Glycoproteomics-based signatures of cancer cell lines
Pirro, M.
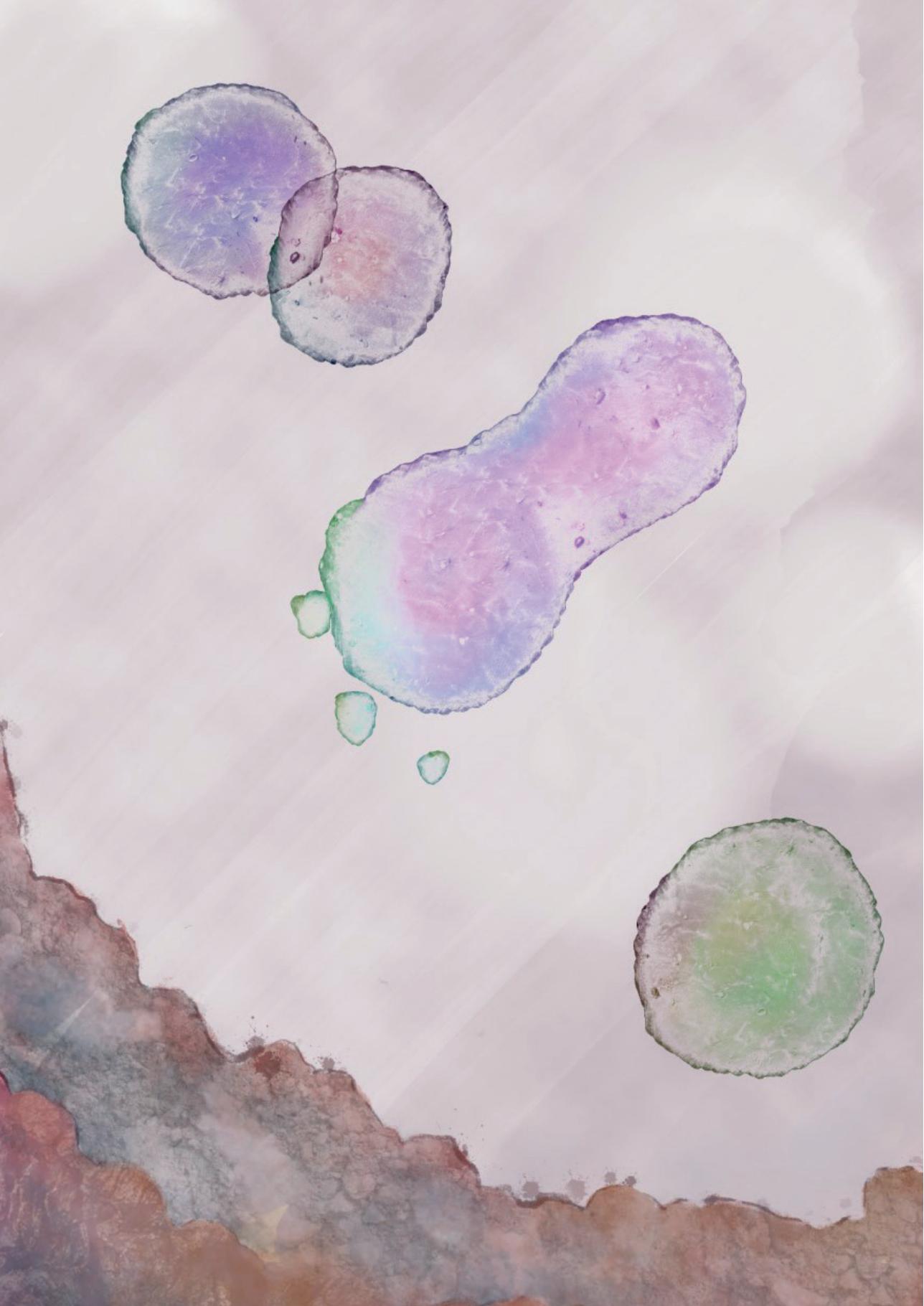
CHAPTER 4

N-GLYCOPROTEINS HAVE A MAJOR ROLE IN MGL BINDING TO COLORECTAL CANCER CELL LINES: ASSOCIATIONS WITH OVERALL PROTEOME DIVERSITY

Martina Pirro[1], Yassene Mohammed[1], Sandra J. van Vliet[2], Yoann Rombouts[3], Agnese Sciacca[1], Arnoud H. de Ru[1], George M. C. Janssen[1], Rayman T. N. Tjokrodirijo[1], Manfred Wuhrer[1], Peter A. van Veelen[1] and Paul J. Hensbergen[1]*

[1] Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands

[2] Amsterdam UMC, Vrije Universiteit Amsterdam, dept. of Molecular Cell Biology and Immunology, Cancer Center Amsterdam, Amsterdam Infection & Immunity Institute, Amsterdam, the Netherlands

[3] Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, Toulouse, France

Abstract

Colorectal cancer (CRC) is the second leading cause of cancer death worldwide due in part to a high proportion of patients diagnosed at advanced stages of the disease. For this reason, many efforts have been made towards new approaches for early detection and prognosis. Cancer-associated aberrant glycosylation, especially the Tn and STn antigens, can be detected using the macrophage galactose-type C-type lectin (MGL/CLEC10A/CD301), which has been shown to be a promising tool for CRC prognosis. We had recently identified the major MGL binding glycoproteins in two high MGL binding CRC cells lines, HCT116 and HT29. Yet, we failed to detect the presence of *O*-linked Tn and STn glycans on most CRC glycoproteins recognized by MGL. We therefore investigated here the impact of *N*-linked and *O*-linked glycans carried by these proteins for the binding to MGL. In addition, we performed quantitative proteomics to study the major differences in proteins involved in glycosylation in these cells. Our results showed that *N*-glycans have a significant, previously underestimated, importance in MGL binding to CRC cell lines. Finally, we highlighted both common and cell-specific processes associated with a high MGL binding phenotype, such as differential levels of enzymes involved in protein glycosylation, and a transcriptional factor (CDX-2) involved in their regulation.

Introduction

Glycosylation is one of the most frequent post-translational modifications on proteins and lipids [16]. In eukaryotic cells, protein glycosylation can be mainly grouped into *N*- and *O*-linked glycans, which are synthetized in the endoplasmic reticulum (ER) and/or Golgi apparatus through the sequential/competitive action of numerous glycosyltransferases and glycosidases, encoded by over 200 genes [1]. The expression, activity and subcellular location of these enzymes dictate the overall glycosylation profile within different cells or tissues [16]. Protein glycosylation has a high impact on a wide range of cell biological processes such as proliferation, adhesion, differentiation, cell-cell interactions and immune responses [3]. Accordingly, changes in glycosylation are observed in various diseases, and aberrant glycosylation is recognized for decades as a hallmark of cancer [1, 5, 12, 15, 16].

