice during the lyophilization process.

PGC cleanup was performed as described in a previous study [22]. Briefly, the dried reduced sample was resuspended in 40  $\mu\text{L}$  of  $\text{H}_2\text{O}$  with 0.1% TFA ( $v/v$ ) with shaking at RT. The empty TopTip containing 50  $\mu\text{L}$  of PGC slurry was preconditioned sequentially with 60  $\mu\text{L}$  of MeOH, 3  $\times$  60  $\mu\text{L}$  of 80% MeCN with 0.1% TFA ( $v/v$ ) and 3  $\times$  60  $\mu\text{L}$  of  $\text{H}_2\text{O}$  with 0.1% TFA ( $v/v$ ), followed by sample loading. The loaded column was washed with 2  $\times$  60  $\mu\text{L}$  of  $\text{H}_2\text{O}$  with 0.1% TFA ( $v/v$ ). The sample was eluted with 2  $\times$  40  $\mu\text{L}$  of 60% MeCN with 0.1% TFA ( $v/v$ ); after pooling the two eluates, the sample was dried under vacuum.

### 3.2.6 Measurement using PGC-nanoLC-MS/MS

The purified GSL glycans alditols were reconstituted in 20  $\mu\text{L}$  of  $\text{H}_2\text{O}$ . A fraction of the sample (2  $\mu\text{L}$ ) was loaded onto a Hypercarb PGC trap column (5  $\mu\text{m}$  Hypercarb Kappa, 320  $\mu\text{m}$   $\times$  30 mm, in house packed) and separated on a Hypercarb PGC nano-column (3  $\mu\text{m}$  Hypercarb Kappa, 75  $\mu\text{m}$   $\times$  100 mm, inhouse packed) within a Dionex Ultimate 3000 nanoLC system and analyzed with an amaZon speed ion trap MS (Bruker Daltonics). Solvent A consisted out of 10 mM ammonium bicarbonate and solvent B out of 60% MeCN in 10 mM ammonium bicarbonate. The separation was performed over a linear nanoLC gradient of solvent B ranging from 1% up to 65% in 80 min with flow rate of 0.5  $\mu\text{L}/\text{min}$ . The column was washed with 95% solvent B for 10 min. The GSL glycans were analyzed with capillary voltage of 1000 V, 280 °C of drying gas at 5 L/min and the nebulizer was set at 3 psi. To enhance sensitivity, 2-propanol enriched nitrogen gas was used for enhancing sensitivity [23]. MS spectra were obtained within the  $m/z$  range 340-2000 in negative ion mode with  $m/z$  1200 as the target mass of the smart parameter settings. MS/MS spectra were produced by collision-induced dissociation on the top three abundance precursors.



### 3.2.7 Exoglycosidase digestion

Prior to exoglycosidase digestions, a pooled sample was generated by combining 30  $\mu$ L of each cell line sample. As a positive control, GSL glycans released from two million PaTu-S cells was used [21]. The reaction was performed according to the manufacturer's instructions. Briefly, the corresponding reagents and samples were added into the reaction system and incubated at 37 °C overnight, followed by purification and drying under vacuum. More details about each exoglycosidase experiment can be found in **Supplementary Table S2**.

### 3.2.8 Data analysis

3

Structural assignments were based on accurate mass, elution properties during the separation on a PGC-nanoLC column, the manual inspection of MS/MS spectra according to previously identified diagnostic ions in negative mode for GSLs, exoglycosidase digestion as well as known biosynthetic pathways of GSL glycans (<https://rb.gy/mt4miw> (accessed 10 May)) [24-27]. Bruker Data analysis software (version 5.0) was used for data analysis. In detail, extracted ion chromatograms (EIC) were generated by extracting the theoretical mass of the first three isotopes of GSL glycans of the observed (singly and doubly charged) species. The peak area under the curve was generated by integrating each peak, which had a signal to noise ratio  $\geq 6$  for all the technical and biological replicates. Relative quantification was conducted on the total area of all GSL glycans detected in one sample which was normalized to 100%. Further data analysis and visualization were performed in "R" software (version 4.0.5) using packages "tidyverse", "Rcpm", "pcaMethods", "stringi", "readxl", "ggplot2", "ggrepel", "reshape2", "ggpubr", "tidyHeatmap", "corrplot", "ComplexUpset" and "data.table".

### 3.2.9 Experimental Design and Statistical Rationale

We performed an in-depth analysis of GSL glycans in 22 CRC cell lines by PGC-nanoLC-electrospray ionization-MS/MS in negative ionization mode. Full technical replicates were performed for all CRC cell lines. In addition, for the cell lines Colo320 and HCT15, biological replicates were performed (i.e. analyses of independent cell cultures). More information about the technical replicates of each sample is given in **Supplementary Tables S3.1 - S3.22**. The minimum positive number (0.00001) as an imputation was used for the application of statistical tools such as principal

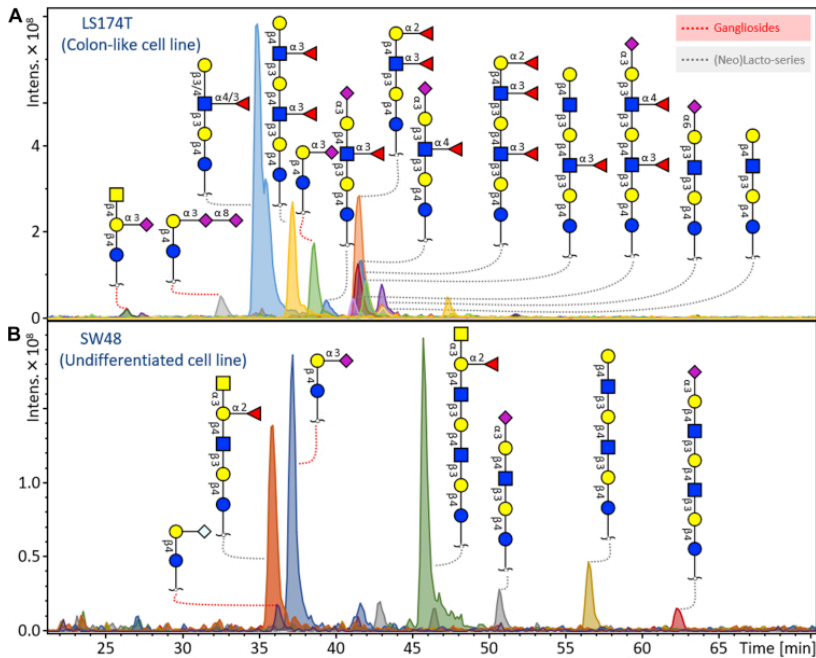
component analysis (PCA), which are sensitive to the missing values. Spearman correlations were assessed between the GSL glycan data of this study and published transcriptomics data of the same cell lines obtained from public repositories [19].

