

# Altered glycosylation of several metastasis-associated glycoproteins with terminal GalNAc defines the highly invasive cancer cell phenotype

Elham Khosrowabadi<sup>1</sup>, Tomasz Wenta<sup>1</sup>, Salla Keskitalo<sup>2</sup>, Aki Manninen<sup>1</sup> and Sakari Kellokumpu<sup>1</sup>

<sup>1</sup>University of Oulu, Faculty of Biochemistry and Molecular Medicine, Oulu, Finland

<sup>2</sup>Institute of Biotechnology, University of Helsinki, Helsinki, Finland

**Correspondence to:** Elham Khosrowabadi, **email:** elham.khosrowabadi@oulu.fi

**Keywords:** cancer; invasion; glycosylation; glycode; terminal GalNAc; Helix Pomatia agglutinin (HPA)

**Received:** September 30, 2021

**Accepted:** December 08, 2021

**Published:** January 10, 2022

**Copyright:** © 2022 Khosrowabadi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

Several distinct metastasis-associated glycosylation changes have been shown to promote cancer cell invasion and metastasis, the main cause of death of cancer patients. However, it is unclear whether their presence reflects cell- or tissue-specific variations for metastasis, or species needed to drive different phases of the metastatic cascade. To address this issue from a different perspective, we investigated here whether different cancer cell lines share any glycotopes that are common and important for their invasive phenotype. By using lectin microarray glycan profiling and an established myoma tissue-based 3D invasion assay, we identified a single glycotope recognized by *Helix Pomatia* agglutinin (HPA), whose expression level in different cancer cells correlated significantly with their invasive potential. Lectin pull-down assay and LC-MS/MS analysis in highly- (A431 and SW-48) and poorly invasive (HepG2 and RCC4) cancer cells revealed ~85 glycoproteins of which several metastasis-promoting members of the integrin family of cell adhesion receptors, the epidermal growth factor receptor (EGFR) and the matrix metalloproteinase-14 (MMP-14) were among the abundant ones. Moreover, we showed that the level of the GalNAc glycotope in MMP-14, EGFR,  $\alpha$ V-,  $\beta$ 1- and  $\beta$ 4 integrin in highly and poorly invasive cancer cells correlated positively with their invasive potential. Collectively, our findings suggest that altered glycosylation of several metastasis-associated glycoproteins with terminal GalNAc drives the highly invasive cancer cell phenotype.

## INTRODUCTION

Glycans are among the fundamental building blocks of life and play a key role in the development and physiology or pathology of multicellular organisms via mediating various cell-cell, cell-extracellular matrix, or cell-pathogen interactions [1–6]. In accord with this, several developmental disorders that are associated with impaired biosynthesis of glycans, have been identified in humans [7]. In all eukaryotes, the Golgi apparatus is the main site of glycosylation of various cell surface proteins and lipids that together form the glycocalyx, the

sugar “coat” on the plasma membrane of all eukaryotic cells [8–9]. In the Golgi, and other organelles (ER and the plasma membrane) where some glycans are also made, their biosynthesis is driven by special enzymes called glycosyltransferases that add specific sugar residues to growing glycan chains in a precise order and linkage type despite of not using any template for synthesis. Because of this, cell surface glycans do not represent random polymers of sugars but rather, a dynamic set of distinct glycans that can be cell-, protein- or lipid-specific [9, 10]. In part, this is due to transcriptional programs that regulate the levels of glycosylation enzymes in the Golgi

**Table 1: Cellular characteristics of the different cancer cell lines**

Cell Type	Gender	Age	Tissue of origin	Cancer type	Cell type	Karyotype (Chr. Number)
SW-48	Female	82 years	colon	adenocarcinoma	epithelial	Diploid 47 (+7)
DLD-1	Male	adult	colon	adenocarcinoma	epithelial	Pseudodiploid (46)
CaCo-2	Male	72 years	colon	adenocarcinoma	epithelial	Tetraploid (96)
HT-29	Female	44 years	colon	adenocarcinoma	epithelial	71
A431	Female	86 years	skin	carcinoma	epithelial	Triploid (74)
HepG2	Male	15 years	liver	carcinoma	epithelial	55
RCC4	?	?	kidney	carcinoma	epithelial	? (VHL <sup>-/-</sup> )
MCF-7	Female	69 years	breast (metastasis)	adenocarcinoma	epithelial	82
MDA-MB231	Female	51 years	breast (metastasis)	adenocarcinoma	epithelial	64

[11–13] and in part to environmental factors such as luminal pH and redox state that are needed to maintain Golgi homeostasis and co-operative functioning of the glycosyltransferases [14, 15].

During the last 3 decades, it has become clear that the Golgi apparatus in cancer cells is both structurally and functionally impaired. It typically has a fragmented morphology and produces shorter, more branched, and differentially fucosylated, sialylated, or sulphated glycans than those produced by non-malignant cells [16–20]. Together with genetic and epigenetic changes, altered glycans have been shown to contribute to cell growth and survival, development of tumors, and their metastatic spread via modulating cell adhesion, migration, invasion, apoptosis, immune evasion, and resistance to chemotherapy [17, 21–30]. Previously, several distinct metastasis-promoting glycotopes have been identified separately by different groups [28, 31–40]. These include increased core fucosylation of N-glycans, increased expression of O-linked GalNAc (Tn) or the sialyl-Lewis A and X (SLea, SLex) antigens, and increased branching of N-glycans with the  $\beta$ -(1,6)-linked GlcNAc. They all seem to have prognostic utility in detecting and treating aggressive cancers, given their prevalence in various cancers and predominant localization at the plasma membrane where they are easily accessible to potential therapeutic drugs.

However, the mechanistic details on how these altered glycotopes enhance metastasis are less clear. Moreover, the fact that there are several distinct metastasis-promoting glycotopes is puzzling. One possibility is that each one has a special purpose during metastasis, be it cell dissemination, migration, degradation of the extracellular matrix (ECM), intra- and extravasation, or colonization at adjacent or remote sites. Alternatively, their presence may reflect cell-, tissue- or cancer type-specific glycoforms to enhance the metastatic potential in different surroundings. Unravelling which one is the case is important but not trivial, given that glycosylation can be both protein- and site-specific. Also, the methodology that is needed to

faithfully mimic metastasis *in vivo* and at the same time allow dissection of different metastatic phases from each other, is not in routine use.

In this study, we decided to take a different approach and investigate whether different cancer cell types share any common glycotopes that are important for their invasive potential. By using an established myoma tissue-based 3D invasion assay [41], lectin microarray glycan profiling, correlation, and multiple linear regression analyses we identified a single GalNAc glycotope that is recognized specifically by the *Helix Pomatia* agglutinin (HPA) and is important for the highly invasive cancer cell phenotype. Moreover, lectin pulldown and LC-MS/MS analyses in highly and poorly invasive cell lines also revealed several distinct and abundant metastasis-promoting glycoproteins that display increased HPA binding in highly invasive cells compared to poorly invasive cells. These findings suggest that altered glycosylation of these metastasis-promoting glycoproteins with a terminal GalNAc is the key to the highly invasive cancer cell phenotype.

## RESULTS

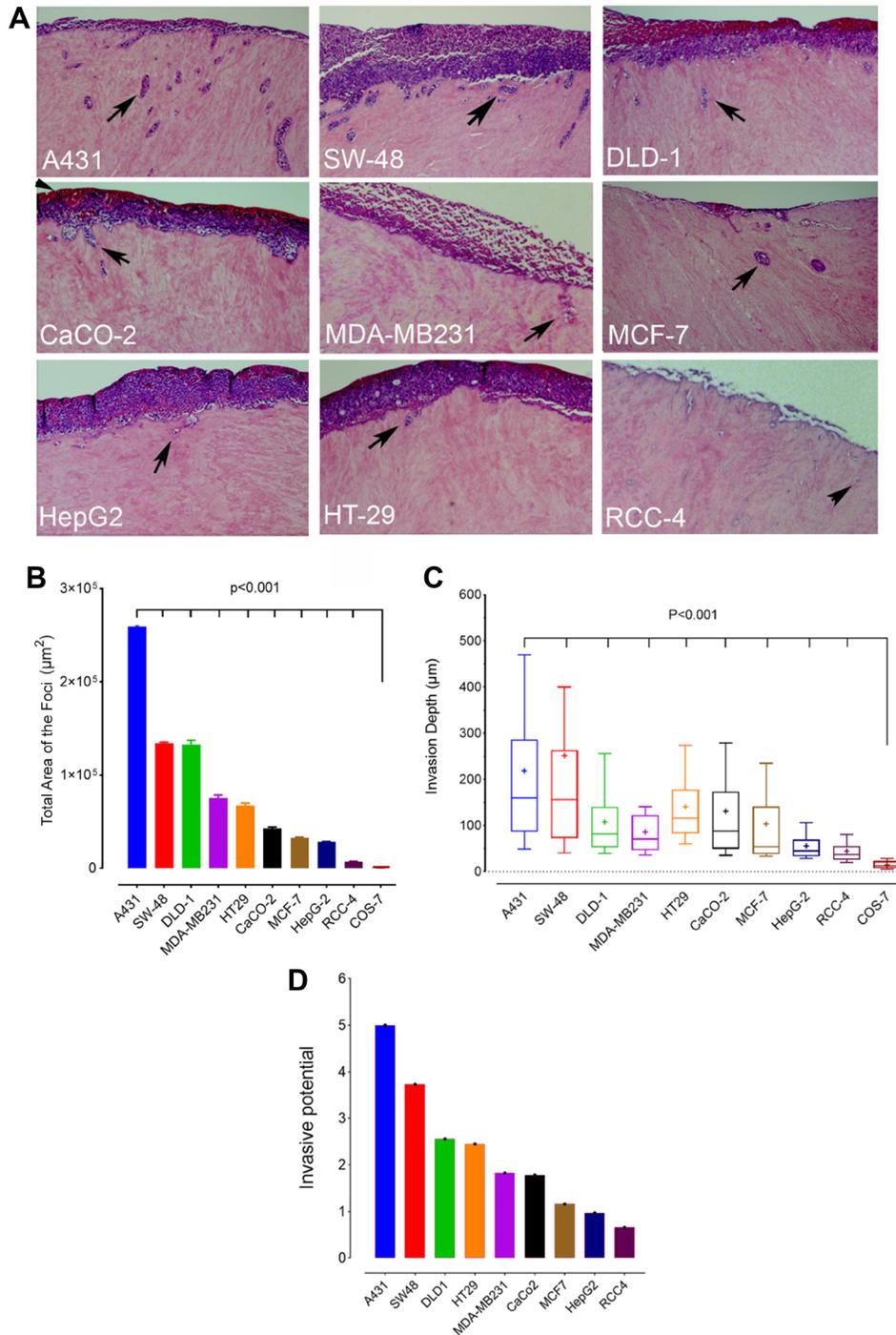
### Cancer cell lines display variable invasive potential in a 3D invasion assay

The geno- and phenotypic characteristics of the nine different cancer cell lines used in this study are depicted in Table 1. Overall, the cells display variable karyotypes and have several different tissue origins. Four of the cell lines are derived from colon adenocarcinomas (SW48, DLD-1, CaCo-2, and HT-29), two from breast cancer metastases (MCF-7, MDA-MB231), and the rest three (A431, RCC4, and HepG2) represent skin, kidney, and liver carcinomas. Except for HepG2, they all form tumors in nude mice. In certain cases, non-malignant COS-7 cells from the kidney of African green monkey were used for comparison.

