Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/45594 holds various files of this Leiden University dissertation

Author: Holst, Stephanie
Title: Glycomic signatures of colorectal cancer
Issue Date: 2017-01-24
Stephanie Holst¹, Jennifer L. Wilding², Kamila Koprowska², Yoann Rombouts¹³, Walter F. Bodmer², Manfred Wuhrer¹

In preparation

¹Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands; ²Department of Oncology, Cancer and Immunogenetics Laboratory, Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom; ³University of Toulouse, Institute of Pharmacology and Structural Biology, Toulouse, France
N-glycomic and transcriptomic changes associated with CDX1 mRNA expression in colorectal cancer cell lines
Abstract

The caudal-related homeobox protein 1 (CDX1) is a transcription factor, which is important in the development, differentiation and homeostasis of the gut. Although the involvement of CDX genes in the regulation of few glycosyltransferase gene expressions has been shown, associations between glycosylation phenotypes and CDX1 mRNA expression have hitherto been not well studied. Here, we characterized the N-glycomic phenotype of 16 colon cancer cell lines, which differed in CDX1 gene expression. We found high CDX1 mRNA expression associated with a higher degree of multifucosylation on N-glycans, which was supported by up-regulated gene expression of fucosyltransferases involved in antenna fucosylation. Interestingly, hepatocyte nuclear factor (HNF)4A, and HNF1A were, among others, positively associated with high CDX1 mRNA expression and were previously proven to regulate antenna fucosylation. Furthermore, the glycomic profiles fit with a role of CDX1 as a tumor suppressor preventing the decrease of bisection as well as increase of sialylation that have previously been found to be associated with more mesenchymal (de-differentiated) phenotypes and epithelial-to-mesenchymal transition.
INTRODUCTION

Glycans form an important part of the outer layer of the cell and are involved in major biological processes including cell differentiation, adhesion, and interactions with other cells, pathogens, or the extracellular matrix as well as cellular transformations such as cancer [1-3]. Glycans occur in several structures as modifications on proteins and lipids. One type of glycosylation on proteins is the N-glycosylation, in which oligosaccharides are attached via an N-glycosidic linkage to an asparagine (N) of the consensus sequence -N-X-S/T- (X any amino acid except proline, S = serine, T- threonine) [4]. These so-called N-glycans share a common pentasaccharide core-structure and form, depending on the elongation, high-mannose type, complex-type, or hybrid-type N-glycans (Fig. 1A) [4]. N-glycans and other glycan classes have been associated with several diseases and cancers, in which they reflect the status of the malignancy as a specific glyco-phenotype [5-7].

Colorectal cancer associated glycan changes have been recently reviewed by us [8]. By screening different colorectal cancer cell lines, we previously showed an association of caudal-related homeobox protein 1 (CDX1) mRNA expression with increased multi-fucosylation (more than one fucose), indicative for antenna fucosylation [9]. Antenna-fucosylation is the result of the activity of fucosyltransferases encoded by the genes FUT1, 2, 3, 4, 5, 6, 7, and 9. Fucosyltransferases are expressed in a cell- and organ-dependent manner and lead to the expression of cancer-associated Lewis antigens (Fig. 1B) which were shown to be elevated in colorectal cancer [10]. While the sialylated variants sialyl Lewis X and sialyl Lewis A were associated with advanced stages, metastatic colon cancer, and poor prognosis [11, 12], the non-sialylated Lewis A antigen was found decreased in poorly differentiated and metastatic colon cancers [13]. The blood group antigens and the di-fucosylated epitopes Lewis Y and Lewis B have been shown to be highly abundant in healthy human proximal colon, whereas the distal colon and rectum showed expression of these antigens under malignant transformation [14]. In colon cancer cell line HT29/M3, transfection with FUT1 increased Lewis Y expression, which resulted in a decreased expression of sialyl Lewis X and a reduced invasiveness and metastatic potential [15]. Reduction of sialyl Lewis antigens and thereby selectin-mediated adhesion and metastasis have been also demonstrated by inhibition of FUT3 in colon cancer cells HT-29LMM [16]. Although increased fucosylation has been described for several cancers, including colorectal cancer [10, 17], different glycans alterations may occur in different stages of the cancer, in different locations of the colon, and may involve different mechanism. A novel metastatic pathway, for example, has been suggested that involves the de-fucosylation of glycoproteins, proposing the loss of fucosylation during cancer progression and metastasis, and leaving enhanced fucosylation a characteristic of early stage cancer [18]. In accordance, a decrease in fucosylation has been observed from colorectal adenomas towards carcinomas in patients [19]. Furthermore, a mutation in the GDP-mannose 4,6-dehydratase (GMDS) gene has been identified especially in metastatic lesions of some colon cancers (10%), which results in reduced Lewis antigens or complete loss of fucosylation [20]. Also the colon cancer cell line HCT116 bears a GMDS mutation and revealed a more aggressive phenotype likely due to apoptosis resistance [21]. These latter results suggest that other mechanisms next to E-selectin mediated extravasation and metastasis formation are involved in colorectal cancer progression.

Cancer progression is further associated with a loss of differentiation. In colon cancer, loss of the differentiation marker villin has been described which resulted in poor differentiation and worse survival [22]. Arango et al. identified the loss of CDX1 expression as cause for the loss of villin and the associated poor differentiation and worse survival [23]. Also others have described the involvement of CDX1 in cell differentiation [24], and the differentiation marker Cytokeratin 20 has been shown to be regulated by CDX1 [25]. CDX1 is a transcription factor which is involved in the modulation of a
variety of processes including proliferation, apoptosis, cell-adhesion, and columnar morphology [24]. Its expression is necessary for a large number of intestine-specific genes [26] and is important for the intestine development, differentiation, and maintenance of the intestinal phenotype [24, 27].