### 3.3 Results

#### 3.3.1 Striking differences in GSL glycan expression in CRC cell lines

GSLs were extracted from harvested cells, followed by enzymatic release of the GSL glycan head groups. In total, 94 GSL glycans could be structurally identified in CRC cell lines with sizes varying from 3 up to 12 monosaccharides (**Supplementary Table S4**). The identified GSL glycans were grouped on the basis of glycan structural features and a highly diverse GSL glycosylation was observed for the different CRC cell lines (**Supplementary Figure S1; Tables S3.1- S3.22 and S5**).

The diversity of GSL glycan expression for two representative CRC cell lines is exemplified in **Figure 1**. The colon-like cell line LS174T was mainly characterized by a relatively high expression of a diverse set of (neo)lacto-series GSL glycans with (sialyl)Lewis-type antigens (see **Table 1** for all the specific GSL glycosylation features). In contrast, the undifferentiated cell line SW48 showed high levels of (neo)lacto-series GSL glycans with blood group A antigens and GM3 ganglioside structures (**Supplementary Table S5**). The overall GSL glycan repertoire of the 22 CRC cell lines revealed that blood group antigens A, B and H were mainly found on (neo)lacto-series glycans, whereas blood group H was also present on some ganglioside and globo-series glycans. Additionally, cell line SW480 was the only cell line that contained a globo-series glycan with blood group A (relative abundance of 4%). Gangliosides were expressed in all CRC cell lines with various abundances, ranging from 0.2% in colon-like cell line SW1463 up to 100% in the neuroendocrine carcinoma-derived cell line Colo320 (unassigned cell line) (**Supplementary Table S6**). GM3 was the main ganglioside detected in most cell lines except for SW620 and SW1463 (**Supplementary Table S5**). Apart from cell line Colo320, all cell lines showed expression of (neo)lacto-series glycans, of which the SW1463 cell line had the highest relative abundance with 98% (**Supplementary Table S6**). The highest abundance of globo-series glycans was detected in SW480 cell line with a relative abundance of 17%. From the 94 detected GSL glycans in this study, only 27 were expressed in all CRC cell line classifications (colon-like, undifferentiated and unassigned; **Supplementary Figure S2 and Table S7**). Many GSL glycans were found



**Figure 1.** GSL glycan profile of two CRC cell lines. **(A)** the GSL glycan profile of the colon-like cell line LS174T displays a high expression of (sialyl)Lewis-type antigens on (neo)lacto-series GSLs. **(B)** undifferentiated human colon adenocarcinoma cell line SW48 is mainly characterized by the high expression of blood group A antigens and ganglioside GM3. CRC, colorectal cancer; GSL, glycosphingolipid.

to be unique for a specific CRC cell line classification (**Supplementary Figure S2** and **Table S7**). Specifically, 23 (neo)lacto-series glycans were only expressed in colon-like cell lines showing glycosylation features such as (A)Le<sup>B/Y</sup>, blood group H and (s)Le<sup>A/X</sup> antigens. The undifferentiated cell lines exhibited 20 unique GSL glycans, with glycosylation features such as blood groups A, B and H as well as I-branched glycans. Only three GSL glycans were found only in the unassigned cell lines, all belonging to the (neo)lacto-series with B-Le<sup>B/Y</sup>, blood group B and I-branched antigens.

### 3.3.2 Association of glycosylation features with cell line differentiation

To further investigate the association of specific GSL glycosylation features with the CRC classifications, a PCA was applied (**Figure 2** and **Supplementary Table S6**). In order to address the sparsity of individual GSL glycans detected in this study, and to account for the shared biosynthetic steps between them, GSL glycans were grouped into gangliosides (a-series and b-series), (neo)lacto-series and globo-series, and further categories were established on the basis of the expression of structural

motifs (**Table 1**), for example, total-,  $\alpha$ 1-2 and  $\alpha$ 1-3/4 fucosylation; total-,  $\alpha$ 2-3 and  $\alpha$ 2-6 sialylation; (s)Le<sup>A/X</sup> and (A)Le<sup>B/Y</sup>; blood group A, B, H as well as I-branching. The analyzed 22 CRC cell lines were classified into two groups based on colon cell differentiation: colon-like cell lines (7) and undifferentiated cell lines (10), and the remaining cell lines were unassigned (5) as there was no conclusive data available that would have allowed their assignment [19].







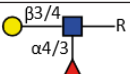
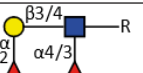
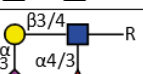
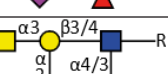
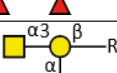


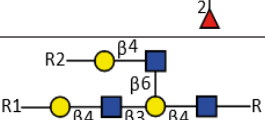
A score plot of the PCA model built on the glycosylation features showed a clear clustering according to cell line differentiation (**Figure 2**). The colon-like cell lines LS174T, LS180, SW948, SW1463 and SW1116 (**Figure 2A**) mainly grouped in the top left of the score plot and showed a higher expression of glycans with (s)Le<sup>A/X</sup>, (A)Le<sup>B/Y</sup> antigens as well as ( $\alpha$ 1-3/4)fucosylation (**Figure 2B**). In all colon-like cell lines sLe<sup>A/X</sup> antigens were detected at high abundance. Specifically, mucin-secreting cell lines LS180 and LS174T were characterized by the very high expression of (s)Le<sup>A/X</sup> and  $\alpha$ 1-3/4 fucosylation. Le<sup>B/Y</sup> antigens showed highest expression in colon-like cell line SW948 while A-Le<sup>B/Y</sup> antigens were found to have the highest expression in colon-like cell line SW1463 (**Supplementary Figure S1** and **Table S8**).

In contrast, undifferentiated cell lines SW620, SW48, HCT15, LOVO, SW480 and the unassigned cell line LS411N showed high expression of globo-series glycans and A, B, H blood group antigens. The highest expression of blood group A and H antigens was detected in the undifferentiated cell lines SW48 (50%) and HCT15 (55%), respectively (**Supplementary Figure S1** and **Table S8**). Only two cell lines (LS411N and LOVO) were found to express GSLs that carry blood group B antigens (23% and 35%, respectively). I-branching played an important role in the clustering of unassigned cell lines HCT8 (2%) and SW1398 (3%) as well as undifferentiated cell lines Caco-2 (5%) and SW837 (9%) (**Figure 2**; **Supplementary Figure S1** and **Table S8**).