To determine first the invasive potential of the different cancer cell lines, we used an established 3D

human myoma tissue-based invasion assay that mimics well the *in vivo*-conditions in that cells need to degrade the myoma tissue before being able to invade into the tissue. In brief, cells were seeded on top of the myoma

slices, allowed to grow for 3 weeks before processing for histochemical staining and quantification of the invasive foci present in each section (Figure 1A, arrows). Because some cancer cell lines displayed few and large



**Figure 1: Invasion potential of different cancer cell lines.** (A) Myoma-tissue-based 3D invasion assay. Cells were seeded on top of myoma discs, allowed to grow for 21 days before fixing and processing for histochemical staining. Sections were cut perpendicularly to the seeded cell layer, stained, and imaged before quantification. Representative figures of the invasive foci inside the myoma tissue are shown (arrowheads). (B) Total invasion area (µm<sup>2</sup>) of the foci in each cell line. (C) The median invasion depth of the foci from the seeded cell layer. Twelve sections from each myoma disc (*n* = 2/24) were used for the quantification with ImageJ software. The whiskers indicate 10th to 90th percentiles. (D) A bar graph showing the relative invasive potential of each cancer cell type. The values were calculated by scaling the medians of the total area and the median depth using scores from 5 (high) to 0 (low). Invasive potential was calculated as the mean of the two scores.

foci deep inside the tissue while in others the foci were small and numerous and near the seeded top cell layer, it was necessary to quantify both the total area (Figure 1B) and the median depth (Figure 1C) of the foci. To get a reliable estimate of the invasion potential, the total area and median depth were scored mathematically using a scale from five (high) to zero (low). The invasive potential (index) was then calculated as the mean of the two (Figure 1D). Based on scoring, skin A431 and colon SW-48 cells displayed the highest invasive potential while liver HepG2 and kidney RCC4 cells had the lowest invasive potential of all the cell types studied. A431 cells were roughly 7.5-fold more invasive than the least invasive RCC4 cells. No invasive foci were detected in COS-7 cells, consistent with their non-malignant phenotype (Figure 1B and 1C).

### **Glycosylation differences between cancer cell lines are both tissue- and cell type-dependent**

Next, we determined glycosylation profiles of the nine cancer cell lines by using lectin microarray glycan profiling. To allow direct comparisons between the cancer cell lines, the calculated medians from three independent samples (36 measurements points) were normalized against  $\alpha$ -tubulin before further analyses. Overall, heat map analysis (Figure 2A) showed that with few exceptions, the same lectins and their specific glycotopes were amongst the most or the least abundant irrespective of the cancer cell line in question, when COS-7 cells were used as a reference cell line. However, principal component analysis (PCA) with SPSS showed marked differences in glycan signatures between the different cancer cell lines, and between non-malignant COS-7 cells and the cancer cell lines (Figure 2B). Interestingly, PCA analysis identified three distinct cell pairs formed by A431 and SW-48 cells, MCF-7 and MDA-MB231 cells, and CaCo-2 and DLD-1 cells that were more closely related to each other than the other cell lines used in the study. In further support, hierarchical clustering with Ward linkage analysis together with Euclidean correlation coefficient as the distance metric showed that the glycosylation profiles (Supplementary Figure 1) of the two cell pairs (MCF-7 and MDA-MB231; CaCo-2 and DLD-1) were the closest homologs in terms of their glycan signatures (Figure 2C) while A431 and SW-48 cells were more distant and formed separate branches in one of the main subclusters. The other main subcluster was formed by the three poorly invasive cell lines: HepG2 (liver), HT-29 (colon), and RCC4 (kidney). Non-invasive COS-7 cells were also classified to this second main subcluster, suggesting their closer relationship with these three cancer cell types than with the other subcluster forming cell lines. Correlation and regression plots between the identified cell pairs confirmed their similar glycosylation patterns in each case (Figure 2D). Given their different tissue origins, it is likely that the variable glycosylation signatures reflect both

tissue- (MCF-7 and MDA-MB231) or cell type-specific (DLD-1 and CaCo-2) glycosylation differences.

Conditional formatting algorithms embedded in Excel were used next to identify glycotopes that are specific for each cell pair. To accomplish this, normalized lectin binding intensities (Figure 3) in each cell pair were classified and determined to be either similar or dissimilar depending on the cell line. For example, by using MCF-7 cells as a reference, we sought lectins whose binding intensities were similar with MDA-MB231 cells but dissimilar in the other cell lines. This approach yielded 6 lectins (PSA, GNL, Calsepa, LCA, PHA-L/E, and SNA) that specifically separate MCF-7 and MDA-MB231 cell pair from the other cell lines studied (Figure 3A and 3B). These lectins recognize various N-glycosylation intermediates, suggesting that MCF-7 and MDA-MB231 cells differ from the other cell lines mostly by their altered N-glycosylation status. The two colorectal cell lines (DLD-1 and CaCo-2) in turn displayed low binding to DBA, CA and HHL lectins (Figure 3A and 3C), suggesting that low levels of certain GalNAc- and mannose-containing glycotopes are typical for this cell pair. In contrast, poorly invasive RCC4 and HepG2 cells as well as moderately invasive HT-29 cells displayed high binding to HHL, GNL and NPA lectins (Figure 3A and 3D). These lectins are specific for various mannose-containing glycotopes and suggest that their high levels in the three cell lines can distinguish them from the other cancer cell types.

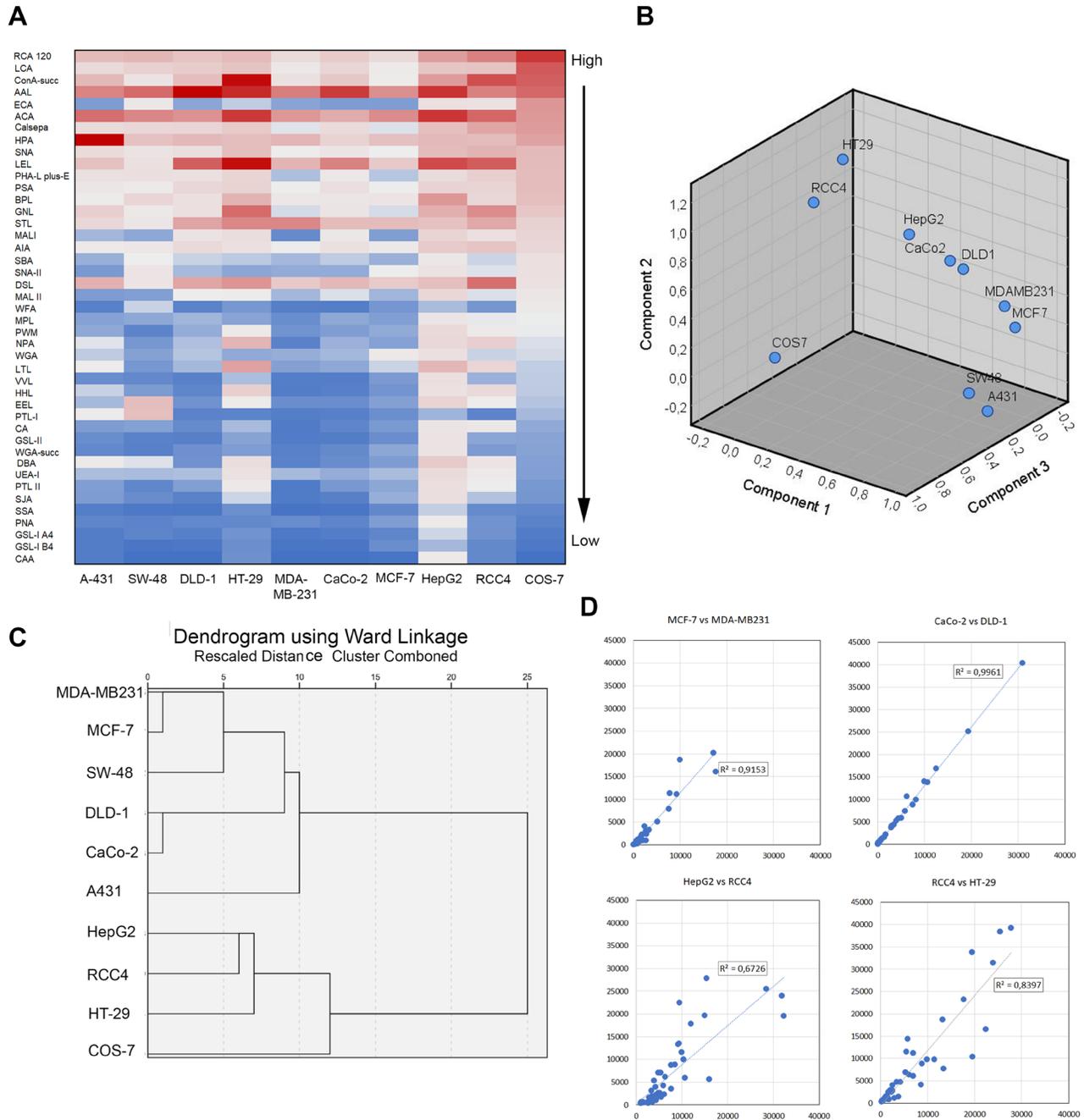
To visualize the main glycosylation changes between RCC4 cell and the other cancer cell lines, we subtracted normalized lectin binding intensities of RCC4 cells from those of the other cell lines to get so called subtracted fingerprints. Statistically significant fingerprints are shown in Supplementary Figure 2. The main differences included increased levels of truncated O-glycans, decreased levels of mannose-containing N-glycans, increased levels of specifically core-fucosylated N-glycans, and decreased levels of terminally glycosylated N-glycans. These changes, therefore, demonstrate that cancer cells differ markedly in their O- and N-glycan signatures.

### **Cancer cell invasive potential correlates with *Helix Pomatia* lectin (HPA) binding**

To find out next whether cancer cells possess any common glycotopes that are important for their invasive phenotype, we performed correlation and multiple linear regression analyses using algorithms embedded in Excel data analysis tool pack. Out of the 43 lectins in the array (Figure 4A), only five lectins (HPA, PTL-1, AJA, MAL I, PWM) were found to correlate either positively or negatively with the invasive potential of the cancer cells (Figure 4B). Multiple linear regression analyses further demonstrated that these five lectins accounted for 97% of the variation observed in the cancer cell invasive

phenotype (Figure 4C). By omitting the least significant contributor from the list after each subsequent round, it was found that HPA (specific for GalNAc glycotope) alone accounted for 58% of the variation (Figure 4C).