CDX1 has been described to act as tumour suppressor, which was supported by several studies showing loss or down-regulation of CDX1 expression in colon cancer tumours [28, 29] and cell lines [30]. Restoring expression of CDX1 mRNA in HT29 resulted in less malignant phenotypes, which was accompanied by an enhancement of the antigen presentation system and increased expression of HLA-I and others, suggesting that the loss of CDX1 expression could help tumour cells to escape the immune system [31]. Interestingly, the CDX1 gene is located on the same chromosomal arm 5q as the APC gene, which is frequently target of genomic rearrangements in colon cancer, and can thereby cause the decrease/loss of CDX1 expression in colon cancer [27].

Although glycosylation as well as CDX1 have both been shown to play a role in cell differentiation and cancer (suppression), hitherto only very little has been described with regard to CDX1-associated glycosylation. Triggered by our previous findings [9], which suggested a relationship between multi-fucosylation and high CDX1 mRNA expression, we here characterized the N-glycosylation phenotype of different colon cancer cell lines with high and low expression of CDX1 mRNA. The glyco-phenotypic data were further correlated with differently expressed (glyco)genes. We confirmed a higher degree of multi-fucosylation in cell lines with high CDX1 gene expression, which was supported by higher mRNA expression of fucosyltransferases (FUT2,3,6) involved in antenna fucosylation. Furthermore, the N-glycomic characterization revealed a decrease in galactosylation and sialylation in cell lines with high CDX1 expression. Interestingly, gene microarray data identified also concerted changes of transcription factors such as hepatocyte nuclear factor (HNF)4A, HNF4G, and HNF1A which likewise showed differential expression between the investigated cell lines with high vs. low CDX1 expression, and have been previously shown to regulate antenna fucosylation [32].

![Fig. 1: N-glycans and Lewis antigens.](image)

**A N-glycans**

- High-mannose
- Complex
- Hybrid

**B Lewis antigens**

- **High-mannose**
- **Complex**
- **Hybrid**

**Legend**

- Man: Mannosues
- Gal: Galactoses
- GlcNAc: Glucosamine
- Fuc: Fucoses
- NeuAc: sialic acid

**Description**

- **A N-glycans**
  - N-glycans are attached to an asparagine (N) in the consensus sequence N-X-S/T with X any amino acid except proline, followed by serine (S) or threonine (T).
  - Main monosaccharides involved in N-glycosylation are mannoses (Man), galactoses (Gal), N-acetylglucosamine (GlcNAc), fucoses (Fuc), and N-acetylmuramic acid (NeuAc).
  - N-glycans share a common core-structure consisting of two GlcNAc and three Man. Depending on the elongation three N-glycan types are differentiated: i) high-mannose type N-glycans, ii) complex-type N-glycans, and iii) hybrid-type N-glycans. The illustrated glycans represent examples; the amount of monosaccharides added can vary and complex type glycans can exhibit more than two antennae.
  - A detailed description of N-glycosylation is given by Stanley [4].

- **B Lewis antigens**
  - **Lewis X/A**
  - **Lewis Y/B**
  - **H-type**

**Activity**

- FUT1/2 attach α1,2-fucosylation to Gal-residues. Activity of several α2,3-sialyltransferases (ST3GAL genes) attach NeuAc residues to galactoses to form sialy Lewis antigens.
MATERIALS AND METHODS

MATERIALS

8M guanidine hydrochloride (GuHCl), 1-hydroxybenzotriazole (HOBt) hydrate, 50% sodium hydroxide (NaOH), and super DHB matrix (2-hydroxy-5-methoxy-benzoic acid and 2,5-Dihydroxybenzoic acid, 1:9) were obtained from Sigma-Aldrich (Steinheim, Germany). HPLC SupraGradient acetonitrile (ACN) was obtained from Biosolve (Valkenswaard, The Netherlands). Dithiothreitol (DTT), ethanol, and sodium bicarbonate (NaHCO₃) were from Merck (Darmstadt, Germany) and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) from Fluorochem (Hadfield, UK). Peptide N-Glycosidase F (PNGase F) was purchased from Roche Diagnostics (Mannheim, Germany). The peptide calibration standard was purchased from Bruker Daltonics (Bremen, Germany). MultiScreen® HTS 96 multiwell plates (pore size 0.45 μm) with high protein-binding membrane (hydrophobic Immobilon-P PVDF membrane) were purchased from Millipore (Amsterdam, The Netherlands), 96-well polypropylene 0.8 mL 96-deepwell plate and 96-well PCR plate polypropylene from Greiner Bio (Alphen a/d Rijn, The Netherlands). All buffers were prepared using ultrapure water generated from a 18.2MΩ-cm Purelab Ultra system from Elga (Ede, The Netherlands). Control Visucon-F plasma pool (citrated and 0.02 M HEPES buffered plasma pool from 20 healthy human donors) was obtained from Affinity Biologicals (Ancaster, Canada). Cell culture media were purchased from Gibco (Paisley, UK), fetal bovine serum (FBS) from Autogen Bioclear (Wiltshire, UK), and penicillin-streptomycin from Invitrogen (Paisly, UK). The RNeasy mini kit was purchased from Qiagen (Manchester, UK).