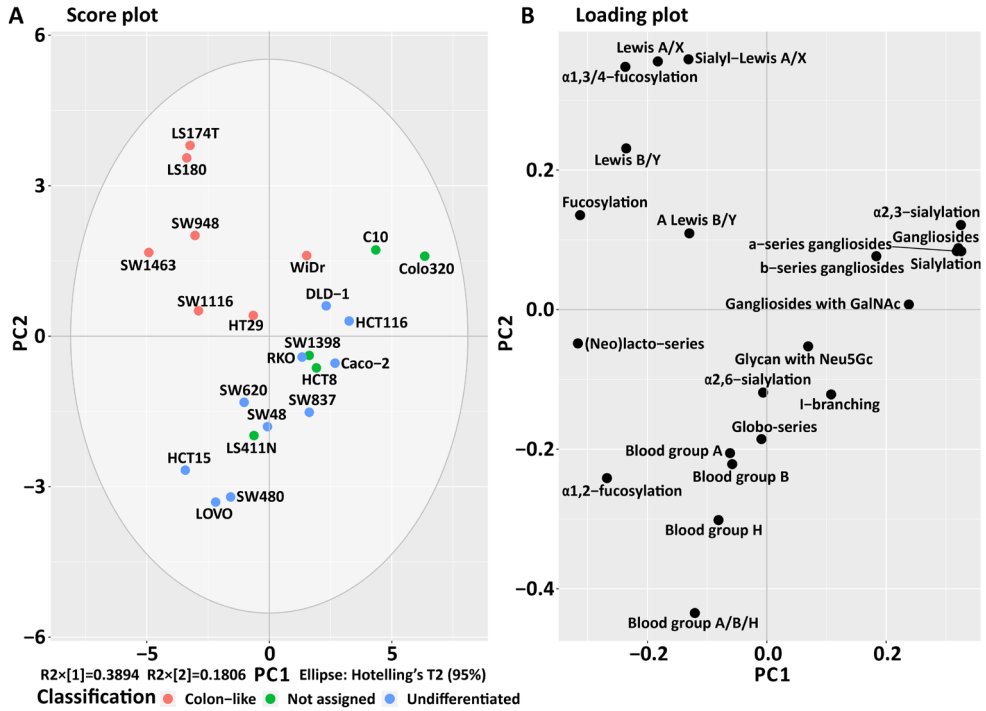
In addition, glycans with *N*-glycolylneuraminic acid (Neu5Gc) were found in most of the undifferentiated cell lines with various abundances ranging from 0.2% (SW837) up to 5% (RKO) (**Supplementary Figure S1** and **Table S8**). It is worth noting that cell lines clustered well based upon consensus molecular subgroup (CMS) classification [28] and their expression of GSL glycans (**Supplementary Figure S3**).

The clustering of CRC cell lines C10, Colo320, WiDr, DLD-1 and HCT116 (**Figure 2A**)

**Table 1. GSL glycosylation features, structure and average relative abundances for different CRC cell line classifications.** Arrows indicate relatively high and low levels of glycosylation features in colon-like versus undifferentiated CRC cell lines. <sup>a</sup>, glycosylation features with relatively higher expression in colon-like cell lines. <sup>b</sup>, glycosylation features with higher abundance in undifferentiated cell lines.

Glycosylation feature	Structure	Average relative abundance (%) in colon-like cell lines	Average relative abundance (%) in undifferentiated cell lines
Fucosylation <sup>a</sup>		105 ↑	41 ↓
$\alpha$ 1-2 fucosylation		38 ↑	37 ↓
$\alpha$ 1-3/4 fucosylation <sup>a</sup>		66 ↑	4 ↓
Sialylation <sup>b</sup>		39 ↓	53 ↑
$\alpha$ 2-3 sialylation <sup>b</sup>		32 ↓	44 ↑
$\alpha$ 2-6 sialylation		4 ↓	5 ↑
Le <sup>A/Xa</sup>		19 ↑	0.2 ↓
Le <sup>B/Ya</sup>		20 ↑	4 ↓
sLe <sup>A/Xa</sup>		9 ↑	0.4 ↓
A-Le <sup>B/Ya</sup>		8 ↑	0 ↓
Blood group A <sup>b</sup>		7 ↓	13 ↑
Blood group B <sup>b</sup>		0 ↓	4 ↑
Blood group H <sup>b</sup>		3 ↓	1 ↑
I-branching <sup>b</sup>		0 ↓	1 ↑

was driven by a high expression of (a- and b-series) gangliosides, gangliosides with GalNAc and ( $\alpha$ 2-3)sialylation (right part of the loading plot of **Figure 2B**). Gangliosides and specifically a-series gangliosides were highly expressed in the unassigned cell lines HCT8, SW1398, C10 and Colo320 as well as undifferentiated cell lines Caco-2,



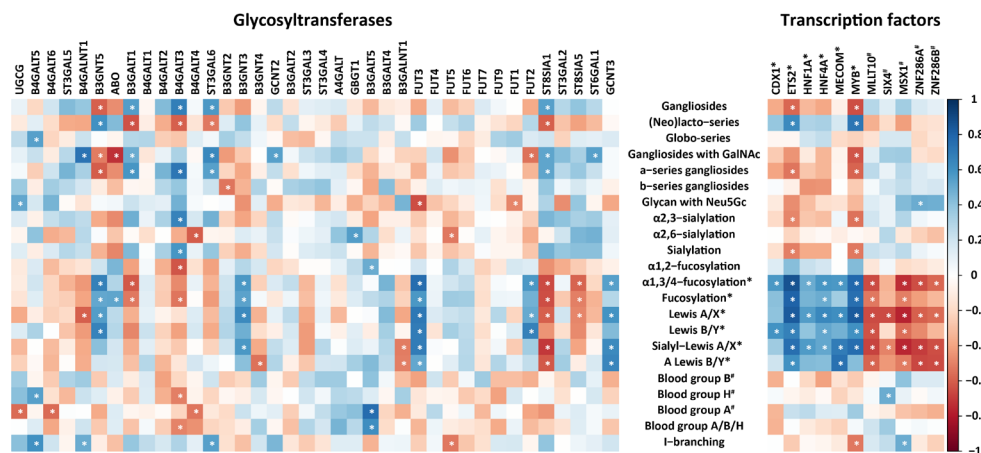
**Figure 2. PCA of glycosylation features of CRC cell lines. (A)** PCA score plot and **(B)** loadings plot of PC1 against PC2. The first two principal components explain 57% of the variation within the data. Technical replicates were averaged per cell line. CRC, colorectal cancer; PC, principal component; PCA, principal component analysis.