Its levels in different cancer cell lines also correlated positively with cancer cell invasive potential ( $R = 0.763$ ;  $p < 0.007$ ; Figure 4B, 4D and 4E). Together with PTL-1, HPA accounted for 76% of the variation, while the rest



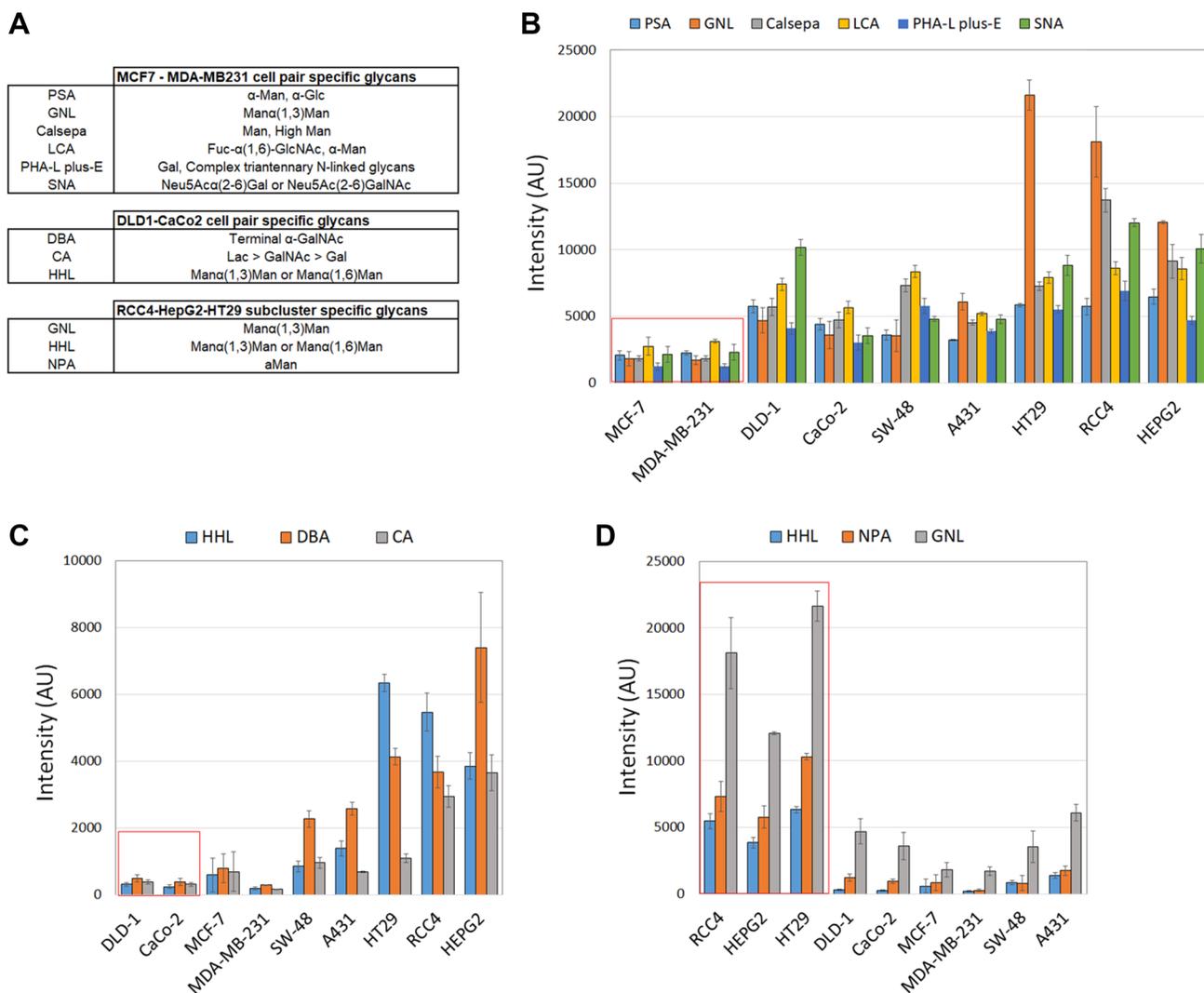
**Figure 2: Comparison of glycosylation profiles between different cancer cell types.** (A) Heat map representation of the lectin microarray glycan profiles. The rows represent normalized median intensity values of each lectin. Three independent samples were used for determination of the intensity values in each cell line. The color scale from blue to red indicates low to high signal intensities. Non-malignant COS-7 cells were used as a reference cell line for ranking. (B) Comparison of the glycan profiles by using principal component analysis tool in Excel. Normalized intensity values for each lectin were used for the analysis. The number of components is determined automatically from the input values. (C) Hierarchical clustering analysis of the glycosylation differences or similarities between different cancers cell type. Normalized median intensity values of each lectin, the SPSS software with Ward linkage analysis and Euclidean correlation coefficient as the distance metric were used for the analysis. (D) Comparison of the glycosylation profiles between subcluster forming cell lines. In each plot, normalized median intensity values were plotted against each other. The lines represent linear correlation equations. The R2 values are shown for each plot.

(AJA, MAL I, PWM) correlated negatively and accounted for 7% each. Thus, increased expression of HPA-binding glycotopes (O-linked GalNAc) in cancer cells appears to be the main factor promoting cancer cell invasive phenotype. Yet, decreased expression of AJA, PWM and MAL I specific glycotopes (O-linked Gal $\beta$ (1,3)GalNAc, N-glycan branching and  $\alpha$ -2,3-sialylation, respectively) also contribute to some extent.

### Identification of HPA binding glycoproteins by lectin blotting and mass spectrometry

Since HPA lectin has been previously suggested to have prognostic utility in detecting metastatic breast and

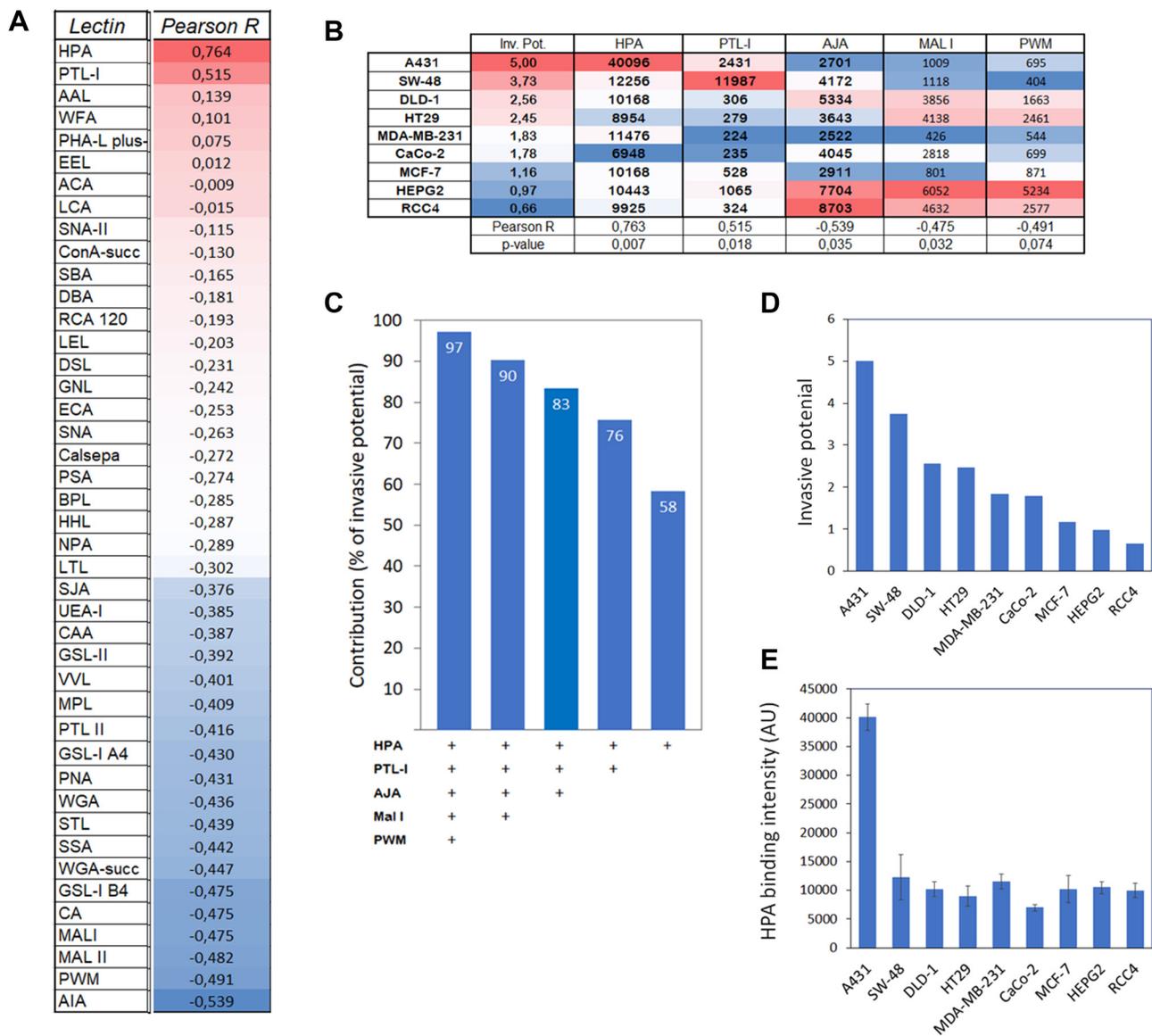
colorectal cancer cell lines [42–47], we decided to identify glycoproteins that carry the glycotope specific for the HPA lectin using highly - (A431 and SW-48) and poorly invasive (RCC4 and HepG2) cells as our targets. It was also anticipated that such identification would also give new insights into mechanistic details for why this glycotope enhances cancer cell invasive potential. To accomplish this, we first used lectin blotting with HPA to visualize HPA binding proteins and their levels in different cell lysates. HPA blotting (Figure 5A) revealed several prominent bands with a MW of ~400–600 kDa, ~240 kDa, ~160 kDa, ~130 kDa, ~80 kDa, 50 kDa, and 45 kDa) in highly invasive A431 and SW-48 cells. Importantly, the bands were almost undetectable in poorly invasive HepG2 and RCC4 cells.



**Figure 3: Identification of lectins that are specific for each clustered cell pair.** Lectins that define clustered cell pairs were identified by using conditional formatting algorithms embedded in Excel. In brief, similar lectin binding intensities were searched for each clustered cell by setting an acceptable intensity limit to  $1.5 \times SD$ . Any values within these limits were designated as similar whereas the ones exceeding these limits were designated as dissimilar. Specific lectins for each cell pair were then selected based on its similarity between the cell pair and dissimilarity in the other cell lines. Normalized median intensity values ( $\pm SD$ ) were used for comparisons. (A) Selected lectins and their sugar specificities specific for each clustered cell pair. (B) Histogram showing lectin binding intensities specific for the clustered MCF-7/MDA-MB231 cell pair (marked by red box). (C) Histogram showing lectin binding intensities specific for the clustered DLD-1/CaCo-2 cell pair. (D) Histogram showing lectin binding intensities specific for the clustered RCC4, HepG2 and HT-29 cell lines.