CELLS AND CELL CULTURE

Details of colorectal cancer cell line origin and characteristics can be found in Table 1 and more detailed in Supplementary Table S1. CC20, COLO678, GP2D, HCT116, ISRECO1, LIM1863, LS174T, PCJW, RCM1 and SW403 cell lines were cultured in Dulbecco Modified Eagle Medium (DMEM). CAR1, HCA46, HDC8, OXCO1 and VACO429 were cultured in Iscove’s Modified Dulbecco’s Medium (IMDM). HCC56 cells were cultured in Roswell Park Memorial Institute (RPMI)-1640 medium. All of the above media were supplemented with 10% FBS and 1% penicillin-streptomycin. Cell lines were grown in a 10% CO₂ incubator to 50% to 80% confluence before next passage or cell pellet preparation. Cell pellets were washed once with PBS, snap-frozen in dry ice and stored at -80°C until further analysis. Cell pellets were dissolved in water and cell counts were estimated using the Countess™ Automated Cell Counter (Invitrogen) based on trypan blue staining.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>CDX1 mRNA [a.u.]</th>
<th>Tumor</th>
<th>Stage/grade</th>
<th>Differentiation</th>
<th>Lumen/Formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCA46</td>
<td>3308.59</td>
<td>sigmoid colon adenocarcinoma</td>
<td>C</td>
<td>poorly</td>
<td>Lumen</td>
</tr>
<tr>
<td>HCC56</td>
<td>4115.35</td>
<td>colon adenocarcinoma/Irresectable</td>
<td>n.a.</td>
<td>moderately</td>
<td>Dense</td>
</tr>
<tr>
<td>GP2D</td>
<td>1058.48</td>
<td>colon adenocarcinoma</td>
<td>B</td>
<td>poorly</td>
<td>Intermediate</td>
</tr>
<tr>
<td>PCJW</td>
<td>2598.75</td>
<td>colon adenocarcinoma</td>
<td>C</td>
<td>poorly</td>
<td>Intermediate</td>
</tr>
<tr>
<td>LS174T</td>
<td>1147.36</td>
<td>colon adenocarcinoma</td>
<td>B/grade I</td>
<td>well</td>
<td>Intermediate</td>
</tr>
<tr>
<td>LIM1863</td>
<td>2862.38</td>
<td>colon adenocarcinoma</td>
<td>C/grade II</td>
<td>poorly</td>
<td>Lumen</td>
</tr>
<tr>
<td>SW403</td>
<td>2964.14</td>
<td>colon adenocarcinoma</td>
<td>C/grade II</td>
<td>n.a.</td>
<td>Intermediate</td>
</tr>
<tr>
<td>RCM1</td>
<td>1152.22</td>
<td>colon adenocarcinoma</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Intermediate</td>
</tr>
<tr>
<td>ISRECO1</td>
<td>52.10</td>
<td>colon adenocarcinoma</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>VACO429</td>
<td>28.69</td>
<td>colon adenocarcinoma</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Dense</td>
</tr>
<tr>
<td>HCT116</td>
<td>59.94</td>
<td>colon adenocarcinoma</td>
<td>grade IV</td>
<td>poorly</td>
<td>Dense</td>
</tr>
<tr>
<td>CC20</td>
<td>59.51</td>
<td>sigmoid colon adenocarcinoma</td>
<td>B</td>
<td>well</td>
<td>Network/aggregate</td>
</tr>
<tr>
<td>CAR1</td>
<td>49.35</td>
<td>colon adenocarcinoma</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Dense</td>
</tr>
<tr>
<td>COLO678</td>
<td>16.21</td>
<td>colon adenocarcinoma, metastatic lymph node</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Dense</td>
</tr>
<tr>
<td>HDC8</td>
<td>29.31</td>
<td>colon adenocarcinoma</td>
<td>C/grade III</td>
<td>n.a.</td>
<td>Intermediate</td>
</tr>
<tr>
<td>OXCO1</td>
<td>49.69</td>
<td>colon adenocarcinoma</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>
**N-glycan release, derivatization, and purification**

N-glycans were released in duplicate from three biological replicates per cell line using a PVDF-membrane based release protocol followed by linkage-specific sialic acid derivatization, purification by cotton-HILIC-SPE, and MALDI-TOF-MS analysis as described previously [9]. Shortly, cell pellets were resuspended in water, disrupted by sonication, and proteins immobilized on 96-well PVDF-filter plates (~0.5x10^6 cells/well; 5µL human plasma control; water as blanks) in presence of chaotropic agents. After washing unbound materials, N-glycans were released overnight at 37°C. Released glycans were derivatized using ethyl esterification allowing for discrimination of N-acetyleneuraminic acid linkages (α2,3 vs. α2,6) [33] in a ratio 20:100 (sample: derivatization reagent), and purified by cotton-thread HILIC-SPE.

**MALDI-TOF-MS analysis**

Released, derivatized, and purified glycans (5 µL sample) were spotted on an anchor chip MALDI target plate (Bruker Daltonics) and co-crystallized with 0.5 µL of 5 mg/mL superDHB in 50% ACN supplemented with 1 mM NaOH. Spectra were recorded in positive-ion reflector mode after calibration with a Bruker peptide calibration kit using a Bruker UltrafleXtreme™ mass spectrometer, controlled by FlexControl 3.4 software Build 119 (Bruker Daltonics). Mass spectra were obtained over an m/z range of m/z 1000 to 5000 for a total of 10 000 shots (1000 Hz laser frequency, 200 shots per raster spot during complete random walk). Tandem mass spectrometry (MALDI-TOF-MS/MS) was performed for structural elucidation via fragmentation in gas-off TOF/TOF mode.