DLD-1, HCT116 and one colon-like cell line (WiDr). A high abundance of b-series gangliosides was detected in undifferentiated cell lines RKO and SW837 as well as for the unassigned cell lines Colo320 and C10. Gangliosides with GalNAc were highly expressed in undifferentiated cell lines SW837, Caco-2 and HCT116 as well as unassigned cell lines SW1398 and Colo320. The high expression of  $\alpha 2$ -3 sialylation, overall sialylation and gangliosides was observed in undifferentiated cell lines Caco-2, DLD-1 and HCT116, colon-like cell line WiDr, as well as unassigned cell lines HCT8, SW1398, C10 and Colo320. On the other hand, globo-series glycans were highly abundant in undifferentiated cell lines HCT15, SW480, Caco-2, HCT116 and colon-like cell line SW1116 as well as the unassigned cell line HCT8. The findings and relative quantification of GSL glycans in CRC cell lines are summarized in **Supplemental Figure S1** and **Table S8**.

### 3.3.3 The correlation between glycosyltransferases and GSL glycans expression

To get further insights into the glycosylation phenotypes, a correlation analysis was performed between the gene expression of GTs and expression of glycosylation features presented in CRC cell lines. A dataset containing mRNA expression of relevant GTs involved in the biosynthesis of GSL glycans was obtained from a previous study [19]. The correlations between GSL glycosylation and relevant GTs are shown in **Figure 3**. Although it did not meet the significance threshold, a trend toward a positive correlation was found between *ST3GAL5* and the expression of gangliosides ( $r = 0.44$ ,  $p = 0.08$ ), specifically the a-series gangliosides ( $r = 0.46$ ,  $p = 0.06$ ). The gene *B4GALNT1*, which encodes the corresponding GT responsible for adding GalNAc to Gal in ganglioside biosynthesis, was positively correlated with gangliosides containing a GalNAc (ganglioside with GalNAc,  $r = 0.73$ ,  $p = 0.001$ ). Moreover, *ST8SIA1*, whose corresponding GT is responsible for the biosynthesis of b-series gangliosides, positively associated with ganglioside expression ( $r = 0.54$ ,  $p = 0.03$ ). In addition, the expression of (neo)lacto-series glycans showed positive correlation with *B3GNT5* ( $r = 0.60$ ,  $p = 0.01$ ), which encodes the key enzyme for the biosynthesis of (neo)lacto-series glycans. No significant correlation was found between globo-series glycans and *A4GALT*, which encodes the GT that catalyzed the initial step of their biosynthesis.

3



**Figure 3. Correlation matrix for glycosylation features of CRC cell lines and the expression of genes encoding relevant GTs (left) and TFs (right).** Glycosylation traits and TFs marked with \* show higher expression in colon-like cell lines, whilst those with # display higher abundance in undifferentiated cell lines. Spearman correlation was performed with the significance cut-off  $p < 0.05$ . White asterisks mark statistically significant correlations. The correlation coefficients are color coded as indicated in the right key bar. CRC, colorectal cancer; GT, glycosyltransferase; TF, transcription factor.

Further correlations were found between several terminal glycosylation motifs and responsible transferases. For example,  $\alpha$ 2-3 sialylation showed positive correlation with *ST3GAL6*, which encodes the enzyme responsible for adding the  $\alpha$ 2-3 linked sialic acid. Additionally, *ST6GAL1* encoding the GT, which adds a sialic acid in  $\alpha$ 2-6 linkage, showed a trend towards a positive correlation with  $\alpha$ 2-6 sialylation. *FUT3* expression had significant positive correlations with fucosylation,  $\alpha$ 3-4 fucosylation,  $\text{Le}^{\text{A}/\text{X}}$ ,  $\text{sLe}^{\text{A}/\text{X}}$ ,  $\text{Le}^{\text{B}/\text{Y}}$  and  $\text{A-Le}^{\text{B}/\text{Y}}$ . *FUT2* showed significant positive correlation with  $\text{Le}^{\text{B}/\text{Y}}$  antigen expression. Moreover, positive correlations were found between *B3GNT5* and  $\alpha$ 3-4 fucosylation, fucosylation,  $\text{Le}^{\text{A}/\text{X}}$  and  $\text{Le}^{\text{B}/\text{Y}}$ . *GCNT2* encoding a GT responsible for synthesis of I-branched glycans showed a trend towards a positive correlation with I-branched glycans ( $r = 0.45$ ,  $p = 0.07$ ), although it did not meet the significance threshold. All significant  $r$ -values are indicated in **Table 2**, for all  $r$ - and  $p$ -values see **Supplementary Table S9**.

### 3.3.4 The association of expression of GSL glycans with TFs

A recent study revealed that gene expression of TFs *CDX1*, *ETS2*, *HNF1A*, *HNF4A*, *MECOM* and *MYB* showed higher expression in colon-like cell lines in comparison to undifferentiated cell lines, which had higher expression of *MLLT10*, *MSX1*, *SIX4*, *ZNF286A* and *ZNF286B* [20]. To get further insights into the potential regulation of GSLs expression in CRC cell lines, associations of glycosylation features with TFs were evaluated (**Figure 3**). Significant correlations were found between glycosylation features and several TFs. The glycosylation feature  $\alpha$ 1-3/4 fucosylation which was high in colon-like cell lines, showed a positive correlation with *CDX1*, *ETS2*, *HNF1A*, *HNF4A*, *MECOM* and *MYB*, whereas it showed negative correlations with TFs *MLLT10*, *MSX1*, *ZNF286A* and *ZNF286B*. Significant correlation was found for  $\text{Le}^{\text{A}/\text{X}}$  with *ETS2*, *HNF1A*, *HNF4A*, *MECOM*, *MYB*, *MLLT10*, *SIX4*, *MSX1*, *ZNF286A* and *ZNF286B*.  $\text{sLe}^{\text{A}/\text{X}}$  correlated positively with *ETS2*, *HNF1A*, *HNF4A*, *MECOM*, *MYB* but negatively with *MLLT10*, *SIX4*, *MSX1*, *ZNF286A* and *ZNF286B*.  $\text{Le}^{\text{B}/\text{Y}}$  showed positive correlations with *CDX1*, *ETS2*, *HNF4A*, *MYB*, *MLLT10* and *MSX1*.  $\text{A-Le}^{\text{B}/\text{Y}}$  correlated positively with *ETS2* and *MECOM* and negatively with *MLLT10*, *MSX1*, *ZNF286A* and *ZNF286B*. Additionally, a positive correlation of (neo)lacto-series was found with *ETS2* and *MYB*. Meanwhile, *SIX4* showed high expression in undifferentiated cell lines and revealed showed a positive correlation with blood group H that was highly expressed in undifferentiated cell lines (**Figure 2** and **Supplementary Figure S1**).