Certain glycans that show increased expression in tumor tissues as compared to normal tissues are referred to as Tumor Associated Carbohydrate Antigens (TACAs) [11]. Many TACAs are in fact truncated glycan structures resulting from incomplete glycan biosynthesis. One of the most common process is represented by the dysregulation of the initial steps of *O*-glycan biosynthesis mediated by polypeptide α-N-acetylgalactosaminyl-transferases (ppGalNAcTs) that transfer UDP-GalNAc to Ser/Thr residues in a polypeptide, forming Tn antigen (GalNAcα1-Ser/Thr) [5, 108]. Humans have 20 genes encoding ppGalNAcTs, and an overexpression, a higher activity and/or a modified subcellular location of these transferases can lead to higher Tn antigen expression, frequently decorating mucins on epithelial tumors [109]. The high levels of GalNAcα2,6-sialyltransferase (ST6GalNAc-I) in tumor tissues gives rise to the Sialyl-Tn epitope (STn, NeuAcα2,6-GalNAcα1-Ser/Thr) and prevents further *O*-glycan elongation. The expression of Tn and STn can also be enhanced by defects in the activity of T-synthase or mutations in its chaperone (Cosmc), which together mediate the synthesis of core 1 *O*-glycans (T antigen, Galβ1,3-GalNAcα1-O-Ser/Thr) [40]. Conversely, some tumors may induce the activation of certain glycosyltransferases involved in the synthesis of neo-antigens decorating glycan termini. For example, the overexpression of fucosyltransferases (Fuc-T 1/3) increases terminal fucosylation and consequently the expression of specific Lewis-blood group antigens (Le^x, Le^y). Enhanced expression of sialyltransferases modifies the latter antigens into the sialylated species SLe^x, and SLe^y. Higher activity of the mannoside acetyl-glucosaminyltransferase 5 (MGAT5) increases *N*-glycan branching, while dysregulation

of the β4-N-acetylgalactosaminyltransferase 3 (B4GALNT3) may lead to the expression of terminal LacdiNAc (GalNAcβ1,4GlcNAcβ1-) on both *N*-linked[13] and *O*-linked glycans [14]. During tumor development, the appearance of TACAs can be detected by cells of the immune system, which can either enhance or dampen the immune response depending on the nature of the interaction [5]. The recognition of TACAs is mediated by glycan binding proteins, i.e. lectins, whose binding specificities depend on their carbohydrate recognition domains (CRD) [109]. Within the large family of lectins, the C-type macrophage galactose-type lectin (MGL/CLEC10/CD301) exclusively recognizes, in a Ca^{2+} dependent manner, terminal GalNAc residues (α- or β-linked) found in the (S)Tn antigens and LacdiNAc epitopes. MGL is primarily expressed by tumor-associated or tolerogenic dendritic cells (DCs) and macrophages. In these cells, ligand binding to MGL activates molecular pathways, resulting in dampening T cell immunity [39]. Besides the primary binding site involved in the recognition of the terminal GalNAc residue, a secondary binding site was found in MGL, which is involved in the binding to the peptide backbone carrying the glycosylated epitope (glycotope) [36]. This suggests that ligand-specific binding may affect different signaling pathways and biological processes.

Since increased expression of MGL ligands is used to discriminate healthy from colorectal cancer (CRC) tissues, and is an independent prognostic marker for CRC stage III patients with lower disease-free survival [45], we recently investigated the identity of cell surface MGL binding proteins on CRC cells lines [110]. However, due to the relatively low number of identified glycopeptides carrying (S)Tn antigens or LacdiNAc epitope, the underlying differences that could explain the differential MGL binders expression remained unclear. Here our aim was to get a broader understanding of the mechanism behind different MGL binding phenotypes of these cell lines. For this purpose, we studied the (glyco-)proteins of CRC cell lines using a combination of lectin staining, MGL pull-downs in the absence and presence of PNGase F treatment and overall comparative quantitative proteomics using Tandem Mass Tag (TMT) labeling.

Materials and Methods

Cell lines culture and lysis

HCT116 and HT29 were provided by the Department of Surgery of the Leiden University Medical Center (Leiden, the Netherlands), whereas LS174T was obtained from the

Amsterdam UMC (Amsterdam, the Netherlands). Cell line authentication was performed using short-tandem repeat (STR) profiling at the forensic laboratory for DNA-research (ISO 17025) and all cell lines matched for 100% with the known profile [90]. All cell lines were cultivated in RPMI-1640 medium containing L-glutamine, 10% fetal bovine serum (FBS) (Invitrogen) and streptomycin/penicillin (Sigma-Aldrich) at 5% CO_2 and 37°C. Cells were maintained till approximately 80% confluence under sterile conditions. For harvesting, cells were washed twice with 1x PBS and incubated for approximately 5 minutes in 1x trypsin/EDTA solution in 1x PBS, whose activity was inhibited by the addition of serum containing medium following visual cell detachment. Cells were subsequently harvested and counted using the CountessTM Automated Cell Counter (Invitrogen, Paisley, UK). Aliquots of 2×10^7 or 4×10^6 cells were washed with 1x PBS and centrifuged at 1500 rpm to obtain cell pellets. Cell pellets were stored at -20°C until use for MGL pull downs or TMT labelling, respectively.

Lectins and Antibodies

Chimeric MGL-Fc was prepared as described previously [38]. For western blot staining the following antibody were used: c-Met mouse mAb (3D4, Thermo Fisher Scientific, concentration 1.5 ug/ml), phospho-Met rabbit mAb (Tyr1234/1235) (D26, Cell Signaling Scientific, dilution 1:1000). For detection of c-Met, a secondary antibody goat anti-mouse immunoglobulins/HRP (Agilent Dako, dilution 1:2000) was used, while for phospho-Met, swine anti-rabbit Immunoglobulins/HRP (Agilent Dako dilution 1:2000). For lectin blot, MGL-Fc was targeted with a secondary antibody peroxidase conjugated goat anti-human IgG-Fcγ (Jackson Immuno Research, dilution 1:1500).