**Data processing and analysis of MALDI-TOF-MS spectra**

Spectra were smoothed (Savitzky Golay algorithm, peak width: m/z 0.06, 4 cycles), baseline corrected (Tophat algorithm) and exported to xy-files using FlexAnalysis 3.4 (Stable Build 76). Mean average spectra were generated per technical replicate, which were summed to one spectrum using the open-source software mMass (http://www.mmass.org; [34]). The spectrum was internally re-calibrated using glycan peaks of known composition (H5N2 at m/z 1257.423, H6N2 at m/z 1419.476, H7N2 at m/z 1581.529, H8N2 at m/z 1743.581, H5N4F1 at m/z 1809.639, H5N4F2 at m/z 1955.697, H5N4E1 at m/z 1982.709, H10N2 at m/z 2067.687, H6N5F1 at m/z 2174.7715, H5N4L1E1 at m/z 2255.793, H5N4E2 at m/z 2301.835, H6N5E1 at m/z 2347.8403, H7N6F1 at m/z 2539.904, H6N5F4 at m/z 2612.945, H6N5L1E1 at m/z 2620.925, H7N6E1 at m/z 2712.973, H7N6L2 at m/z 2940.016, H9N8 at m/z 3124.111, H7N6L4F1 at m/z 3632.243) as calibrants (minimum five used), followed by peak picking in mMass with cut-off signal-to-noise (S/N) 3. The peaklist was manually revised, and analysed in GlycoWorkbench 2.1 stable build 146 (http://www.eurocarbdb.org/) using the Glyco-Peakfinder tool (http://www.eurocarbdb.org/ms-tools/) for generation of a glycan compositions list. Our novel in-house software developed for automated data processing, MassyTools version 0.1.8.0 [35], was used for calibration using a 3rd degree function and targeted data extraction of the area under the curve for each individual mass spectrum. To prevent over-estimation of overlapping glycan species, only the first three isotopes were extracted and the area under the curve corrected based on the theoretical isotopic pattern. The quality of the data was assessed using several quality parameters calculated within the software. Only good quality spectra (total intensity > 1x10^5 and fraction of analyte area with S/N>9 is more than 50%) as well as analytes (S/N > 6, ppm < 20, quality score > 0.10) were included for analyses. Finally, the corrected area-under-the-curve values were rescaled to a total relative intensity of 100% for each spectrum. Selected glycan compositions were confirmed by MS/MS and a final peak list as well as MS/MS data is given in Supplementary Table S2. MS/MS spectra were manually interpreted and fragment ions annotated using GlycoWorkbench 2.1 according to the nomenclature of Domon and Costello [36]. Averages of direct traits per cell line were used to build a principle component analysis model in SIMCA Version 13.0 (Umetrics AB, Umea, Sweden) with seven random cross-validation (CV)
For increased robustness, derived glycan traits such as galactosylation, fucosylation, sialylation, and others were calculated in SPSS Version 23 (IBM Corp, Armonk, NY). The formulas for calculation are given in Supplementary Table S4, average relative abundances are given in Supplementary Table S3. A two-tailed Mann-Whitney test was performed in GraphPad Prism Version 6 (GraphPad Software, Inc., La Jolla, CA) with significance level $\alpha=0.05$ to assess differences in N-glycosylation between CDX1 high and low expressing cells as well as CDX2 high and low expressing cells. Boxplots for visualization were generated in Rstudio statistical software environment (Version 0.99.892, http://www.r-project.org/) and show the median with interquartile range.

**Gene Expression Microarrays and data analysis**

Total RNA was extracted by using the RNeasy mini kit according to the manufacturer’s instructions. Twenty micrograms of RNA of each sample were sent to the Molecular Biology Core Facility of the Paterson Institute for Cancer Research, Manchester, UK, for gene expression microarray analysis using the Human genome U133+2 chips following the manufacturer’s instructions (Affymetrix, High Wycombe, UK). Microarray data were analyzed using Partek Genomics Suite software. The data were log2-transformed, RMA-normalized (with GC correction), using quantile normalization with Median Polish for Probeset summarization as optional settings in the software.

Glycogenes were selected based on the microarray Glyco_v4_Hs from the Consortium for Functional Glycomics (CFG; http://www.functionalglycomics.org/glycomics/publicdata/microarray.jsp) as well as genes of interest based on literature, resulting in a list of 843 genes. Differentially expressed genes were identified by low probabilities from a t test comparing mean-expression levels of the top tercile CDX1 high expressing cell lines (n=32) versus the bottom tercile CDX1 low expressing cell lines (n=32). The significant level was adjusted for multiple testing. For significantly different expressed glycosyltransferases, fold changes were calculated for CDX1$^{\text{high}}$ and CDX1$^{\text{low}}$ cell lines and data was visualized as boxplots in Rstudio showing the median and interquartile range. To evaluate the correlation between CDX1 gene expression and glycosyltransferase gene expression in the investigated cell lines, Pearson and Spearman correlation analysis was performed in GraphPad Prism Version 6 (GraphPad Software).

**RESULTS**

Based on our previous findings of increased fucosylation with CDX1 mRNA expression in a set of colorectal cancer cell lines [9], we characterized the N-glycosylation of a different set of colorectal cancer cell lines with high (8 cell lines; CDX1$^{\text{high}}$) vs. low (8 cell lines; CDX1$^{\text{low}}$) expression of CDX1 mRNA. The CDX1$^{\text{high}}$ cell lines have on average 65-fold higher mRNA expression of CDX1 as compared to CDX1$^{\text{low}}$ cell lines (Table 1, Supplementary Table S1).

**Mass spectrometric N-glycan profiling**

As previously observed, N-glycomic profiles of all cell lines were dominated by high-mannose type N-glycans, but differed in the ratios of these glycans as seen in the exemplary spectra in Fig. 2 showing the cell lines Colo978 (CDX1$^{\text{low}}$) and HCA46 (CDX1$^{\text{high}}$). With regard to complex type N-glycans, those derived from Colo678 exhibited more sialic acid residues, especially in $\alpha_{2,3}$-linkage, while glycans from HCA46 were characterized by the presence of many fucoses, low sialylation level, and additional N-acetylhexosamines (HexNAc). In total, 221 individual glycan species were identified (Supplementary Table S2). In order to evaluate whether CDX1$^{\text{high}}$ and CDX1$^{\text{low}}$ expressing cells can be distinguished based on their N-glycomic signature, a principal component analysis (PCA) was performed on the relative expression levels of all individual glycans. This resulted in a model with four principle
Fig. 2: Exemplary MALDI-TOF-MS spectra. MALDI-TOF-MS spectrum (A) of CDX1-low expressing cell line COLO678, and (B) of CDX1-high expressing cell line HCA46. Spectra were recorded in positive ion reflectron mode on a Bruker UltrafleXtreme mass spectrometer and show a zoom in the higher m/z range. Main peaks were annotated with glycan cartoons reflecting compositions. The presence of structural isomers cannot be excluded. Differences in N-acetylneuraminic acid linkages are indicated using angles.
components (PC) explaining 68.9% of the variation in the data. The score plot of PC1 vs. PC2 showed clear separation of CDX1<sup>high</sup> and CDX1<sup>low</sup> cell lines along PC1 (Fig. 3A) in this unsupervised model, suggesting a different N-glycan profile between the two groups. To assess which glycans drive the separation, the corresponding loading plot was investigated and showed various multi-fucosylated glycans (Fig. 3B, coloured in green) associated with CDX1<sup>high</sup> cell lines, whilst sialylated glycans (Fig. 3C, coloured in blue and red) were found to largely mark CDX1<sup>low</sup> cell lines.