Next, HPA-binding proteins were pulled down with the lectin before their identification by liquid chromatography and tandem mass spectrometry (LC-MS/MS). Altogether, we identified ~85 glycoproteins that potentially bind HPA (Supplementary Table 1). After classifying the proteins by their abundance (peak heights) in highly invasive A431 cells, we selected the 60 most abundant ones and tested how well their levels correlate with the invasive potential of these same cells.

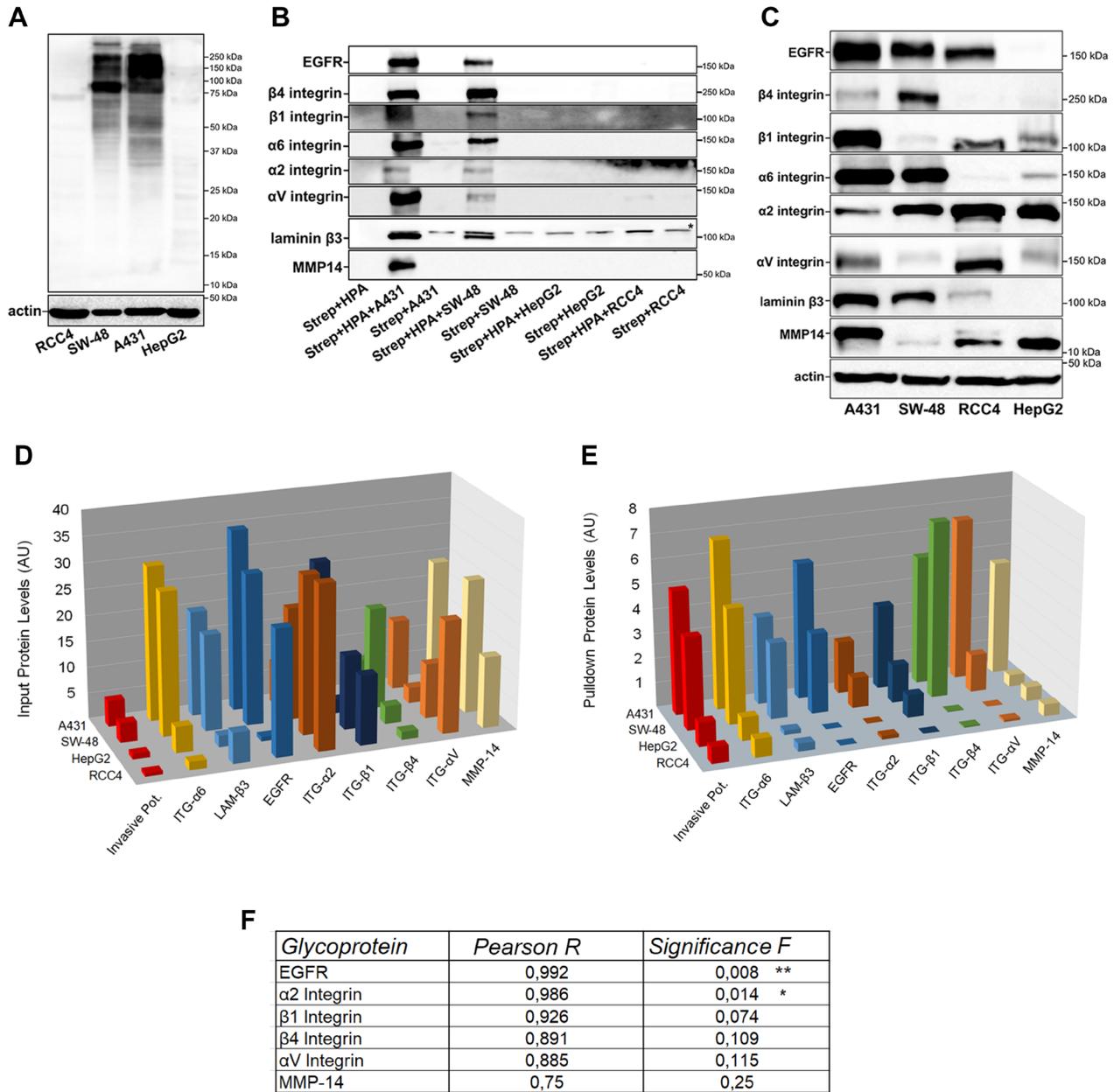
Using a Pearson R cutoff of >0.77, we were left with 35 glycoproteins of which more than half represented glycoproteins with a known role in cell adhesion, migration, cell signaling, or metastasis (Table 2). Of these, we selected eight glycoprotein candidates (epidermal growth factor receptor (EGFR), matrix metalloproteinase-14 (MMP-14),  $\beta$ 4-,  $\beta$ 1-,  $\alpha$ 6-,  $\alpha$ 2- and  $\alpha$ V-integrin and laminin  $\beta$ 3 for further analyses. Quantification of the bands after western blotting showed first that all these proteins were



**Figure 4: Correlation and impact of HPA binding to cancer cell invasive potential.** (A) Correlation analysis between lectin binding and invasive potential of different cancer cell types. Normalize median intensity values for each lectin and relative invasive potential (Figure 1D) were used for the analysis with Excel's data analysis tool pack. Pearson R values are shown and ranked from high to low and marked with red (high) and blue (low) colors. (B) Correlation analyses of selected lectins with their normalized median intensity values in relation to invasive potential of different cancer cell types. Pearson R values and statistical significance of the correlation are shown at the bottom of the table. Regression analysis with ANOVA was used for testing the statistical significance. (C) Multiple linear regression analysis showing the contribution of each lectin to cancer cell invasive potential. Each bar represents the contribution (as percentages) of the lectins used for the analysis (bottom). (D and E) Comparison of the cancer cell invasive potential (D) with the level of HPA binding glycotopes (E) in different cancer cell lines. Correlation and regression analyses showed a significant correlation between these two variables (R-value of 0.76 and p-value of 0.007\*\*).

expressed at higher levels in highly invasive A431 and SW-48 cells than in poorly invasive HepG2 and RCC4 cells, as expected (Figure 5B, 5D). However, when compared to protein input levels (Figure 5C, 5E), we noticed that out of these eight glycoproteins,  $\alpha 6$  integrin and laminin  $\beta 3$  levels

in pull-down and input samples matched well with each other, and in fact, did show a significant correlation with each other (Supplementary Figure 3A). Similar results were also obtained by calculating their ratios (Supplementary Figure 3B). These data suggested that these two proteins



**Figure 5: Identification of HPA binding glycoproteins and their comparison between highly and poorly invasive cancer cell types.** (A) HPA-binding proteins in highly (A431 and SW-48) and poorly (RCC4 and HepG2) invasive cancer cell samples as revealed by lectin blotting with biotinylated HPA. Complexes were visualized using HRP-conjugated streptavidin. A representative blot is shown. (B) Immunoblotting of selected HPA-binding glycoproteins after lectin pull-down with relevant antibodies in highly (A431, SW-48) and poorly (RCC4 and HepG2) invasive cancer cells. The star (\*) in laminin  $\beta 3$  blot denotes a non-specific band. A representative blot is shown. (C) Immunoblotting of the HPA binding glycoproteins in input samples of the same cell types. A representative blot is shown. (D) Quantification of HPA pull-down protein levels in the immunoblot (B). ImageJ software was used for quantification of the band intensities as arbitrary units (AU). (E) Quantification of protein input levels in the immunoblot (C). ImageJ software was used for quantification of the band intensities as arbitrary units (AU). (F) Correlation analysis between HPA pull-down protein levels with invasion potential. Excel data analysis tool pack's correlation and regression analyses with ANOVA were used to get the Pearson R and significance values. The stars (\*) denote the statistical significance of  $p < 0.01^{**}$  and  $p < 0.05^*$ , respectively.

**Table 2: The list of invasion-promoting glycoprotein candidates identified by LS-MS/MS from HPA pull-down samples**

Gene	Glycans	ID (Uniprot)	Protein name	Function (based on Uniprot ID)	Pearson R
<i>ITGB4</i>	N+O	P16144	<b>Integrin beta-4</b>	<b>Cell adhesion, migration, motility, signaling</b>	<b>0,997</b>
<i>ITGA6</i>	N+O	P23229	<b>Integrin alpha-6</b>	<b>Cell adhesion, migration, motility, signaling</b>	<b>0,997</b>
<i>ASPH</i>	N+O	Q12797	Aspartyl/asparaginyl beta-hydroxylase	Hydroxylation of Asp or Asn (EGF-like domains)	0,995
<i>GOLM1</i>	N+O	Q8NBJ4	Golgi membrane protein 1	Regulation of lipid/protein metabolism	0,992
<i>HSP90AB1</i>	N+O	P08238	Heat shock protein HSP 90-beta	Protein folding	0,989
<i>TFRC</i>	N+O	P02786	Transferrin receptor protein 1	Iron transport and homeostasis	0,971
<i>BSG</i>	N+O	P35613	<b>Basigin</b>	<b>Cell adhesion, migration, motility, signaling</b>	<b>0,964</b>
<i>ST14</i>	O	Q9Y5Y6	Suppressor of tumorigenicity 14 protein	Degradation of the ECM	0,959
<i>ATP5F1B</i>	O	P06576	ATP synthase subunit beta, mitochondrial	ATP synthesis	0,957
<i>PTGFRN</i>	N+O	Q9P2B2	Prostaglandin F2 receptor negative regulator	Regulates Prostaglandin receptor activity	0,956
<i>EGFR</i>	N+O	P00533	<b>Epidermal growth factor receptor</b>	<b>Migration, signaling, malignant transformation</b>	<b>0,945</b>
<i>SPINT1</i>	N+O	O43278	Kunitz-type protease inhibitor 1	Endopeptidase inhibitor	0,934
<i>PRKDC</i>	N+O	P78527	DNA-dependent protein kinase catalytic subunit	DNA damage sensor	0,924
<i>ITGA2</i>	N+O	P17301	<b>Integrin alpha-2</b>	<b>Cell adhesion</b>	<b>0,924</b>
<i>PLXNB2</i>	N+O	O15031	<b>Plexin-B2</b>	<b>Invasive growth, migration, signaling</b>	0,921
<i>DSG2</i>	N+O	Q14126	<b>Desmoglein-2</b>	<b>Cell adhesion</b>	<b>0,918</b>
<i>F11R</i>	N+O	Q9Y624	<b>Junctional adhesion molecule A</b>	<b>Cell adhesion, tight junction formation, migration</b>	<b>0,915</b>
<i>PLXNA1</i>	N+O	Q9UIW2	<b>Plexin-A1</b>	<b>Cell adhesion (homotypic)</b>	<b>0,890</b>
<i>EPCAM</i>	N+O	P16422	<b>Epithelial cell adhesion molecule</b>	<b>Cell adhesion</b>	<b>0,890</b>
<i>ITGB1</i>	N+O	P05556	<b>Integrin beta-1</b>	<b>Cell adhesion</b>	<b>0,885</b>
<i>LRP1</i>	N+O	Q07954	Pro-low density lipoprotein receptor-related protein 1	Endocytosis	0,877
<i>GLG1</i>	N+O	Q92896	<b>Golgi apparatus protein 1</b>	<b>Cell adhesion</b>	<b>0,875</b>
<i>GALNT3</i>	N+O	Q14435	Polypeptide N-acetylgalactosaminyltransferase 3	O-Glycosylation	0,870
<i>LGALS3BP</i>	N+O	Q08380	<b>Galectin-3-binding protein</b>	<b>Cell adhesion</b>	0,847
<i>PSAP</i>	N+O	P07602	<b>Prosaposin</b>	<b>Lipid metabolism</b>	<b>0,838</b>
<i>LAMA3</i>	N+O	Q16787	<b>Laminin subunit alpha-3</b>	<b>Adhesion, migration, differentiation</b>	<b>0,818</b>
<i>ITGB6</i>	N+O	P18564	<b>Integrin beta-6</b>	<b>Cell adhesion</b>	<b>0,813</b>
<i>IGSF8</i>	N+O	Q969P0	<b>Immunoglobulin superfamily member 8</b>	<b>Cell motility</b>	<b>0,797</b>
<i>ITGAV</i>	N+O	P06756	<b>Integrin alpha-V</b>	<b>Cell adhesion</b>	<b>0,777</b>
<i>MMP14</i>	O	P50281	<b>Matrix metalloproteinase-14</b>	<b>Degradation of the ECM</b>	0,774

The proteins were classified first by their abundance and then, by their correlation (Pearson R) with cancer cell invasive potential. Bold text refers to proteins which have a known role in cell migration, adhesion, and metastasis according to the Uniprot database.

have the same amount of GalNAc glycocone in highly and poorly invasive cancer cells.