Pull-down assay and PNGase F treatment

Protein extracts were obtained as described before [91]: cell pellets were incubated for 20 min on ice in lysis buffer (10 mM triethanolamine pH 8.2, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$ and 1% (volume/volume) Triton X-100, containing EDTA-free protease inhibitor (Roche Diagnostics)). Protein quantification was performed using the BCA assay (BCA Protein Assay Kit, Pierce™), following the manufacturer's instructions. MGL ligands were pulled down from 1 mg of protein extracts with or without prior treatment with PNGase F PRIME (N-Zyme Scientifics, concentration 0.1 ug/ml) overnight at 37°C. Two µg of chimeric MGL-Fc, coupled

to 50 µl Dynabeads protein G (Invitrogen) were used as previously described [91]. Following washing, the elution of specific ligands was performed using 100 mM EDTA.

SDS-PAGE and western/lectin blot

Protein extracts and MGL pull-down samples were separated by SDS-PAGE (4–15% Mini-PROTEAN® TGX Stain-Free™ Protein Gels, Biorad) and transferred to a PVDF membrane (Trans-Blot Turbo Mini PVDF Transfer Packs, Biorad). 5 % bovine serum albumin (Sigma-Aldrich) in 0.1% phosphate-buffered saline with Tween-20 (Sigma-Aldrich) (PBS-T) was used to block the blots for 1 h. Immunoblotting was performed with specific antibodies in BSA 1% PBS-T, followed by peroxidase-conjugated secondary antibodies. For lectin blots, 5 % BSA in TSM buffer (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM $CaCl_2$ and 2 mM $MgCl_2$) was used as blocking buffer, and incubation with the lectin was performed in BSA 1 % TSM. The following washes and incubation with peroxidase-conjugated secondary antibodies were performed in TSM with 0.1 % Tween-20. Immunodetection was done by enhanced chemiluminescence (ECL) using Clarity Western ECL substrate (Bio-Rad) and an Amersham Imager 600 (Cytiva, Marlborough, United States).

SDS-PAGE and NanoLC-MS/MS analysis

For sample clean-up, a short SDS-PAGE run (NuPAGE™ 4-12% Bis-Tris Protein Gels, Thermo Fisher Scientific) of the samples obtained from the MGL pull-downs after PNGase F treatment was performed. Gels were stained with SimplyBlue™ Safe Stain (Invitrogen) for 1 hour at room temperature (RT) and washed with distilled water for 3 hours. Bands corresponding to the whole lane were cut from the gel, and the proteins were then subjected to reduction with dithiothreitol (10 mM), alkylation with iodoacetamide (50 mM) and in-gel trypsin digestion with trypsin (Worthington Enzymes), using a Proteineer DP digestion robot (Bruker) [91].

Tryptic peptides were separated by online C18 nano-HPLC MS/MS with an Easy nLC 1000 gradient HPLC system (Thermo, Bremen, Germany) coupled to a Orbitrap Fusion Lumos mass spectrometer (Thermo), as previously described [91]. Briefly, fractions were loaded onto a homemade precolumn and eluted via a homemade analytical nano-HPLC column (for 20 min), followed by electrospray injection into the Mass Spectrometer. The MS was operated in data-dependent MS/MS (top-10 mode) with a normalized collision energy of 32% and recording of

the MS^2 spectrum in the Orbitrap (parameters specified in [91]). For protein identification, raw data was converted to mzXML using Proteowizard. Peptide and protein identification as well as the after statistical validation were performed in Trans Proteomics Pipeline version 5.1.0 using included software pipeline: X! Tandem Jackhammer TPP (2013.06.15.1-LabKey, Insilicos, ISB) search engine, PeptideProphet and ProteinProphet. The parameters were set as follows: precursor mass error of 10 ppm, fragment mass error of 0.04 Da, carbamidomethyl (Cys) and oxidation (Met) as fixed and variable modifications respectively. All results were filtered for FDR threshold of 1% as well as a minimum of two per protein. Data extraction and table generation was done using R version 3.4.4.

Quantitative proteomics using TMT labelling

Cell lysis, digestion and TMT labelling was performed as described [111].Cellular extract from 4 x 10^6 HCT116, HT29 and LS174T cells were prepared in triplicate by a 4 min incubation at 95°C in SDS lysis buffer (5% SDS, 100 mM Tris-HCl pH 7.6). Protein concentration was determined by BCA assay. 100 μg of protein was used for subsequent reduction with 5 mM TCEP, alkylation with 15 mM iodoacetamide and quenching with 10 mM DTT. Protein lysates were cleaned by methanol-chloroform precipitation. The resulting protein pellets were resuspended in 40 mM Hepes (pH 8.4) and incubated with 10 μg trypsin O/N at 37°C. Peptide concentration was measured with BCA assay. 10 μg of each of the 9 peptide preparations was dissolved in 25 μl of 40 mM Hepes (pH 8.4) and incubated with 40 μg of one of the 9 amino reactive TMT10plex Label Reagents (126 to 130, Thermo Scientific, Lot #UG282327) for 1 h at room temperature. Excess TMT label was quenched by incubation with 6 μl 5 % hydroxylamine for 15 min at room temperature. The 9 labelled peptide samples were then mixed, freeze-dried and measured using MultiNotch MS^3 procedure [25].

TMT-labelled peptides were dissolved in 0.1% formic acid and subsequently analyzed by on-line C18 nanoHPLC MS/MS with a system consisting of an Easy nLC 1200 gradient HPLC system (Thermo, Bremen, Germany), and an Orbitrap Fusion LUMOS mass spectrometer (Thermo). Fractions were injected onto a homemade precolumn (100 μm × 15 mm; Reprosil-Pur C18-AQ 3 μm, Dr. Maisch, Ammerbuch, Germany) and eluted via a homemade analytical nano-HPLC column (50 cm × 75 μm; Reprosil-Pur C18-AQ 1.9 μm). The analytical column temperature was maintained at 50°C with a Sonation PRSO-V2 column oven. The gradient