In order to facilitate the analysis, glycan-derived traits were calculated by grouping glycan species (direct traits) into classes according to glycosylation characteristics such as sialylation and fucosylation. Derived traits gave insight into general structural differences and showed increased robustness as compared to individual glycans. Observations from the example spectra and the PCA plot were confirmed by comparing derived glycan traits of CDX1<sup>high</sup> and CDX1<sup>low</sup> expressing cell lines and differences were evaluated using a Mann-Whitney test. The relative abundances of derived traits as well as the calculations are given in Supplementary Table S3+S4 and differences between the two groups of cells are discussed in the following.

![Principle component analysis](image)

**Fig. 3: Principle component analysis.** Principle component analysis (PCA) was performed to explore differences in N-glycosylation between CDX1-high and CDX1-low expressing cells. (A) Score plot of principal components (PC) 1 vs. 2 explaining 25.8% and 19.4% of the data, respectively. (B) Corresponding loading plot with color indications for multi- (green) and mono-fucosylated (purple) individual N-glycans, and (C) the same loading plot with color indications for α<sub>2</sub>,3-sialylated (rose) and α<sub>2</sub>,6-sialylated (blue) N-glycans.
**Differences in Fucosylation**

Differences were found in multi-fucosylation (more than one fucose), indicative for antenna fucosylation (Lewis X/A, Y/B), which was with $\varnothing$ 54% significantly higher in CDX1 high expressing cell lines as compared to $\varnothing$ 33% in CDX1 low expressing cell lines (Fig. 4A, Supplementary Table S3), which is in line with our previous findings [9]. In order to map the N-glycomic phenotype to transcriptomic data, a gene microarray was performed and filtered for glycan-related genes based on the glycan-gene microarray from the Consortium for Functional Glycomics (Glyco_v4_Hs; not specific for N-glycans only) as well as relevant genes based on literature, resulting in a selection of 843 genes (Supplementary Table S6). Gene expression data on the 16 cell lines investigated by MALDI-TOF-MS were compared to genes differentially expressed between CDX1$^{\text{high}}$ and CDX1$^{\text{low}}$ cell lines in a larger set of samples (32 CDX1$^{\text{high}}$ and 32 CDX1$^{\text{low}}$). The significance level for differential gene expression was adjusted for multiple testing and $p$-values $<$5.9x10E-05 were considered significant. In accordance with the MS-based glyco-phenotype, gene microarray data showed significantly increased expression (3- to 8-fold) of the corresponding fucosyltransferase genes FUT2, involved in Lewis Y/B synthesis, and FUT3 and 6, involved in Lewis X/A, in the eight CDX1$^{\text{high}}$ cell lines compared to eight CDX1$^{\text{low}}$ cells (Fig. 3B-D, Supplementary Table S5). This was confirmed in the larger set of cell lines (Supplementary Table S6) with fucosyltransferases FUT2, 3, 4, 6 up-regulated in CDX1$^{\text{high}}$ (fold change of 2.5-4.5 for CDX1$^{\text{high}}$ vs. CDX1$^{\text{low}}$, $p$-value $<$8.0x10E-07). Furthermore, the GDP-L-fucose precursor GMDS and the GDP-fucose transporter solute carrier family 35 member C1 (SLC35C1) were 2.2-fold ($p$-value 4.6x10E-08) and 1.6-fold ($p$-value 3.1x10E-06), respectively, increased with high CDX1 expression in the larger sample set (Supplementary Table S6). The correlation between CDX1 gene expression and glycosyltransferase mRNA expression was further evaluated using Pearson and Spearman correlation analyses and showed significant correlations for FUT2, 3, 4, and 6 (Supplementary Table S7).

**Differences in Hex/HexNAc-ratio and Galactosylation**

Glycan structures with an equal amount or more HexNAc than hexoses (Hex) were 2-fold increased (Fig. 4E; Supplementary Table S3) and represent glycan structures with bisecting N-acetylgalactosamine (GlcNAc), non-galactosylated antennae, or addition of N-acetylgalactosamine (GalNAc). Gene expression of the glycosyltransferase involved in the formation of bisecting GlcNAc (MGAT3) showed a trend towards increased expression in CDX1$^{\text{high}}$ cells compared to CDX1$^{\text{low}}$ (Fig. 4F) cells and was significantly up-regulated in CDX1$^{\text{high}}$ cells of the larger set of cell lines with differential CDX1 expression ($p$-value 1.1x10E-05; Supplementary Table S6). MGAT3 expression also showed significant Pearson correlation with CDX1 expression (Supplemental Table S7).

Other glycan epitopes contributing to the HexNAc≥Hex class are the LacdiNAc epitope and the SdA antigen (NeuAcα2-3GalNAcβ1-4Galβ1-4GlcNAc-R; NeuAc=N-acetylneuraminic acid, Gal=galactose). Accordingly, glycosyltransferases adding a GalNAc-residue to a GlcNAc to form LacdiNAc structures on N-glycans, B4GALNT3 and 4, showed a trend towards higher expression with high CDX1 gene expression with B4GALNT3 being significantly higher expressed in CDX1$^{\text{high}}$ versus CDX1$^{\text{low}}$ cells (Fig. 4G, Supplementary Table S3+S6). Gene expression of B4GALNT2, coding for the enzyme which adds the GalNAc for the SdA epitope, and ABO blood group α1,3-N-acetylgalactosaminyltransferase showed likewise a trend towards up-regulated expression in CDX1$^{\text{high}}$ cell lines.