Additionally, *MSX1* showed a trend towards positive correlation with blood group B ( $r = 0.41$ ,  $p = 0.10$ ), which was also highly expressed in the undifferentiated cell line LOVO (**Supplementary Figure S1**). All significant  $r$ -values are presented in **Table 2**, for all  $r$ - and  $p$ -values see **Supplementary Table S9**.

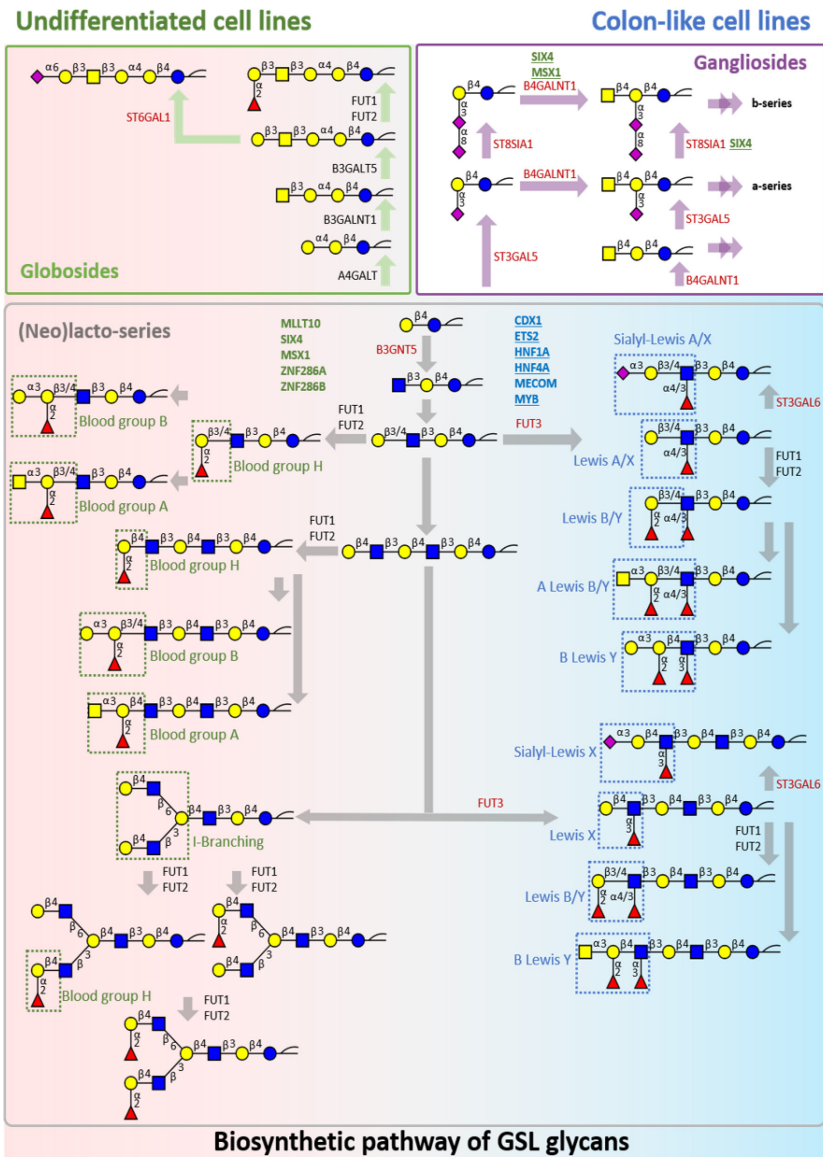
3 In order to investigate the role of TFs in relation to GTs, a correlation analysis was performed between TFs and GTs that are known to be involved in the biosynthesis of GSL glycans. *FUT3*, which showed positive correlations with Lewis antigens expression and  $\alpha$ 1-3/4 fucosylation, correlated positively with *CDX1*, *ETS2*, *HNF1A*, *HNF4A* and *MYB*, which are all highly expressed in colon-like cell lines (**Supplementary Figure S4**). *B3GNT5* showed positive correlation with TFs that were highly expressed in colon-like cell lines and presented a negative correlation with the TFs that showed high expression in the undifferentiated cell lines. *B4GALNT1*, encoding a key GT for gangliosides series, showed positive correlations with *SIX4* and *MSX1*, both highly expressed in undifferentiated cell lines. Additionally, *ST8SIA1* revealed a negative association with *ETS2*, *HNF1A*, *MECOM*, *MYB* and correlated positively with *MLLT10*, *SIX4*, *MSX1*, *ZNF286A* and *ZNF286B*, which were highly expressed in the undifferentiated cell lines. All  $p$ - and  $r$ -values can be found in **Supplementary Table S10**.

### 3.3.5 The association of GSL glycan and O-glycan in CRC cell lines

We recently performed an in-depth *O*-glycosylation analysis of different CRC cell lines [20] which allowed us to compare the expression patterns of glycosylation features shared between *O*-glycans and GSL glycans. Positive correlations between the same glycosylation features of the GSL glycome and the *O*-glycome (**Supplementary Figure S5** and **Table S11**) included Le<sup>A/X</sup>, sLe<sup>A/X</sup>, blood group B, blood group H and  $\alpha$ 2-6 sialylation. Moreover, the Le<sup>A/X</sup> antigens of the GSL glycome showed significant positive correlations with specific *O*-glycan features such as sLe<sup>A/X</sup> structures, sulfo-Le<sup>A/X</sup> and (sialyl)dimeric Le<sup>A/X</sup>. Similarly, the sLe<sup>A/X</sup> (GSL glycome) had positive correlations with the additional *O*-glycan features such as Le<sup>A/X</sup>, sulfo-Le<sup>A/X</sup> and (sialyl) dimeric Le<sup>A/X</sup>. All  $p$ - and  $r$ -values can be found in **Supplementary Table S11**.

## 3.4 Discussion

In this study, the GSL head groups of 22 CRC cell lines were investigated using a



**Figure 4. Differences in the expression of GSL glycan between colon-like and undifferentiated CRC cell lines including the relevant genes involved in the biosynthetic pathway.** The main subgroups of GSL glycan globo-series, gangliosides and (neo)lacto-series are presented in green, purple and grey rectangles, respectively. GTs that correlated positively with corresponding glycosylation features or subgroups of GSL glycan are marked in red. TFs upregulated in colon-like cell lines or undifferentiated cell lines are indicated in bold blue or bold green, respectively. Underlined TFs indicate a significant positive correlation with the expression of certain GTs. Dashed green and blue rectangles indicate the high expression of corresponding glycosylation feature in undifferentiated cell lines and colon-like cell lines, respectively. The display of gangliosides on a white background indicates a similar expression for both classifications. Double arrows display that glycan can

be further elongated. CRC, colorectal cancer; GSL, glycosphingolipid; GT, glycosyltransferase; TF, transcription factor.