was run from 5% to 30% solvent B (20/80/0.1 water/acetonitrile/formic acid (FA) v/v) in 240 min. The nano-HPLC column was drawn to a tip of ~5 µm and acted as the electrospray needle of the MS source. The LUMOS mass spectrometer was set to use the MultiNotch MS[3]-based TMT method [25]. The MS spectrum was recorded in the Orbitrap (resolution 120,000; m/z range 400–1500; automatic gain control (AGC) target 2×10^5; maximum injection time 50 ms). Dynamic exclusion was after n=1 with an exclusion duration of 60 s with a mass tolerance of 10 ppm. Charge states 2-4 were included Precursors for MS[2]/MS[3] analysis were selected using a TopSpeed of 3 sec. MS[2] analysis consisted of collision-induced dissociation (quadrupole ion trap analysis; AGC 1×10^4; normalized collision energy (NCE) 35; maximum injection time 50 ms). The isolation window for MS/MS was 0.7 Da. Following acquisition of each MS[2] spectrum, the MultiNotch MS[3] spectrum was recorded using an isolation window for MS[3] of 2 Da. MS[3] precursors were fragmented by high energy collision-induced dissociation (HCD) and analyzed using the Orbitrap (NCE 65; AGC 1×10^5; maximum injection time 105 ms, resolution 60,000). In a post-analysis process, raw data were first converted to peak lists using Proteome Discoverer version 2.4 (Thermo Electron), and then submitted to the Uniprot Homo sapiens minimal database (20205 entries), using Mascot v. 2.2.04 (www.matrixscience.com) for protein identification. Mascot searches were done with 10 ppm and 0.02 Da deviation for precursor and fragment mass, respectively, and trypsin enzyme was specified. Methionine oxidation and Acetyl (Protein N-term) were set as variable modifications and Carbamidomethyl (C) was set as a static modification. Peptides with an FDR<1% were accepted. The TMT ratio from the MultiNotch MS[3] spectra were used for quantification using Proteome Discoverer 2.4.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [64] partner repository with the dataset identifier PXD020344.

Results and discussion

N-glycans are important for MGL binding in CRC cell lines

We recently demonstrated differential binding of the C-type lectin MGL to the colorectal cancer cell lines HCT116, HT29 and LS174T [110]. While high MGL binding was observed to

HCT116 and HT29 cells, binding to LS174T cells was negligible. The comprehensive characterization of *N*- [53] and *O*-glycans [54] expressed by the three CRC cell models used in our study, did not provide an explanation for the difference in MGL binding to these cell lines. We identified the major cell surface proteins binding to MGL in HCT116 and HT29 cells and, for some of these, found a glycopeptide with the MGL specific glycotope (e.g. LacdiNAc (on a *N*-glycan) and Tn antigen), but for many the glycotope remained elusive [110]. In order to gain more insight into the relative contribution of *N*- and *O*-glycans to MGL binding to glycoproteins of CRC cells, we performed MGL pull-down experiments in combination with PNGase F digestion and lectin (MGL) blots. For this purpose, two different types of experiments were performed. In the first experiment, PNGase F treatment was performed **A**fter the MGL pull-down (A, Figure 1A), in the second **B**efore the capturing with MGL (B, Figure 1A). As expected, MGL pull-downs with HCT116 and HT29 cells showed a few intense bands corresponding to MGL ligands in the high molecular weight range, which were absent in LS174T cells (Figure 1A, PNGase F untreated (-) samples) in all three biological replicates (Figure 1A and Supplemental Figure S1). With HCT116 cells, these bands disappeared using PNGase F treatment either after (A) or before (B) the MGL pull-down, indicating that the MGL binding is mainly due to *N*-glycans. On the other hand, PNGase F treatment could not fully abrogate the binding of the major MGL binding proteins from HT29 cells, even though some major bands disappeared when the cell lysate was treated with PNGase F before the MGL pull-down experiments (B, Figure 1A). These data suggest that *N*-glycans have a substantial role in the binding of proteins to MGL in HCT116 and HT29.
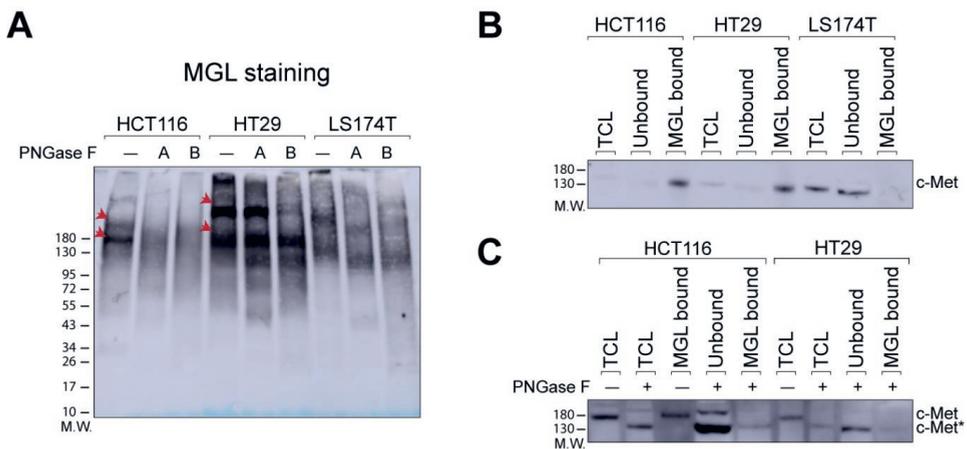
*Figure 1. Release of N-glycans reduces the MGL-binding of proteins from CRC cell lines. A) MGL pull-downs were performed without (-) or with PNGase F treatment (A and B) and captured proteins were subsequently analyzed by MGL-lectin blot. PNGase F treatment was performed either **A**fter (A) or **B**efore (B) performing the MGL pull-down. The lectin blot is a representative of 3 biological replicates (see Figure S1). Red arrows indicate major stained bands. B) MGL-binding of c-Met in HCT116, HT29 and LS174T cells. MGL pull-downs were performed and bound and unbound proteins were analyzed by western blot using a c-Met antibody. TCL: total cell lysate. C) Influence of PNGase F treatment on the binding of c-Met to MGL in HCT116 and HT29 cells. MGL pull-downs were performed with or without prior treatment of the total cell lysate (TCL) with PNGase F. Samples were analyzed by western blot using a c-Met antibody. c-Met* represents the protein with released N-glycans. M.W.: Molecular weight.*

One of the major cell surface MGL binding proteins in both HT29 and HCT116 cell lines is the receptor tyrosine kinase c-Met [110]. However, in our previous study we could not identify the MGL specific glycotope on the c-Met glycopeptides. Therefore, we also studied here the influence of PNGase F treatment on the binding of c-Met to MGL.