Accordingly, overall galactosylation showed the opposite results and was found to be significantly lower in CDX1$^{\text{high}}$ cells than in CDX1$^{\text{low}}$ cell lines ($\varnothing$ 85% vs. $\varnothing$ 72%; Fig 4H, Supplementary Table S3). Corresponding galactosyltransferases showed the same trend (Supplementary Tables S5+S6).
Fig. 4: Differentially expressed N-glycan traits and glycosyltransferase genes. Derived N-glycan traits were calculated to evaluate glycosylation characteristics associated with CDX1 mRNA expression. N-glycomic phenotypes of the eight CDX1-high vs. eight CDX1-low cell lines were further compared to transcriptomic data of the same cell lines obtained from a gene expression microarray analysis using the Human genome U133+2 Affimetrix chips. Differential expression was observed for (A) Multi-fucosylation (more than one fucose) indicative for antenna fucosylation, and associated gene expression of (B) FUT2, (C) FUT3, and (D) FUT6, for (E) N-glycans with more or equal N-acetylhexosamines than hexoses (HexNAc≥Hex) and associated glycosyltransferase gene expressions of (F) MGAT3 (biseecting GlcNAc) and (G) B4GALNT3 (LacdiNAc), as well as (H) overall galactosylation. Larger N-glycans with (I) seven or more HexNAc and corresponding genes (J) B3GNT3 (poly-LacNAc), (K) B3GNT8 (poly-LacNAc), (L) MGAT4A (branching), as well as (M) overall sialylation, (N) α2,3-sialylation, and (O) α2,3-sialyltransferase gene ST3GAL6 showed differential expression in CDX1-high vs. CDX1-low expressing cell lines. Differences were evaluated by Mann-Whitney test for derived traits and t-test for LOG2-transformed mRNA expression data. Boxplots consistently indicate the median and interquartile range. Significances after multiple testing correction are indicated and correspond to * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.
Differences in antennarity and Poly-LacNAc expression

N-glycan structures with seven or more HexNAcs, indicative for branched structures or (poly-)LacNAc repeats (-Galβ1-4GlcNAcβ1-3Galβ1-3/4-GlcNAc-), were higher in CDX1high expressing cell lines as compared to CDX1low expressing cell lines with ∅ 27% vs. ∅ 20.0% (Fig. 4I, Supplemental Table S3). Since our MS data could not sufficiently differentiate between LacNAc-repeat and additional antenna, glycosyltransferase expression data could give more detailed insight. Genes encoding for beta-1,3-N-acetylglucosaminyltransferase 3 (B3GNT3) and B3GNT8, both involved in the synthesis of type-1 chains and poly-LacNAc repeats, were around 2-fold up-regulated with high CDX1 expression (Fig. 4J+K, Supplemental Table S5) and showed significant correlation for Pearson and Spearman (Supplemental Table S7), suggesting the presence of LacNAc-repeat structures. This could be confirmed in the larger set of cell lines, where B3GNT3 (p-value 2.5x10E-05) and B3GNT8 (p-value 3.1x10E-14) showed a 2-fold increase in CDX1 high expressing cell lines compared to CDX1low cell lines (Supplemental Table S6). Also the glycosyltransferase encoded by gene MGAT4A which is involved in the branching on the 1,3-arm of N-glycans to form tri- and tetra-antennary N-glycans, was significantly higher expressed in CDX1high cells as compared to CDX1low cells (Supplemental Table S5). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines (Supplemental Table S6).

Differences in sialylation

As indicated by the PCA analysis, overall sialylation was lower in CDX1 high cell lines compared to CDX1 low cell lines (∅ 36% vs. ∅ 21%; Fig. 4M, Supplementary Table S3). Most predominant differences were found for α2,3-sialylated N-glycans with ∅ 11% in CDX1high vs. ∅ 26% in CDX1low cell lines (Fig. 4N, Supplementary Table S3). The corresponding mRNA levels of sialyltransferases ST3GAL3,4, and 6 were decreased in CDX1 high cell lines as compared to CDX1 low cell lines, though not significantly (Fig. 4O, Supplementary Table S5). However, the Spearman correlation between ST3GAL6 mRNA and CDX1 gene expression was significant (Supplementary Table S7) and also in the larger set of samples, α2,3-sialyltransferases ST3GAL1, 3, 4, and 6, involved in sialylation of different glycan classes were decreased in CDX1high cell lines, though only ST3GAL1 significantly (p-value 4.1x10E-05, 3.2-fold decrease). Notably, sialylation per galactose was likewise decreased from ∅ 48% to ∅ 31% in association with high CDX1 expression (Supplementary Fig. S3), suggesting the decrease in sialylation being an ‘independent’ event and not a mere result of substrate limitation through decreased galactosylation.

Differences in gene expression involved in other glycan classes

Sialyltransferase ST6GALNAC1, mainly involved in the expression of O-glycan Sialyl-Tn was highly (8.8-fold) increased with high CDX1 expression (p-value 1.8x10E-011) and also O-glycan related N-acetylgalactosaminyltransferases GALNT4 and -6 showed increased expression with high CDX1 mRNA expression (p-value < 4.6x10E-06; Supplementary Table S6). On the other hand, POMGNT1, involved in O-mannosylation was decreased in high CDX1 expressing cells as compared to CDX1 low expressing cells. Differential glycosylation of glycan classes other than N-glycans were, however, outside the scope of this study and need further investigations.
DIFFERENCES IN GENE EXPRESSIONS INVOLVED IN EPITHELIAL DIFFERENTIATION