PGC-nanoLC-MS/MS platform, showing a highly diverse GSL glycome. The correlations of specific glycosylation features were investigated with corresponding GTs involved in the biosynthesis of GSL glycans as well as with certain TFs. In addition, this study revealed the associations between TFs and GTs related to the GSL biosynthetic pathway.

### 3

The differences in the expression of GSL glycome, relevant GTs and TFs of colon-like and undifferentiated CRC cell lines can be interpreted on the basis of the known GSL biosynthetic pathway (**Figure 4**). Overall, blood group A/B/H antigens and globo-series glycans were highly expressed in undifferentiated cell lines while colon-like cell lines showed a high abundance of (s)Le<sup>A/X</sup>, Le<sup>B/Y</sup> antigens and  $\alpha$ 1-3/4 fucosylation (**Table 1**). Especially, Le<sup>B/Y</sup> was highly expressed in colon-like cell lines and significantly positively correlated with the gene expression of *FUT2*, which has previously shown to be increased in colon tumors [29]. Moreover, our data revealed strong associations between (s)Le antigen expression and *FUT3* expression, which is one of GTs essential for the biosynthesis of sLe antigens [30]. Specifically, (s) Le<sup>A/X</sup> antigens were expressed in high abundance by the mucin-secreting cell lines LS180 and LS174T, which was consistent with the recent finding on *O*-glycans [20]. Whereas, a higher amount of Lewis antigens on GSL glycans was also detected in SW1463 cell line, which was not the case in the *O*-glycome [20]. In our study, we discovered significant correlations of (s)Le<sup>A/X</sup> expression between the GSL glycome and *O*-glycome (**Supplementary Figure S5** and **Table S11**), and similar correlations were found for the Le<sup>A/X</sup> and <sup>B/Y</sup> antigens. Moreover, previous *N*-glycosylation studies of the same set of cell lines showed increased antenna fucosylation, particularly in the colon-like cell lines LS180, LS174T, HT29, WiDr and SW1116, indicative of the presence of Lewis type antigens [31]. Previous studies have revealed that an increased expression of sLe<sup>A/X</sup> antigens is associated with poor prognosis and sLe<sup>X</sup> could be a potential prognostic marker in CRC [6, 7]. This is in line with our finding that a high expression of sLe<sup>A/X</sup> in colon-like cell lines belonged to either CMS2 or CMS3 [19], suggesting that sLe<sup>A/X</sup> might be a marker for CMS2/3 CRC tumors.

In contrast, the undifferentiated cell lines were characterized by a high abundance

of blood group A, B and H antigens. These findings were consistent with the expression of blood group antigens in *O*-glycans (**Supplementary Figure S5**) [20]. The LOVO cell line obtained from an individual with blood type B expressed particularly high levels of blood group B antigens on the GSL glycans [32]. Similarly, cell lines derived from patients with blood type A (SW480, SW620 and HT29) revealed a relatively high abundance of blood group A antigens on GSL glycans [20, 32]. The Caco-2 cell line (*O*-blood group individual) expressed blood group H antigens for both GSL and *O*-linked glycans [20]. Surprisingly, blood group A antigens were found in cell line SW1116 which is derived from an *O*-blood type individual; however, this is in agreement with previous findings [32]. Interestingly, no blood group antigens on GSL glycans were detected for the LS174T cell line (individual with blood group O). In the SW48 cell line, derived from an individual with blood type AB, only blood group A antigens were found for both the GSLs and *O*-glycome (**Supplementary Table S6**) [20]. Whilst cell lines HCT15, HCT8 and DLD-1 were obtained from the same patient, they showed profound and diverse glycosylation profiles: high expression of blood group H antigens was found in HCT15 and HCT8, while no expression of blood group H antigens were expressed in cell line DLD-1 which was consistent with the results studying the *O*-glycome [20]. While good associations were found between the GSL and *O*-glycome, further studies are needed to investigate the *N*-glycome of these CRC cell lines. This can further confirm the specificity of expression of blood group antigens in certain CRC cell lines, especially for the undifferentiated cell lines which correspond to CMS1/4 CRC tumors [19].

Increased sialylation of glycans is observed in various cancers including colon cancer and especially, the  $\alpha$ 2-3 sialylation is suggested to play an essential role in metastasis with regards to EMT [4]. Moreover, undifferentiated cell lines show an upregulation of transforming growth factor  $\beta$ -induced genes and EMT signatures and EMT signatures [19]. In our study, a high expression of ( $\alpha$ 2-3)sialylation was found for the undifferentiated cell lines DLD-1 and HCT116 as well as colon-like cell line WiDr. The aberrant expression of sialylation may result from the dysregulation and/or altered activity of sialyltransferases. Compared to  $\alpha$ 2-3-sialylation, the  $\alpha$ 2-6 sialylation is critical for the malignant and metastatic progression of which *ST6GAL1* plays an important role [33, 34]. This is supported by a study in which *ST6GAL1* knock down resulted in a significant suppression in metastasis of the colon-like cell line HT29 [35]. Moreover, the upregulation of  $\alpha$ 2-6 sialylation was found in colon

cancer tissue [36]. Correlation analysis showed that the  $\alpha$ 2-6 sialylation on GSL glycans had a strong association with  $\alpha$ 2-6 sialylation found on the *O*-glycome of CRC cell lines (**Supplementary Figure S5**). In our study, a high expression  $\alpha$ 2-6 sialylation on GSL glycans was detected in cell lines SW1398, LS411N and SW620, which can be selected as suitable models for studying the function of  $\alpha$ 2-6 sialylation and *ST6GAL1* during CRC invasion.

### 3

A previous study indicated that *GCNT2* and its corresponding I-branched glycans accelerate EMT in CRC [37]. I-branched glycans are considered as emerging effectors of malignant progression and were found to be relatively highly expressed in the undifferentiated cell lines SW873 and Caco-2 [33]. Moreover, this feature correlated positively with *GCNT2* gene expression (**Figure 3**). Neu5Gc is normally not found in humans attributable to a deletion of the gene *cmah* [38]. Previous studies have indicated that the detection of Neu5Gc in human cells is most likely due to the dietary incorporation of glycoproteins carrying Neu5Gc [39]. It is highly likely that the found expression of Neu5Gc in the CRC cell lines of this study is due to uptake and incorporation from the culture medium, which contains 10% FBS.