Importantly, all the other glycoproteins (EGFR, MMP-14,  $\beta$ 4-,  $\beta$ 1-,  $\alpha$ 2- and  $\alpha$ V integrin) in HPA pulldown samples were markedly enriched in highly invasive cells relative to poorly invasive cells (Figure 5D), indicating that these proteins carry increased levels of a terminal

GalNAc glycocone in their glycans. Of these same glycoproteins, only the EGFR and  $\alpha$ 2 integrin correlated significantly with cancer invasive potential (Figure 5F). Thus, these two proteins might be the most important for highly invasive cancer cell phenotype. However, because pulldown experiments are only semi-quantitative, the impact of the other GalNAc-carrying proteins (MMP-14,

$\beta$ 4,  $\beta$ 1, and  $\alpha$ V integrins) on cancer cell invasion cannot be excluded.

## DISCUSSION

By using lectin microarray glycan profiling and a 3D myoma tissue-based invasion assay, we identified here one single glycotope, a terminal GalNAc, that is recognized by Helix Pomatia agglutinin (HPA) and that correlated significantly with the invasive potential of the nine different cancer cell types used in this study. Moreover, by using lectin pulldown and proteomics tools, we uncovered ~85 glycoproteins with either known or potential binding site for the lectin. We also identified several metastasis-associated glycoproteins including MMP-14, EGFR,  $\alpha$ V,  $\beta$ 1, and  $\beta$ 4 integrins that displayed higher levels of the GalNAc glycotope in lectin pull-down samples of highly invasive cancer cell lines, in contrast to poorly invasive cells. Altogether, these findings suggest that altered glycosylation with a terminal GalNAc of these glycoproteins in highly invasive A431 and SW-48 cancer cells is the key to their high invasive potential when compared to poorly invasive HepG2 and RCC4 cancer cells.

Increased HPA binding has been previously shown to be associated with poor prognosis and development of metastases *in vivo* both in cancer patients and mouse models [42–48]. For example, Schumacher and Adam [42] showed that inoculation of HPA positive colon (HT29) or breast cancer cell lines (MCF-7, T47D) into immunodeficient mice resulted in the development of metastases in 23 out of 26 cases while metastases were non-existent or rare in HPA negative cell lines. On the other hand, previous breast cancer studies by Milde-Langosch et al. [47] suggested that HPA binding *in vivo* was associated with increased invasion. By using a Matrigel invasion assay, Rye et al. [43] in turn showed that an HPA positive melanoma cell line was more invasive than an HPA-negative control cell line. Thus, our observations are fully consistent with these findings, but they also emphasize that increased HPA binding is likely a general feature of all cancer cell types with high invasive potential. It is also of note that despite the use of different methodologies and approaches used in the above studies, the outcomes are similar.

Together, the above studies provide strong support for the view that HPA binding and the presence of a GalNAc glycotope in specified glycoproteins is an important determinant of highly invasive cancer cells. Yet, because HPA binds in addition to terminal GalNAc, a broader array of glycotopes including GalNAc $\alpha$ 1,3Gal and GlcNAc $\beta$ 1,4Gal [43, 47], the exact identity of the HPA-specific glycotope remains to be elucidated. Moreover, it is also unreasonable to expect that all lectins with a similar nominal binding specificity will give identical results, as lectin specificity can depend also on sub-terminal as well

as 3D structural motifs. This phenomenon likely explains why HPA, in contrast to DBA and PTL-1, did exhibit a significant correlation with cancer cell invasive potential. In accord with this, Laferte et al. [49] have shown that glycoproteins from both colon cancer tissue and HT-29 colon cancer cells bound HPA while they displayed poor binding to DBA.

To gain insight into why HPA binding is associated with high invasive potential, it was also important to identify glycoproteins that carry the glycotope(s) for HPA. Previous work from Dwek's Laboratory [50–52] has indicated that in metastatic tumor cells from breast and colon, HPA binding correlated with levels of  $\alpha$ 6 integrin, HnRNP family of transcription factors (heterogeneous nuclear ribonuclear proteins H1, D-like, and A2/B1), heat shock protein 27 (Hsp27), glial fibrillary acidic protein and enolase 1 (ENO1). None of these proteins were detected in the non-metastatic breast cancer cell lines. Interestingly, we also identified  $\alpha$ 6 integrin as one of the most abundant glycoproteins in highly invasive cells. Its levels together with laminin  $\beta$ 3 in highly and poorly invasive cells also correlated with HPA binding. However, when input levels were taken into account, this correlation was lost, suggesting that the level of HPA binding sites in either glycoprotein was not different between highly and poorly invasive cells. In contrast to earlier studies, we detected several other cell adhesion receptors (including  $\alpha$ 2-,  $\beta$ 1-,  $\alpha$ V- and  $\beta$ 4 integrin) in addition to the EGF receptor and MMP-14 that were expressed in both highly and poorly invasive cells but also, were markedly enriched by HPA pull-down from highly invasive A431 and SW-48 cells, in comparison to poorly invasive HepG2 and RCC4 cells. This finding suggested these proteins were differentially glycosylated with terminal GalNAc between highly and poorly invasive cells.

Interestingly, all the above glycoproteins have been previously implicated to enhance invasion and/or metastasis [53–56]. Moreover, some metastasis suppressor genes (KAI1, also known as CD82) interact with both integrins and the EGFR, and abrogate their signaling, thereby attenuating metastasis [57–59]. Although mechanistic details on how altered glycosylation of these invasion/metastasis promoting glycoproteins enhance cancer cell invasive potential remain unclear, previous studies have shown that glycosylation can modulate the activity of several of these proteins. For example, the EGFR, a transmembrane tyrosine kinase and a therapeutic target, which upon dimerization, activates several signaling cascades including MAPK-, Akt- and JNK-kinase pathways [53]. It also modulates transcriptional program downstream of the EGFR [60]. EGFR is upregulated in many types of cancers [53] and its ectodomain is heavily N-glycosylated [61] and its intracellular domain was shown to be a substrate for O-GlcNAc transferase [62]. Importantly, it is also O-glycosylated by a polypeptide

GalNAc-transferases (GALNT2 and GALNT6) that add the first GalNAc to serine or threonine amino acids giving rise to tumor-associated Tn-antigen [63]. Upregulation of GALNT2 and GALNT6 resulted in enhanced migration and invasion of oral and ovarian cancer cells, in part by increasing O-glycosylation and activation of the EGFR [62–64]. Another metastasis-associated glycoprotein whose activity is known to be regulated by altered O-glycosylation is matrix metalloproteinase-14 (MMP-14). Nguyen et al. showed that GALNT1-dependent O-GalNAc glycosylation markedly increased MMP-14 activity, ECM degradation, tumor growth, and invasiveness in a mouse xenograft model [65].

Members of the integrin family of cell adhesion receptors also carry both N- and O-glycans which, when altered, can influence migration and adhesion of tumor cells, a prerequisite for their invasiveness [66–68]. However, direct evidence on the role of O-glycans in integrin-mediated cell adhesion or signaling is almost non-existent.  $\beta$ 1 integrin subunit is an exception, and it has been shown to carry core 1 O-glycan(s) a product of the C1GALT1 transferase [63], which is typically overexpressed in hepatocellular carcinomas and HCC cells. Importantly, overexpression of C1GALT1 was found to enhance HCC cell adhesion to ECM proteins, their migration, and invasion, whereas RNAi-mediated C1GALT1 knockdown suppressed this phenotype. By using a mouse xenograft model, the authors also showed that C1GALT1 promotes HCC cell metastasis. These effects were strictly dependent on  $\beta$ 1 integrin since the C1GALT1-mediated phenotypic changes were suppressed by the anti- $\beta$ 1-integrin antibody that blocks its activity. An O-GalNAc/GlcNAc modification of  $\alpha$ V- and  $\alpha$ 6 integrin has been reported [51, 52] but their functional significance remains currently unclear.

To summarize, these findings suggest that altered glycosylation of several distinct metastasis-associated glycoproteins, including integrins, EGFR, and MMP-14, is a key to the highly invasive cancer cell phenotype. Yet, further studies are warranted to confirm whether this holds true for all different integrin subunits including those that are overexpressed in cancers but do not display enhanced HPA binding. Our data also highlight the fact that cancer cell invasive potential depends not on a single protein, but rather a compilation of GalNAc-glycosylated proteins, each of them having a special role in invasion, be it cell adhesion, migration, or degradation of the extracellular matrix. Our findings also emphasize the prognostic and therapeutic utility of these invasion-promoting glycoproteins bearing a GalNAc glycotope. However, better identification of the glycotope(s) that bind HPA, and a better understanding of how altered glycosylation regulates the activity of the above metastasis-promoting proteins are also needed before we can rationally address and prevent cancer cell invasiveness in a clinical setting.

## MATERIALS AND METHODS

### Cell culture and sample preparation

All cell lines (COS-7, HepG2, HT-29, SW-48, CaCo-2, DLD-1, MCF-7, MDA-MB 231, RCC4, and A431) were from ATCC (Manassas, VA). Cells were grown in high glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with Glutamax (Gibco BRL, Grand Island, NY, USA), 10% fetal bovine serum (HyClone, Cramlington, United Kingdom), and antibiotics (100 U/ml Penicillin and 100  $\mu$ g/ml Streptomycin: Sigma-Aldrich, St. Louis, MO) in humidified conditions at +37°C and 5% CO<sub>2</sub>.

### Reagents and antibodies

All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless stated otherwise. The list of antibodies used in this study is present in (Supplementary Table 2).