Villin 1 (VIL1) as differentiation marker was associated with CDX1 expression [23] and showed differential expression in the small sample set (Fig. 5A) as well as 18-fold higher expression in CDX1\(^{\text{high}}\) cells as compared to CDX1\(^{\text{low}}\) cells in the larger sample set (\(p\)-value 4.6x10\(^{-17}\); Supplemental Table S6). In contrast, vimentin (VIM; \(p\)-value 2.1x10\(^{-05}\)) and Zinc Finger E-Box Binding Homeobox 1 (ZEB1; \(p\)-value 2.2x10\(^{-08}\)), markers for a more undifferentiated cell phenotype were 4-5-fold down-regulated in CDX1\(^{\text{high}}\) cells compared to CDX1\(^{\text{low}}\) cells (Supplemental Table S6). Interestingly, HNF4A (\(p\)-value 1.5x10\(^{-13}\)), HNF4G (\(p\)-value 2.3x10\(^{-08}\)), HNF1A (\(p\)-value 2.0x10\(^{-08}\)), and GATA transcription factor 6 (\(p\)-value 2.5x10\(^{-12}\)) were 2-7-fold up-regulated in CDX1\(^{\text{high}}\) expressing cell lines (Supplemental Table S6) and were previously associated with CDX1 as well as intestinal development [37-39]. Already in the small subset of cell lines, HNF1A and HNF4A showed increased expression with high CDX1 mRNA levels (Fig. 5B+C). Moreover, galectin 4 (LGAL4), a target gene of HNF4A, was 47-fold up-regulated in CDX1\(^{\text{high}}\) cells (\(p\)-value 3.5x10\(^{-16}\)). Other galectins showed as well differential expression between CDX1\(^{\text{high}}\) and CDX1\(^{\text{low}}\) cells with galectin 3 (LGAL3) increased in CDX1\(^{\text{high}}\) cells (\(p\)-value 4.1x10\(^{-08}\)), whereas galectin 1 (LGAL1) was reduced (\(p\)-value 2.3x10\(^{-08}\); Supplemental Table S6). On the other hand, transforming growth factor (TGF) B1 (\(p\)-value 1.9x10\(^{-07}\)) and B2 (\(p\)-value 1.4x10\(^{-07}\)), which can induce epithelial-to-mesenchymal transition (EMT) and were associated with reduced antenna fucosylation [40], were significantly down-regulated with high CDX1 mRNA (Supplemental Table S6).

**Fig. 5:** Other differentially expressed genes. (A) The differentiation marker villin, (B) transcription factors hepatocyte nuclear factor (HNF) 1A, and (C) HNF4A showed increased expression with high CDX1 expression in a set of 16 colorectal cancer cell lines.

DISCUSSION

In two independent sets of colorectal cancer cell lines we observed increased multi-fucosylation in CDX1 high expressing cell lines, next to increased terminal HexNAc epitopes as well as decreased galactosylation and sialylation. Changes in glycosylation have mainly been attributed to alterations in the corresponding glycosyltransferases [17, 41]. Predictions on the glycan phenotype are, however, challenging since several biological effectors can influence not only the expression of glycan-initiating, -elongating, and -degrading enzymes, but also their activity [42]. Guo and Pierce recently reviewed the involvement of transcription factors in glycan expression and reported that they regulate the expression of glycosyltransferase genes in a tissue- and cell-specific manner [42].

Interestingly, we found transcription factors HNF1A and HNF4A as well as fucosyltransferase genes FUT2, 3, 4, 6, 9, as well as the GMDS gene - which is involved in GDP-L-fucose synthesis – and the GDP-fucose transporter SLC35C1 all positively associated with CDX1 expression. In line, cell
lines expressing high mRNA levels of CDX1 showed an N-glycan phenotype with increased multifucosylation, indicative for antenna-fucosylation. Colon cancer cell line HCT116 is mutated in the GMDS gene, leading to low levels of fucosylation, and has been associated with a more aggressive phenotype, whereas restoration of GMDS and therefore enhanced fucosylation resulted in suppressed tumour formation and a less aggressive phenotype [43]. Additionally, HCT116 cells have shown higher levels of cancer stem cells and loss of CDX1 expression, whereas induced expression of CDX1 has likewise resulted in a less invasive and aggressive cancer formation [44]. Taking our findings and the reports from literature one may speculate that CDX1 expression regulates (antenna-) fucosylation in some way, leading to a more differentiated and less invasive phenotype in colorectal cancer cell lines. Supporting this hypothesis and our findings, Lauc et al. identified in a genome-wide association study that fucosyltransferases FUT3, 5 and 6, all involved in antenna-fucosylation, are positively regulated by HNF1A and its downstream target HNF4A, while FUT8, initiating core-fucosylation, is inhibited [32]. At the same time, several groups reported on interaction between CDX1 and the highly related CDX2 with other transcription factors, including HNF1A and HNF4A. Boyd et al. identified CDX2 binding sites on CDX1 and HNF1B, both being positively regulated by CDX2, while HNF1B is needed for activation of HNF4A, HNF1A, and HNF3G [39]. Notably, also direct binding of CDX2 to HNF4A was shown, which positively influenced transcriptional activity [39]. HNF4A was also identified as transcriptional activator for intestinal differentiation and gene expression of HNF4A and its target genes was upregulated in colorectal cancer cell lines with a more epithelial phenotype as compared to cells with mesenchymal phenotype [45]. One target gene of HNF4A is galectin 4 which was described as tumour-suppressor in colon cancer [46, 47], but also pancreatic duct adenocarcinoma [48], and was accordingly positively associated with high CDX1 mRNA expression.

Interestingly, in our investigated cell lines, CDX2 mRNA expression correlated with CDX1 mRNA expression (both high or both low) with exception of cell line RCM1. Consequently, similar observations as for CDX1-associated glycosylation were made for CDX2, but correlations with the N-glycan phenotype and glycosyltransferase expressions were less pronounced for CDX2 as compared to CDX1. In contrast, a direct involvement of CDX2 in the regulation of FUT2, which revealed potential binding sites for CDX1 and CDX2 and is involved in generation of LeY/B antigens, was shown in colon cancer cell lines HT29 and DLD-1 [49]. Expression of CDX2 and thereby FUT2 could be reduced through treatment with epidermal growth factor (EGF)/bFGF [49]. On the other hand, the sialyl Lewis antigen promoting glycosyltransferases ST3GAL1/3/4 and FUT3 were transcriptionally up-regulated by c-Myc [49]. Sialyl Lewis antigens were further affected by FUT1 gene expression and induced FUT1 expression in HT-29 and HepG2 cells resulted in increased expression of Lewis Y/B, whereas sialyl Lewis X antigens were reduced [50]. Fig. 6 summarizes our hypothesis together with reports from literature [24, 32, 49, 51].