The findings in our study highlight the value of an in-depth glycomic study as different glycan motifs may present different functions. For example, it is known from previous studies that specific glycosylation features such as  $\alpha$ 2-3 sialylation and I-branching play an essential role in the EMT and  $\alpha$ 2-6 sialylation and sLe<sup>A/X</sup> in the metastasis of CRC. Nevertheless, the specific mechanism of how cancer-associated GSL-glycome features are involved in the progress of CRC remains unknown. Therefore, specific CRC cell line models can be selected for future studies on the basis of their glycome.

To gain insights into the potential regulation of GSL glycans, a correlation study of GSL glycans with TFs was performed. Our study showed that high expression of (s)Le<sup>A/X</sup>, Le<sup>B/Y</sup> antigens and fucosylation (specifically  $\alpha$ 1-3/4 fucosylation) in colon-like cell lines were significantly correlated with expression of *FUT3* and presented positive correlations with TFs that are upregulated in colon-like cells including *CDX1* (**Figure 3**). *CDX1* as an intestine-specific TF is involved in the regulation of different biological processes including proliferation, apoptosis and cell adhesion and associated with oncogenesis of the gastrointestinal tract [40, 41]. An association of

highly fucosylated *N*-glycans with *CDX1* in CRC cell lines was revealed and validated by Holst *et al.*, with the finding of *CDX1*-high expressing CRC cell lines characterized by higher expression of antenna fucosylation (Le<sup>A/X</sup> and <sup>B/Y</sup>) on *N*-glycans [31, 41]. Meanwhile, a positive correlation between *FUT3* and high expression of *CDX1* in CRC cell lines was discovered [41]. Besides, previous studies indicated that TFs *HNF1A* and *HNF4A* associated with intestinal development [42] and played an essential role in the expression of antenna fucosylation via the regulation of *FUT3* [43], which are consistent with our findings of correlations between *FUT3* and *CDX1*, *HNF1A* and *HNF4A* (**Supplementary Figure S4**). Furthermore, undifferentiated cell line HCT116 lacked *CDX1* expression while showing more aggressive behavior. Interestingly, *CDX1* expression in HCT116 cells induced intestinal epithelial differentiation to crypt-forming colonies [44]. Taking our findings and results from previous publications, we speculate that the aforementioned TFs, particularly *CDX1*, may play roles in the expression of Lewis type glycosylation features via the upregulation of *FUT3*, which may contribute to the differentiation of CRC cell lines. Our findings pave the way for future functional studies validating the generated hypotheses of the regulation of GSL glycan expression in CRC cell lines.

### 3.5 Conclusions

Our work revealed a strikingly diverse expression of GSL glycans in 22 CRC cell lines and provides an insight into the alteration of glycosylation during the development of CRC. The specific GSL glycome signatures were found for each CRC cell line classification. Additionally, correlations were found between GSL glycans and their corresponding GTs, which are known to be involved in the GSL biosynthesis providing valuable information on the regulation of GSL glycan expression in CRC cell lines. Besides, we hypothesize that certain TFs might play a role in the expression of fucosylated GSL glycans through the regulation of *FUT3* on the basis of the discovered correlations between GSL glycans and selected GT and TF genes as well as between GTs and certain TFs. However, mechanistic studies are needed to confirm the role of specific TFs such as *CDX1* in the regulation of GSL glycans expression. In conclusion, our in-depth analysis of GSL glycans profiling in CRC cell lines provides an informative resource for future research on biomarkers for CRC and paves the way for further studies on the regulation of GSL glycans in CRC.

### 3.6 Data availability

The raw mass spectrometric data files supporting the findings of this study are available in GlycoPOST in mzXML format, with the identifier GPST000239, accessible via the following link: <https://glycopost.glycosmos.org/preview/164822151461d35202311d4> with pin code: 5211.

### 3.7 Supplementary data

This article contains supplemental data

### 3.8 Acknowledgements

3

We kindly thank Y. van Kooyk, P. ten Dijke and N. G. Dekker-Ensink for providing the cell lines. We also thank N. G. Dekker-Ensink for the support with cell culturing.

### 3.9 References

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2018. **68**(6): p. 394-424.
2. Dekker, E., et al., *Colorectal cancer*. The Lancet, 2019. **394**(10207): p. 1467-1480.
3. Tikhonov, A.A., et al., *Analysis of Anti-Glycan IgG and IgM Antibodies in Colorectal Cancer*. Bull Exp Biol Med, 2019. **166**(4): p. 489-493.
4. Holst, S., M. Wuhrer, and Y. Rombouts, *Glycosylation characteristics of colorectal cancer*. Adv Cancer Res, 2015. **126**: p. 203-56.
5. Ugorski, M. and A. Laskowska, *Sialyl Lewis(a): a tumor-associated carbohydrate antigen involved in adhesion and metastatic potential of cancer cells*. Acta Biochim Pol, 2002. **49**(2): p. 303-11.
6. Nakayama T, W.M., Katsumata T, Teramoto T, Kitajima M. , *Expression of sialyl Lewis(a) as a new prognostic factor for patients with advanced colorectal carcinoma*. Cancer, 1995 Apr 15. **75**(8):2051-6
7. Schiffmann, L., et al., *A novel sialyl Le(X) expression score as a potential prognostic tool in colorectal cancer*. World J Surg Oncol, 2012. **10**: p. 95.
8. Russo, D., et al., *Glycosphingolipid metabolism in cell fate specification*. J Cell Sci, 2018. **131**(24): p. jcs219204.
9. Ajit Varki, R.K., Bryan Toole, and Pamela Stanley, *Glycosylation Changes in Cancer*, in *Essentials of Glycobiology, 3rd edn*, C.R. Varki A, Esko JD et al (ed), Editor. 2017, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
10. Holst, S., et al., *Investigations on aberrant glycosylation of glycosphingolipids in colorectal cancer tissues using liquid chromatography and matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS)*. Mol Cell