### Lectin microarray printing and labelling

#### Sample labelling

Cells cultivated on plates to 70–80% confluency were lysed for 30 min on ice with lysis buffer (50 mM sodium tetraborate buffer, pH 8.5, 150 mM NaCl, 1% Triton X-100) supplemented with the protease inhibitor cocktail (Complete Mini, Roche) and clarified by centrifugation (12,000  $\times$  g at 4°C for 15 min). 12  $\mu$ g of total protein of each cell lysate was labeled with 6  $\mu$ g of NHS activated DyLight 633 dye (Thermo Scientific, Waltham, USA) in 50  $\mu$ l of labeling buffer (50 mM boric acid/150 mM NaCl, pH 8.5) for 1 h at RT with constant agitation (600 rpm). Any extra label was then quenched at RT for 1 h by adding 50  $\mu$ l of quenching buffer (75 mM ethanolamine in 200 mM Tris-HCl/150 mM NaCl, pH 8.5) before dilution (1:12) with the assay buffer (50 mM Tris/300 mM NaCl/2 mM MgCl<sub>2</sub>/2 mM MnCl<sub>2</sub>/2 mM CaCl<sub>2</sub>, pH 7.1) to give the final pH 7.4 to the labeled sample mix.

#### Staining of microarray slides and quantification

After clearing by centrifugation (12,000  $\times$  g for 10 min at RT), 400  $\mu$ l of the labeled sample was applied to each well on a pre-printed (see below) Nexterion H microarray slide (Schott, Germany) embedded in a rubber housing with six separate wells. After further incubation in a humidified chamber with constant agitation for 2 h in RT, wells were washed five times for 5 min each with the washing buffer (50 mM phosphate buffer/0.05% Tween) and finally by dipping the whole slide in washing buffer and water to remove any remaining salt before drying. Array images were generated using the Genepix 4200AL laser scanner (Axon Instruments) with an appropriate filter

set for the DyLight 633™ dye. Quantification of the median intensities of the bound label in each spot was done by using the GenePix Pro® microarray analysis software array before calculating the median intensities from three separate wells each having four parallel arrays and three separate lectin spots/array (36 measurement spots/sample). Glycan profiles were then generated by calculating the mean ( $\pm$ SD) from three different samples per cell line.

### Printing

A Microgrid II array printer equipped with a four pin printer head and in-build software package were used for printing 10  $\mu$ M lectin stocks (diluted in printing buffer: 50mM phosphate buffer, 1% glycerol and 0, 05% Tween, pH 7.6) on microarray slides. After printing, the slides were allowed to equilibrate for 4 h at RT and 70% humidity before pre-quenching the activated slides with 50 mM ethanolamine and labeling. The lectins (43 different) were purchased from either EY Laboratories (San Mateo, CA, USA) or the Vector Laboratories (Youngstown, OH, USA). All the lectins used are listed in (Supplementary Table 3) with their sugar specificities according to manufacturers' data sheets. Optimization of the binding specificity and signal-noise ratio was performed using labeled fetuin and asialofetuin as markers. Different lectin-sample ratios were also tested to guarantee unsaturated binding capacity of the lectin spots.

### 3D invasion assay

The invasive properties of each cell type were investigated using an organotypic 3D-myoma-invasion model, as described earlier [41]. In brief, myoma discs pre-equilibrated at +4°C in DMEM were placed in tightly fitted Transwell® inserts (Corning, Inc., Corning, NY, USA) after which  $5 \times 10^5$  cells (in 50  $\mu$ l of DMEM) were added on top of each disc. After attachment, myoma discs with cells were transferred onto uncoated nylon discs placed on curved steel grids ( $3 \times 12 \times 15$  mm) in 12-well plates, each well-containing 1ml of DMEM. The myoma organotypic cultures were maintained for 21 days with daily media changes. Each assay was performed in triplicate. The specimens were fixed in 4% formalin overnight, dehydrated, and embedded in paraffin. Finally, 6  $\mu$ m thick sections were cut and deparaffinized before staining with Mayer's Hematoxylin-Eosin. After digitalization, the total area and median invasion depth of the invasive foci with cells from each microscopic field were determined by measuring the area and distance of the invasive cell foci from the top cell layer on each disc using the Image J (Fiji) v1.46o software (National Institute of Health, USA).

### Lectin/western blotting

Biotinylated HPA (Sigma-Aldrich) was dissolved in  $1 \times$  PBS containing 1mM  $\text{CaCl}_2$  and 0.33 mM  $\text{MgCl}_2$  to 1 mg/mL. Cells were lysed in 50 mM Tris-HCl, pH 7.5, 150

mM NaCl, 1% Triton X-100, 1 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$  buffer supplemented with the protease inhibitor cocktail (Sigma-Aldrich) on ice for 1 h. Total protein concentration was estimated using BCA Kit (Pierce). Thirty micrograms of total cell lysate protein were resolved in 10% acrylamide gel by SDS-PAGE and then transferred onto the PVDF membrane. To avoid unspecific binding the membrane was incubated in a blocking buffer containing 5% BSA (Sigma-Aldrich) in TBST (TBS with 0.1% Tween-20 (Sigma-Aldrich)) for 3 h at RT. Next, the membrane was probed with 20  $\mu$ g/mL HPA in 5% BSA, 1 mM  $\text{CaCl}_2$  in TBST solution for 1h followed by  $3 \times 10$  minutes washing in TBST and incubation with HRP-conjugated streptavidin in 5% BSA, 1 mM  $\text{CaCl}_2$  in TBST solution for 1 h at RT. After washing ( $5 \times 10$  minutes each) the positive signal was revealed using SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific).

For the western blotting assay, membranes with transferred proteins were incubated for 1 h in 5% skimmed milk and probed with specific primary antibodies (Supplementary Table 2) overnight at 4°C. Secondary antibodies conjugated with HRP and Lumi-Light Western Blotting Substrate (Roche) were used to visualize specific protein bands. The bands were detected using Fujifilm LAS-3000 bioimaging and scientific research imaging equipment (FUJI Photo film Co., LTD.).

### Pull-down assay

Cells were incubated in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 1 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$ ) supplemented with the protease inhibitor cocktail (Sigma-Aldrich) on ice for 1 h and centrifuged at  $17000 \times g$  (4°C) for 15 min. Protein concentration was estimated using BCA Kit (Pierce). To avoid unspecific binding, cell lysates were precleared using Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen). 70  $\mu$ g of biotinylated HPA was incubated with 0.3 mg of the total cell lysate proteins in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Triton X-100, 1 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$  supplemented with protease inhibitor cocktail for 3 h at 4°C. After this time, 100  $\mu$ l Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen) was added to the mixture for 1 h at 4°C. Finally, the beads were washed  $4 \times 1$  ml of 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Triton X-100, 1 mM  $\text{CaCl}_2$ , 1 mM  $\text{MgCl}_2$ , after which proteins were harvested by adding  $2 \times$  Laemmli buffer and incubating the beads at 96°C for 5 minutes. Samples were resolved in 10% acrylamide gel followed by silver staining using Silver Staining Kit (Pierce) or probed with a specific antibody in the western blotting assay.

### LC-MS/MS analysis

Silver-stained (Silver Staining Kit (Pierce)) protein bands were cut out of the polyacrylamide gel and digested by adding 0.75  $\mu$ g trypsin (Sequencing

Grade Modified Trypsin, Promega) overnight at 37°C. Cysteine bonds were reduced with 0.045 M dithiothreitol (Sigma-Aldrich) for 20 min at 37°C and alkylated with 0.1 M iodoacetamide (Sigma-Aldrich) at room temperature. Peptides were purified with C18 micro-spin columns (Harvard Apparatus) according to manufactures protocol and dried ones were reconstituted in 40 µl 0.1% trifluoroacetic acid (Sigma-Aldrich) in 1% acetonitrile (Sigma-Aldrich). In the next step, each sample was diluted further 1:4 in 1% acetonitrile, and then 4 µl was injected per LC-MS/MS run. Liquid chromatography coupled with mass spectrometry analysis was carried out on a nanoElute (Bruker Daltonics) coupled online to a hybrid trapped ion mobility spectrometry – quadrupole time of flight mass spectrometer (timsTOF Pro, Bruker Daltonics). Liquid chromatography was performed at 50°C with a constant flow of 400 nL/min using a two-column setup consisting of a 5 mm Acclaim™ PepMap™ 100 C18 trap column (Thermo Fisher Scientific), followed by 25 cm × 75 µm ID, 1.6 µm C18-Aurora emitter column with nanoZero and CaptiveSpray Insert (IonOptics). As the mobile phases, water with 0.1% formic acid (vol/vol; VWR) and acetonitrile with 0.1% formic acid (vol/vol) were applied. Peptides were separated with a linear gradient from 2 to 17% of formic acid within 60 min, followed by an increase to 25% of formic acid within 30 min and further to 37% within 10 min, followed by a 10 min washing step from 37% to 95% and another 10 min wash with 95% formic acid. The timsTOF Pro mass spectrometer was operated in positive PASEF mode using the DDA standard\_1.1 sec\_cycletime method in acquisition. In brief, MS and MS/MS spectra were recorded from m/z 100 to 1700 and acquired with 100 ms ramp time, 100% duty cycle and 10 PASEF MS/MS scans with precursor target value set to 20 000 a.u. For precursor ions, charge minimum and maximum of 0 and 5 were used, respectively. Range for ion mobility coefficient (1/K0) –0.60–1.60 Vs/cm<sup>2</sup> was applied. For the number of distinct peptides assigned for each protein by HPA lectin pull-down and LC-MS/MS (Supplementary Table 4).

## Data analysis

Mass spectrometry raw files were processed with FragPipe v15.0 using the protein sequence database of reviewed Human proteins (UniProtKB release 2021\_03, Human Proteome UP000005640). Decoy sequences and common contaminants were generated and added to the original database as part of the FragPipe workflow as described in [69, 70]. Trypsin was selected as the cleavage specificity and methionine oxidation and N-terminal acetylation were set as variable modifications. Static residue modification was set for carbamidomethylation of cysteines. The allowed peptide length and mass ranges of 5–50 residues and 200–5000 Da, respectively. Within FragPipe all peptide-spectrum matches (PSMs),

peptides, and proteins were filtered to 1% PSM and 1% protein FDR. FDR was calculated based on the hits to decoy database. MSFragger 3.2 precursor and fragment tolerance was set to 20ppm with mass calibration and parameter optimization enabled. Two missed cleavages were allowed, and two enzymatic termini were specified. Isotope error was set to 0/1/2. The minimum number of fragment peaks required to include a PSM in modeling was set to two and the minimum number required to report the match was four. The top 150 most intense peaks and a minimum of 15 fragment peaks required to search a spectrum were used according to recommended settings. Philosopher 3.4.13 was applied for data analysis [71]. Glycosylation status was analyzed using the UNIPROT database and specific for O-glycosylation data collected in [67, 72, 73].

## Statistical analyses

Statistical analysis was performed using either Excel or GraphPad Prism Software (GraphPad Software Inc., La Jolla, CA, USA). Unless stated otherwise, the comparison of medians (±SD) between two groups was done by a two-tailed Student's *t*-test, whereas multiple groups were compared by one-way ANOVA. All error bars represent standard deviation (SD) and the *p* values <0.05 were considered statistically significant.