First of all, using western blot, we could confirm the specific MGL binding of c-Met from HT29 and HCT116 cells (Figure 1B). Importantly, the overall c-Met level in LS174T cells was similar to that in HCT116 and HT29 cells (Suppl. Figure S2), but in MGL pull-down experiments with LS174T cells, c-Met was found in the unbound fraction (Figure 1B), confirming that in these cells the glycosylation of c-Met is different. Next, we tested the influence of PNGase F treatment on the MGL pull-down of c-Met from HT29 and HCT116 cells. In both cell lines, *N*-glycan release prior to MGL pull-down strongly reduced the MGL-binding of c-Met, which is evident from the high amount of unbound c-Met compared to the bound fraction under this condition (Figure 1C). The fraction of unbound c-Met is negligible without PNGase F treatment (Figure 1B). A concomitant shift in the apparent molecular weight was observed. These results demonstrate that c-Met binding to MGL can, to a large extent, be attributed to N-glycans in both HCT116 and HT29 cells.

c-Met is a tyrosine kinase receptor, whose activation is mediated by dimerization following binding to the hepatocyte growth factor (HGF) [112]. This results in phosphorylation of intracellular tyrosines and subsequent activation of downstream pathways involved in cell survival proliferation, migration and invasion [112]. Since HGF-independent hyperactivation of c-Met is involved in the carcinogenesis of CRC, as well as in many other human malignancies [112], we investigated the activation of the receptor in the three CRC cell lines by western blot. This showed c-Met phosphorylation in the two high MGL binding cell lines, but not in the

low MGL binding cell line (Supplemental figure S2B). However, the role of the differential glycosylation in c-Met signaling remains to be determined.

To obtain a broader understanding on the role of *N*- and *O*-glycans in the binding of proteins to MGL, we also performed MGL pull-down experiments using HCT116 and HT29 cells after *N*-glycan release, followed by LC-MS/MS analysis of the bound proteins. We focused our qualitative comparison on the top 20 MGL binding proteins that we previously identified [110]. We assigned a protein as a MGL binder when it was identified in two out of three biological replicates. In Figure 2 these are colored in green, while non MGL binders after *N*-glycan release are colored in red. In line with the data shown in Figure 1C, c-Met lost its capability to bind to MGL after *N*-glycan release in HCT116 cells (Figure 2). On the other hand, binding of c-Met to MGL was still observed after PNGase F treatment, in HT29 cells, even though the number of peptides observed was low (Suppl. Table S1). Of note, SORL1, PTK7 and GOLM1, which were previously shown to carry a LacdiNAc epitope on an *N*-glycan in HT29 cells [110], lost the binding to MGL after PNGase F release. On the contrary, proteins such as integrins (ITGB1 and ITGA3) and TFRC still bind to MGL despite the *N*-glycan release. Overall, the MGL binding of more than half of major MGL binding proteins was affected by *N*-glycan release.

Altogether, our results demonstrate a noteworthy contribution of *N*-glycans to the MGL binding in CRC cell lines. This can most probably be explained by the presence of the LacdiNAc epitope.

UniProt Entry	Protein names	Gene names	HCT116		HT29	
			PNGase F -	PNGase F +	PNGase F -	PNGase F +
Q92896	Golgi apparatus protein 1 *	GLG1	green	green	green	green
P08581	Hepatocyte growth factor receptor	MET	green	red	green	green
Q92673	Sortilin-related receptor *	SORL1	green	green	green	red
P05556	Integrin beta-1	ITGB1	green	green	green	green
Q13308	Inactive tyrosine-protein kinase 7 *	PTK7	green	green	green	red
Q9P2B2	Prostaglandin F2 receptor negative regulator	PTGFRN	green	green	green	green
O75054	Immunoglobulin superfamily member 3	IGSF3	green		green	green
P26006	Integrin alpha-3 *	ITGA3	green	green	green	green
P10586	Receptor-type tyrosine-protein phosphatase F	PTPRF		green	green	green
Q6ZRP7	Sulfhydryl oxidase 2	QSOX2	green	red	green	green
Q8NBJ4	Golgi membrane protein 1 *	GOLM1	green	green	green	red
P02786	Transferrin receptor protein 1	TFRC	green	green	green	green
P20023	Complement receptor type 2	CR2	green		green	green
P05023	Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	green	green	green	green
P51572	B-cell receptor-associated protein 31	BCAP31			green	red
O75369	Filamin-B	FLNB			green	green
Q15262	Receptor-type tyrosine-protein phosphatase kappa	PTPRK	green		green	green
P08195	4F2 cell-surface antigen heavy chain	SLC3A2	green	green	green	green
P15529	Membrane cofactor protein	CD46			green	red
P11279	Lysosome-associated membrane glycoprotein 1	LAMP1			green	green

Figure 2. Release of N-glycans reduces the MGL-binding of proteins from CRC cell lines. MGL pull-downs were performed after N-glycan release of the total protein extracts of HCT116 and HT29 cells. Bound proteins were analyzed by SDS-PAGE, trypsin digestion and LC-MS/MS. The top 20 MGL-binding proteins that we previously identified [110] are shown (in green, PNGase F - (white indicates not identified in that cell line)). After N-glycans release (PNGase F +) MGL binders maintained (green) or lost (red) the ability to bind to MGL. *: proteins with previous [110] identification of glycopeptides with a LacdiNAc epitope on an N-glycan. See table S1 for further details.

Quantitative proteomics provides insights into glycosylation mechanisms involved in high MGL binding

We next investigated the potential mechanism(s) responsible for the differential MGL binding to CRC cell lines, such as i) variable expression of proteins carrying the MGL epitope ii) alteration in _N-/O_-glycosylation pathways (e.g. levels of glycosyltransferases or transcription factors). Therefore, we performed comparative bottom-up quantitative proteomics. For this purpose, protein extracts from HCT116, HT29 and LS174T cells (three biological replicates for each) were digested with trypsin, isotopically labelled (9-plex TMT-labelling (Tandem Mass Tags)), mixed, fractionated and analyzed by LC-MS/MS. Overall, this approach resulted in the identification of 6126 proteins. For 5141 of these, quantitative data was obtained (Table S2). Binary comparisons of the cell lines showed between 175 and 303 proteins to be differentially abundant (Table S2, Figure S3). Among these, we observed several proteins that have previously been found to be drivers of initiation and progression of cancer. For example, we found higher levels of TP53 in HT29 cells compared to the other two, in line with the reported overexpression of TP53 in this cell line (Table S2) [52].

We next focused on the differences that might explain the differential MGL binding between these cell lines. First of all, we checked the relative abundances of the top-20 cell surface MGL binding proteins that we previously identified [110] (Table S3). In line with the experiments shown above (Figure S2A), the overall proteomics data showed similar levels of c-Met in all three cell lines, irrespective of the MGL binding to these cells (Table S3). This was also true for most of the other major MGL binding proteins that we previously identified. Hence, the MGL binding of specific proteins in HT29 and HCT116 cells compared to LS174T cannot solely be explained by the higher levels of these proteins in these cells, even though some differences were observed. For example, HCT116 and HT29 cells have higher levels of the MGL binder ITGA3 in comparison to LS174T.