- Proteomics, 2013. **12**(11): p. 3081-93.
11. Misonou, Y., et al., *Comprehensive Clinico-Glycomic Study of 16 Colorectal Cancer Specimens: Elucidation of Aberrant Glycosylation and Its Mechanistic Causes in Colorectal Cancer Cells*. J. Proteome Res., 2009. **8**(6): p. 2990-3005.
  12. Birklé, S., et al., *Role of tumor-associated gangliosides in cancer progression*. Biochimie, 2003. **85**(3): p. 455-463.
  13. Heimbürg-Molinaro, J., et al., *Cancer vaccines and carbohydrate epitopes*. Vaccine, 2011. **29**(48): p. 8802-8826.
  14. Desselle, A., et al., *Anti-Gb3 monoclonal antibody inhibits angiogenesis and tumor development*. PLoS One, 2012. **7**(11): p. e45423.
  15. Falguieres, T., et al., *Human colorectal tumors and metastases express Gb3 and can be targeted by an intestinal pathogen-based delivery tool*. Mol Cancer Ther, 2008. **7**(8): p. 2498-508.
  16. Distler, U., et al., *Shiga toxin receptor Gb3Cer/CD77: tumor-association and promising therapeutic target in pancreas and colon cancer*. PLoS One, 2009. **4**(8): p. e6813.
  17. Mouradov, D., et al., *Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer*. Cancer Res, 2014. **74**(12): p. 3238-47.
  18. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity*. Nature, 2012. **483**(7391): p. 603-7.
  19. Berg, K.C.G., et al., *Multi-omics of 34 colorectal cancer cell lines - a resource for biomedical studies*. Mol Cancer, 2017. **16**(1): p. 116.
  20. Madunic, K., et al., *Colorectal cancer cell lines show striking diversity of their O-glycome reflecting the cellular differentiation phenotype*. Cell Mol Life Sci, 2021. **78**(1): p. 337-350.
  21. Zhang, T., et al., *Differential O- and Glycosphingolipid Glycosylation in Human Pancreatic Adenocarcinoma Cells With Opposite Morphology and Metastatic Behavior*. Front Oncol, 2020. **10**: p. 732.
  22. Jensen, P.H., et al., *Structural analysis of N- and O-glycans released from glycoproteins*. Nat Protoc, 2012. **7**(7): p. 1299-310.
  23. Madunic, K., et al., *Dopant-Enriched Nitrogen Gas for Enhanced Electrospray Ionization of Released Glycans in Negative Ion Mode*. Anal Chem, 2021. **93**(18): p. 6919-6923.
  24. Ruhaak, L.R., A.M. Deelder, and M. Wührer, *Oligosaccharide analysis by graphitized carbon liquid chromatography-mass spectrometry*. Anal Bioanal Chem, 2009. **394**(1): p. 163-74.
  25. Karlsson, H., A. Halim, and S. Teneberg, *Differentiation of glycosphingolipid-derived glycan structural isomers by liquid chromatography/mass spectrometry*. Glycobiology, 2010. **20**(9): p. 1103-16.
  26. Chai, W., V. Piskarev, and A.M. Lawson, *Negative-ion electrospray mass spectrometry of neutral underivatized oligosaccharides*. Anal Chem, 2001. **73**(3): p. 651-7.



27. Anugraham, M., et al., *A platform for the structural characterization of glycans enzymatically released from glycosphingolipids extracted from tissue and cells*. Rapid Commun Mass Spectrom, 2015. **29**(7): p. 545-61.
28. Becht, E., et al., *Immune and Stromal Classification of Colorectal Cancer Is Associated with Molecular Subtypes and Relevant for Precision Immunotherapy*. Clin Cancer Res, 2016. **22**(16): p. 4057-66.
29. Muinelo-Romay, L., E. Gil-Martín, and A. Fernández-Briera,  *$\alpha$ (1,2) fucosylation in human colorectal carcinoma*. Oncol Lett, 2010. **1**(2): p. 361-366.
30. Hirakawa, M., et al., *Fucosylated TGF-beta receptors transduces a signal for epithelial-mesenchymal transition in colorectal cancer cells*. Br J Cancer, 2014. **110**(1): p. 156-63.
31. Holst, S., et al., *N-glycosylation Profiling of Colorectal Cancer Cell Lines Reveals Association of Fucosylation with Differentiation and Caudal Type Homeobox 1 (CDX1)/Villin mRNA Expression*. Mol Cell Proteomics, 2016. **15**(1): p. 124-40.
32. Dahiya R, I.S., Byrd JC, Kim YS *ABH blood group antigen expression, synthesis, and degradation in human colonic adenocarcinoma cell lines*. Cancer Res, 1989. **49**: p. 4550-4556.
33. Dimitroff, C.J., *I-branched carbohydrates as emerging effectors of malignant progression*. Proc Natl Acad Sci U S A, 2019. **116**(28): p. 13729-13737.
34. Lu, J. and J. Gu, *Significance of  $\beta$ -Galactoside  $\alpha$ 2,6 Sialyltransferase 1 in Cancers*. Molecules, 2015. **20**(5): p. 7509-27.
35. Yingting Zhu, U.S., Asad Ullah, Harish Gagneja, Charles S. Berenson, Peter Lance, *Suppression of a sialyltransferase by antisense DNA reduces invasiveness of human colon cancer cells in vitro*. Biochimica et Biophysica Acta 2001. **1536**: p. 148-160.
36. Dall'Olio F, T.D., *Expression of alpha 2,6-sialylated sugar chains in normal and neoplastic colon tissues. Detection by digoxigenin-conjugated Sambucus nigra agglutinin*. Eur J Histochem 1993. **37**: p. 257-65.
37. Chao, C.C., et al., *Downregulation of miR-199a/b-5p is associated with GCNT2 induction upon epithelial-mesenchymal transition in colon cancer*. FEBS Lett, 2017. **591**(13): p. 1902-1917.
38. Irie, A., et al., *The molecular basis for the absence of N-glycolylneuraminic acid in humans*. J Biol Chem, 1998. **273**(25): p. 15866-71.
39. Banda, K., et al., *Metabolism of vertebrate amino sugars with N-glycolyl groups: mechanisms underlying gastrointestinal incorporation of the non-human sialic acid xeno-autoantigen N-glycolylneuraminic acid*. J Biol Chem, 2012. **287**(34): p. 28852-64.
40. Guo, R.J., E.R. Suh, and J.P. Lynch, *The role of Cdx proteins in intestinal development and cancer*. Cancer Biol Ther, 2004. **3**(7): p. 593-601.
41. Holst, S., et al., *N-Glycomic and transcriptomic changes associated with CDX1 mRNA expression in colorectal cancer cell lines*. Cells, 2019. **8**(3): p. 273.
42. Boudreau, F., et al., *Hepatocyte nuclear factor-1 alpha, GATA-4, and cauda*

- related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene.* J Biol Chem, 2002. **277**(35): p. 31909-17.
43. Lauc, G., et al., *Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation.* PLoS Genet, 2010. **6**(12): p. e1001256.
  44. Yeung, T.M., et al., *Cancer stem cells from colorectal cancer-derived cell lines.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3722-7.