## Ethics approval/consent to participate

All uterine leiomyoma tissues were obtained from routine surgeries of otherwise healthy donors after their informed consent. The study protocol was reviewed and approved by the Regional Ethics Committee of the Northern Ostrobothnia Hospital District (license number 2/2017).

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author (elham.khosrowabadi@oulu.fi) on reasonable request.

## Author contributions

E.K did lectin microarray analysis, 3D myoma invasion assay, prepared the figures, planned, and wrote the manuscript together with S.K. T.W, S.K, and A.K were responsible for proteomics, identification of glycoproteins and reviewing the manuscript.

## ACKNOWLEDGMENTS

The authors wish to thank Dr. Maija and Prof. Tuula Salo for providing help in myoma tissue samples.

Dr. Markku Varjosalo (Institute of Biotechnology Proteomics Unit at Helsinki Institute of Life Science (HiLIFE)) is acknowledged for helping with the proteomics analysis.

## CONFLICTS OF INTEREST

Authors have no conflicts of interest to declare.

## FUNDING

We acknowledge the University of Oulu Graduate School, the Finnish National Agency for Education (CIMO), the Thelma Mäkikyrö Foundation, and the Academy of Finland for funding.

## REFERENCES

1. Kohyama-Koganeya A, Nabetani T, Miura M, Hirabayashi Y. Glucosylceramide synthase in the fat body controls energy metabolism in *Drosophila*. *J Lipid Res*. 2011; 52:1392–99. <https://doi.org/10.1194/jlr.M014266>. [PubMed]
2. Ryczko MC, Pawling J, Chen R, Abdel Rahman AM, Yau K, Copeland JK, Zhang C, Surendra A, Guttman DS, Figeys D, Dennis JW. Metabolic Reprogramming by Hexosamine Biosynthetic and Golgi N-Glycan Branching Pathways. *Sci Rep*. 2016; 6:23043. <https://doi.org/10.1038/srep23043>. [PubMed]
3. Akintayo A, Stanley P. Roles for Golgi Glycans in Oogenesis and Spermatogenesis. *Front Cell Dev Biol*. 2019; 7:98. <https://doi.org/10.3389/fcell.2019.00098>. [PubMed]
4. Bishop JR, Schuksz M, Esko JD. Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature*. 2007; 446:1030–37. <https://doi.org/10.1038/nature05817>. [PubMed]
5. Varki A. Biological roles of glycans. *Glycobiology*. 2017; 27:3–49. <https://doi.org/10.1093/glycob/cww086>. [PubMed]
6. Russo D, Della Ragione F, Rizzo R, Sugiyama E, Scalabri F, Hori K, Capasso S, Sticco L, Fioriniello S, De Gregorio R, Granata I, Guarracino MR, Maglione V, et al. Glycosphingolipid metabolic reprogramming drives neural differentiation. *EMBO J*. 2018; 37:e97674. <https://doi.org/10.15252/embj.201797674>. [PubMed]
7. Chang IJ, He M, Lam CT. Congenital disorders of glycosylation. *Ann Transl Med*. 2018; 6:477. <https://doi.org/10.21037/atm.2018.10.45>. [PubMed]
8. Stanley P. Golgi glycosylation. *Cold Spring Harb Perspect Biol*. 2011; 3:a005199. <https://doi.org/10.1101/cshperspect.a005199>. [PubMed]
9. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*. 2012; 13:448–62. <https://doi.org/10.1038/nrm3383>. [PubMed]
10. Maccioni HJ, Giraudo CG, Daniotti JL. Understanding the stepwise synthesis of glycolipids. *Neurochem Res*. 2002; 27:629–36. <https://doi.org/10.1023/a:1020271932760>. [PubMed]
11. Rudd P, Karlsson NG, Khoo KH, Packer NH. Glycomics and Glycoproteomics. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, et al. *Essentials of Glycobiology*. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015–2017. <https://doi.org/10.1101/glycobiology.3e.051>. [PubMed]
12. Nairn AV, Aoki K, dela Rosa M, Porterfield M, Lim JM, Kulik M, Pierce JM, Wells L, Dalton S, Tiemeyer M, Moremen KW. Regulation of glycan structures in murine embryonic stem cells: combined transcript profiling of glycan-related genes and glycan structural analysis. *J Biol Chem*. 2012; 287:37835–56. <https://doi.org/10.1074/jbc.M112.405233>. [PubMed]
13. Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW. Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem*. 2008; 283:17298–313. <https://doi.org/10.1074/jbc.M801964200>. [PubMed]
14. Varki A. Factors controlling the glycosylation potential of the Golgi apparatus. *Trends Cell Biol*. 1998; 8:34–40. [https://doi.org/10.1016/s0962-8924\(97\)01198-7](https://doi.org/10.1016/s0962-8924(97)01198-7). [PubMed]
15. Kellokumpu S. Golgi pH, Ion and Redox Homeostasis: How Much Do They Really Matter? *Front Cell Dev Biol*. 2019; 7:93. <https://doi.org/10.3389/fcell.2019.00093>. [PubMed]
16. Roth J, Zuber C, Wagner P, Taatjes DJ, Weisgerber C, Heitz PU, Goridis C, Bitter-Suermann D. Reexpression of poly(sialic acid) units of the neural cell adhesion molecule in Wilms tumor. *Proc Natl Acad Sci U S A*. 1988; 85:2999–3003. <https://doi.org/10.1073/pnas.85.9.2999>. [PubMed]
17. Hakomori S. Aberrant glycosylation in tumors and tumor-associated carbohydrate antigens. *Adv Cancer Res*. 1989; 52:257–331. [https://doi.org/10.1016/s0065-230x\(08\)60215-8](https://doi.org/10.1016/s0065-230x(08)60215-8). [PubMed]
18. Yang JM, Byrd JC, Siddiki BB, Chung YS, Okuno M, Sowa M, Kim YS, Matta KL, Brockhausen I. Alterations of O-glycan biosynthesis in human colon cancer tissues. *Glycobiology*. 1994; 4:873–84. <https://doi.org/10.1093/glycob/4.6.873>. [PubMed]
19. Kumamoto K, Goto Y, Sekikawa K, Takenoshita S, Ishida N, Kawakita M, Kannagi R. Increased expression of UDP-galactose transporter messenger RNA in human colon cancer tissues and its implication in synthesis of Thomsen-Friedenreich antigen and sialyl Lewis A/X determinants. *Cancer Res*. 2001; 61:4620–27. [PubMed]
20. Brockhausen I. Glycodynamics of mucin biosynthesis in gastrointestinal tumor cells. *Adv Exp Med Biol*. 2003; 535:163–88. [https://doi.org/10.1007/978-1-4615-0065-0\\_11](https://doi.org/10.1007/978-1-4615-0065-0_11). [PubMed]

21. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*. 2015; 15:540–55. <https://doi.org/10.1038/nrc3982>. [PubMed]
22. Greville G, McCann A, Rudd PM, Saldova R. Epigenetic regulation of glycosylation and the impact on chemoresistance in breast and ovarian cancer. *Epigenetics*. 2016; 11:845–57. <https://doi.org/10.1080/15592294.2016.1241932>. [PubMed]
23. Dall'Olio F, Trinchera M. Epigenetic Bases of Aberrant Glycosylation in Cancer. *Int J Mol Sci*. 2017; 18:998. <https://doi.org/10.3390/ijms18050998>. [PubMed]
24. Oliveira-Ferrer L, Legler K, Milde-Langosch K. Role of protein glycosylation in cancer metastasis. *Semin Cancer Biol*. 2017; 44:141–52. <https://doi.org/10.1016/j.semcancer.2017.03.002>. [PubMed]
25. Magalhães A, Duarte HO, Reis CA. Aberrant Glycosylation in Cancer: A Novel Molecular Mechanism Controlling Metastasis. *Cancer Cell*. 2017; 31:733–35. <https://doi.org/10.1016/j.ccell.2017.05.012>. [PubMed]
26. Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. *Annu Rev Pathol*. 2015; 10:473–510. <https://doi.org/10.1146/annurev-pathol-012414-040438>. [PubMed]
27. Kannagi R, Yin J, Miyazaki K, Izawa M. Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants—Hakomori's concepts revisited. *Biochim Biophys Acta*. 2008; 1780:525–31. <https://doi.org/10.1016/j.bbagen.2007.10.007>. [PubMed]
28. Varki A, Kannagi R, Toole B, Stanley P. Glycosylation Changes in Cancer. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH, et al. *Essentials of Glycobiology*. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015–2017. [PubMed]
29. Taniguchi N, Kizuka Y. Glycans and cancer: role of N-glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv Cancer Res*. 2015; 126:11–51. <https://doi.org/10.1016/bs.acr.2014.11.001>. [PubMed]
30. Tsuboi S, Sutoh M, Hatakeyama S, Hiraoka N, Habuchi T, Horikawa Y, Hashimoto Y, Yoneyama T, Mori K, Koie T, Nakamura T, Saitoh H, Yamaya K, et al. A novel strategy for evasion of NK cell immunity by tumours expressing core2 O-glycans. *EMBO J*. 2011; 30:3173–85. <https://doi.org/10.1038/emboj.2011.215>. [PubMed]
31. Sperandio M. Selectins and glycosyltransferases in leukocyte rolling *in vivo*. *FEBS J*. 2006; 273:4377–89. <https://doi.org/10.1111/j.1742-4658.2006.05437.x>. [PubMed]
32. Machida E, Nakayama J, Amano J, Fukuda M. Clinicopathological significance of core 2 beta1,6-N-acetylglucosaminyltransferase messenger RNA expressed in the pulmonary adenocarcinoma determined by *in situ* hybridization. *Cancer Res*. 2001; 61:2226–31. [PubMed]
33. St Hill CA, Baharo-Hassan D, Farooqui M. C2-O-sLeX glycoproteins are E-selectin ligands that regulate invasion of human colon and hepatic carcinoma cells. *PLoS One*. 2011; 6:e16281. <https://doi.org/10.1371/journal.pone.0016281>. [PubMed]
34. Hatakeyama S, Kyan A, Yamamoto H, Okamoto A, Sugiyama N, Suzuki Y, Yoneyama T, Hashimoto Y, Koie T, Yamada S, Saito H, Arai Y, Fukuda M, Ohyama C. Core 2 N-acetylglucosaminyltransferase-1 expression induces aggressive potential of testicular germ cell tumor. *Int J Cancer*. 2010; 127:1052–59. <https://doi.org/10.1002/ijc.25117>. [PubMed]
35. Chen Z, Gulzar ZG, St Hill CA, Walcheck B, Brooks JD. Increased expression of GCNT1 is associated with altered O-glycosylation of PSA, PAP, and MUC1 in human prostate cancers. *Prostate*. 2014; 74:1059–67. <https://doi.org/10.1002/pros.22826>. [PubMed]
36. Fernandes B, Sagman U, Auger M, Demetrio M, Dennis JW. Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer Res*. 1991; 51:718–23. [PubMed]
37. Takano R, Nose M, Nishihira T, Kyogoku M. Increase of beta 1-6-branched oligosaccharides in human esophageal carcinomas invasive against surrounding tissue *in vivo* and *in vitro*. *Am J Pathol*. 1990; 137:1007–11. [PubMed]
38. Cuello HA, Segatori VI, Albertó M, Gulino CA, Aschero R, Camarero S, Mutti LG, Madauss K, Alonso DF, Lubieniecki F, Gabri MR. Aberrant O-glycosylation modulates aggressiveness in neuroblastoma. *Oncotarget*. 2018; 9:34176–88. <https://doi.org/10.18632/oncotarget.26169>. [PubMed]
39. Seelentag WK, Li WP, Schmitz SF, Metzger U, Aeberhard P, Heitz PU, Roth J. Prognostic value of beta1,6-branched oligosaccharides in human colorectal carcinoma. *Cancer Res*. 1998; 58:5559–64. [PubMed]
40. Leathem AJ, Brooks SA. Predictive value of lectin binding on breast-cancer recurrence and survival. *Lancet*. 1987; 1:1054–56. [https://doi.org/10.1016/s0140-6736\(87\)90482-x](https://doi.org/10.1016/s0140-6736(87)90482-x). [PubMed]
41. Åström P, Heljasvaara R, Nyberg P, Al-Samadi A, Salo T. Human Tumor Tissue-Based 3D In Vitro Invasion Assays. *Methods Mol Biol*. 2018; 1731:213–21. [https://doi.org/10.1007/978-1-4939-7595-2\\_19](https://doi.org/10.1007/978-1-4939-7595-2_19). [PubMed]
42. Schumacher U, Adam E. Lectin histochemical HPA-binding pattern of human breast and colon cancers is associated with metastases formation in severe combined immunodeficient mice. *Histochem J*. 1997; 29:677–84. <https://doi.org/10.1023/a:1026404832394>. [PubMed]
43. Rye PD, Fodstad O, Emilsen E, Bryne M. Invasion potential and N-acetylgalactosamine expression in a human melanoma model. *Int J Cancer*. 1998; 75:609–14. [https://doi.org/10.1002/\(sici\)1097-0215\(19980209\)75:4<609::aid-ijc19>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0215(19980209)75:4<609::aid-ijc19>3.0.co;2-3). [PubMed]
44. Brooks SA. The involvement of Helix pomatia lectin (HPA) binding N-acetylgalactosamine glycans in cancer progression. *Histol Histopathol*. 2000; 15:143–58. <https://doi.org/10.14670/HH-15.143>. [PubMed]