Next, we compared the levels of proteins involved in *N*- and *O*-glycosylation in our cells lines. We limited the analysis to 245 glycosylation-related proteins as annotated in Gene Ontology (filtering for the term "protein glycosylation"). Overall, we covered 30% of these in our dataset (Figure 3), probably indicating that many others are present, but at very low levels.

The initial step of mucin type O-glycosylation is mediated by a family of 20 polypeptide N-acetylgalactosaminyltransferases (GALNTs). Seven of these (GALNT 1-5, 7 and 12) were found in our dataset, four of which could be reliably quantified. While the levels of GALNT2, 4 and 7 were comparable in the three cell lines, GALNT3 was found at higher levels in HT29 compared to HCT116 and LS174T (Figure 3). This is in line with recent evidence of higher GALNT3 mRNA expression in BRAFV600E mutated cell lines, as HT29, which could be responsible for higher Tn expression in this cell line [83], and the relatively high MGL binding, even after *N*-glycan release (Figure 1A). The elongation of Tn is mediated by T-synthase (C1GALT1) and its chaperone Cosmc (C1GALT1C1). These two proteins were not observed in our dataset. However, their mRNA expression was not reported to be different in BRAFV600E CRC cell lines in the study mentioned above [83]. I-branching (GlcNAcβ1-6Gal-R) of mucin type *O*-glycans is mediated by β1,6-N-acetylglucosamine transferase 3 (GCNT3). The quantitative data we obtained for this enzyme revealed higher levels in HT29 and LS174T, in comparison to HCT116 (Figure 3), in accordance with transcriptomic and glycomic data from literature [52, 54]. Another characteristic of HCT116 is the overall low level of fucosylation, associated with a more aggressive phenotype [53, 54]. In HCT116, this feature can partially be explained by a deletion of 142 amino acids of the GDP-mannose-4,6-dehydratase (GMDS) involved in GDP-L-fucose synthesis, which may lead to misfolding and degradation of the enzyme [113]. Indeed, GMDS protein levels are much higher in HT29 and LS174T in comparison to HCT116 (Table S2).

In the *N*-glycosylation pathway, most of the identified and quantified proteins involved have comparable levels in the three cell lines (Figure 3). An exception is represented by the Dolichyl-phosphate β-glucosyltransferase (ALG5) and UDP-N-acetylglucosamine-dolichyl-phosphate N-acetylglucosaminephosphotransferase (DPAGT1), both involved in the initial steps of oligosaccharide biosynthesis linked to the dolichol molecule. In fact, the two enzymes were found at higher levels in HT29 and LS174T compared to HCT116 (Figure 3).
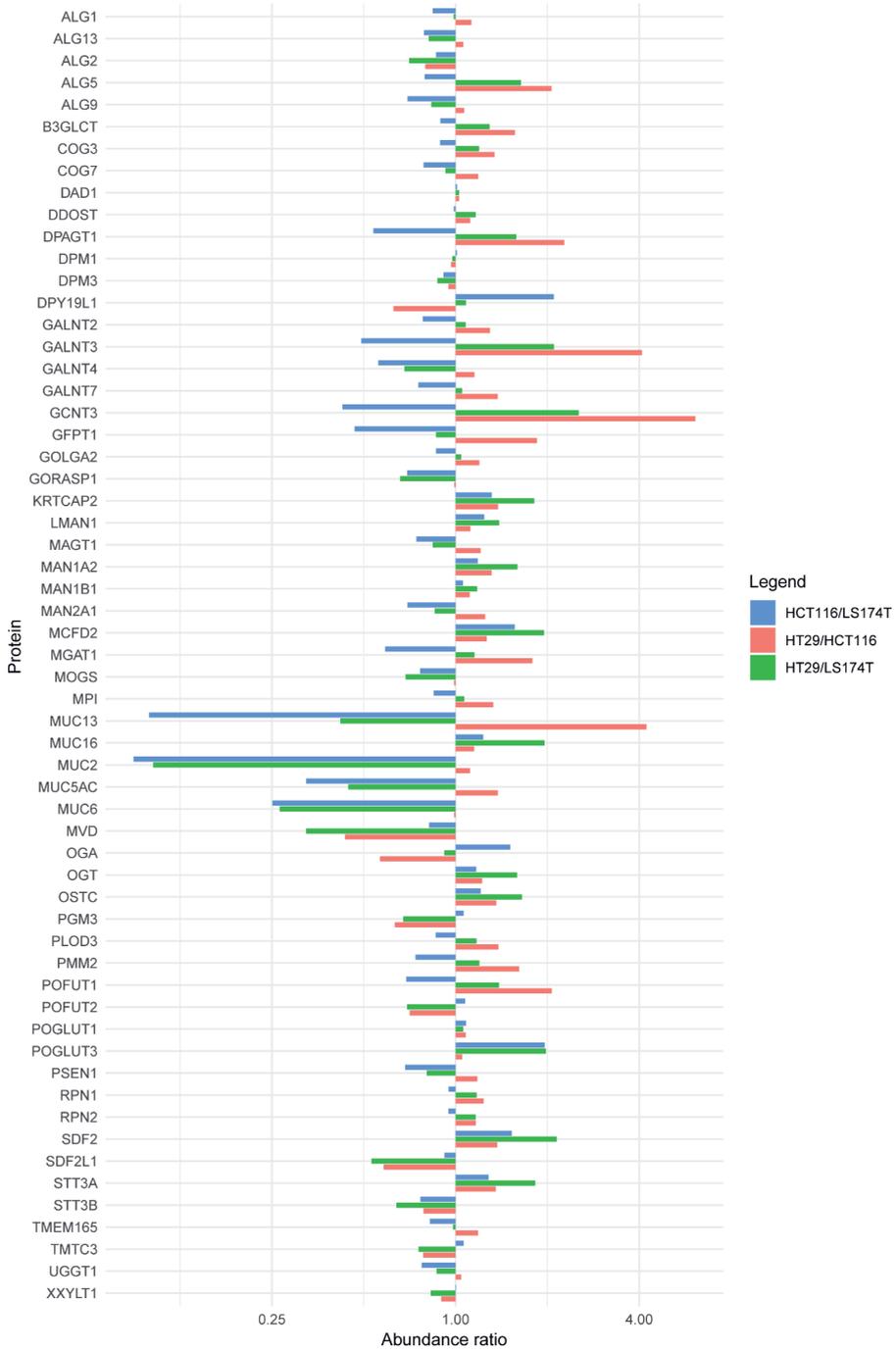
Figure 3. Quantitative proteomics analysis of proteins involved in glycosylation in CRC cell lines. The y-axis shows proteins involved in the glycosylation machinery covered in our study. The x-axis shows

the ratio of protein abundance in HCT116 versus LS174T (blue), HT29 versus HCT116 (pink) and HT29 versus LS174T (green).