We further observed a decrease in galactosylation with high CDX1 mRNA expression. Previous reports showed that CDX1, CDX2, HNF1A, and HNF1B are involved in the transcriptional regulation of B3GALT5, the gene encoding for the enzyme responsible for type-1 chain (-Galβ1–3GlcNAcβ-) expression on glycolipids and glycoproteins with preferences for O-glycans and glycosphingolipid-glycans [52]. B3GALT5 was found to be down-regulated in colon cancer cell lines, but up-regulated in CaCo2-differentiation [52]. However, our mass spectrometric approach does not allow distinction between type-1 and type-2 chains (-Galβ1–4GlcNAc-), but enzyme levels showed an up-regulation of type-1 chain glycosyltransferases B3GALTs as well as B3GNTs according to observations by Isshiki et al., while type-2 chain enzymes were decreased with high CDX1 expression and might explain the decreased galactosylation we observe. Furthermore, many glycosyltransferases involved in glycan elongations act on N-glycans, O-glycans, glycosphingolipid-glycans, and/or others and changes on different glycan classes can be differently regulated. Therefore, it is striking that the N-glycomic data
obtained by mass spectrometry are well in accordance with the transcriptomic data, suggesting that observed changes in, for example, fucosylation and sialylation do not only occur in the here investigated N-glycans, but also in O-glycans and glycosphingolipid-glycans.

Our results further revealed the presence of terminal HexNAc residues to be increased in CDX1 high expressing cells. Glycan motifs that may contribute to this increase include LacdiNAc structures as well as the Sda antigen; the latter was shown to be expressed in normal colon tissues and decreased during colon cancer progression [53]. LacdiNAc, on the other hand, was associated with differentiation of mammary epithelial cells and tumour suppression in neuroblastoma, whereas its expression was increased in human prostate, ovarian, and pancreatic cancers [54]. Furthermore, bisecting GlcNAc containing glycans contribute to this group and corresponding MGAT3 gene expression was up-regulated in CDX1\textsuperscript{high} cells. Several reports describe reduced bisection in several tumours [55] and it was shown that it suppresses N-glycan branching by MGAT5 [56] as well as \(\alpha_{2,3}\)-sialylation [57].

Further, TGFB1 and 2 were significantly reduced with high CDX1 mRNA levels. TGFB1 is known to induce EMT which is promoting malignant phenotypes in epithelial cancer cells and has previously been associated with enhanced levels of high-mannose type glycans, core-fucosylation and corresponding FUT8 expression, as well as increased \(\alpha_{2,6}\)-sialylation and ST6GAL1 gene expression, whereas antennarity of N-glycans, antenna-fucosylation, and bisection were suppressed [40, 58], which is in line with our findings. Inhibition of FUT3 and FUT6 has been shown to affect TGF-\(\beta\) receptor glycosylation, resulting in decreased fucosylation, and FUT3/6-associated sialyl Lewis antigens as well as TGF-\(\beta\)-mediated EMT and invasion in colorectal cancer cell [59]. Furthermore, loss of CDX1 and/or CDX2 was shown to impact TGF-\(\beta\) signalling and tumour invasion in murine APC mutant colon cancer models: Upon loss of CDX1/2, cells were poorly differentiated, invasive, and developed a villous morphology, which was accompanied by loss of the epithelial marker E-cadherin, whereas expression of VIM, TWIST1, ZEB1, and ZEB2 was induced [60]. Interestingly, genes associated with EMT such as ZEB1 and VIM were negatively associated with high CDX1 mRNA expression in our study and also TWIST1 and 2 showed a trend towards lower expression in CDX1\textsuperscript{high} cells, whereas E-cadherin (CDH1) showed a trend towards higher expression in CDX1\textsuperscript{high} cells as compared to CDX1\textsuperscript{low} and is characteristic for differentiated epithelial cells.

Although O-glycosylation was not investigated in this work, the gene microarray data showed differential expression of sialyltransferases involved in O-glycan sialylation. One of them, ST6GALNAC1,
ACKNOWLEDGEMENTS

We thank Gabi W. van Pelt for support with cell counting. This project was supported by the European Union (Seventh Framework Programme HighGlycan project, grant number: 278535).

ASSOCIATED CONTENT

Supporting Tables are available and will be accessible online.

REFERENCES


showed putative CDX-binding sites and transcriptional regulation of the gene by CDX2 has been demonstrated in preneoplastic intestinal metaplasia lesions [61]. However, we are not aware of any reports showing an involvement of CDX1 or CDX2 with sialyltransferases acting on N-glycans which could explain the observed N-glycosylation phenotype.

While the involvement of transcription factors was shown for fucosyltransferases, reports on sialyltransferases and galactosyltransferases involved in the elongation of N-glycans are still lacking and more research on the integrated regulation of glycosyltransferases is needed. Also the role of LacdiNAc structures in differentiation and colorectal cancer needs further investigations.

Our data, in combination with reports from literature suggest that CDX1 (and CDX2) are involved in the regulation of multiple glycosyltransferases. However, CDX-genes may (additionally) influence the expression of fucose-carrying glycoproteins themselves and thereby leading to an N-glycan phenotype with enhanced multi-fucosylation. Taking into account the interaction of the two CDX-proteins with HNF1A, HNF4A and HNF1B and together with the proven role of HNF1A and HNF4A in the regulation of fucosylation, we hypothesize a cooperation of these transcription factors being involved in the expression of FUT genes and thereby increasing fucosylation of glycoproteins with high CDX1/CDX2 expression (summarized in Fig. 6). Certainly, more functional studies are needed to prove the involvement of CDX1 in glycosyltransferase regulations and also its role in the EMT progress should be further elucidated.
N-GLYCOSYLATION ASSOCIATED WITH CDX1 mRNA-EXPRESSION