45. Mitchell BS, Schumacher U. The use of the lectin Helix pomatia agglutinin (HPA) as a prognostic indicator and as a tool in cancer research. *Histol Histopathol.* 1999; 14:217–26. <https://doi.org/10.14670/HH-14.217>. [PubMed]
46. Schumacher U, Brooks SA, Mester J. The lectin Helix pomatia agglutinin as a marker of metastases--clinical and experimental studies. *Anticancer Res.* 2005; 25:1829–30. [PubMed]
47. Milde-Langosch K, Schütze D, Oliveira-Ferrer L, Wikman H, Müller V, Lebok P, Pantel K, Schröder C, Witzel I, Schumacher U. Relevance of  $\beta$ Gal- $\beta$ GalNAc-containing glycans and the enzymes involved in their synthesis for invasion and survival in breast cancer patients. *Breast Cancer Res Treat.* 2015; 151:515–28. <https://doi.org/10.1007/s10549-015-3425-0>. [PubMed]
48. Parameswaran R, Tan WB, Nga ME, Soon GST, Ngiam KY, Brooks SA, Sadler GP, Mihai R. Binding of aberrant glycoproteins recognizable by Helix pomatia agglutinin in adrenal cancers. *BJS Open.* 2018; 2:353–59. <https://doi.org/10.1002/bjs5.70>. [PubMed]
49. Laferté S, Prokopishyn NL, Moyana T, Bird RP. Monoclonal antibody recognizing a determinant on type 2 chain blood group A and B oligosaccharides detects oncodevelopmental changes in azoxymethane-induced rat colon tumors and human colon cancer cell lines. *J Cell Biochem.* 1995; 57:101–19. <https://doi.org/10.1002/jcb.240570111>. [PubMed]
50. Dwek MV, Ross HA, Streets AJ, Brooks SA, Adam E, Titcomb A, Woodside JV, Schumacher U, Leatham AJ. Helix pomatia agglutinin lectin-binding oligosaccharides of aggressive breast cancer. *Int J Cancer.* 2001; 95:79–85. [https://doi.org/10.1002/1097-0215\(20010320\)95:2<79::aid-ijc1014>3.0.co;2-e](https://doi.org/10.1002/1097-0215(20010320)95:2<79::aid-ijc1014>3.0.co;2-e). [PubMed]
51. Saint-Guiron J, Zeqiraj E, Schumacher U, Greenwell P, Dwek M. Proteome analysis of metastatic colorectal cancer cells recognized by the lectin Helix pomatia agglutinin (HPA). *Proteomics.* 2007; 7:4082–89. <https://doi.org/10.1002/pmic.200700434>. [PubMed]
52. Rambaruth NDS, Greenwell P, Dwek MV. The lectin Helix pomatia agglutinin recognizes O-GlcNAc containing glycoproteins in human breast cancer. *Glycobiology.* 2012; 22:839–48. <https://doi.org/10.1093/glycob/cws051>. [PubMed]
53. Uribe ML, Marrocco I, Yarden Y. EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance. *Cancers (Basel).* 2021; 13:2748. <https://doi.org/10.3390/cancers13112748>. [PubMed]
54. Appert-Collin A, Hubert P, Crémel G, Bennisroune A. Role of ErbB Receptors in Cancer Cell Migration and Invasion. *Front Pharmacol.* 2015; 6:283. <https://doi.org/10.3389/fphar.2015.00283>. [PubMed]
55. Desgrosellier JS, Cheresh DA. Integrins in cancer: biological implications and therapeutic opportunities. *Nat Rev Cancer.* 2010; 10:9–22. <https://doi.org/10.1038/nrc2748>. [PubMed]
56. Zhang K, Myllymäki SM, Gao P, Devarajan R, Kytölä V, Nykter M, Wei GH, Manninen A. Oncogenic K-Ras upregulates ITGA6 expression via FOSL1 to induce anoikis resistance and synergizes with  $\alpha$ V-Class integrins to promote EMT. *Oncogene.* 2017; 36:5681–94. <https://doi.org/10.1038/onc.2017.177>. [PubMed]
57. Khan I, Steeg PS. Metastasis suppressors: functional pathways. *Lab Invest.* 2018; 98:198–210. <https://doi.org/10.1038/labinvest.2017.104>. [PubMed]
58. Odintsova E, Sugiura T, Berditchevski F. Attenuation of EGF receptor signaling by a metastasis suppressor, the tetraspanin CD82/KAI-1. *Curr Biol.* 2000; 10:1009–12. [https://doi.org/10.1016/s0960-9822\(00\)00652-7](https://doi.org/10.1016/s0960-9822(00)00652-7). [PubMed]
59. Bienstock RJ, Barrett JC. KAI1, a prostate metastasis suppressor: prediction of solvated structure and interactions with binding partners; integrins, cadherins, and cell-surface receptor proteins. *Mol Carcinog.* 2001; 32:139–53. <https://doi.org/10.1002/mc.1073>. [PubMed]
60. Tajadura-Ortega V, Gambardella G, Skinner A, Halim A, Van Coillie J, Schjoldager KTG, Beatson R, Graham R, Achkova D, Taylor-Papadimitriou J, Ciccarelli FD, Burchell JM. O-linked mucin-type glycosylation regulates the transcriptional programme downstream of EGFR. *Glycobiology.* 2021; 31:200–10. <https://doi.org/10.1093/glycob/cwaa075>. [PubMed]
61. Zhen Y, Caprioli RM, Staros JV. Characterization of glycosylation sites of the epidermal growth factor receptor. *Biochemistry.* 2003; 42:5478–92. <https://doi.org/10.1021/bi027101p>. [PubMed]
62. Stateva SR, Villalobo A. O-GlcNAcylation of the human epidermal growth factor receptor. *Org Biomol Chem.* 2015; 13:8196–204. <https://doi.org/10.1039/c5ob00443h>. [PubMed]
63. Liu CH, Hu RH, Huang MJ, Lai IR, Chen CH, Lai HS, Wu YM, Huang MC. C1GALT1 promotes invasive phenotypes of hepatocellular carcinoma cells by modulating integrin  $\beta$ 1 glycosylation and activity. *PLoS One.* 2014; 9:e94995. <https://doi.org/10.1371/journal.pone.0094995>. [PubMed]
64. Lin MC, Huang MJ, Liu CH, Yang TL, Huang MC. GALNT2 enhances migration and invasion of oral squamous cell carcinoma by regulating EGFR glycosylation and activity. *Oral Oncol.* 2014; 50:478–84. <https://doi.org/10.1016/j.oraloncology.2014.02.003>. [PubMed]
65. Nguyen AT, Chia J, Ros M, Hui KM, Saltel F, Bard F. Organelle Specific O-Glycosylation Drives MMP14 Activation, Tumor Growth, and Metastasis. *Cancer Cell.* 2017; 32:639–53.e6. <https://doi.org/10.1016/j.ccell.2017.10.001>. [PubMed]
66. Marsico G, Russo L, Quondamatteo F, Pandit A. Glycosylation and Integrin Regulation in Cancer. *Trends Cancer.* 2018; 4:537–52. <https://doi.org/10.1016/j.trecan.2018.05.009>. [PubMed]
67. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen

- S, Pedersen NB, Marcos-Silva L, Gupta R, Bennett EP, Mandel U, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 2013; 32:1478–88. <https://doi.org/10.1038/emboj.2013.79>. [PubMed]
68. The GlycoDomainViewer (<https://glycodomain.glycomics.ku.dk/uniprot/>).
69. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods.* 2017; 14:513–20. <https://doi.org/10.1038/nmeth.4256>. [PubMed]
70. Teo GC, Polasky DA, Yu F, Nesvizhskii AI. Fast Deisotoping Algorithm and Its Implementation in the MSFragger Search Engine. *J Proteome Res.* 2021; 20:498–505. <https://doi.org/10.1021/acs.jproteome.0c00544>. [PubMed]
71. da Veiga Leprevost F, Haynes SE, Avtonomov DM, Chang HY, Shanmugam AK, Mellacheruvu D, Kong AT, Nesvizhskii AI. Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat Methods.* 2020; 17:869–70. <https://doi.org/10.1038/s41592-020-0912-y>. [PubMed]
72. Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, Bennett EP, Mandel U, Wandall H, Lavery SB, Clausen H. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods.* 2011; 8:977–82. <https://doi.org/10.1038/nmeth.1731>. [PubMed]
73. Vakhrushev SY, Steentoft C, Vester-Christensen MB, Bennett EP, Clausen H, Lavery SB. Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Mol Cell Proteomics.* 2013; 12:932–44. <https://doi.org/10.1074/mcp.O112.021972>. [PubMed]