Altogether, we observed several differences in enzymes involved in protein glycosylation but they did not provide a clear picture with common differences between high MGL binding cells on the one hand (HCT116 and HT29) and the low MGL binding cell line (LS174T) on the other hand. Obviously, this could be due to the fact that many proteins involved in protein glycosylation are expressed at low levels and could not be covered within the 6000+ proteins identified here. For example, given the considerable contribution of *N*-glycans for binding to MGL as shown here, and the fact that we previously identified the LacdiNAc epitope on peptides from some of the MGL binding proteins in the CRC cell lines, it would have been interesting to look at β4-N-acetylgalactosaminyltransferase 3 (B4GALNT3). On the other hand, previous experiments with the cell lines used in our study showed that the mRNA levels of B4GALNT3 were similar in these cells [52].

In addition to enzymes directly involved in protein glycosylation, the selection of proteins also included mucins. Five of these were found in our data (MUC2, MUC5AC, MUC6, MUC13 and MUC16) and, with the exception of MUC16, all were found at lower levels in HT29 and HCT116 compared to LS174T (Figure 3). Mucins have been characterized as MGL binding proteins in other cancer cells [42, 43] and Mucin 1 (MUC1) expressed on colorectal cancer tissues can be recognized by MGL [114]. However, none of these mucins was identified as MGL binder in our previous experiments [110] and our current experiments demonstrate an inverse correlation between the expression of these proteins and the degree of MGL binding.

Interestingly, when we looked more in general for proteins that were different between the high binding cell lines (HT29 and HCT116) and the low MGL binding cell line (LS174T), we observed the cluster of mucins (Figure S4) in close proximity to the transcription factor CDX-2. Together with CDX-1, this protein acts as a transcription factor that in cooperation with HNF4A and 1A, is involved in the regulation of multiple intestinal specific genes [115], as well as fucosyltransferases [116] and α2,6-sialyltransferase [117]. CDX-1 is very similar to CDX-2, and these proteins share several tryptic peptides, but we did not observe a unique CDX-1 peptide in our dataset. In line with our proteomics data, higher CDX-2 (and CDX-1) mRNA levels were found in highly differentiated cells together with multi-fucosylated N- and O-glycans, e.g. LS174T [54, 118] and mucin 2 is upregulated by CDX-2 [119]. Hence, our data indicate that downstream targets of CDX-2 (and potentially CDX-1), especially the genes

encoding proteins involved in glycosylation, could play a role in the differential binding of MGL but this warrants further investigation.

Conclusions

Notwithstanding the importance of the Tn antigen for the binding to MGL in CRC [45, 46, 83], our current study demonstrates a hitherto unrecognized notable contribution of protein *N*-glycosylation for the binding of MGL to glycoproteins of CRC cell lines. This should be taken into account in future investigations aiming to understand the responses in immune cells, but also cancer cells, following interaction of MGL with its ligands. In fact, a variety of MGL mediated responses have been described. On the one hand, activation of MGL on DC's by synthetic glycopeptides carrying Tn structures (e.g. from CD45, CD43 or MUC1), showed an immunosuppressive response in cancer [82]. On the other hand, the MGL binding to Tn-bearing CD45 on T cell leukemia cells induced cell death [39]. Moreover, MGL signal transmission and outcome is dependent on the type of glycan structure [35] as well as the peptide backbone binding to the secondary binding site in the MGL CRD [36]. For this reason, we believe that the identification of MGL ligands will help to understand whether MGL binding to cancer cells induce receptor-specific signaling thereby promoting or reducing cell survival.

With the identification of more than 6000 proteins through our proteomics study, we gained more insights into the MGL binding phenotype of HCT116 and HT29 compared to LS174T. First of all, we found the major MGL binding proteins from HT29 and HCT116 cells were found at comparable levels in LS174T cells. Moreover, this analysis ruled out the major role of mucins as MGL binders in CRC cell lines, in contrast with many MGL investigations on CRC tissues [46] and other cancer types [42]. Even though the higher levels of GALNT3 in HT29 could partly explain the high MGL binding to this cell line, the involvement of other glycosylation enzymes in the specific glycotope on the MGL ligands in HT29 and HCT116 warrants further investigation. Our study indicates that downstream targets of CDX-2 could be good candidates.

Acknowledgments

4

Supporting Information

Supporting Information is available free of charge via https://www.mdpi.com/1422-0067/21/15/5522#supplementary
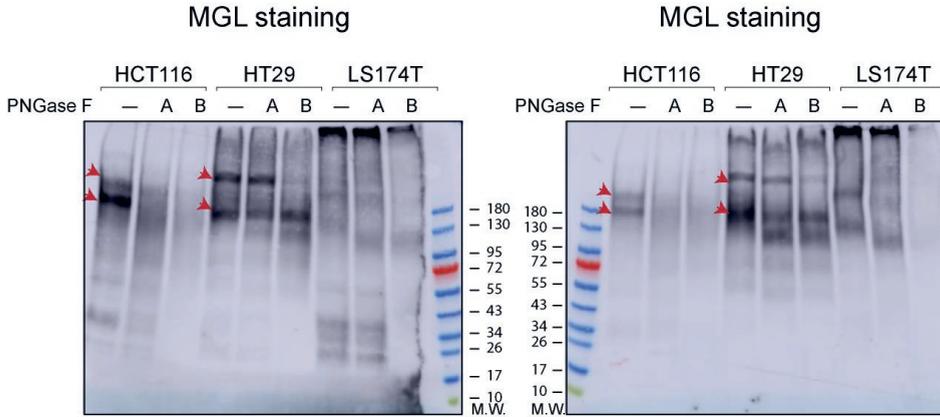
Supplemental figure 1: MGL staining of MGL-binding proteins from HCT116, HT29 and LS174T following N-glycan release. *MGL lectin blot of MGL pull-downs, performed without (-) or with PNGase F treatment (A, B). The N-glycans were released either **A**fter (A) or **B**efore (B) performing MGL pull-downs. The lectin blots represent two out of three biological replicates (The first replicate is shown in figure 1). Red arrows indicate major stained bands. M.W.: Molecular weight.*
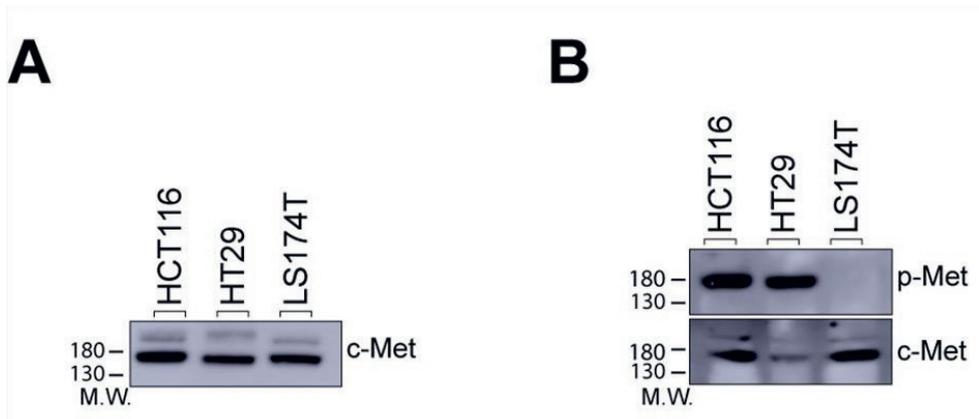
Supplemental figure 2: c-Met levels and activation in HCT116, HT29 and LS174T. A) *Western blot analysis of c-Met protein levels in the three CRC cell lines. **B)** Western blot analysis of c-Met activation (Tyrosine 1234/1235 phosphorylation, p-Met) in the three CRC cell lines under baseline culture conditions. M.W.: Molecular weight.*
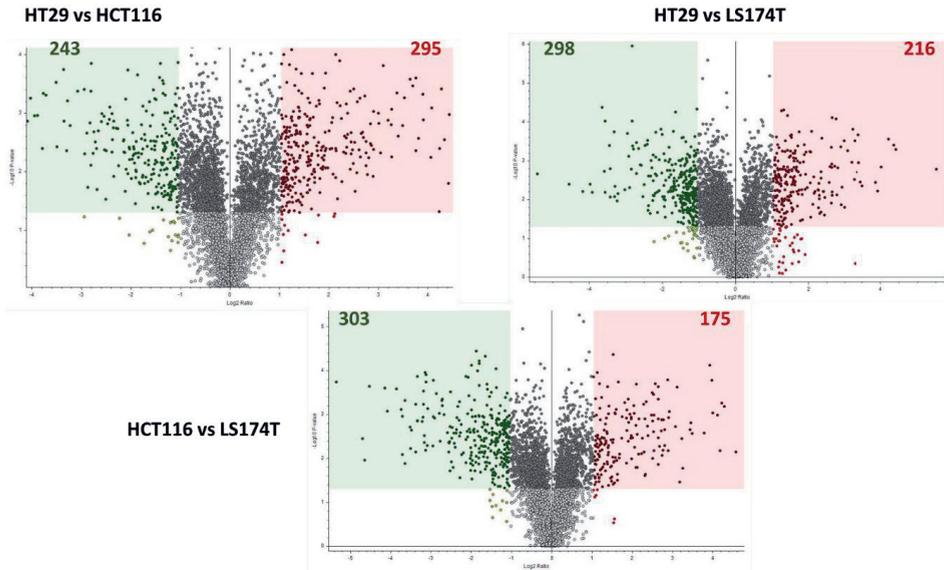
Supplemental figure 3: Volcano plots of binary comparisons of protein abundances in the three CRC cell lines (HCT116, HT29 and LS174T) based on quantitative proteomics analysis. *Proteins at higher abundance (log$_2$ ratio>1, p<0.05) are shown in red, the ones at lower abundance (log$_2$ ratio<-1, p<0.05) in green.*

4
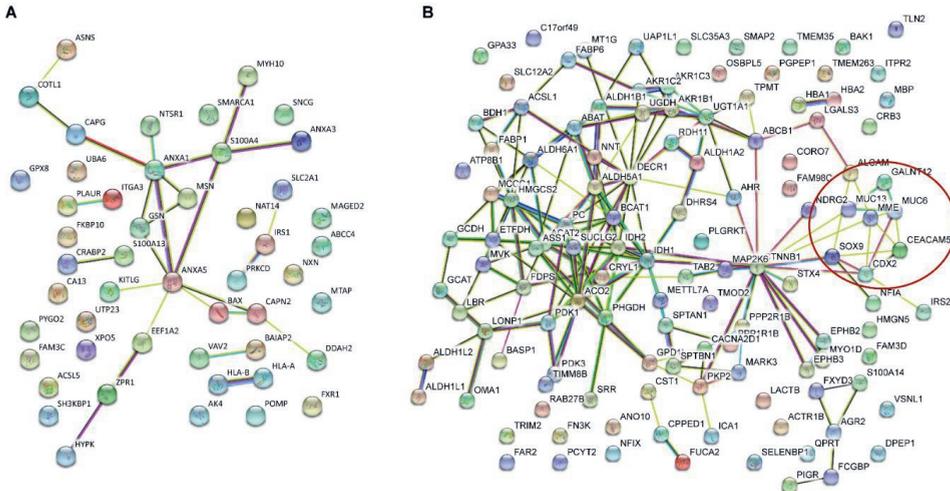
Supplemental figure 4: Proteins observed at different levels in the high MGL binding cell lines (HT29 and HCT116) compared to the low MGL-binding cell line. *Functional protein association networks of proteins higher (A) and lower (B) in the high MGL binding cell lines analyzed by STRING (version 11.0). The cluster of mucins and CDX-2 is highlighted by the red ellipse. The empty nodes represent proteins of unknown 3D structure, while filled nodes have a known or predicted 3D structure. Edges represent protein-protein interactions either as known (light blue = from database, purple = experimental) or predicted (green = gene neighborhood, red = gene fusion, blue = gene co-occurrence). Other associations are either derived from text mining (yellow), co-expression (black) or protein homology (grey).*

Table S1: *Release of N-glycans reduces the MGL-binding of proteins from CRC cell lines.*

Table S2: *Raw data of comparative quantitative proteomics with TMT labeling on HCT116, HT29 and LS174T. The binary comparison of the protein abundance is shown.*

Table S3: *Binary abundance ratio of the major cell surface MGL binding proteins in HCT116 and HT29 cells compared to LS174T.*